Social Push and the Direction of Innovation*

Elias Einiö, *VATT Institute for Economic Research*
Josh Feng, *University of Utah – David Eccles School of Business*
Xavier Jaravel, *London School of Economics* and *CEPR*

March 2025

Abstract

What are the implications of unequal access to innovation careers for the direction of innovation and inequality? Leveraging novel linked datasets in the United States and Finland, we document that innovators create products more likely to be purchased by consumers like them in terms of gender, socioeconomic status, and age. Homophily exists both within narrow product categories and across industries, and has been stable in recent decades. Incorporating this "social push" channel into a growth model, we estimate that unequal access to innovation careers has a large effect on cost-of-living inequality and long-run growth.

*For thoughtful comments, we are grateful to Daron Acemoglu, Ruchir Agarwal, Philippe Aghion, Victor Bennett, Corina Boar, Jen Brown, Raj Chetty, Louphou Coulibaly, Gary Dushnitsky, Louisa Gagliardi, Matthew Jackson, Ben Jones, Kieran Larkin, Katja Mann, Kyle Myers, Tommaso Porzio, Stefanie Stantcheva, Jason Snyder, Johannes Spinnewijn, Peter Thompson, John Van Reenen, Fabian Waldinger, and Heidi Williams. The article also benefited from comments from seminar participants at the AOM meetings, ASSA meetings, BAIC, Bank of Finland, BU Questrom, BYU-Utah Joint Research Day, Female Entrepreneurship Seminar Series, IIOC, Kauffman Showcase, Max Planck Institute, NBER Summer Institute, Society for Economic Dynamics, Society of the Study of Economic Inequality, SMS meetings, Strategy Science, UCLA Anderson, and VMACS. We also thank Gabriel Leite Mariante for excellent research assistance, and PatentsView for providing data on the gender of patent inventors. The results reflect our own analyses, calculated based in part on data from Nielsen Consumer LLC and marketing databases provided through the Nielsen IQ Datasets at the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the Nielsen IQ data are ours 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, the Ewing Marion Kauffman Foundation, and Sorenson Center for Discovery and Innovation at the University of Utah. This paper replaces and supersedes an earlier working paper that was first circulated in May 2019 under the title "What Drives Inclusive Innovation? The Importance of Innovators' Backgrounds".

"If nobody else is going to invent a dishwashing machine, I'll do it myself."

Josephine Cochrane, inventor of the dishwasher (U.S. patent no. 355,139)

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., Linder, 1961; Schmookler, 1966; Aghion and Howitt, 1992; Acemoglu, 2002; Acemoglu and Linn, 2004; Acemoglu, 2007; Jaravel, 2019). At the same time, the discovery and pursuit of entrepreneurial opportunities depends on the distribution of information in society (e.g., Hayek, 1945), often requires engagement with specific real-world problems and users (e.g., Von Hippel, 1986; Shane, 2000), and is incentivized by non-pecuniary benefits (Stern, 2004; Hurst and Pugsley, 2011). Because innovators from different socio-demographic backgrounds are likely to be exposed to different ideas and to be responsive to different intrinsic motivations, they may pursue different types of innovations, independent of financial incentives. Despite the apparent plausibility of this "social push" channel, little is known about its economy-wide importance for the direction of innovation, its relevance in terms of innovators' family backgrounds, and its implications for economic inequality.[1]

Understanding the social push channel is particularly relevant given recent research showing that innovators in many countries are not representative of society at large. Women, minorities and individuals from low-income backgrounds and certain regions are vastly underrepresented among innovation leaders including startup founders, patent inventors, and venture capitalists (e.g., Bell et al., 2019b; Agarwal and Gaulé, 2020; Hvide and Oyer, 2020; Aghion et al., 2023).

Several historical examples suggest that innovators' personal experience may shape their entrepreneurial vision, and in turn inequality across the socio-demographic groups who benefit from these innovations. Despite the lack of opportunities for most American women

[1] Contemporaneous research focusing on gender, specific sectors, and early stages of innovation documents that patents with a female lead inventor are more likely to focus on women's health (Koning et al., 2021), provides evidence of gender homophily between scientific authors and commercializing inventors (Koffi and Marx, 2023), and shows how undergraduate gender diversity affects the direction of scientific research (Truffa and Wong, 2022).

in the 19th century, Josephine Cochrane managed to receive a U.S. patent for her "Dish Washing Machine" in 1886; she wanted to protect her fine china and avoid having to hand-wash them herself.[2] Madam C.J. Walker, an African-American entrepreneur and the first female self-made millionaire in America, made her fortune in the late nineteenth century by developing and marketing a line of cosmetics and hair care products for Black women; she suffered severe dandruff and other scalp ailments herself from a young age. Louis Braille invented a world-famous reading and writing system for visually impaired people; he was himself blinded at the age of three as a result of an accident in his father's workshop.

Similar examples can be found in the modern era in terms gender, socio-economic background, race, and age. For instance, entrepreneur Ida Tin developed an app to track the menstrual cycle and coined the term "femtech" to refer to new health technologies targeting women.[3] Dating apps are another salient example. Bumble, for instance, was founded by a woman and has more female users than other dating apps, such as Tinder. The scholarship app Scholly, which helps students look for private scholarship funds, illustrates the role of socio-economic background: this venture is directly rooted in the life experience of its founder, Christopher Gray, who was the son of a struggling single mother and did not have the money to attend college.[4] Another example is the American hip hop apparel company FUBU – standing for "for us, by us" –, a clothing company founded by Daymond John at the age of 23, initially targeting young men in his neighborhood.

Going beyond anecdotal evidence, the goal of this paper is to systematically characterize the importance of innovators' background and social experience for the direction of modern innovations, and to assess the implications for inequality. We do so in two steps. First, we uncover new stylized facts on the relationship between innovators' backgrounds and the

[2] Josephine Cochrane was posthumously inducted into the National Inventors Hall of Fame in 2006 for patent 355,139 issued on December 28, 1886, for her invention of the first hand-powered dishwasher. The United States Patent Office writes: "she struggled against society's limits on women, working tirelessly to build a successful prototype, sell her invention, and ultimately turn a tedious task into an iconic American appliance."

[3] Another example in this space is the work of Cindy Eckert, an American entrepreneur known for founding Sprout Pharmaceuticals, which introduced the first FDA-approved drug designed to enhance female libido. Eckert subsequently founded The Pink Ceiling, a venture capital investment firm focusing on products targeting women. Conversely, the company Ro, which was started by a male founder, focuses on health issues affecting men such as erectile dysfunction.

[4] Students can use the app to search for private scholarship funds based on various criteria, including their major and state of residence, making the process more accessible.

direction of innovation. Second, we assess the quantitative importance of the social push channel for economic growth and cost-of-living inequality across consumer groups in general equilibrium. Overall, we find that innovators bring to market new goods that are significantly more likely to be purchased by consumers like themselves, which, combined with the under-representation of certain groups in the innovation system, leads to reduced growth and greater cost-of-living inequality.

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. The key challenge is to push the data frontier to paint a comprehensive portrait of the relationship between consumer and innovator characteristics. We do so by building several datasets linking consumer characteristics to innovators' gender, parental income, and age. Consumer characteristics are measured in comprehensive consumption surveys, in detailed scanner data for consumer packaged goods, and in a new data set covering mobile phone applications. Innovators and their backgrounds are identified from patent records, start-up and venture capital databases, registries of firms, and administrative tax records.

Using this comprehensive dataset, we document that innovator-consumer homophily is a very common feature of modern innovation systems, holding in both the United States and Finland for all types of innovations we study (new phone applications, new consumer goods, patents, and new firms). Female entrepreneurs create products that have 18 percent higher female customer share than products in the same category created by male entrepreneurs. A one-year increase in an entrepreneur's age at founding is associated with a 0.14 year increase in average consumer age. We also find that innovators from a high-income family are more likely to create products purchased by high-income consumers: they 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.[5]

We then assess heterogeneity and potential mechanisms driving consumer-innovator homophily. We examine in turn heterogeneity by industry, product, consumer, and innovator characteristics. Overall, we find that homophily is a broad-based phenomenon, which is not

[5] We also find strong geographic homophily. Phone apps created by venture-backed companies are three times more likely to be used by consumers in the same state as their headquarters, even after controlling for category-by-state fixed effects.

driven by a small number of products, industries, consumer, or innovator types. We also document some meaningful heterogeneities across categories, such as stronger homophily for health and household products. Finally, we find that vertical quality differentiation cannot explain observed homophily patterns.[6]

To assess whether innovator-consumer homophily is likely to persist as under-represented groups gain better access to the innovation system, we examine the relationship between the secular increase in the share of female inventors and homophily. We find that gendered inventor-consumer homophily has remained stable over the past 25 years, even though the share of women among patent inventors has more than doubled during this period. Overall, the first part of the paper establishes that innovator-consumer homophily is a striking empirical regularity.

In the second part of the article, we investigate the relevance of the social push channel for cost-of-living inequality and economic growth. Our model extends Romer (1990) to include multiple sectors, heterogeneity in consumer tastes, and barriers to entering the innovation system that vary across socio-demographic groups, including by gender. Importantly, this modeling framework allows for directional differences in the choice of target markets by innovators' socio-demographic backgrounds. Modeling choices and model parameters are disciplined by both the stylized facts about innovator-consumer homophily and other findings in the literature. We analyze counterfactual scenarios reducing access barriers, thus changing the equilibrium size and composition of the pool of innovators.

We find that, in a world with social push, access barriers for potential female innovators create a 18.20% difference in cost-of-living between men and women, which is almost as large as the gender pay gap. Lowering access barriers induces some highly productive women to start innovating, and reduces the cost-of-living gap due to the disproportionate impact on female consumers through homophily. We also find large impacts on overall growth rates: eliminating access barriers leads to an increase in long-run growth rates from 2% to 3.4% per year.[7] We obtain comparably large effects, for both inequality and growth, when analyzing

[6] In our working paper (Einiö et al., 2023), we also show that the socio-demographic composition of study peers influences the types of businesses subsequently started by entrepreneurs, with no change in profits. This suggests that the direction of innovation is influenced by social factors independent of financial incentives.

[7] In Section IV.E we assess the plausibility of the model's predicted growth effects with auxiliary tests and comparisons to relevant findings in the literature, suggesting that the orders of magnitude are plausible.

counterfactuals by income groups. In particular, we find that innovator-consumer homophily by socioeconomic status has a larger impact on cost-of-living inequality across the income distribution than market size effects.

Together, our analysis addresses the question of whether the "social push" channel is an important driver of the direction of innovation. The descriptive homophily evidence provides support to the idea that the social push channel operates at a large scale, both across and within industries, in countries as different as the United States and Finland, and for many different types of innovations. The structural model shows that the economic implications of the social push channel can be significant even in general equilibrium.

Prior work. This article primarily contributes to the literature on unequal access to innovation. Recent research documents that certain groups are underrepresented in the innovation system, particularly in terms of gender, race, parental background, and geography (Bell et al., 2019b; Toole et al., 2019; Cook et al., 2022). Several articles have studied potential mechanisms that can explain this under-representation, including funding barriers (Brooks et al., 2014; Malmström et al., 2017; Kanze et al., 2018), preferences (Thébaud, 2010; Bönte and Piegeler, 2013; Caliendo et al., 2015), intergenerational transmission (Dunn and Holtz-Eakin, 2000; Mishkin, 2021; Hvide and Oyer, 2020), social exposure (Markussen and Røed, 2017; Calder-Wang and Gompers, 2021), and frictions in social networking (Howell and Nanda, 2023). Consistent with our findings, Koning et al. (2021) document that biomedical patents with female first authors are more likely to mention female medical conditions.

Our work builds on and extends this literature in two ways: (i) we provide comprehensive, economy-wide evidence on the homophily between innovators and their consumers by several socio-demographic factors (including but not limited to gender); and (ii) we quantify the macroeconomic importance of this channel in general equilibrium using a structural model. Furthermore, we focus on products that make it to market (i.e., are economically viable), which complements contemporaneous work documenting gender homophily at earlier stages in the research process (Koning et al., 2021; Koffi and Marx, 2023; Truffa and Wong, 2022).

We also contribute to the long literature on the determinants of the direction of innovation. As noted earlier, the economics literature has focused on market size as the key driver of the direction of innovation (Schmookler, 1966; Acemoglu, 2002, 2007), with implications for unequal gains from innovation across consumer groups (Jaravel, 2019). A related concept in the innovation literature is user innovation: Von Hippel (1986) and a body of subsequent work has noted the importance of users in driving the direction of innovation. Our paper provides evidence for a distinct mechanism, showing the importance of innovators' social background and gender in determining the overall direction of innovative activity in an economy with unequal access to innovation.

Finally, this article contributes to the literature on endogenous growth and inequality. Our model builds on the product variety literature, starting with Romer (1990) and 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. Hsieh et al. (2019) propose an analysis of the impact of misallocation of talent on welfare, but without entrepreneurs and innovators; our framework extends this analysis by developing an endogenous growth model. Our results show that misallocation in the innovation sector can have a sizable impact on long-run growth *rates* and cost-of-living inequality, while the reallocation channel studied by Hsieh et al. (2019) affects the *level* of GDP.

The remainder of the article is organized as follows. Section II presents the data. Section III documents the homophily between innovators and their consumers. Section IV examines the implications of these findings for growth and inequality with a quantitative growth model.

II Data and Summary Statistics

This section presents our data sources, samples, key variables, and summary statistics.

II.A Data Sources and Variable Descriptions

To document the extent and generality of the "social push" channel, we compile complementary datasets in the United States and Finland, allowing us to study different types of innovations. We use two "micro" datasets for the United States. We start with sector-specific analysis employing respondent-level information on usage of phone applications and purchases of consumer packaged goods. We then provide economy-wide evidence for the full consumption basket based on measures of consumer characteristics at a detailed industry level. We draw socio-demographic information on innovators – including gender, parent income, and age – from startup databases (Crunchbase), patent records (USPTO and PATSTAT databases) and administrative data (for inventors and entrepreneurs in Finland).

Phone applications. The first micro dataset is the Nielsen's Electronic Mobile Measurement (EMM) panel, which tracks the mobile phone application usage of a representative sample of ten thousand U.S. consumers in every month. To our knowledge, this database has never been used for research purposes. The data contains detailed information on the gender, income, and state of residence for each panelist from April 2017 to June 2019. Apps are classified into 58 categories (e.g., gaming, social networking, photography, etc.). A key advantage of this dataset is that it measures consumption at the individual level.

The data provide the company name 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. It contains data on the name, location, and founders of each startup. For each founder, it also records gender and LinkedIn URL.[8] We clean company names in both the phone app and Crunchbase datasets prior to merging. Because Crunchbase attempts to track both legacy companies and startups, we define startups as firms founded after 2007, primarily to reflect the start of the smartphone era.

Consumer packaged goods. To track consumption patterns for consumer packaged goods in the United States, we rely on Nielsen's Homescan Consumer Panel.[9] The data are based on a panel of 40k-60k consumers and provide information on household-level expenditure by products identified by barcodes from 2004 to 2016. Products are classified into nine departments (e.g., non-food groceries), 118 groups (e.g., alcoholic beverages), and 1305 modules (e.g., light beer). These categories account for about 15% of aggregate consumption

[8] Crunchbase uses first names to guess the gender of each person, and then manually checks for errors. In our linked dataset, the number of available phone app categories falls to 53, compared to 58 in the overall phone application data.

[9] These data have been widely used in economic research (for an overview, see, e.g., Dubois et al., 2022).

expenditures, and about 40% of expenditures on goods. For each household, Nielsen also records income, family structure, and household type (single female, female-led, married, male-led, single male).

To identify the manufacturer associated with each bar code in the Nielsen Consumer Panel, we use manufacturer prefix data provided by GS1, the organization in charge of allocating barcodes. The data contains the universe of barcode prefixes as of February 2016, with information on the current owner of the prefix. The GS1 data links almost all purchased goods to a manufacturer (97.5 percent of total revenue and 98.5 percent of total quantities).

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 additional steps, we also manually check unmatched companies in GS1 and Crunchbase that share the same city and first word of the company name and manually collect additional demographic information on each founder. Finally, we also match all GS1 companies to patent data by company name. This allows us to measure the gender and age composition of the inventors who patent at a given company. As the linked Nielsen-GS1-Crunchbase dataset focuses on venture-backed startups, it excludes the large mass of established consumer packaged goods.

Industry-level data in the United States. To extend our analysis to cover the whole consumption basket, we leverage several data sources in the United States. To characterize consumers, we employ the Consumer Expenditure Survey (CEX), which is administered by the U.S. Bureau of Labor Statistics and provides information on consumption patterns by U.S. households at a detailed industry level. We use these data to construct industry-level measures for the gender, socio-economic, and age composition of consumers. To characterize patent inventors, we use individual-level data on gender from PatentsView and age from Jones (2009).[10] We link these datasets by six-digit NAICS industry. Specifically, we use the CEX category to NAICS industry crosswalk and the data processing steps in Borusyak and Jaravel (2018). For inventors, we link the NAICS code by primary patent class, using the concordance created by Lybbert and Zolas (2014).[11] Furthermore, we obtain information on

[10]Patentees in our dataset are disambiguated by PatentsView following Monath et al. (2021).

[11]Their text-based crosswalk maps each U.S. Patent Class to a set of six-digit NAICS codes, each with some probability weight. Using these weights, we compute a weighted average of consumption statistics for

entrepreneurship and parental income from the Panel Survey of Income Dynamics (PSID), which we link to the CEX at the level of the broader industries available in PSID.

As a summary measure of consumer income, we use an industry's income elasticity, which captures the relationship between a household's expenditure share on an industry and the household's total expenditures (the Engel curve). Specifically, we rely on the industry income elasticities estimated by Borusyak and Jaravel (2018) in the Consumer Expenditure Survey.[12]

Industry-level data in Finland. Finally, we use Finnish administrative data covering the full working-age population. The dataset is based on administrative registers compiled by Statistics Finland. It provides individual-level information on income, entrepreneurship, and industry. The data also include information on family links from the Finnish Population Information System, which allows us to measure the parental income of individuals. We study in turn entrepreneurs and patent inventors.

Information on entrepreneurship status is based on pension contribution and tax records. We use the status for the last week of the year, which allows for temporal consistency across variables.[13] The other key variable is the unique company identifier, which 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.

To identify patent inventors, we link individuals in the PATSTAT database to the Finnish population panel by first name, family name, postcode, and company identifier. The company identifiers are drawn from the Business Information System web interface maintained by the Finnish Patent and Registration Office. We also use different combinations of the match variables to include inventors who are not associated with a company in the popu-

each patent class.

[12]Borusyak and Jaravel (2018) split households in the CEX sample into 11 bins by pre-tax household income and compute expenditure shares across all spending categories for each bin. They regress, across income bins, spending shares on the log of total expenditure in this income group, averaged across households. This approach recovers income elasticities without making any parametric assumptions on the utility function or estimating demand structurally. Intuitively, higher-income consumers have larger expenditure shares on income-elastic products.

[13]An individual is defined as an entrepreneur if she 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.

lation panel, have different spelling of the first or family name in the two datasets, or have missing location information in PATSTAT. We include only exact unique matches.

To study consumption patterns and the direction of innovation, we use measures of consumer characteristics from the U.S. CEX because it is more granular than available Finnish consumption surveys.[14] We link the CEX to the Finnish population panel by the industry code of the company an individual is associated with in the last week of the year. The industry codes available in Statistics Finland are NACE codes, which are standard in Europe; to link the consumption data, we use a crosswalk between NACE and NAICS.[15]

II.B Summary Statistics

Panel A of Table 1 provides summary statistics on the characteristics of innovators and consumers for the two sector-specific micro datasets available in the United States, phone applications and consumer packaged goods.

We find that female founders are underrepresented in both sectors. Companies with at least one female founder represent only 14 percent of venture-backed startups in the phone app industry and 24 percent of startups in consumer packaged goods. The rates of female venture capital partner involvement are 15 percent for phone applications and 4 percent for consumer packaged goods. There is substantial variation in consumer gender composition across startups. For phone applications, the average time share of female users is 54% and the standard deviation is 38pp across applications. For consumer packaged goods, the average share of purchases coming from families with a female head-of-household is 26%, with a standard deviation of 25pp. Heterogeneity by age is lower.

[14]The Finnish consumption survey describes expenditure patterns at the level of 89 categories by the purpose of consumption (COICOP), which is not sufficiently detailed to characterize innovator-consumer homophily accurately. Linking the Finnish and U.S. consumption surveys at this level of product aggregation, we find that the income elasticities and consumption shares are very similar in both datasets. Specifically, we rank the broad COICOP categories by average consumer income, and we find that the rank-rank correlation coefficient between the Finnish and U.S. datasets is 0.73 (s.e. 0.12). Furthermore, using OECD data on household consumption at the 3-digit level of the COICOP classification, we find that U.S. consumption shares predict Finnish shares well: in a regression of Finnish on U.S. consumption shares, the slope is 0.83 (s.e. 0.27).

[15]The match rate for the sample of individuals with industry codes in the population panel is 80%, of which 52% are matched by 4-digit, 15% by 3-digit , and 33% by 2-digit industry. We focus on the 2007–2015 period, because the match rate is poor before 2007 due to a change in the industry classification; 2015 is the last year for our linked inventor data.

Panel B of Table 1 reports the industry-level patterns from the inventor-CEX data covering the full consumption basket, for both the United States and Finland. The first part of the panel reports statistics on innovators, considering both entrepreneurs (Col. (1) and (3)) and patent inventors (Col. (2) and (4)). The table shows that there is substantial variation in terms of innovators' gender, age, and parent income. The second part of the table documents that there is also significant heterogeneity in consumer characteristics, including age, gender, and income elasticities – our summary measure of the variation in consumer incomes. In sum, we find large variation in the characteristics of both innovators and consumers, at all levels of analysis. Our goal in the next section is to estimate the extent to which these characteristics covary.

Prior work has widely documented the under-representation of women and individuals from low-income backgrounds in innovation careers (e.g., Thursby and Thursby, 2005; Ding et al., 2006; Akcigit et al., 2017; Kahn and Ginther, 2017; Bertrand, 2020; Hvide and Oyer, 2020; Feng et al., 2022; Aghion et al., 2023; Koffi and Marx, 2023). For instance, Bell et al. (2019b) report that, in the 1980s in the United States, eight out of 1,000 children born to parents in the top 1% of the income distribution became inventors, 10 times higher than the rate among those with below-median-income parents. They document that the fraction of women among patent inventors rises slowly over time, from 7% in the 1940 birth cohort to 18% for the 1980 birth cohort. We find similar patterns of under-representation in our data, as reported in Figure C1. Gender gaps among entrepreneurs and patent inventors close slowly over time. The gap in inventorship by parental income also narrows slowly. At these rates, it will take at least 50 years to reach parity by gender or parental income.

III Estimating Innovator-Consumer Homophily

This section estimates the extent of innovator-consumer homophily, which we find to hold across all measures of innovations and sectors we consider, in both the United States and Finland.

We report the homophily estimates separately for phone applications (Subsection III.A), consumer packaged goods (Subsection III.B), and at the industry level (Subsection III.C),

Table 1: Summary Statistics

Panel A: Phone Applications and Consumer Packaged Goods in the United States

		Phone Applications	Consumer Packaged Goods
Innovator statistics	# VC-backed startups	1,578	158
	# Manufacturers with patents	N/A	1,191
	Female founders ≥ 1	0.14 (0.35)	0.24 (0.43)
	Female VC partner ≥ 1	0.15 (0.36)	0.04 (0.19)
	Founder age at founding	N/A	35.56 (9.00)
	Female patent inventor ratio	N/A	0.11 (0.13)
Consumer statistics	# Startup Products	3,211	4,058
	# Product categories	53	294
	Female consumer share	0.54 (0.38)	0.26 (0.25)
	Consumer average age	41.36 (10.84)	47.18 (7.62)
	# Panelists	50,725	168,775
	Data sources	Nielsen EMM, Crunchbase	Nielsen Homescan, Crunchbase, USPTO
	Timeframe	2017–2019	2007-2016

Panel B: Industry-level Data in the United States and Finland

		United States		Finland	
		Entrepreneurs (1)	Patent Inventors (2)	Entrepreneurs (3)	Patent Inventors (4)
Innovator Statistics	# Innovators	325	2,219,193	344,698	9,643
	Fraction female	0.27 (0.30)	0.12 (0.32)	0.35 (0.48)	0.078 (0.26)
	Fraction parent income in top 20%	0.345 (0.21)	N/A	0.11 (0.31)	0.26 (0.44)
	Age	43.03 (6.2)	47.0 (13.9)	46.5 (10.4)	41.7 (9.4)
	# Industries	19	325	476	342
Consumer Stats.	Female consumer share	0.60 (0.071)	0.57 (0.09)	0.63 (0.10)	0.56 (0.09)
	Industry income elasticity	1.21 (0.30)	1.07 (0.36)	1.12 (0.53)	1.27 (0.35)
	Consumer average age	48.9 (5.1)	47.4 (2.2)	51.19 (6.19)	49.11 (4.13)
	# Panelists	20,700		20,700	
	Data source	CEX, PSID	CEX, USPTO	CEX, Admin. data	CEX, Admin. data, PATSTAT
	Timeframe	2017	1976-2015	2007-2015	

Notes: This table provides summary statistics on innovators and consumers for the micro-datasets in the U.S. (Panel A) and the industry-level analysis in both the U.S. and Finland (Panel B). In Panel B, the first row shows the number of innovators in the largest available samples. For some variables the sample is restricted to a smaller number of individuals due to data limitations. The number of innovators with available information on parent income is 275 in Column 1; 99,189 in Column 3; and 3,812 in Column 4. Age in Column 2 uses data from Jones (2009), available for 48,156 inventors. Standard deviations are reported in parentheses and computed across the most detailed available product category and across industries.

using regression specifications of the form:

$$\text{ConsumerType}_{ij} = \alpha + \beta \cdot \text{InnovatorType}_{ij} + \mu_k + \varepsilon_{ij}, \tag{1}$$

where i indexes a product sold by company j, $\text{InnovatorType}_{ij}$ is a characteristic of the innovator, and ConsumerType_{ij} is a measure of consumer characteristics in the firm's market for product i. We study alternative socio-demographic characteristics to separately estimate homophily by gender, parent income, and age. For example, for gender homophily, "InnovatorType" is a binary indicator for female innovator and "ConsumerType" is the share of sales of product i to women in the innovator's company j. μ_k is a fixed effect for product category k, which we use in certain specifications to assess whether homophily arises primarily within or across product categories.

As discussed in Section IV, the regression coefficient β has a direct consumer welfare interpretation under the assumption of Constant Elasticity of Substitution (CES) consumer demand, allowing us to assess the magnitude of the distributional effects across consumers that arise from unequal access to innovation careers through innovator-consumer homophily. Our homophily estimates will thus discipline the structural analysis of cost-of-living inequality in the last part of the paper.

III.A Innovator-Consumer Homophily for Phone Applications

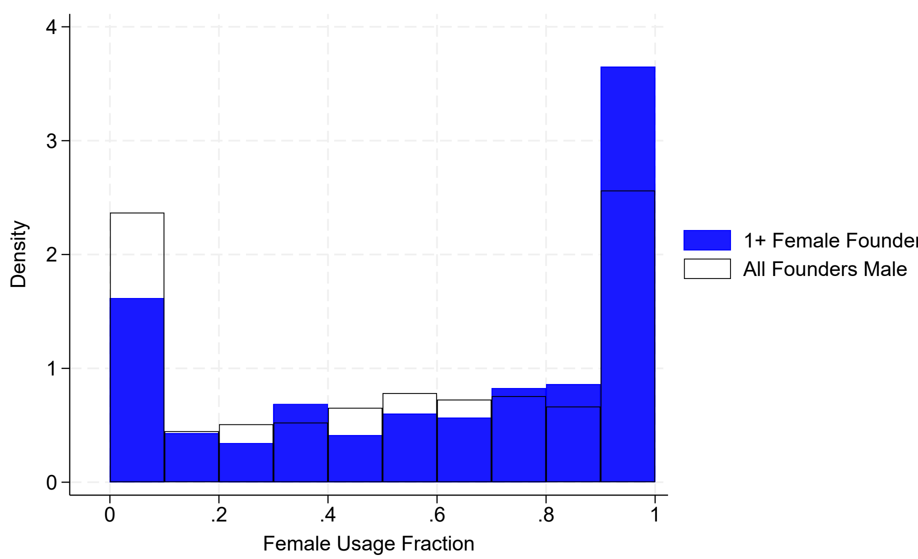
We now estimate innovator-consumer homophily for phone applications, which are one of the iconic forms of innovation of the past decade. Table 2 presents the estimates.

Table 2: Innovator-Consumer Homophily for Phone Applications

	Female User Share		
	(1)	(2)	(3)
Female Founder Fraction	0.112***	0.0955***	0.0929**
	(0.0358)	(0.0355)	(0.0399)
Female VC Fraction			0.149*
			(0.0793)
Fixed Effects	None	Category	None
Sample Size	3,211	3,211	1,391

Notes: Estimates of Equation 1 based on all phone applications created by VC-backed startups. The outcome variable is the fraction of time usage by female users, with a sample mean of 0.542. The level of observation is a firm-application. Female VC fraction is measured based on the gender of venture capital partners involved in funding rounds. Funder information is unavailable for about half of the data points in the core sample. Standard errors are clustered at the startup level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Share of Female Usage of Phone Applications by Founder Gender Composition



Notes: The sample used in this figure includes all phone applications for VC-backed startups. The histograms depict the distribution of time use by gender for phone apps from a VC-backed startup with either at least one female founder (blue histogram) or all founders being male (white histogram). For example, a value above 0.9 for "Female usage fraction" on the x-axis covers apps for which more than 90% of time use is accounted for by female users.

We focus on gender in a regression of the fraction of time usage of an application accounted for by female users on measures for the gender composition of the startup founders as well as of the venture capitalists who fund the startup. Female users account for 54%

of total time spent on phone applications created by VC-backed startups, and the share is 11.2pp higher (or 20 percent of the baseline rate) when moving from an all-male founding team to an all-female founding team (Column 1). Figure 1 illustrates these homophily patterns non-parametrically by showing the distribution of female user shares separately for all-male and at-least-one-female founder teams. The figure shows that homophily primarily stems from gender-specialized applications that have more than 90% or less than 10% female users.

Column (2) of Table 2 shows that the homophily estimate remains similar in magnitude when the specification includes fixed effects for the 53 application categories (e.g., productivity, social media, etc.). The stability of the coefficient across Columns (1) and (2) indicates that gender homophily occurs to a similar degree within and across these detailed application categories. In Column (3), we augment the specification to study gender homophily between consumers and both founders and venture capitalists. The level of founder homophily remains stable, while homophily by the gender composition of venture capitalists is even stronger. For applications created by companies funded only by female venture capitalists, female users account for 69 percent of total time usage, which is 15pp higher (or 27.7 percent of the baseline rate) relative to companies funded only by male venture capitalists. With both an all-female founding team and a female venture capitalist, the share of female users is 24pp higher (or 44.4 percent of the baseline rate).

Table C1 documents homophily by place of residence, studying the fraction of time usage of an application by users located in the same U.S. state as the founder of the application. We estimate Equation (1) with "ConsumerType$_{ij}$" being the share of application i's usage by consumers in state j and "InnovatorType$_{ij}$" being an indicator of whether the startup that created the application is based in state j. We find a large "home bias": the time share of users in the same state as the founder is 8.8pp larger than for users from other states, which is five times the average state share. The magnitude reduces to three times the average state share when including category-by-state fixed effects.

III.B Innovator-Consumer Homophily for Consumer Packaged Goods

We now turn to product innovations within consumer packaged goods sector, a segment which has been widely examined in the literature on creative destruction at the micro level (e.g., Broda and Weinstein, 2010). Our specification includes fixed effects for detailed product categories (product modules), so that we isolate homophily arising at a granular level, before turning to differences arising across industries in the next section.

We first analyze gender homophily, measuring consumer gender composition by the share of sales to households with a female head. Column (1) of Table 3 reports that startups founded by female entrepreneurs are more likely to sell to female consumers, with a 4.7pp higher share of sales to households with a female head for startups with all-female founders, compared to those with all-male founders. This constitutes a difference of 18.8% relative to the baseline rate of 25.6%. Founder-consumer gender homophily for consumer packaged goods is thus quantitatively similar to our estimates for phone applications.

Table 3: Innovator-Consumer Homophily for Consumer Packaged Goods

	Share of Sales to Women		Average Consumer Age, Sales-weighted
	(1)	(2)	(3)
Female Founder Fraction	0.047**		
	(0.021)		
Female Patent Fraction		0.025*	
		(0.015)	
Founder Age			0.135**
			(0.052)
Product Module F.E.	Yes	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 295,292$	Startups, $N = 4,058$

Notes: Estimates of Equation 1 based on all products created by VC-backed startups that are tracked by Nielsen Homescan. In columns (1) and (2), the outcome variable is the fraction of sales to households with a female head of household. The sample means are 0.256 in column (1) and 0.265 in column (2). In column (3), the outcome variable is the average age of consumers, using sales weights, with a sample mean of 47.2. The level of observation in this table is a firm-product. Standard errors are clustered at the firm level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Column (2) conducts a similar analysis in the sample of manufacturers with at least one patent. Manufacturers with a higher share of female patent inventors sell more to female consumers: a change from all-male to all-female inventor composition in a firm is associated with a 2.5pp increase in sales to households with a female head (or 9.6% of the baseline rate). This homophily estimate is smaller than for founders in column (1), consistent with the hypothesis that the founder may set the "entrepreneurial vision" of the startup, whereas patent inventors implement firm-level goals.

In addition, the consumer packaged goods dataset allows us to study homophily by age, as reported in column (3) of Table 3. For each startup in our sample, we compute the average, sales-weighted age of consumers. We find that entrepreneurs that are one year older (at the time of founding) sell to consumers that are on average 0.135 years older. Thus, a one standard deviation increase in founder's age (9 years) is associated with an increase in average consumer age of 1.22 years, about 2.7% of the average consumer age of 47.2 years.

Several robustness checks are reported in the Online Appendix. First, we obtain similar results when we repeat the analysis with an alternative measure of consumer gender, weighting sales by the fraction of female members in the household (Table C2). Second, we also present an alternative analysis of age-based homophily, using age group bins, in Table C3. Finally, we present complementary analyses suggesting that the inventor results in the baseline sample are not driven by established products (Tables C4 and C5).

III.C Industry-Level Estimates of Innovator-Consumer Homophily

We now turn to the industry-level homophily estimates, which are reported in Table 4 for both the United States and Finland. In these homophily regressions, innovator characteristics vary at the individual level, whereas the outcomes (e.g., the share of sales to female consumers) only vary at the industry level. The industry-level homophily estimates thus only capture the "between-industry" component of homophily and are additive to the "within-industry" homophily patterns documented in the previous sections.[16]

[16]With Y_i the outcome for innovator i, we can write $Y_i \equiv Y_{j(i)} + \Delta Y_i$, where $Y_{j(i)}$ is the average outcome in i's industry, indexed by j, and ΔY_i is the deviation between i's outcome and the industry average. By linearity of OLS, the overall homophily regression coefficient, obtained by regressing Y_i on innovator characteristic X_i, can be equivalently obtained by adding the industry-level regression coefficient (which we analyze in Table 4) to the within-industry coefficient (which we analyze in Tables 2 and 3 for specific industries).

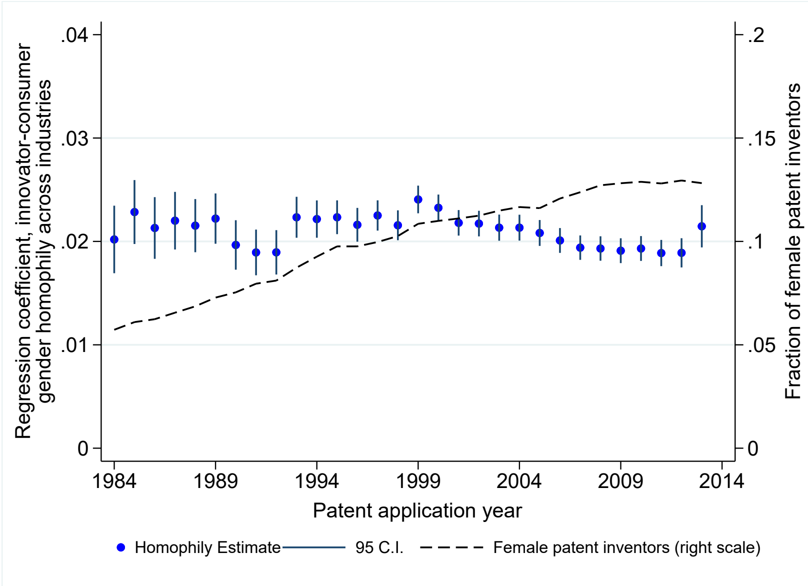
Table 4: Innovator-Consumer Homophily across Industries

	Share of Industry Sales to Women			Industry Income Elasticity			Average Consumer Age, Sales-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Patent Inventor	0.0222***	0.0484**							
	(0.000194)	(0.0037)							
Female Entrepreneur			0.0302***						
			(0.0004)						
Patent Inventor's Log Parent Income					0.0304**				
					(0.0125)				
Entrepreneur's Log Parent Income				0.0394***		0.1416**			
				(0.0129)		(0.0034)			
Patent Inventor Age							0.00502***	0.0535***	
							(0.000739)	(0.0050)	
Entrepreneur Age									0.0122***
									(0.0010)
Country	U.S.	Finland	Finland	U.S.	Finland	Finland	U.S.	Finland	Finland
Mean	0.573	0.569	0.6367	1.2259	1.2766	1.1205	47.39	49.11	51.19
N industries	325	342	476	17	253	441	323	342	476
N individuals	2,219,193	9,643	344,698	275	3,812	99,189	48,156	9,643	344,698

Notes: Estimates of Equation 1 at the level of an individual innovator, with outcomes measured at the industry level based on CEX data. Columns (1) and (7) are based on all patents granted by the U.S. Patent Office. Columns (2), (5), and (8) are based on Finnish inventors in the PATSTAT database. Columns (3), (6), and (9) are based on administrative data from Statistics Finland. Column (4) is based on the Panel Survey of Income Dynamics (PSID). Standard errors are clustered at the individual level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

We first analyze gender homophily patterns, finding again that female innovators are more likely to cater to female consumers. Column (1) of Table 4 shows that, within the set of patents granted by the U.S. Patent Office, female patent inventors work in industries where the share of sales to households with a female head is on average 2.2pp higher, or a 3.9% increase relative to the baseline rate. This "between-industry" homophily adds up to the within-industry homophily, which we summarize as a 17.9% increase relative to the baseline rate by taking the average of our results for phone applications and consumer packaged goods. Thus, our estimate of the overall gender homophily coefficient for the United States is 21.8% of the baseline rate.

Figure 2: Innovator-Consumer Gender Homophily across Industries over Time



Notes: The figure reports estimates of gender homophily at the industry level (Equation 1), estimated in each year for granted patents submitted between 1984 and 2014 (left scale). The figure also reports the fraction of female patent inventors over time (right scale). The vertical lines represent 95 percent confidence intervals, with heteroskedasticity-robust standard errors.

Figure 2 shows that gender homophily across industries has been very stable over time, despite the large increase in the share of female patent inventors. The figure reports year-specific homophily coefficients, based on the application date of patents. The homophily coefficient hovers close to 2pp from 1984 to 2014. During this period, the fraction of female patent inventors more than doubled, rising from 5% in 1984 to 13% in 2014. This finding supports the idea that gender homophily would persist if access barriers to the innovation system were significantly reduced for women. Our results stand in contrast to those found in Koning et al. (2021), who find that homophily has weakened over time in life sciences patenting. They note that their results may be driven by the fact that scientific funding has increasingly promoted research into women's health, a feature not present in the broader economy.

Columns (2) and (3) of Table 4 document industry-level gender homophily for Finland. The patterns are similar to the United States, with slightly larger magnitudes. Finnish female patent inventors work in industries where the share of sales to women is 8.5% larger than average (col. (2)). The corresponding increase is 4.7% for Finnish female entrepreneurs (col. (3)).

Next, we document homophily between innovators' parent income and consumers' income. We find that innovators from more affluent backgrounds cater to richer consumers. For our baseline specification, we use an industry's income elasticity as our summary measure of consumer income. Column (4) reports the patterns for entrepreneurs in the United States, showing that an increase in parent income of one log point is associated with an increase of 0.039 in the industry's income elasticity. Column (5) documents similar patterns for patent inventors in Finland, with an increase of 0.0304 in the industry's income elasticity when parent income increases by one log point. Column (6) shows that the relationship is stronger for entrepreneurs in Finland, with an increase of 0.14 in the industry's income elasticity. We find that this large coefficient is driven by entrepreneurs in agriculture; when excluding agriculture, the coefficient falls to 0.0119 (Table C6).

While the income elasticity is a convenient summary measure for depicting the patterns across the income distribution, we also provide result for the share of sales by household income groups (Table C7). We examine whether an entrepreneur's family background systematically varies with the ratio of sales to households earning over $100k ("high-income") to sales to those earning less than $30k ("low-income"). We find that, in the United States, the fraction of sales to high-income households increases by 4.15% relative to the baseline when an entrepreneur comes from the top 20% of the family income distribution, compared with the bottom 20%. Importantly, this number only reflects the "between industry" component of homophily by social background, ignoring any within-industry patterns. This magnitude of "between-industry" homophily by income groups is very similar to gender homophily, as we found a 3.9% increase relative to the baseline rate for gender.[17] Using sales fractions as outcomes in Finland yields similar magnitudes: the fraction of sales to high-income households increases by 3.85% relative to the baseline when an entrepreneur comes from a family in the top income quintile instead of the bottom quintile. Using a threshold of $60k for both low- and high-income households leads to similar conclusions.

[17]If we assume that the relative magnitudes of gender homophily and income homophily are the same within and between industries, then by rescaling our overall gender homophily estimate (= 21.8) by the ratio of between-industry income and gender homophily estimates (= 4.15/3.9), we obtain that the overall income homophily coefficient for the United States is 23.8%.

Finally, columns (7) to (9) of Table 4 show that there is also homophily by age across industries. Older patent inventors are more likely to work in industries selling to older consumers in both the United States (col. (7)) and Finland (col. (8)). In column (9), we report that older entrepreneurs are also more likely to cater to older consumers.

Several robustness results are reported in the Online Appendix. First, in Figure C2 we depict graphically the main regressions from Table 4, showing that the linear specifications provide a good fit to the data and that the estimates are not driven by outliers. Second, for completeness, in Table C8 we run industry-level regressions after averaging worker characteristics by industry; as expected, the coefficients are much larger than in Table 4, where the independent variables vary within industries while the dependent variables do not. Third, we obtain similar results when restricting the sample to patents with U.S.-based inventors only, instead of all patents filed at the United States Patent Office (Table C9 and Figure C3).

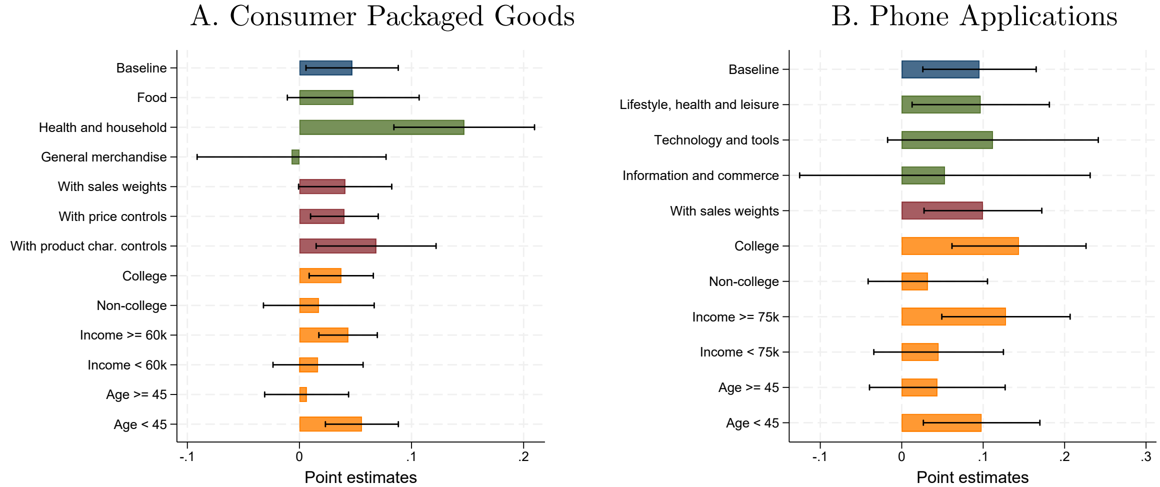
III.D Heterogeneity and Mechanisms

In this section, we provide suggestive evidence on where homophily is strongest and how it might arise. Overall, we find that homophily is a broad-based phenomenon, which is not driven by a small number of products, industries, or consumer types — even though meaningful heterogeneity exists across categories. Figure 3 summarizes the main heterogeneity patterns. Panel A focuses on consumer packaged goods, for which we are able to analyze heterogeneity across the largest number of dimensions – by product categories, product features, and consumer characteristics. Panel B reports the results for phone applications. We discuss these results in the remainder of this section, along with evidence from our other datasets.[18]

Heterogeneity across product categories. We start by reporting innovator-consumer homophily estimates across the product space at a relatively granular level in the U.S. micro data (second to fourth rows in Panels A and B of Figure 3). For consumer packaged goods, we find that homophily is particularly strong for health and household products. Similarly,

[18] Detailed estimates and standard errors are presented in Tables C10 to C22.

Figure 3: Heterogeneity in Innovator-Consumer Gender Homophily



Notes: The figure depicts heterogeneity in the estimated innovator-consumer gender homophily for consumer packaged goods and phone applications, considering in turn heterogeneity across product categories (green bars), by product features (red bars), and by consumer characteristics (orange bars). 95 percent confidence intervals are reported with standard errors clustered at the firm level.

when we segment the product space into three broad categories for phone applications – "lifestyle, health, and leisure", "technology and tools", and "information and commerce" – homophily estimates are positive everywhere and appear larger for the first two groups. These results suggest that homophily operates broadly and that it may be of particular relevance in health- and household-related areas.

Heterogeneity by product characteristics. Next, we run three complementary analyses that characterize heterogeneity by product characteristics, considering in turn the roles of product popularity, vertical differentiation, and horizontal differentiation.

First, we analyze whether homophily is also present for the highest-selling products. Indeed, it could be that only "marginal" products, with lower commercial impact, might be subject to homophily. This is an important source of heterogeneity to uncover to discipline the model in the next section: the welfare impact of homophily would be limited if homophily were relevant only for products with lower sales. To assess this, we repeat our analysis with sales weights for consumer packaged goods and time use weights for phone applications, reflecting the importance of the product. The results are reported in Figure 3 for consumer packaged goods and phone apps (fifth rows in Panels A and B): the estimates are very similar compared to the baseline results without weights.[19] Similarly, using the Finnish data, we show that the estimates remain similar with weighted regressions, giving higher weights to more impactful innovations (using patent counts or proxies for firms' private returns), i.e., the results are not driven by marginal innovators (Table C14).

Second, we analyze whether vertical differentiation could help explain homophily. We use the price as a proxy for vertical differentiation – i.e., "quality" – within detailed product categories (e.g., Moshary et al., 2023). We conduct the analysis in the sample of consumer packaged goods, where we have detailed price information. To start, we show that female-founded teams tend to have significantly higher unit prices (Table C15), about a 24 percent difference between all-female and all-male teams within detailed product categories. Next, we estimate gender homophily while controlling for prices. As reported in the sixth row of Panel A in Figure 3 and detailed in Table C16, controlling non-parametrically for prices (with a fourth-order polynomial) leaves the homophily coefficient almost unaffected; if anything, it increases slightly, suggesting that homophily would be even stronger if unit prices were similar. Thus, homophily is not driven by vertical differentiation.

Third and finally, we examine whether product characteristics (i.e., observable horizontal product differentiation) can account for homophily. We first add three basic controls, which are available for many products: the size of the item, whether it is part of a "multipack", and whether it is an organic product. These controls leave the homophily coefficient almost unchanged (Table C17). Next, we consider all product characteristics available in the Nielsen dataset, selecting the relevant variables using LASSO. The estimate in the seventh row of Panel A in Figure 3 shows that the homophily coefficient remains similar as we add characteristics under a LASSO specification.[20] These results suggest that gender homophily operates through characteristics that are not easily captured by traditional characteristics

[19]Tables C11 and C12 provide detailed estimates. Although we focus on expenditure shares in the main text, as this specification can be linked to welfare (see Section IV), in Table C13 we analyze how the *level* of usage from female users depends on innovator gender. We find that the level of female usage is positively correlated with the female founder indicator in our sample, but the estimates are imprecise.

[20]For the full sample, we draw a random subset of 50% of all firms – given the very large number of characteristics in the Nielsen data, we run into computational issues with a larger sample. When using control variables selected for the 50% sample, results are similar in the full sample, although less precise because the set of control variables is not optimized for this sample. Table C18 shows that the results are similar in the CPG startup sample.

data.

Consumer characteristics. Next, we document whether homophily varies depending on consumer type. To do so, we compute female consumer shares within subgroups of consumers (e.g., families with college-educated heads of household). The results are reported in the last six rows of Panels A and B in Figure 3. For consumer packaged goods, higher education, higher income, and younger consumers appear to exhibit more homophily. However, these differences are difficult to estimate precisely given our limited statistical power. The results for phone applications are similar, with higher gender homophily for more educated, higher income, and younger consumers. Although statistical power is limited in some specifications, the point estimates remain positive in all categories.[21]

Heterogeneity across industries. We use the Finnish data to further characterize the patterns that drive our between-industry homophily estimates. Specifically, we estimate the strength of homophily by broad industry sectors (manufacturing vs. other sectors), and between or within 2-digit industries. Results are presented in Table C21. We observe prevalent homophily patterns by gender and income for patent inventors and entrepreneurs within both broad sectors and finer 2-digit industries, with homophily primarily occurring between rather than within 2-digit industries.

Innovator characteristics. The Finnish data offer rich information on innovator's characteristics, so we proceed to study whether these characteristics relate to the degree of innovator-consumer homophily. Results are presented in Table C22. We do not detect any statistically significant differences in gender or income homophily by patent inventors' education (college vs. non-college), age (below or above 40 years), or family background (high vs. low parental income). In the entrepreneur sample, with a substantially larger number of observations, we find weaker gender homophily among highly-educated entrepreneurs in Finland. While the results suggest some nuanced differences in the degree of gender and income homophily by innovator type, nearly all subgroup-specific homophily coefficients turn

[21] Tables C19 and C20 report all estimates and standard errors.

out to be positive (21 out of 24), indicating the prevalence of gender and income homophily across innovator groups.

Team composition. Given the increased importance of teams for innovation (Jones, 2009; Jaravel et al., 2018), it is instructive to investigate potential non-linear homophily patterns within teams. Figure C4 shows that there is a steady increase in team size for startup founders and patent inventors over time. While there is a long-term downward trend in the prevalence of all-male teams, such teams remain very common by the end of the study period (Figure C5).

We document two facts on team formation and non-linear effects. First, there is significant gender clustering in teams compared to random assignment (Figure C6). Second, we find that it is generally sufficient to have one female founder or inventor in the team to increase the share of sales to female consumers (Table C23 for the United States and Table C24 for Finland).[22] Statistical imprecision makes it challenging to assess the extent to which a larger number of female founders or inventors within the team further impacts homophily.

Social mechanisms independent of market size. In our working paper (Einiö et al., 2023), we provide direct evidence that a mechanism for innovator-consumer homophily is that innovators' social experiences have a causal impact on the direction of innovation, independent of financial incentives. Using a quasi-experimental study-peer design in Finnish education programs, we examine the impacts of the gender and socioeconomic composition of the peer group of a student who later becomes an entrepreneur on the types of consumers they cater to. The results provide direct evidence that social factors can affect the direction of innovation, which stands in contrast with mechanisms based on financial incentives, such as market size.

III.E Extensions

Appendix A reports two extensions: i) homophily between the socio-demographic characteristics of entrepreneurs and their employees; and ii) the relationship between innovators'

[22]For income homophily, the relationship between the parental income composition of the team and the direction of innovation appears to be more linear (Table C24).

backgrounds and the environmental and social impacts of their innovations.

IV Implications for Cost-of-Living Inequality

Guided by the empirical estimates from the previous sections, we now develop a quantitative model to assess the equilibrium implications of the "social push" channel for cost-of-living inequality between men and women, due to unequal access to innovation careers. We also highlight implications for overall growth rates.

IV.A Motivation

Before presenting the full-fledged macro analysis, we motivate the potentially large impact of broadening the pool of talent on cost-of-living inequality by gender with a simple love-of-variety framework (Appendix B.A provides details). Using a simple static framework with CES consumer preferences, we find that the ratio of sales from each innovator type (male vs. female) to each consumer type (male vs. female) is a sufficient statistic for the relative welfare effect of innovations created by male and female innovators, $\Delta \mathrm{W}$, with:

$$\Delta \mathrm{W} = \frac{\lambda^F/(1-\lambda^F)}{\lambda^M/(1-\lambda^M)},$$

where λ^F denotes the share of sales to women in a startup with a female founder, while λ^M denotes this sales share in a startup with a male founder. This sufficient statistic is directly obtained from our homophily regressions in Section III: with the notation from Equation (1), we have $\lambda^F = \alpha + \beta$ and $\lambda^M = \alpha$. The result is intuitive: when agents have CES preferences with similar elasticities of substitution, welfare differences can be reduced to differences in spending shares, and in turn to differences in the firm's revenue shares from each type of innovator.

Using this insight, we find that our homophily estimates from Section III imply that the welfare gains from female-founded startups are 27 percent larger for the representative female household, relative to the representative male household. Considering a scenario where the fraction of female innovators increases from 12% to parity, we find that welfare increases by 10% more for female consumers, compared to male consumers. This back-of-the-envelope calculation highlights the possibility that access barriers combined with innovator-consumer homophily may have a significant impact on cost-of-living inequality between men and women, which we now investigate formally in a quantitative growth model.[23]

IV.B Model and Estimation

Model. We build a model staying as close as possible to workhorse models of endogenous growth (e.g., Romer, 1990) but allowing for unequal access to innovation, heterogeneity in tastes, and differences in the direction of innovation stemming from social push rather than financial incentives. The quantities of interest are the long-run growth rate and steady-state inequality across groups along a balanced growth path. We consider an economy with a unit mass of agents indexed by i belonging to either of two equally-sized groups, men and women, indexed by $g \in \{M, W\}$.[24]

Preferences and consumption. Agents in the economy maximize lifetime discounted utility $\int_0^\infty e^{-\rho t} \log(C_i(t))\, dt$, where the aggregate utility is a composite over the two sectors, $C_i(t) = C_{1i}(t)^{\alpha_{g(i)}} \cdot C_{2i}(t)^{1-\alpha_{g(i)}}$. Preference parameters α_g are specific to each group, allowing for potential cost-of-living inequality in equilibrium. The consumption indices for each sector are determined by the consumption of sectoral varieties, with $C_{ji}(t) = \left(\int_0^{N_j(t)} c_{ji}(\nu, t)^{(\varepsilon-1)/\varepsilon} d\nu\right)^{\varepsilon/(\varepsilon-1)}$, where $N_j(t)$ denotes the number of varieties available in sector j at time t and ε is the elasticity of substitution between varieties. The corresponding price index for each sector is $P_j(t) = N_j(t)^{1/(1-\varepsilon)}$.[25] Because preferences only vary based on group membership g, an agent's price index is given by $P_{it} = P_{1t}^{\alpha_g} \cdot P_{2t}^{1-\alpha_g}$, i.e., the cost of living is lower for consumers with a stronger taste for sectors where product variety is higher.

Agents allocate consumption optimally across sectors to maximize lifetime discounted utility subject to their intertemporal budget constraint, $\int_0^\infty e^{-rt} P_{it} C_i(t)\, dt \leq \int_0^\infty e^{-rt} w_{it}\, dt$, where r denotes the interest rate and w_{it} earnings per period.

Technology and production. Agents supply labor inelastically and decide whether to work in the production of existing varieties or innovate, i.e., create new varieties. All labor

[23] We find larger effects in the full-fledged growth model, because the back-of-the-envelope calculation from the simple love-of-variety framework does not take into account that reducing access barriers may bring into the innovation system some highly productive female innovators, who will have a disproportionate impact on female consumers through homophily.

[24] We focus on gender in Section IV.C, and report a complementary analysis by socio-economic status in Section IV.D.

[25] Without loss of generality, we normalize the price of individual varieties, $p(\nu)$, to be one.

is allocated to production or innovation and the market clears: $\sum_j (L_{jI}(t) + L_{jM}(t)) \leq 1$. $L_{jI}(t)$ denotes the quantity of labor allocated to innovation in sector j, while $L_{jM}(t) = \int_0^{N_j(t)} l_j(\nu, t)d\nu$ gives the quantity of labor allocated to production. $l_j(\nu, t)$ denotes the quantity of labor allocated to the production of existing variety ν in sector j, which is paid at the wage rate $w(t)$. For tractability we assume that the wage rate is the same in all sectors, as when production workers are perfectly mobile.

Agents differ in their innovation productivity η_i, which follows a Pareto distribution with scale parameter $\bar{\eta}$ and shape parameter λ and is identical across groups.[26] The innovation production function, featuring knowledge spillovers, takes the form $\dot{N}_{ji}(t) = \eta_i N(t)$, with $N(t) = \sum_j N_j(t)$.[27] Agents inventing new varieties in sector j receive perpetual patents generating profits $\pi_j(t)$ in each period, with a total value $V_j(t) = \frac{\pi_j(t)}{r}$ based on the balanced growth path interest rate r.

Occupational choice and frictions. Two factors govern each agent's choice of occupation: financial payoffs and frictions. Absent frictions, the financial incentives are standard: the agent is indifferent between innovating in sector j and producing varieties when $w(t) = \eta_i N(t) V_j(t)$. There are many potential frictions that could explain why individuals from certain socio-demographic backgrounds are under-represented in innovation careers: i) social factors, including "exposure" to science and innovation careers, as in Bell et al. (2019a), Breda et al. (2023), Mertz et al. (2024), and Calaway (2025); ii) preferences, including family and fertility decisions, as in Rutigliano (2024); and iii) access to financial resources, as in Brooks et al. (2014).

While identifying the specific mechanism at play is outside of the scope of this paper, our goal is to set up a tractable model that can accommodate a wide range of mechanisms. Indeed, through the lens of a model, these mechanisms can all enter the agent's decision problem as "wedges." The key modeling choice is to determine which part of the agent's decision problem or production function is affected by wedges, as it will determine which

[26]The assumption of a Pareto distribution for productivity is in line with the empirical evidence on the citations and wages of inventors in Bell et al. (2019a).

[27]The use of total varieties when specifying spillovers in the innovation production function is a necessary assumption to avoid explosive growth in a single sector. This assumption is similar to the trade and innovation model of Rivera-Batiz and Romer (1991). Intuitively, this means that innovators' research productivity benefits from all innovations in the economy, not just those in the sector on which they focus.

part of the distribution of talent is missing in the innovation sector – and, in particular, whether we may be missing out on some of the most talented potential innovators. We now introduce and discuss three potential wedges, which affect different parts of the agent's decision problem and production function:

Type 1 Wedge ("Untapped Marie Curies"): a binary wedge τ_{gi} determining whether the agent is able to enter the innovation sector, regardless of ability. Specifically, all agents in group M are able to enter the innovation sector ($\tau_{Mi} = 1$), while those in group W draw a binary variable τ_{Wi} that follows a Bernoulli distribution, $\tau_{Wi} \sim B(\tau)$. Agents who draw $\tau_{gi} = 0$ never pursue innovation, whereas those who receive $\tau_{gi} = 1$ can (and will decide whether or not to pursue such career based on financial incentives). We further assume that τ_{gi} is uncorrelated with agents' abilities to innovate. This can be interpreted as social exposure factors in career choice or a lack of resources precluding agents from pursuing these careers regardless of their underlying productivity (e.g., lacking initial funding or the necessary social connections to start a venture). With this type of wedge, some of the most talented potential female innovators are mechanically screened out of innovation careers – even agents whose productivity η_i is very high will be unable to pursue an innovation career if they have $\tau_{gi} = 0$, raising the potential for "Lost Marie Curies."

Type 2 Wedge ("Untapped Marginal Female Inventors"): a wedge $\tilde{\tau}_{gi}$ that affects the agent's utility when entering the innovation sectors but not their innovation productivity. Formally, the agent compares the returns to innovating in sector j, $(1 - \tilde{\tau}_{gi})\eta_i N(t)V_j(t)$,[28] to compensation when producing varieties, $w(t)$; but the rate of successful innovation remains unaffected, i.e., the innovation production function remains $\dot{N}_{ji}(t) = \eta_i N(t)$. This formulation covers various mechanisms, including a "preference" against innovation careers (including fertility and family factors), certain forms of discrimination (e.g., discrimination during studies that affect utility but not productivity, and discrimination on the job market that affect compensation but not productivity), or lack of access to resources (affecting the decision to enter an innovation career but not overall productivity, e.g., because liquidity constraints can be overcome through sufficient effort). In this case, agents whose produc-

[28] The wedge $\tilde{\tau}_{gi}$ is thus a reduced-form way of capturing factors that affect the financial returns or the utility of being an innovator. Agents who belong to the same group g all receive the same value of $\tilde{\tau}_{gi}$. In the calibration below, we set $\tilde{\tau}_{gi} > 0$ for women and $\tilde{\tau}_{gi} = 0$ for men.

tivity η_i is very high will always pursue an innovation career because, even with the wedge $\tilde{\tau}_{gi}$, their utility in the innovation sector is significantly higher than the outside option. The wedge $\tilde{\tau}_{gi}$ only affects the decision of "marginal" innovators, who were close to the cutoff $w(t)$. Furthermore, productivity in research is unaffected. As a result, there are no "Lost Marie Curies".

Type 3 Wedge ("Stifled Marie Curies"): a wedge $\grave{\tau}_{gi}$ that affects both the decision to enter the innovation sector and productivity conditional on innovation. This is identical to the previous case except that the wedge also affects innovation productivity: the agent's return in the innovation sector is $(1-\grave{\tau}_{gi})\eta_i N(t) V_j(t)$ and the innovation production function is $\dot{N}_{ji}(t) = (1-\grave{\tau}_{gi})\eta_i N(t)$. This formulation covers mechanisms that simultaneously impact utility in the innovation sectors and research productivity, which can cover certain group-specific shocks (e.g., family and fertility events may have a disproportionate impact on female innovators because of social norms implying unequal task sharing within the household, which could affect both preferences for innovation careers and productivity)[29] and forms of discrimination with both a utility and productivity costs (e.g., biased teaching during studies, discrimination in the labor market), or lack of access to certain resources (e.g., liquidity constraints reducing both expected compensation and output). As in the previous case, agents whose productivity η_i is very high will always pursue an innovation career; however, their productivity is now reduced. Therefore there are no "Lost Marie Curies" *per se*, as the highest ability innovators do enter the innovation sector (unlike case 1), but their productivity is reduced, so society does lose out on the output of some of the most talented innovators (like with type 1 wedges and unlike with type 2 wedges).

Homophily. Motivated by the evidence on innovator-consumer homophily presented earlier in this article, we specify that agent $i \in g$ is assigned with probability $\phi \in (0,1)$ to the sector for which group g has a stronger relative taste preference, as governed by α_g in the agent's utility function, and with probability $1-\phi$ to the other sector.[30] This sectoral

[29]The wedge $\grave{\tau}_{gi}$ scales down the innovation productivity parameter η_i in the innovation production function, which is a reduced-form way of capturing several potential underlying mechanisms – e.g., because of unequal sharing of family duties, women may not be able to work long hours, or may be less productive per hour worked as their surrounding may make it harder to concentrate, etc.

[30]This structure greatly simplifies the solution algorithm but does not drive our quantitative results; similar results can be obtained in a structure where innovators choose which sector to enter and innovators have a large enough comparative advantage in one sector.

assignment mechanism generates "social push" towards a sector to which an individual is most exposed to and governs the strength of innovator-consumer homophily. It incorporates any factor driving homophily. Without loss of generality, we assume that men have a stronger relative preference for sector 1. Agents can decide whether to innovate only in the specific sector they were assigned to. They choose whether to innovate in this sector or to produce existing varieties by maximizing expected lifetime utility, comparing the returns to innovation in the sector they are exposed to, $V_j(t)$, with production earnings, w_{it}, subject to the wedges described previously.

Inequality. We introduce a parameter, δ, to model the gender pay gap. Specifically, while $w(t)$ is the average wage in the economy, we assume that women earn $w_{Wt} = \delta \cdot w(t)$, while men earn $w_{Mt} = (1 - \delta) \cdot w(t)$. This parameter allows us to match the observed median earnings ratio across genders.[31] Consumption inequality is given by:

$$\underbrace{\frac{C_{Mt}}{C_{Wt}}}_{\text{gender consumption gap}} = \underbrace{\frac{w_{Mt}}{w_{Wt}}}_{\text{gender pay gap}} \times \underbrace{\frac{P_{Wt}}{P_{Mt}}}_{\text{gender cost-of-living gap}}. \tag{2}$$

The gender wage gap is given by the ratio $(1-\delta)/\delta$.[32] Using the expression for sectoral price indices and consumer preferences, we obtain that the cost-of-living gender gap is $\frac{P_{Wt}}{P_{Mt}} = \left(\frac{N_1(t)}{N_2(t)}\right)^{\frac{\alpha_M - \alpha_W}{\varepsilon - 1}}$.

Equilibrium and counterfactuals. We solve for the steady-state equilibrium along a balanced growth path, where consumption growth, wage growth, the interest rate, and the growth rate of varieties in both sectors are constant, along with the amount of labor allocated to each sector, and to research and production within each sector. The economy is closed, i.e., $y(v, t) = c(v, t)$, and the equilibrium interest rate is consistent with the Euler equation, the financial incentives governing the optimal career choice between production work and innovation, and zero net savings in equilibrium. All first-order conditions and constraints characterizing the equilibrium are provided in Appendix B.B.1, along with our numerical solution algorithm. In particular, we solve for the research productivity cutoffs in

[31] We assume the parameter δ scales in the same way the returns to innovation $V_j(t)$ for each group g, so that it leaves occupational choice unaffected.

[32] For tractability, in the baseline model we compute the gender pay gap ignoring entrepreneurial earnings. We include entrepreneurial earnings in an extension presented in Appendix B.B.5: when relaxing wedges, the counterfactuals for growth and cost-of-living inequality are similar to our baseline model, while the impacts on earnings inequality are larger.

each sector j, above which agents exposed to innovation in that sector decide to pursue an innovation career.

In equilibrium, the allocation of innovators across sectors is governed by two forces, leading to cost-of-living gender inequality. First, profit incentives must be equalized across sectors, such that there are more innovators in larger sectors. Due to the nominal gender wage gap, the sector preferred by men is larger, which attracts more innovators. Second, women are less likely to work as innovators, which has a disproportionate impact on innovation in the sector preferred by women due to innovator-consumer homophily. In equilibrium, the productivity cutoff in the female-preferred sector is lower, in order to create enough new goods to satisfy equilibrium conditions. Despite the lower cutoff, there are fewer new goods created in this sector and, therefore, a higher effective price. This creates the cost-of-living gap in Equation 2.

Finally, one feature is worth highlighting about the counterfactual equilibrium we study, relaxing wedges. A key offsetting effect in general equilibrium operates through interest rates. When no individual faces barriers, there is additional entry into innovation careers. The resulting increase in the growth rate of varieties requires an increase in interest rates in equilibrium so that the Euler equation holds, which reduces the value of innovation as discounted profits are now reduced; this mechanism dampens the general equilibrium impact of reducing entry barriers in the innovation sector.[33]

Estimation. Table 5 summarizes our baseline parameters for the analysis by gender.[34] Using values from the literature, in Panel A we directly set five parameters of the model: the expenditure dissimilarity index by gender $|\alpha_M - \alpha_W|$, the elasticity of substitution between varieties ε, the discount rate ρ, the Pareto shape parameter λ for the innovators' productivity distribution, and the male to female earnings ratio $\frac{1-\delta}{\delta}$. In Panel B, we report the last three parameters of the model, which jointly target the three moments reported in Panel C: the

[33]We follow endogenous growth models (e.g., Romer, 1990; Aghion and Howitt, 1992), and allow long-run growth rates to vary in our counterfactuals. With a semi-endogenous growth model (Jones, 1995), there would be no impact on long-run growth, but we conjecture that: (i) the effects on steady-state cost-of-living inequality would be similar; (ii) the impacts on growth rates would be similar over the transition to the new steady state, which could take several decades given that we are far from parity among innovators.

[34]This table uses the version of the model with Type 1 wedges. We discuss in the next subsection why this is our baseline specification.

Table 5: Baseline Parameters

Panel A: Parameters calibrated outside of the model			
Model Parameter	Parameter Definition	Source	Value
$\lvert \alpha_M - \alpha_W \rvert$	Expenditure dissimilarity index by gender	Consumer Expenditure Survey, Nielsen data, phone applications data (cf. Table C25, first row)	0.24
ε	Elasticity of substitution between varieties	DellaVigna and Gentzkow (2019)	1.9
ρ	Discount rate, annual	Kaplan et al. (2018)	0.051
λ	Pareto shape parameter of innovators' productivity	Bell et al. (2019a)	1.26
$\frac{1-\delta}{\delta}$	Male to female earnings ratio	U.S. Department of Labor	1.205

Panel B: Jointly estimated model parameters		
Model Parameter	Parameter Definition	Value
τ	Barrier to enter innovation careers	0.105
ϕ	Sectoral assignment	0.732
$\bar{\eta}$	Pareto scale parameter of of innovators' productivity	0.011

Panel C: Targeted moments and model fit		
Targeted Moment [Source]	Data	Model
Share of female patent inventors [Toole et al. (2019)]	0.128	0.128
Gender homophily regression coefficients [2nd col. of Table 2, 1st col. of Tables 3 and 4]	0.218	0.218
Annual growth rate of labor productivity, 1990-2020 [Saint Louis Fed]	0.02	0.02

Notes: This table presents the baseline parameters of the growth model, for the analysis by gender. Panel A shows the model parameters which are directly set to the value observed in data or taken from the literature. In Panel B, the three parameters are estimated jointly to match the moments from the model with moments observed in the data, displayed in Panel C. The gender homophily coefficient is computed by taking an average of the within-industry estimates from phone apps and consumer packaged goods (normalized by baseline rates) and adding it to the across-industry estimate (again normalizing by the baseline rate).

observed fraction of female innovators, the empirical innovator-consumer gender homophily documented in Section III, and the average growth rate of labor productivity. The model matches these three moments exactly. Next, we use these parameters to analyze several counterfactuals and document the sensitivity of the results to alternative parameters.

Adjudicating between Wedges. Before turning to counterfactuals, we provide a brief summary of how our model's predictions under different wedges compare to empirical evidence within our analysis samples and in the literature. Our overall takeaway is that Type 1 wedges best capture the body of empirical evidence. Appendix B.B.3 provides a more detailed discussion.

First, we note that the model with Type 2 wedges makes stark predictions that run counter to empirical patterns in our sample. Under Type 2 wedges, women should be much more likely to be top innovators conditional on pursuing an innovation career, whereas under the other two wedge types, the model predicts slightly lower probabilities. Within our analysis samples (phone apps, consumer packaged goods, citation-weighted patents), we consistently find that female innovators are slightly less likely to be in the top 5 and

10 percent of all innovators in terms of usage, revenue, and patenting output. Table C26 presents both model predictions and empirical estimates.

Second, we note that evidence in the literature is more consistent with Type 1 wedges than Type 3. A model with Type 1 wedges predicts that the distribution of *latent* innovator productivity should be similar for men and women, whereas a model with Type 3 wedges predicts that women who become innovators should come overwhelmingly from the top of the latent innovation productivity distribution. Although the latent innovation productivity distribution is unobserved, a simple measure is math test scores in childhood.[35] Bell et al. (2019a) and Calaway (2025) both find that women are much less likely to become inventors or pursue STEM careers regardless of their math test scores; Bell et al. (2019a) report similar patterns for low SES individuals. These patterns are inconsistent with models with Type 3 wedges – where the female and low SES inventors with the highest latent productivity should always enter innovation, and do less well than men and high SES inventors *ex-post* due to productivity barriers.

Based on these patterns, we proceed with Type 1 wedges for the baseline analysis in the next subsection. We present the results with Type 2 and Type 3 wedges as extensions.

IV.C Counterfactuals by Gender

Main results. Figure 4 present our main results on the effects of the under-representation of women on cost-of-living inequality and economic growth. Panel A of Figure 4 shows that the cost-of-living gender gap is 18.7% in our baseline model, which is close to the nominal pay gap between men and women of 20.5%. Therefore, per Equation (2), the real consumption of men is 43% larger compared to women; without the cost-of-living gender gap, the consumption gender gap would only be half as large.

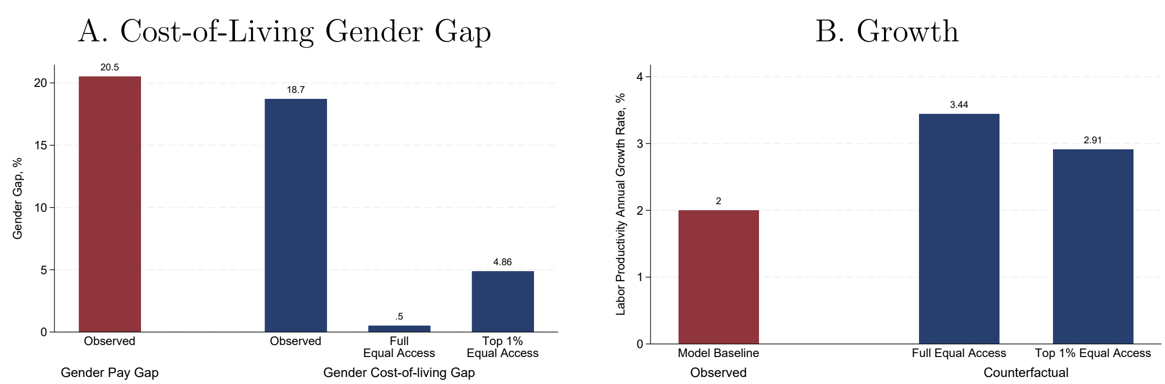
The large cost-of-living gender gap uncovered by our model is consistent with reduced-form evidence about product variety across gender groups. In the consumer packaged goods context, we find that goods created by female entrepreneurs are 24.3% higher in unit prices

[35] Several paper have shown that high mathematical ability in childhood strongly correlates with knowledge production in adulthood (e.g., Aghion et al., 2023; Akcigit et al., 2017; Bell et al., 2019a; Agarwal and Gaulé, 2020).

relative to goods in the same product category created by male entrepreneurs.[36] Measuring time use by gender for phone applications, we find that the Herfindahl–Hirschman Index (HHI) is 60% larger for women. Specifically, the HHI is 0.034 for women and 0.021 for men, consistent with there being a much larger variety of phone applications catering to the tastes of men.

We start by studying a counterfactual equilibrium in which women are not under-represented among innovators. Specifically, we first examine an "equal access" counterfactual, in which women no longer face wedges to enter innovation careers, i.e., we set $\tau_i = 1$ for all women. Panel A shows that, in this counterfactual scenario, the cost-of-living gender gap falls to 0.50%. As reported in the figure, we find that the fall is also large when considering an alternative "equal access" policy targeting only women in the top 1% of the innovator productivity distribution, i.e., we set $\tau_i = 1$ for all women in this group. In this case, the cost-of-living gender gap falls to 4.86%. Intuitively, given the skewness of the distribution of innovation abilities, it is sufficient to attract the most talented individuals to generate most of the gains.

Figure 4: Main Counterfactual Estimates



Notes: This figure reports counterfactuals in the model with Type 1 wedges, bringing the wedges to zero. Panel A focuses on cost-of-living gender gaps in different scenarios. The baseline specification uses the model parameters summarized in Table 5. The "full equal access" counterfactual sets $\tau_i = 1$ for all women. The "top 1% equal access" scenario sets $\tau_i = 1$ for all women in the top 1% of the innovation productivity distribution. The observed gender gap is shown for comparison. Panel B reports labor productivity growth at the observed equilibrium and in the two counterfactual scenarios, with full equal access or top 1% equal access.

[36]Table C15 reports the results. Some of the difference could be due to product quality, as documented by Moshary et al. (2023) in a subsample of Nielsen CPG products.

Even with equal access of women to innovation careers, the cost-of-living gender gap does not completely disappear because of market size effects. Due to the gender pay gap, it is more profitable for innovators to target men instead of women. However, Panel A shows that this effect is relatively small quantitatively. The cost-of-living gender gap falls from 0.50% in the "full equal access" counterfactual with market size effects to 0 in the scenario eliminating differences in market size (i.e., $\delta = 0.5$).

Thus, in a model with social push we find that access frictions for female innovators explain 97% (= 18.2/18.7) of the cost-of-living gender gap, and the implication for consumption inequality by gender are just as large as the gender pay gap, around 20% of consumption.[37] In contrast, the market size channel — which has been the focus of the literature on the direction of innovation — plays a limited role, representing about 0.5% of consumption.[38] This is the first key takeaway of our quantitative analysis.[39]

Panel B of Figure 4 turns to the growth of labor productivity across counterfactual scenarios. At baseline, our estimated model matches the 2 percent observed growth rate of labor productivity in the United States. The panel shows that growth increases to 3.44% per year under the "full equal access" counterfactual, in which women are fully exposed to innovation careers and account for 50% of innovators. Such a large increase in growth rates (+72% relative to baseline) comes from the facts that our counterfactual relaxes barriers to innovation for half of the population.[40] In this counterfactual, the fraction of individuals who

[37]Without cost-of-living inequality, consumption inequality between men and women is $\frac{C_{Mt}}{C_{Wt}} = \frac{w_{Mt}}{w_{Wt}} = 1.205$. With cost-of-living inequality stemming from innovator-consumer homophily, our estimates yield $\frac{C_{Mt}}{C_{Wt}} = 1.205 \times 1.182 = 1.424$.

[38]The finding that market size effects have a relatively small impact on cost-of-living inequality by gender is intuitive and can be seen in a back-of-the-envelope calculation, stepping back from our model. Reduced-form work has documented that a 1% increase in market size leads to a fall in the price index ranging between 0.1% and 0.3% (Jaravel, 2019; Costinot et al., 2019; Faber and Fally, 2022). Due to the pay gap between men and women, market size is 17.1% smaller for women (= 1/1.205). Because there is a 76% overlap in the expenditure patterns of men and women (cf. Table 5, first row), a back-of the-envelop calculation suggests market size effects lead to a modest increase in the price index for women ranging between 1.2% $(= (-17.1\%) \times 0.24 \times (-0.3))$ and 0.41% $(= (-17.1\%) \times 0.24 \times (-0.1))$.

[39]The relative magnitude of the social push vs. market size channels is driven by the productivity of new entrants (Table C27). When access barriers fall, highly productive women enter innovation careers, especially the sector catering to female customer because of homophily. These highly productive female innovators can have large impacts on price indices in the new long-run equilibrium. In contrast, when market size changes, the flows only involve innovators who are close to the productivity cutoff (in either sector), with smaller impacts on price indices.

[40]We further discuss the plausibility of the growth impacts in Section IV.E.

are now able to enter the innovation sector increases by 79% relative to baseline; however, the equilibrium number of innovators increases by 37% only (looking at entry net of exit), and the average productivity of innovators increases by 23%, as less productive ones are displaced.[41] This large growth impact is the second key takeaway of our quantitative analysis.

Panel B also reports the results when access barriers are removed only for the top 1% of women: the growth rate increases to 2.91 percent, a 46% increase relative to the baseline. In this case, the equilibrium number of innovators falls by 9% and the average productivity of inventors increases by 66%, as high productivity female innovators displace less productive ones.

Thus, our quantitative model shows that policies expanding the pool of talent can have a very large macroeconomic impact, even when they focus on a relatively small number of top-talent individuals. This macroeconomic quantification complements microeconomic evidence suggesting that simple policies providing role models could help attract women into innovation careers in practice (e.g., Del Carpio and Guadalupe, 2022; Breda et al., 2023; Calaway, 2025). Our results suggest that such policies may be a powerful approach both to increase growth rates and reduce consumption inequality between men and women.

Robustness and extensions. Next, we analyze the extent to which our quantitative counterfactual results are sensitive to parameter choices. The results are reported in Table 6. Row A summarizes the results from our baseline specification, which are identical to those in Panel A of Figure 4.

Focusing on the "equal access" counterfactual, we assess in turn the sensitivity of our results to changes in the elasticity of substitution between varieties, ε, and to alternative target moments for innovator-consumer gender homophily, β.[42] Rows B to E show that the magnitudes remain large across all specifications, demonstrating the robustness of our results. Rows F to H illustrate the key role of the skewness of the ability distribution, parametrized by λ. The effects become smaller with less skewed ability distributions (with a lower λ), because female innovators displace male innovators who are closer in ability.

[41] For completeness, Table C28 reports the number of innovators and their productivity at baseline and in the counterfactual equilibrium.

[42] For ε we use the range from DellaVigna and Gentzkow (2019); for β we take the range from Tables 2, 3, and 4.

Table 6: Alternative Parameters and Scenarios

Specification	Cost-of-Living Gender Gap			Change in Labor
	Before Policy Shock	After	Change	Productivity Growth
	(1)	(2)	(3)	(4)
A. Main: Type 1 wedge model, $\varepsilon = 1.9$, $\beta = 0.218$, $\lvert\alpha_M - \alpha_F\rvert = 0.24$	18.70%	0.50%	-18.20pp	+1.44pp
B. Lower epsilon, $\varepsilon = 1.34$	57.50%	1.33%	-56.17pp	+1.44pp
C. Higher epsilon, $\varepsilon = 2.5$	10.86%	0.30%	-10.56pp	+1.44pp
D. Lower innovator-consumer homophily, $\beta = 0.169$	12.25%	0.36%	-11.89pp	+1.37pp
E. Higher innovator-consumer homophily, $\beta = 0.24$	22.58%	0.44%	-22.14pp	+1.50pp
F. Less skewed ability distribution, $\lambda = 1.5$	15.51%	0.80%	-14.71pp	+1.33pp
G. Less skewed ability distribution, $\lambda = 2$	11.73%	0.61%	-11.12pp	+1.10pp
H. Less skewed ability distribution, $\lambda = 3$	8.10%	0.80%	-7.30pp	+0.80pp
I. Type 3 Wedge ("Stifled Marie Curies")	18.44%	0.41%	-18.03pp	+1.45pp
J. Type 2 Wedge ("Untapped Marginal Female Inventors")	4.44%	0.57%	-3.87pp	+0.38pp
K. Equal access for top 0.1%	18.70%	9.17%	-9.53pp	+0.53pp
L. Equal access for top 0.5%	18.70%	6.27%	-12.43pp	+0.78pp
M. Equal access for top 1%	18.70%	4.86%	-13.84pp	+0.91pp

Notes: This table presents the results of the models for alternative specifications, considering alternative values for certain parameters of the model, or alternative counterfactual scenarios. Unless otherwise noted, the parameters are identical to the baseline specification in the first row. Unless specified otherwise, we study the "full equal access" counterfactual, setting $\tau_i = 1$ for all women.

We also present results obtained with the alternative wedge types. With Type 3 wedges, the wedge reducing entrepreneurial income for women also applies to their productivity for innovation. This specification thus reduces the aggregate innovation productivity in the economy, and row I shows that the counterfactual with no barriers for women yields large impacts for inequality (-18.03pp) and growth (+1.45pp). These magnitudes are close to our baseline model with Type 1 wedges. We also consider a specification with Type 2 wedges, where the wedge only reduces entrepreneurial income for women but leaves their innovation productivity unaffected (in the spirit of Hsieh et al. (2019)), i.e., they get lower private returns but generate as impactful innovations for society. With this specification, row J shows that relaxing frictions yields smaller effects for inequality and growth.

Returning to the model with Type 1 wedges, rows K to M in Table 6 also report the results with targeted counterfactuals, eliminating access barriers for women in the top 0.1%

and 0.5% of the innovation ability distribution. The effects are smaller than in the "equal access" or "top 1% equal access" counterfactuals, but remain sizable.

IV.D Counterfactuals by Income Groups

We now repeat the analysis by considering low- and high-income groups, defined as individuals in the top and bottom income quintiles. The model is unchanged, except that g now indexes these two income groups. The estimation strategy is similar to our approach for the model by gender; the parameters are reported and discussed in Table C29.

Table 7 reports our main result. We find that cost-of-living inequality between the top and bottom income quintiles is 14.25% at baseline. Row A shows that, in a counterfactual imposing equal access to innovation careers for the agents from the bottom income quintile, cost-of-living inequality falls by 9.74pp to 4.51%. Thus, innovator-consumer homophily by socioeconomic status explains 68% of cost-of-living inequality across the distribution in our model. In contrast with the model focusing on gender, sizable cost-of-living inequality remains in the counterfactual scenario, because of market size effects. Indeed, the difference in market size between the top and bottom income quintiles is much larger than between genders, and leads to a larger set of varieties preferred by the rich in equilibrium.[43]

Thus, equalizing access to innovation across the income distribution would lead to a 11.3% increase in purchasing power for the bottom income quintile, i.e., about \$4,400 per year for each household in this group. This represents a transfer of about \$135 billion,[44] which comes close to the budget of the Supplemental Nutrition Assistance Program (\$182.5 billion) and represents about 2% of total U.S. federal spending in 2021 (\$6.4 trillion). The macroeconomic relevance of the channel we study is also shown by considering growth rates, which increase from 2% a year to 3.34%, a 67% increase. This finding highlights again the large macroeconomic potential of policies broadening access to innovation careers.

For completeness, rows B to K of Table 7 document the sensitivity of the equal access

[43] Following the same steps as in footnote 38, a back-of the-envelope calculation based on reduced-form estimates suggests market size effects alone lead to an increase in the price index for the bottom income quintile ranging between 8.6% ($= (-84.5\%) \times 0.34 \times (-0.3)$) and 2.9% ($= (-84.5\%) \times 0.34 \times (-0.1)$). The result from the model, 4.51%, is in the middle of this range.

[44] For the 30.75 million households in the bottom income quintile, we apply the 11.3% increase in purchasing power to their average income of \$39,000 per household.

Table 7: Counterfactuals for Top and Bottom Income Quintiles

Specification	Cost-of-Living Inequality			Change in Labor
	Before Policy Shock	After	Change	Productivity Growth
	(1)	(2)	(3)	(4)
A. Main: full access counterfactual policy, $\varepsilon = 1.9$, $\beta = 0.238$, $\lvert\alpha_H - \alpha_L\rvert = 0.34$	14.25%	4.51%	-9.74pp	+1.34pp
B. Lower homophily, $\beta = 0.184$	11.75%	4.30%	-7.45pp	+1.34pp
C. Industry-level homophily, $\beta = 0.0415$	5.93%	4.48%	-1.45pp	+1.35pp
D Lower epsilon, $\varepsilon = 1.34$	42.27%	12.38%	-29.89pp	+1.34pp
E. Higher epsilon, $\varepsilon = 2.5$	8.32%	2.68%	-5.64pp	+1.34pp
F. Less skewed ability distribution, $\lambda = 1.5$	15.11%	6.73%	-8.38pp	+1.21pp
G. Less skewed ability distribution, $\lambda = 2$	16.57%	10.10%	-6.47pp	+0.98pp
H. Less skewed ability distribution, $\lambda = 3$	18.08%	13.67%	-4.41pp	+0.70pp
I. Equal access for top 0.1%	14.25%	8.81%	-5.44pp	+0.50pp
J. Equal access for top 0.5%	14.25%	6.86%	-7.39pp	+0.74pp
K. Equal access for top 1%	14.25%	5.92%	-8.33pp	+0.86pp

Notes: This table presents the results of the models when studying high- and low-income groups, defined as individuals in the top and bottom income quintiles, considering alternative values for certain parameters of the model. Unless otherwise noted, the parameters are identical to the baseline specification in the first row. In all rows, we study the "full access" counterfactual in the model with Type 1 wedges, setting $\tau_i = 1$ for all agents from the low-income group.

counterfactuals to alternative parameter values, with effects similar to the robustness analysis by gender.

IV.E Discussion of the Model's Predicted Growth Effects

The effects documented in the previous section are large, especially for growth. To assess their plausibility, we conduct complementary analyses.

First, we present complementary quantitative results, studying in our model the growth impact of the observed increase in the representation of women in innovation careers over the last several decades. The share of women among patent inventors increased from 5% in 1985 to to 12.8% in 2014. We use our model to estimate the impact of increased female access on growth, backing out in the model the required fall in the Type 1 wedge to produce the observed increase in the share of female inventors. The results are reported in Table C30: we find an increase in annual growth of 0.12pp only. The predictions of our model are therefore not inconsistent with the path of growth in the last decades.[45]

This modest increase may seem surprising compared to the much larger effect we document in our main counterfactual, with an increase in annual growth of 1.44 percentage points. The difference stems from two reasons. First, our main counterfactual constitutes a much bigger shock to women's representation among innovation, going from 12.8% to 50% – a 37 percentage point increase, about five times larger than the 7.8 percentage point increase observed from 1985 to 2014. Second, the effect of reducing wedges on growth is non-linear and depends on the starting point in terms of female representation: the more we reduce access barriers, the stronger the competition becomes among innovators, i.e., we only retain the most talented innovators and less productive innovators exit in equilibrium.[46]

Next, as additional tests of the plausibility of the quantitative magnitudes in our model, we compare them to correlational evidence from recent decades and to the results from two historical quasi-experiments. The ideal experiment to test the predictions of our model would be an exogenous shift in the supply of female innovators at the country level. Absent such variation in the data, we conduct indirect and imperfect tests, with the goal of assessing whether the order of magnitude found in our model is plausible.

First, we conduct an analysis across technology classes. We assess whether and how much patenting increases in technology classes where female inventors became better represented over time, using USPTO data from 1983 to 2014. We run Poisson regressions to relate the change in the number of patents and total stock market responses (Kogan et al., 2017) at the class level (882 U.S. patent classes) to the fraction of women in a technology class. With class fixed effects, the model is identified in changes. Although the baseline specification

[45] The increase in growth of 12 basis points predicted by our model is small relative to the observed fall in labor productivity growth of about 1.5pp in the last decades (U.S. Bureau of Labor Statistics), which is driven by other factors in the context of secular stagnation (e.g., Bloom et al., 2020). We also note that the increased supply of female inventors over the past decades may have been partly offset by a fall in the supply of inventors from low-income backgrounds, given recent evidence of growing gaps by social class in economic opportunities (Chetty et al., 2024).

[46] Figure C7 show this result, plotting average innovator productivity, growth, and cost-of-living inequality against the fraction of female inventors in equilibrium, as we reduce the entry wedge. The growth and productivity graphs are convex: the marginal impact of increased female representation on productivity and growth become larger as female representation increases. Intuitively, for low representation of female innovators (such as the observed equilibrium, at 12%), the female innovators entering the innovation system for a small change in access barriers may not be particularly productive; as we get close to 50% female representation, the new entrants are much more productive (because they are selected against a higher productivity cutoff), with a larger impact on growth.

only has class and year fixed effects, we also report the results with detailed technology subcategory-year fixed effects (37 subcategories), so that the estimate is only identified from residual variation in relatively homogeneous parts of the technology space.

Table 8 reports the results. Columns (1) and (2) show that total patent production increases in technology classes where the fraction of female inventors increases. To assess magnitudes, we use the point estimate to predict the increase in patents from a 8pp change in the fraction of female inventors (as from 1984 to 2014) or from a 37pp change (as in our main "equal access" counterfactual). The empirical estimates are similar to those predicted by the model. With detailed technology subcategory by year fixed effects, the increase in patenting is 69% for a 37pp increase in female inventor fraction, which is in line with the model prediction (+72%).

Columns (3) and (4) present complementary results, using the measure of Kogan et al. (2017) to value the economic impact of patents by the change in the stock market valuation of the assignee around patent grants. We also find results of the same order of magnitude as the change in growth rates predicted by our model.

As a second step, we use evidence from a historical quasi-experiment studied by Moser et al. (2014). They study an important immigration shock in the 1930s, when high-performing scientists fleeing Nazi Germany arrived in the United States. They document that the arrival of these prominent emigres in the 1930s had a large impact on subsequent patenting and entry in the emigres' fields over the next forty years. They estimate large impacts on patents which are entirely explained by entry of new inventors to the field, with no change to the productivity of incumbent inventors. This historical quasi-experiment can thus serve as an "entry shock" and we can compare the magnitudes they estimate to the predictions of our model, again to assess whether orders of magnitudes are similar. While this shock does not specifically affect female inventors, it can be useful to assess the relationship between inventor entry and the overall innovation response.

Moser et al. (2014) document that there is a 61% (gross) increase in the number of inventors in emigres' fields, and a 71% increase in patents. In our "equal access" counterfactual, there is a 65% (gross) increase in the number of inventors, and a 72% increase in the growth rate of productivity. Thus, the historical quasi-experiment of Moser et al. (2014) parallels

Table 8: Stylized Fact: The Relationship b/w Patenting and Female Inventor Fraction, 1984-2014

	Patents		Firm's Stock Market Response	
	(1)	(2)	(3)	(4)
Female Inventor Fraction	2.648***	1.873***	1.453***	1.060**
	(0.479)	(0.377)	(0.464)	(0.442)
Predicted innovation increase from:				
8pp increase in Female inventor fraction (model prediction: +6.4%)	+21%	+15%	+12%	+8%
37pp increase in Female inventor fraction (model prediction: +72%)	+98%	+69%	+54%	+39%
Technology Class Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓		✓	
Technology Subcategory by Year Fixed Effects		✓		✓
N	14,662	11,997	10,504	9,428

Notes: This table estimates the relationship between patenting in a technology class and the fraction of female inventors in that class. The level of observation is a technology class in each year, in a sample running from 1984 to 2014. Specifications include either technology class fixed effects and year fixed effects alone, or also include technology subcategory by year fixed effects. The sample size is slightly reduced for the latter specification, as technology subcategories (Hall et al., 2001) are missing for a few technology classes. While columns (1) and (2) use the count of patents in a year as outcome, columns (3) and (4) sum up the stock market reactions to patent grants (following Kogan et al. (2017)), as a proxy for the economic impact of patents. All columns use a Poisson estimator. To compare magnitudes, the table also reports the predictions of the model using Type 1 wedge (simply scaling the point estimates by 0.08 in row 3 or 0.37 in row 4).

our setting well, suggesting the magnitudes of our model's prediction are not implausible. Of course, this comparison is not a direct test of the model – our model makes a general equilibrium prediction, whereas Moser et al. (2014) estimate differences across fields. But their evidence is consistent with the idea that large changes in inventor entry can lead to large changes in overall innovation rates, and hence productivity growth.

It is also instructive to mention historical evidence from a recent paper by Gozen (2024). This paper analyzes the impacts of strengthened property rights for women on patenting, leveraging a staggered difference-in-differences across U.S. states. The paper documents a 40% increase in patenting activity by women, including for the most novel patents, and documents no crowding-out effects on patenting by men. These results are broadly consistent with our model, with a large impact on innovation dynamics from better inclusion of female inventors.

V Conclusion

In this paper, we have established two complementary results to characterize how the "social push" channel determines the direction of innovation. First, we documented a widespread pattern of innovator-consumer homophily. Second, we developed a quantitative growth model that incorporates innovator-consumer homophily to assess the macroeconomic relevance of broadening access to innovation, documenting meaningful effects both for the overall growth rate and for cost-of-living inequality.

These findings highlight the importance of policies and initiatives aimed at promoting access to entrepreneurship for women and individuals from disadvantaged socio-economic backgrounds. Such policies have the potential to lead to a more diverse set of new goods and services, and to yield a double dividend by simultaneously increasing growth and reducing 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.

— **and Joshua Linn**, "Market size in innovation: theory and evidence from the pharmaceutical industry," *The Quarterly journal of economics*, 2004, *119* (3), 1049–1090.

— **, 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 and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

— **, 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**, "Parental education and invention: The Finnish enigma," *International Economic Review*, 2023.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas, "The rise of american ingenuity: Innovation and inventors of the golden age," *National Bureau of Economic Research Working Paper*, 2017.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Do tax cuts produce more Einsteins? The impacts of financial incentives versus exposure to innovation on the supply of inventors," *Journal of the European Economic Association*, 2019.

_ , _ , _ , _ , and _ , "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 2019, *134* (2), 647–713.

Bertrand, Marianne, "Gender in the twenty-first century," in "AEA Papers and proceedings," Vol. 110 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2020, pp. 1–24.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb, "Are ideas getting harder to find?," *American Economic Review*, 2020, *110* (4), 1104–1144.

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," *Working paper*, 2018.

Breda, Thomas, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre, "How Effective are Female Role Models in Steering Girls Towards Stem? Evidence from French High Schools," *The Economic Journal*, 2023, *133* (653), 1773–1809.

Broda, Christian and David E Weinstein, "Product creation and destruction: Evidence and price implications," *American Economic Review*, 2010, *100* (3), 691–723.

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.

Calaway, Ian, "Early Mentors for Exceptional Students," 2025.

Calder-Wang, Sophie and Paul A Gompers, "And the children shall lead: Gender diversity and performance in venture capital," *Journal of Financial Economics*, 2021, *142* (1), 1–22.

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.

Carpio, Lucia Del and Maria Guadalupe, "More women in tech? Evidence from a field experiment addressing social identity," *Management Science*, 2022, *68* (5), 3196–3218.

Chetty, Raj, Will S Dobbie, Benjamin Goldman, Sonya Porter, and Crystal Yang, "Changing opportunity: Sociological mechanisms underlying growing class gaps and shrinking race gaps in economic mobility," *National Bureau of Economic Research Working Paper*, 2024.

Cook, Lisa D, Janet Gerson, and Jennifer Kuan, "Closing the Innovation Gap in Pink and Black," *Entrepreneurship and Innovation Policy and the Economy*, 2022, *1* (1), 43–66.

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.

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

DellaVigna, Stefano and Matthew Gentzkow, "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Ding, Waverly W, Fiona Murray, and Toby E Stuart, "Gender differences in patenting in the academic life sciences," *science*, 2006, *313* (5787), 665–667.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell, "The Use of Scanner Data for Economics Research," *Annual Review of Economics*, 2022, *14* (1), 723–745.

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, Josh Feng, and Xavier Jaravel, "Social push and the direction of innovation," *SSRN Working Paper 3383703*, 2023.

Faber, Benjamin and Thibault Fally, "Firm heterogeneity in consumption baskets: Evidence from home and store scanner data," *The Review of Economic Studies*, 2022, *89* (3), 1420–1459.

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

Feng, JJ, Xavier Jaravel, and Eleonore Richard, "Pour une stratégie nationale dâinnovation par tous," *Focus du CAE*, 2022, (089-2022).

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.

Gozen, Ruveyda Nur, "Property rights and innovation dynamism: The role of women," 2024.

Guzman, Jorge, Jean Joohyun Oh, and Ananya Sen, "What motivates innovative entrepreneurs? evidence from a global field experiment," *Management Science*, 10 2020, *66*, 4808–4819.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, "The NBER patent citation data file: Lessons, insights and methodological tools," 2001.

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

Howell, Sabrina T. and Ramana Nanda, "Networking Frictions in Venture Capital, and the Gender Gap in Entrepreneurship," *Journal of Financial and Quantitative Analysis*, 2023, pp. 1–56.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Hurst, Erik and Benjamin Wild Pugsley, "What do small businesses do?," *Brookings Papers on Economic Activity*, 2011, pp. 73–143.

Hvide, Hans K. and Paul Oyer, "Who Becomes a Successful Entrepreneur? The Role of Early Industry Exposure," *Working paper*, 2020.

Jaravel, Xavier, "The unequal gains from product innovations: Evidence from the us retail sector," *The Quarterly Journal of Economics*, 2019, *134* (2), 715–783.

_ , Neviana Petkova, and Alex Bell, "Team-specific capital and innovation," *American Economic Review*, 2018, *108* (4-5), 1034–1073.

Jones, Benjamin F., "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?," *Review of Economic Studies*, 2009, *76* (1), 283–317.

Jones, Charles I, "R & D-based models of economic growth," *Journal of political Economy*, 1995, *103* (4), 759–784.

Kahn, Shulamit and Donna Ginther, "Women and STEM," *National Bureau of Economic Research Working Paper*, 2017.

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.

Kaplan, Greg, Benjamin Moll, and Giovanni L Violante, "Monetary policy according to HANK," *American Economic Review*, 2018, *108* (3), 697–743.

Koffi, Marlène and Matt Marx, "Cassatts in the Attic," *Working paper*, 2023.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological innovation, resource allocation, and growth," *The quarterly journal of economics*, 2017, *132* (2), 665–712.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson, "Who do we invent for? Patents by women focus more on women s health, but few women get to invent," *Science*, 2021, *372* (6548), 1345–1348.

Linder, Staffan Burenstam, *An essay on trade and transformation*, Almqvist & Wiksell Stockholm, 1961.

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 Differently They Talk About Female Entrepreneurs," *Harvard Business Review*, 2017, pp. 1–6.

Markussen, Simen and Knut Røed, "The gender gap in entrepreneurship - The role of peer effects," *Journal of Economic Behavior & Organization*, 2017, *134*, 356–373.

Mertz, Mikkel, Maddalena Ronchi, and Viola Salvestrini, "Female representation and talent allocation in entrepreneurship: The role of early exposure to entrepreneurs," *Available at SSRN 4920176*, 2024.

Mishkin, Elizabeth, "Gender and sibling dynamics in the intergenerational transmission of entrepreneurship," *Management Science*, 2021, *67* (10), 6116–6135.

Monath, Nicholas, Christopher Jones, and Sarvo Madhavan, "PatentsView: Disambiguating Inventors, Assignees, and Locations," *American Institutes For Research, of Technological Capability: Koreaâs DRAM and TFT-LCD Industriesâ, World Development*, 2021, *36* (12), 2855–2873.

Moser, Petra, Alessandra Voena, and Fabian Waldinger, "German Jewish émigrés and US invention," *American Economic Review*, 2014, *104* (10), 3222–3255.

Moshary, Sarah, Anna Tuchman, and Natasha Vajravelu, "Gender-based pricing in consumer packaged goods: A pink tax?," *Marketing Science*, 2023.

Rivera-Batiz, Luis A and Paul M Romer, "Economic integration and endogenous growth," *The Quarterly Journal of Economics*, 1991, *106* (2), 531–555.

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

Rutigliano, Valentina, "Minding Your Business or Minding Your Child? Motherhood and the Entrepreneurship Gap," *Motherhood and the Entrepreneurship Gap (March 15, 2024)*, 2024.

Schmookler, Jacob, *Invention and economic growth*, Harvard University Press, 1966.

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

Stern, Scott, "Do scientists pay to be scientists?," *Management Science*, 2004, *50* (6), 835–853.

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

Thursby, Jerry G and Marie C Thursby, "Gender patterns of research and licensing activity of science and engineering faculty," *The Journal of Technology Transfer*, 2005, *30* (4), 343–353.

Toole, Andrew A., Stefano Breschi, Ernest Miguelez, Amanda Myers, Edoardo Ferrucci, Valerio Sterzi, Charles A.W. DeGrazia, Francesco Lissoni, and Gianluca Tarasconi, "Progress and Potential: A Profile of Women Inventors on U.S. Patents," *USPTO, Office Of The Chief Economist. IP Data Highlights*, 2019.

Truffa, Francesca and Ashley Wong, "Undergraduate gender diversity and direction of scientific research," *Working paper*, 2022.

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

For Online Publication

Appendix to "Social Push and the Direction of Innovation"

Elias Einio, *VATT Institute for Economic Research*
Josh Feng, *University of Utah – David Eccles School of Business*
Xavier Jaravel, *London School of Economics*
March 2025

A Extensions

In this appendix we report two extensions, documenting homophily between the socio-demographic characteristics of entrepreneurs and their employees, and relating innovators' backgrounds to the environmental and social impacts of their innovations.

Entrepreneur–employee homophily. Using the Finnish administrative dataset, we document strong patterns of homophily between entrepreneurs and their employees. Figure C8 depicts these patterns for gender, parent income, and age. The share of female employees is 35pp larger in a firm headed by a female entrepreneur. The elasticity of the parent income of employees to the entrepreneur's parent income is 0.15, while the employee-entrepreneur age elasticity is 0.19. These findings suggest an additional channel whereby broadening access to innovation can reduce inequality, by stimulating labor demand for women and individuals from low-income backgrounds.

Environmental Impacts Prior work has emphasized the role of financial incentives to steer innovation toward "green innovations" (e.g., Acemoglu et al., 2012; Aghion et al., 2016). Motivated by the idea that an innovator's social identity and intrinsic motivations matter for the direction of innovation, we document whether female inventors have a different propensity to invent "green" patents, with positive environmental externalities. Specifically, using the data in Aghion et al. (2016), we study the differences in characteristics of inventors of "clean" versus "dirty" patents.[1] In this sample of energy patents, where 13.3 percent of energy patents are classified as "clean", we find that moving from an all-male energy patent to an all-female patent is associated with a 32.6pp increase in the probability of the patent being clean, as reported in Table C31. This difference is large: women are 2.5 times more likely to work on green patents than men. Expanding access to innovation careers for women may thus be a powerful tool to direct innovation toward cleaner technologies.[2]

B Theory Appendix

In this appendix, we first present a simple love-of-variety framework (Section B.A) and then proceed to the full-fledged growth model (Section B.B).

B.A A Simple Love-of-Variety Framework

We present a simple love-of-variety framework to assess the effects of unequal access to the innovation system on cost-of-living inequality.

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

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

where σ is the elasticity of substitution between products, $q_{k,i}$ 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.

In this setting, we model innovation as the introduction of new goods, i.e., an increase in the set of available products Ω_t. Following Feenstra (1994), the welfare gains as a percentage

[1] Following Aghion et al. (2016), we classify patents as "green" depending on their their international patent classification (IPC). We then compute the fraction of female inventors at the patent level.

[2] Using the Nielsen CPG-Crunchbase sample and the full Crunchbase sample, we also find that female-led startups are more likely to mention "healthy", "kids", "sustainability", and B-Corporation certification in their company descriptions (not reported). These results are consistent with recent evidence that women entrepreneurs are motivated by social impact (Guzman et al., 2020).

of i's current income, i.e., the equivalent variation for household i, are given by:

$$\pi_i = \frac{1}{\sigma - 1} \log \left(\frac{1 + \text{Growth of spending on continued goods}_i}{1 + \text{Growth of spending on all goods}_i} \right).$$

Assuming inelastic labor supply and taking the wage as the numeraire,[3] the equilibrium growth of total spending must be zero (by normalization), and the growth in spending on continued products is mechanically related to the share of spending on new goods, S_i^N, with

$$\text{Growth of spending on continued goods}_i = -S_i^N.$$

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 10% is equivalent to a welfare increase of 2% (= 10/5). This number can equivalently be interpreted as the fall in the cost-of-living brought about by innovation.

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."[4] We derive the welfare comparison between these two households when transitioning from the "Baseline" to the "Equal" distribution of innovator backgrounds. We then discuss how to bring these formulas to the data, computing the distributional effects between high- and low-income households, as well as between male and female consumers.

[3] We assume there is only one wage rate in the economy but that different households are endowed with different efficiency units of labor, such that they can have different income and spending levels.

[4] 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 terms of their consumer base. In other words,through preference parameters $\omega_{i,k}$, different groups of consumers may have different spending shares on the new goods introduced by different startups. S_1^N denotes the spending share of the "Type 1" representative agent from the bottom income decile on the 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.

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:

$$\frac{\pi_1}{\pi_2} \approx \frac{S_1^N}{S_2^N} = \frac{S_1^N \cdot Y_1}{S_2^N \cdot Y_2} \cdot \frac{Y_2}{Y_1} = \frac{R_1/R_2}{Y_1/Y_2}, \tag{A1}$$

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, Y_1/Y_2. 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.

A simple calibration. We wish to examine the extent to which one of the household types may benefit more from transitioning to a new distribution of innovator background. When moving from the baseline distribution of entrepreneur background ("Baseline") to a counterfactual distribution ("Equal"), from Equation (A1) we obtain that the unequal welfare effect across household types can be expressed as:

$$\Delta W \equiv \frac{\pi_1^{\text{Equal}}/\pi_1^{\text{Baseline}}}{\pi_2^{\text{Equal}}/\pi_2^{\text{Baseline}}} = \frac{R_1^{\text{Equal}}/R_2^{\text{Equal}}}{R_1^{\text{Baseline}}/R_2^{\text{Baseline}}}. \tag{A2}$$

The relative welfare effect is thus governed by the share of sales to household groups of different types. We can directly connect this expression to the homophily regression coefficients from specification (1) in the main text. Denoting by λ the share of sales to "Type

1" households, we can write $R_1/R_2 = \lambda/(1-\lambda)$. Tables 3 and 4 are directly informative about λ for startups with different innovator backgrounds.

Consider female ("Type 1") and male ("Type 2") entrepreneurs. Using the notation for the regression coefficients in Equation (1), then for a female entrepreneur we have $\lambda^F = \alpha + \beta$, while for a male entrepreneur $\lambda^M = \alpha$. For example, for consumer packaged goods, $\lambda^F = 0.297$ and $\lambda^M = 0.25$ (Table 3, col. (1)). Our homophily estimates are thus directly informative about changes in revenue shares, which govern the welfare gains from different types of startups across consumers. For consumer packaged goods, Equation (A2) yields:

$$\Delta W = \frac{\lambda^F/(1-\lambda^F)}{\lambda^M/(1-\lambda^M)} = \frac{0.297/(1-0.297)}{0.25/(1-0.25)} \approx 1.27$$

Thus, in the consumer packaged goods sample our preferred specification indicates that the welfare gains from female-founded startups are 27 percent larger for the representative female household, relative to the representative male household.[5] This number increases to 46 percent in the sample of phone applications.[6]

Using these estimates, we can assess the relative effect of having a more representative pool of innovators. We start from the observed distribution of entrepreneur backgrounds, where women represent about 12 percent of innovators (Toole et al. (2019)), i.e., $\lambda^{\text{Baseline}} = 0.12 \times \lambda^F + 0.88 \times \lambda^M = 0.2556$. We consider a counterfactual distribution with parity between male and female inventors, where $\lambda^{\text{Counterfactual}} = 0.5 \times \lambda^F + 0.5 \times \lambda^M = 0.2735$. According to Equation (A2), moving to a world where fifty percent of innovators are women would yield 9.6% larger relative benefits to female consumers ($= \frac{0.2735/(1-0.2735)}{0.2556/(1-0.2556)}$), i.e., cost-of-living inequality would fall by this amount.[7]

[5] Note that this difference is larger than the difference in revenue shares from female households that arises between female-founded and male founded startup. As shown in Column (1) of Table 3, the revenue share from female-led households is 18.8% larger for female-founded startups compared with male-founded startups (29.47% 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.

[6] 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.

[7] We find larger effects in the full-fledged growth model, because the back-of-the-envelope calculation from the simple love-of-variety framework does not take into account that reducing access barriers may bring into the innovation system some highly productive female inventors, who will have a disproportionate impact on female consumers through homophily.

Note that this simple framework is easy to take to the data to document the relative welfare effects from changes in the distribution of innovator characteristics, but it does not provide estimates of the welfare effects in levels. This limitation is addressed by the endogenous growth model we develop.

B.B Endogenous Growth Model

This section presents the main derivation steps for the growth model, as well as our numerical solution algorithm and complementary results.

B.B.1 Derivations

In this section, we present derivations related to the model setup in Section IV.B.

Our goal is to derive the first-order conditions that characterize the research productivity cutoff in each sector in a balanced growth path equilibrium. Given the setup in the main text, we have two sets of conditions that pin down the cutoffs. First, we have indifference conditions at the cutoff in each sector, where the marginal agent is indifferent between production work and innovation work. Second, we have the Euler equation, which ensures that the allocation of labor to innovation in the economy accords with consumer preferences. For example, if consumers are very impatient, then the economy cannot sustain high rates of innovation in equilibrium.

We make one simplifying assumption that allows us to vary market size in a tractable way. We assume that the earnings of inframarginal entrepreneurs is negligible and therefore do not impact market size. Under this assumption, δ governs the relative market size of the two sectors.

Given preference parameters $(\rho, \varepsilon, \alpha_M, \alpha_W)$ and research-related parameters $(\bar{\eta}, \lambda, \tau, \phi)$, the model is solved as follows:

1. We express the distribution of research productivity in each sector, $f_1(x)$ and $f_2(x)$, as a function of the exogenous parameters of the model:

    - Let $f(x)$ be the research productivity distribution in the population, a Pareto distribution with scale parameter $\bar{\eta}$ and shape parameter λ.

- Taking into account frictions and sector assignment, we can write the two sector-level innovation productivity distributions in terms of parameters as follows:
  - $f_1(x) = \frac{1}{2}(\phi \cdot f(x) + (1-\phi) \cdot \tau \cdot f(x))$ and $f_2(x) = \frac{1}{2}((1-\phi) \cdot f(x) + \phi \cdot \tau \cdot f(x))$.
  - Note that at $\tau = 1$, when there are no frictions, the two distributions will be identical. The distributions will also be equal at $\phi = \frac{1}{2}$ for any value of τ.

2. We define some additional variables:

   - Let $\hat{\eta}_1$ and $\hat{\eta}_2$ represent the cutoffs in each sector. Agents assigned to the sector pursue an innovation career if their productivity exceeds the sector-specific cutoff.
   - Let $\tilde{\alpha} = \delta\alpha_W + (1-\delta)\alpha_M$ represent the market-size-weighted preference, or "effective taste", for sector 1 across the whole population.
   - Let L_{sM} denote the measure of agents devoted to production work in sector s.

3. Using this notation, we have the following in equilibrium:

   (a) Production workers will be split between sectors based on effective tastes in the economy: $\frac{L_{1M}}{L_{2M}} = \frac{\tilde{\alpha}}{1-\tilde{\alpha}}$.

   (b) The labor marker clears, so with a mass one of agents the number of production workers relates to the number of innovators as follows: $L_{1M} + L_{2M} = 1 - \int_{\hat{\eta}_1}^{\infty} f_1(x)dx - \int_{\hat{\eta}_2}^{\infty} f_2(x)dx$

4. In equilibrium, the marginal agent in each sector is indifferent between production work and innovation:

   (a) The indifference condition is $w(t) = \hat{\eta}_s N(t) V(t) = \hat{\eta}_s N(t) \frac{\pi(t)}{r^*} = \frac{1}{\varepsilon - 1} \cdot \hat{\eta}_s N(t) \frac{w(t)}{r^*} \cdot \frac{L_{sM}}{N_s(t)}, \forall s$.

   (b) This indifference condition implies $\hat{\eta}_1 L_{1M}/N_1(t) = \hat{\eta}_2 L_{2M}/N_2(t)$.

   (c) Furthermore, we know that varieties grow at the same rate across the two sectors in a balanced growth path equilibrium: $\frac{N(t)}{N_1(t)} \int_{\hat{\eta}_1}^{\infty} x f_1(x)dx = \frac{N(t)}{N_2(t)} \int_{\hat{\eta}_2}^{\infty} x f_2(x)dx$

(d) Combining steps 3(a), 4(b), and 4(c), we obtain the first equilibrium condition **(FOC1)**:

$$\frac{\hat{\eta}_1}{\hat{\eta}_2} = \frac{1-\tilde{\alpha}}{\tilde{\alpha}} \cdot \frac{\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx}{\int_{\hat{\eta}_2}^{\infty} x f_2(x)dx}.$$

This first-order condition pins down the relative levels of innovation productivity cutoffs across sectors.

5. Next, we use the Euler equation to pin down the level of these cutoffs:

- The standard Euler equation holds: $\frac{\dot{C}(t)}{C(t)} = r(t) - \rho = \frac{\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_1(t)}{N_1(t)} + \frac{1-\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_2(t)}{N_2(t)}$.
- From step 4(a) we have $r^* = \frac{1}{\epsilon-1} \cdot \hat{\eta}_s N(t) \frac{L_{sM}}{N_s(t)}, \forall s$. Plugging this expression on the LHS of the Euler equation yields: $\frac{1}{\varepsilon-1}\hat{\eta}_1 \frac{N(t)}{N_1(t)} L_{1M} - \rho = \frac{\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_1(t)}{N_1(t)} + \frac{1-\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_2(t)}{N_2(t)}$.
- On the RHS of the Euler equation, we plug in the growth in varieties using the research production function: $\frac{1}{\varepsilon-1}\hat{\eta}_1 \frac{N(t)}{N_1(t)} L_{1M} - \rho = \frac{\tilde{\alpha}}{\varepsilon-1}\frac{N(t)\cdot\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx}{N_1(t)} + \frac{1-\tilde{\alpha}}{\varepsilon-1}\frac{N(t)\cdot\int_{\hat{\eta}_2}^{\infty} x f_2(x)dx}{N_2(t)}$.
- Using the equations in steps 3(a) and 4(b), we obtain the ratio of varieties across sectors, $\frac{N_1(t)}{N_2(t)} = \frac{\int_{\hat{\eta}_1} x f_1(x)dx}{\int_{\hat{\eta}_2} x f_2(x)dx}$. We use this expression to plug in for $\frac{N(t)}{N_1(t)}$ and $\frac{N(t)}{N_2(t)}$ on the RHS of the Euler equation, and we arrive at the second equilibrium condition **(FOC2)**:

$$\frac{1}{\varepsilon-1}\hat{\eta}_1 \frac{\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx + \int_{\hat{\eta}_2}^{\infty} x f_2(x)dx}{\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx} L_{1M} - \rho = \frac{1}{\varepsilon-1}\left(\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx + \int_{\hat{\eta}_2}^{\infty} x f_2(x)dx\right).$$

Together, FOC1 and FOC2 pin down the two innovation productivity cutoffs. Note that in FOC2, L_{1M} also depends on $\hat{\eta}_1$, so the entire system is non-linear and does not permit a closed form solution.

Having solved the model, we compute the cost-of-living gender gap as $\frac{P_{Wt}}{P_{Mt}} = \left(\frac{N_1(t)}{N_2(t)}\right)^{\frac{\alpha_M - \alpha_W}{\epsilon-1}}$, using the fact that the sector CES price index is $P_s = \left(\int_0^{N_s} p(x)^{1-\varepsilon}dx\right)^{1/(1-\varepsilon)} = N_s^{1/(1-\varepsilon)}$, which we plug into each consumer group's price index, $P_g = P_1^{\alpha_g} P_2^{1-\alpha_g}$.

B.B.2 Simulation Algorithm

We now describe our numerical simulation algorithm. The estimation of the model proceeds in three steps:

First, for each agent we draw a research productivity value $n_i \sim PI(1, 1.26)$ (type-I Pareto distribution). We also draw a sector assignment number$\widetilde{\phi}_i \sim U[0,1]$ for all individuals. For agents in the minority group, we draw an access friction value from $\widetilde{\tau}_i \sim U[0,1]$. All individuals in the majority group (i.e., men or the top income quintile) are assumed to face no barriers. We thus obtain draws from uniform distributions once at the beginning of the simulation, avoiding us to redrawing random variables under each new parameter guess.

Second, for a given parameter vector guess $(\tau, \overline{\eta}, \phi)$, we convert the numerical draws described above into Bernoulli draws. Agents from the minority group are exposed if $\widetilde{\tau}_i < \tau$, and we assign agents to the sector for which they have a taste preference if $\widetilde{\phi}_i < \phi$. We also scale the research productivity draws by $\overline{\eta}$ to control the shape of the innovation productivity distribution, resulting in a distribution with CDF $F(x) = 1 - \left(\frac{\overline{\eta}}{x}\right)^{\alpha}$. This allows us to then compute the empirical distributions $f_1(x), f_2(x)$, corresponding to the distribution of productivity of individuals who are exposed and assigned to sector 1 or sector 2.

Finally, we numerically search for the cutoffs $\hat{\eta}_1$ and $\hat{\eta}_2$ that satisfy the two first-order conditions, FOC1 and FOC2, derived in the previous subsection. Given these cutoffs, we then compute the growth rate, the propensity to pursue an innovation career for each group, and homophily. We perform a derivative-free search ("fminsearch" in MATLAB) over the set of parameters $(\tau, \overline{\eta}, \phi)$ to match the observed moments, which we are able to match exactly.

Once we have calibrated the baseline economy, we run counterfactual analyses. Specifically, we vary the parameter values for τ, and recompute the BGP growth rate and cost-of-living inequality. We either run a "full equal access" counterfactual, where all agents can enter innovation, or "targeted exposure" counterfactuals where we give access to all individuals in the top x percent of the productivity distribution (among those who previously faced barriers).

To check that the results are not sensitive to the specific draws of the random variables, we consider an economy with a large number of agents, 8 million. We thus have 4 million agents from each group, where groups refer to gender or top/bottom income quintiles. Furthermore, we repeat the analysis one hundred times and report values for median draws.

B.B.3 Evidence on Wedges

Here, we provide additional details related to the discussion in the main text on empirical evidence related to access wedges.

First, we study innovation productivity, comparing model predictions, depending on the wedge, to the patterns in the data. Panel A of Table C26 reports the probability of being a "top innovator" (top 10% or top 5% of the innovation productivity distribution) conditional on innovating, across models.[8] The model with wedge 2 makes stark predictions: women should be highly likely to be top innovators conditional on innovating. Indeed, in this model only the most productive female innovators find it worthwhile to enter the innovation sector (given the cost $\tilde{\tau}_{gi}$ affecting their returns to innovation), while their productivity is unaffected. In contrast, the models with wedges of type 1 or 3 both predict that women should have lower probabilities of being a top innovator.[9]

Panel B of Table C26 presents the patterns in the data. We examine the probability of being in the top 10% or top 5% of innovators for phone applications (ranking by time usage of apps), consumer packaged goods (ranking by sales), and patents (ranking by citations). In all cases, we find that female innovators are *under-represented* at the top of the production. These empirical patterns reject the model with Type 2 wedge and favor models with Type 1 or Type 3 wedges.[10]

[8]Innovation productivity is governed by η_i for models with wedges of Type 1 and 2, and by $(1 - \grave{\tau}_{gi})\eta_i$ for the model with the Type 3 wedge.

[9]This under-representation at the top comes from the fact that, in equilibrium, the productivity cutoff for entry of innovator is slightly lower in the sector focusing on female customers (as noted above), where there are relatively more female innovator due to innovator-consumer homophily. This results from the interaction of demand and supply forces in equilibrium. On the demand side, customer demand is a bit lower for the female-focused sector because of the gender wage gap, which reduced the relative purchasing power of female customers. On the supply side, there is much lower supply of innovators in the female-focused sectors, because of the frictions female innovators face. To ensure that supply meets demand in equilibrium, the productivity cutoff must be lower and the price index is higher in the sector focusing on female customers. This mechanism applies to the three types of wedges and push for slightly lower productivity for women, who disproportionately enter the female-focused sector through homophily. With wedge 2, the selection effect (only the most productive female innovators decide to enter innovation) vastly offsets the heterogeneous cutoffs effect. With wedge 1 and 2, there is no offsetting selection effect, so female innovators are slightly less productive in equilibrium and, in particular, are a bit under-represented among top innovators. Finally, it is worth noting that the heterogeneous cutoffs mechanism is consistent with the finding that, in the consumer packaged goods data, female entrepreneurs start companies with higher unit prices.

[10]As a complement to the simulation results in Table C26, we show analytically in Appendix B.B.4 that models with Type 1 and Type 3 wedges make very similar predictions for the probability of becoming a top inventor by gender.

While a model with Type 1 wedges predicts that the distribution of *latent* innovator productivity should be similar for men and women, a model with Type 3 wedges predicts that women who becomes innovators come overwhelmingly from the top of the latent innovation productivity distribution. Although the latent innovation productivity distribution is unobserved, a simple test is to use math test scores in childhood.[11] Focusing on patent inventors, Bell et al. (2019a) report that, compared to men, women and minorities remain much less likely to become inventors even when they have math test scores two standard deviations above the mean (see Panel A of Figure C9). Focusing on exceptional math students in the American Mathematics Competition (AMC), Calaway (2025) shows that even among these exceptional math students in middle school or high school, women and students from certain demographic groups remain less likely to attend selective universities and major in STEM fields. Calaway (2025) highlights that the gender gap is particularly striking: exceptional female students pursue STEM majors at a rate that is 19pp/20pp lower than male students (Panel B of Figure C9). These patterns are inconsistent with models with Type 3 wedges – where the female inventors with the highest productivity should always enter innovation, and do less well than men *ex-post* due to productivity barriers.

B.B.4 Gender Ratios among Top Inventors

In this appendix, we derive simple formulas for the gender ratio among the top X percent of inventors under different types of wedges. The formulas show that, under our calibrated parameters, the gender ratio among top inventors is almost identical under Type 1 and Type 3 wedges.

We work with the CDF of the basic Pareto distribution, $Pr(X > x) = \left(\frac{x_m}{x}\right)^{\alpha}$. When comparing the female fraction in the top X percent of inventors for the Type 3 wedge model, female inventors have their productivities multiplied by τ_{real} (a penalty), so the CDF for the productivity of female inventors becomes $Pr(\tau_{real}X > x) = \tau_{real}^{\alpha}\left(\frac{x_m}{x}\right)^{\alpha}$, while men keep the basic CDF, $Pr(X > x) = \left(\frac{x_m}{x}\right)^{\alpha}$. Taking ratios of CDFs, the fraction of female inventors above any cutoff x (greater than the entry cutoff in the male sector) is given by $\frac{\tau_{real}^{\alpha}}{1+\tau_{real}^{\alpha}}$. In

[11] Several paper have shown that high mathematical ability in childhood strongly correlates with knowledge production in adulthood (e.g., Aghion et al., 2023; Akcigit et al., 2017; Bell et al., 2019a; Agarwal and Gaulé, 2020).

our calibration, $\tau_{real} = 0.1683, \alpha = 1.26$, so $\tau_{real}^{\alpha} \approx 0.106$.

In the Type 1 wedge model, the probability of being above a certain innovator productivity cutoff shrinks uniformly for women, $Pr(X > x) = \tau_{exposure} \left(\frac{x_m}{x}\right)^{\alpha}$. Therefore, the gender ratio is $\frac{\tau_{exposure}}{1+\tau_{exposure}}$. In our calibration, $\tau_{exposure} = 0.111$, which is very close to the value of τ_{real}^{α} found above. Thus, the gender ratio is similar at each percentile cutoff in models with Type 1 and Type 3 wedges.

B.B.5 Endogenizing the Gender Wage Gap

Our baseline model features an exogenous gender earnings gap, governed by the exogenous parameter δ. This parameter scales in the same way production worker wages $w(t)$ and the returns to innovation $V_j(t)$ for each group g (men or women), so that occupational choice is left unaffected.[12] This parameter remains identical when we relax access barriers to the innovation sector for women. In this version of the model, for tractability we ignore differences in the share of women in entrepreneurship when computing earnings gap by gender such that, given our calibration of the earnings penalty parameter δ, men remain 20.5% richer than women regardless of the counterfactual.

In this appendix, we introduce a model with an endogenous gender earnings gap. The model is identical to the baseline model except that we now compute the difference in total earnings by adding up earnings in production work and earnings in entrepreneurship. In this version, differences in earnings between men and women account for the fact that being under-represented in innovation careers reduces average earnings for women, as these careers tend to be more lucrative. In this version of the model, the endogenous gender earnings gap is 1.568, compared to 1.205 in the baseline model. This increase in the gender earnings gap comes from the fact that innovators earn on average 4.93 more than production workers in the calibrated model.[13] In this version of the model, we solve for a fixed point equilibrium such that the market size of each sector is consistent with total earnings of each group.

[12]This assumption is consistent with the fact that certain factors that lead to lower earnings for the median female worker, such as the child penalty, also operate for entrepreneurs with similar magnitudes (e.g., Rutigliano (2024)).

[13]Production worker wages act as the numeraire, so the gap between innovators and production workers should be close to the mean of the Pareto distribution for inventor productivity. Indeed, with $\lambda = 1.26$, as in our calibration, the Pareto mean is $\lambda/(\lambda - 1) = 4.85$.

Panel A of Table C32 shows that the two versions of the model deliver similar results for growth and cost-of-living counterfactuals. There is a fall in cost-of-living inequality of 16.3 percentage points in the model with endogenous wage gaps, compared to 18.20pp in the baseline model. The growth impacts of equal access to innovation careers is 1.38pp with endogenous wage gaps, compared to 1.44pp in the baseline model. The differences across the two models in terms of growth and cost-of-living inequality are small because the endogenous gender earnings gap affects innovation dynamics through market size, which only affects "marginal" inventors close to the productivity cutoff.

Panel B of Table C32 provides the calibrated parameters in both versions of the model, which explain why the cost-of-living and growth impacts are slightly smaller in the model with endogenous wages. The calibrated parameters are different in the model with endogenous gender wage gaps because of market size effects at the initial equilibrium. The relative earnings of men compared to women are larger in the endogenous gender wage gap model than in the baseline model because entrepreneurial income is taken into account. As a result, the market size of the sector targeting female customers, relative to the market size of the sector targeting male customers, is smaller in the endogenous gender wage gap model. This ends up changing several aspects of the baseline calibration. First, there is an impact on the calibrate homophily parameter, ϕ. As noted in the main text, we calibrate the model to target our empirical estimates of the average difference in female sales shares between male and female entrepreneurs, divided by the baseline rate of female consumption – namely, female entrepreneurs should sell about 25% more to female customers. Because the baseline sales share of female customers (relative to male customers) is smaller in the endogenous model, we no longer need as large an absolute difference in sales shares to male vs. female customers to match our regression coefficient. Consequently, the parameter ϕ in the model, which captures an absolute difference in the tendency to sell to male vs. female customers, is calibrated to be smaller in the endogenous wage model than in the baseline model. Therefore, the cost-of-living impacts of equal access to innovation by women turns out to be slightly smaller, at -16.3pp in the endogenous wage model compared to -18.2pp in the baseline model. Second, the wedge parameter τ is also affected. The smaller value of ϕ means that there is less sectoral specialization, i.e. that women are less shielded from competition from men (by focusing on a different sector) than they were in the baseline model. Consequently, matching the overall female inventor fraction requires a higher τ (lower barriers) in the endogenous wage model than in the baseline model. The counterfactual bringing τ to one (i.e., eliminating barriers) therefore constitutes a smaller change in the endogenous gender wage gap model, with slightly lower growth impacts, at 1.38pp instead of 1.44pp.

While the impacts on growth and cost-of-living inequality are limited, there is a large impact on relative earnings between men and women. The results are reported in Panel C of Table C32. In our baseline model with exogenous wage gap, the earnings ratio remain fixed at 1.205. In the model taking into account endogenous entrepreneurial income, the male to female ratio falls from 1.57 to 1.212 when relaxing barriers to innovation for women.[14] Indeed, relaxing barriers leads to an increase in the share of women in the innovation sectors, which raises the average earnings of women as this sector offers higher compensation. The level of inequality in the counterfactual without access barriers equilibrium thus falls back to the level of inequality set by the penalty parameter δ.[15] The fall in overall inequality is thus much stronger than in the baseline model.[16]

Because the expanded model takes twenty times longer to run than the baseline model and that the gender wage gap is not our focus, it is not used in our baseline analysis, for tractability.

[14] The latter number would converge to 1.205 with an infinite sample; the reported result differs due to sampling error in our 100 bootstrap runs.

[15] The inequality patterns in our calibrate model are broadly consistent with external benchmarks. First, innovators earn 22.2 % of aggregate income in the model, similar to the share of income earned by the top 1 % in the United States (World Income Inequality Database). Second, the gender wage gap for the median worker is 20.5 % according to the U.S. Bureau of Labor Statistics, which we match exactly by setting δ. Third, when accounting for access barriers for innovation, the total gender earnings gap is such that men earn 57 % more than women, similar to the overall earnings disparities by gender documented by Andrew, Bandiera, Costa Dias, and Landais (2024).

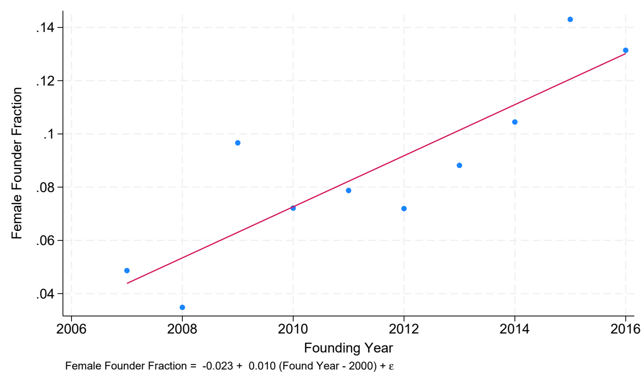
[16] Panel C of Table C32 also reports the change in the earnings ratios of entrepreneurs to production workers with and without innovation barriers for women. In the counterfactual without barriers, the earnings ratio falls from 4.93 to 4.63. This change occurs because there are now relatively more female entrepreneurs among innovators and they are subject to the earnings penalty δ.

C Additional Figures and Tables

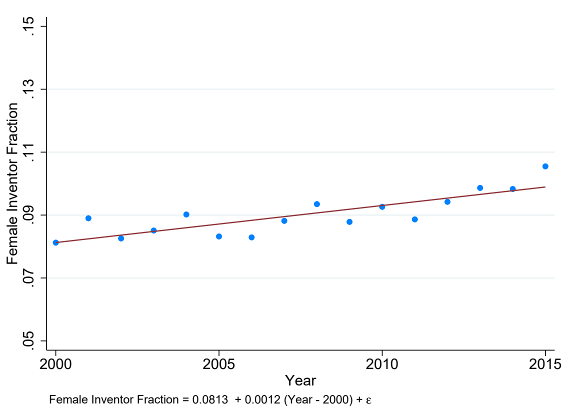
Figure C1: Gender Gaps in Entrepreneurship in the United States

A. Fraction of Female Startup Founders within Consumer Packaged Goods

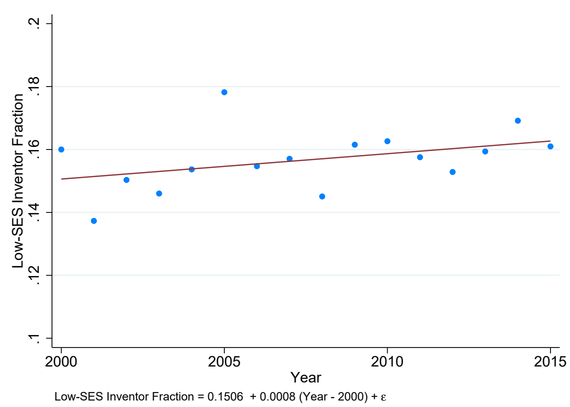
B. Fraction of Female Startup Founders within Phone Applications



C. Fraction of Female Patent Inventors



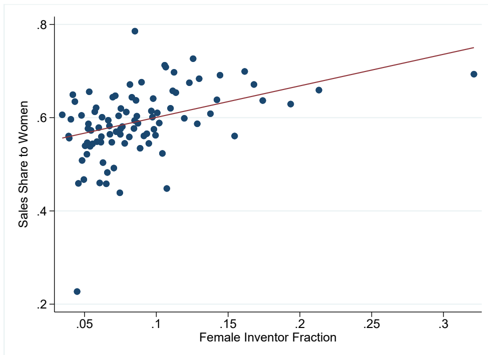
D. Fraction of Low-SES Patent Inventors



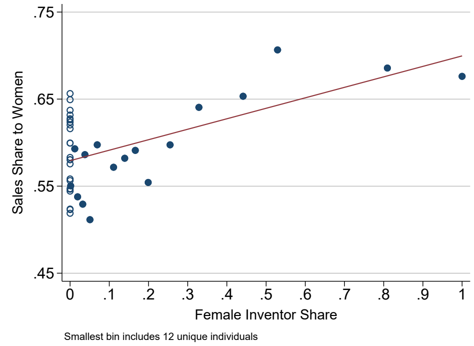
Notes: This figure presents the fraction of female startup founders in the United States for consumer packaged goods (Panel A) and phone applications (Panel B); the fraction of female patent inventors in Finland (Panel C); and the fraction of low-SES patent inventors, whose parental income is in the lowest quintile, in Finland (Panel D). The trend lines show that, at this rate, it will take 50 years (from the start of the sample) to reach gender parity among founders in the United States; it will take over 60 years to reach parity by gender and parental income among patent inventors in Finland.

Figure C2: Binned Scatter Plots for Innovator-Consumer Homophily across Industries

A. Gender

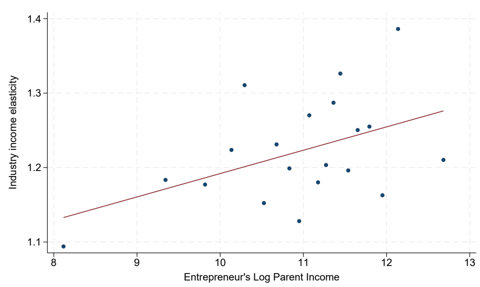


United States



Finland

B. Parent Income

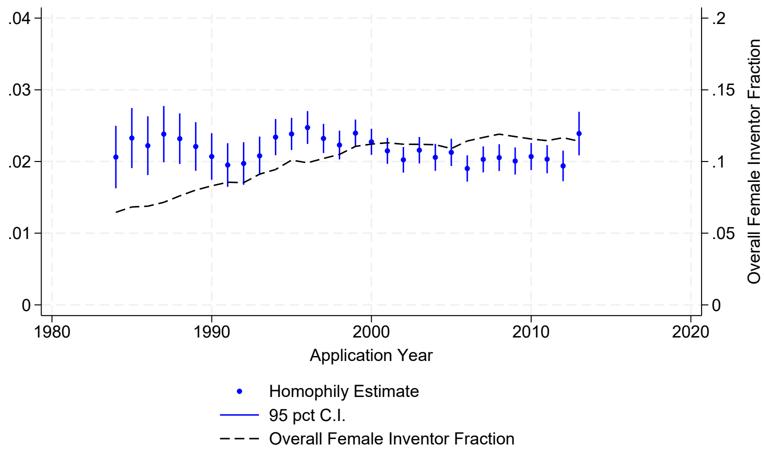


United States

Finland

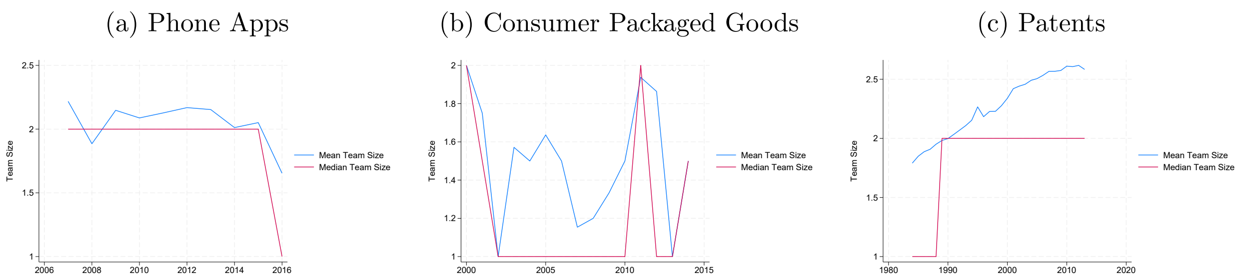
Notes: This figure presents binned scatter plots depicting innovator-consumer homophily by gender (Panel A) and social class (Panel B), for both the United States and Finland. The samples are described in Section II in the main text. Due to data limitations, the sample for the analysis by parent income in the United States only includes 300 observations. We start from the full PSID sample in 2017, which includes 26,445 individuals. We then link these individuals to their parents in earlier waves of the PSID: we can link 13,172 individuals. We only keep individuals between the ages of 21 and 69, reducing the sample to 5,257 individuals. In this sample, we have 325 entrepreneurs, defined as self-employed business owners. For the regression, we further exclude outliers – industries with the maximum or minimum industry income elasticities, and observations in the top 1% or bottom 1% of the parent income distribution.

Figure C3: Homophily Trends – U.S.-Based Inventors



Notes: The figure displays the estimates of the inventor-level homophily regression by filing year with industry-level consumption measures, based only on patents with all U.S.-based inventors instead of all U.S. patents (about 48 percent of the overall sample). The outcome is female consumer fraction, based on CEX data. We show 95 percent confidence intervals using robust standard errors, and also report the female inventor fraction in each year.

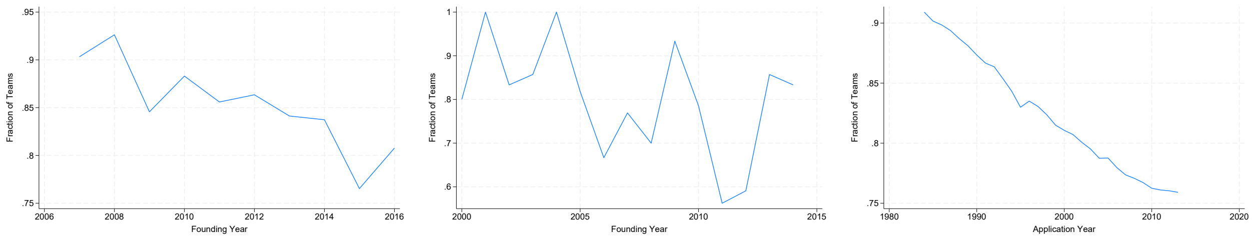
Figure C4: Team Size over Time



Notes: The panels of this figure report the mean and median time sizes over time for founder teams of phone apps startups (panel a) and of consumer packaged goods startups (panel b), as well as inventor teams in the U.S. patent data (panel c). The patterns for consumer packaged goods are noisy due to the limited sample size in each year.

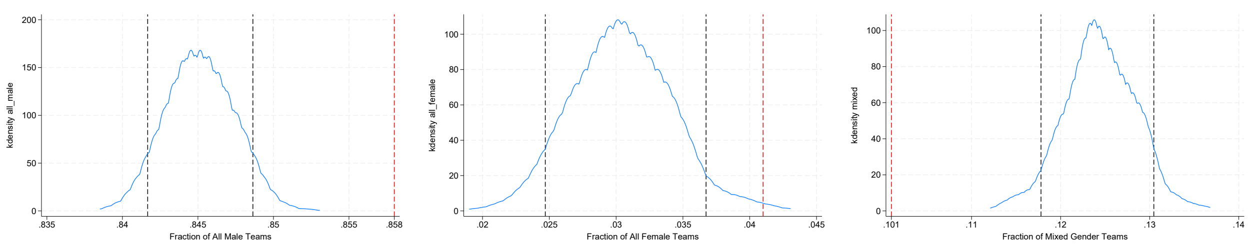
Figure C5: Fraction of All-Male Innovator Teams over Time

(a) Phone Apps (b) Consumer Packaged Goods (c) Patents

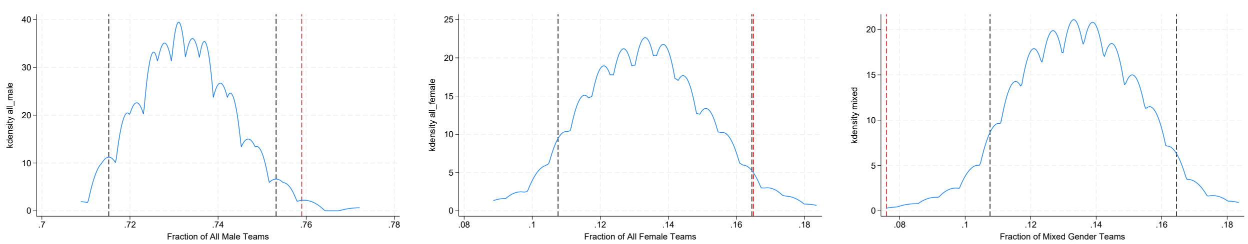


Notes: The panels of this figure report the fraction of teams with only male founders in phone apps startups (panel a) and consumer packaged goods startups (panel b), and with only male inventors in the U.S. patent data (panel c). The patterns for consumer packaged goods are noisy due to the limited sample size in each year.

Figure C6: Gender Clustering in Teams



(a) All Male Teams, Phone Apps (b) All Female Teams, Phone Apps (c) Mixed Gender Teams, Phone Apps

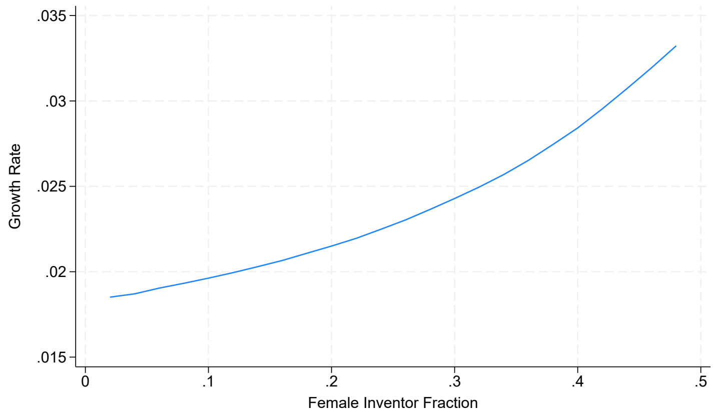


(d) All Male Teams, CPG (e) All Female Teams, CPG (f) Mixed Gender Teams, CPG

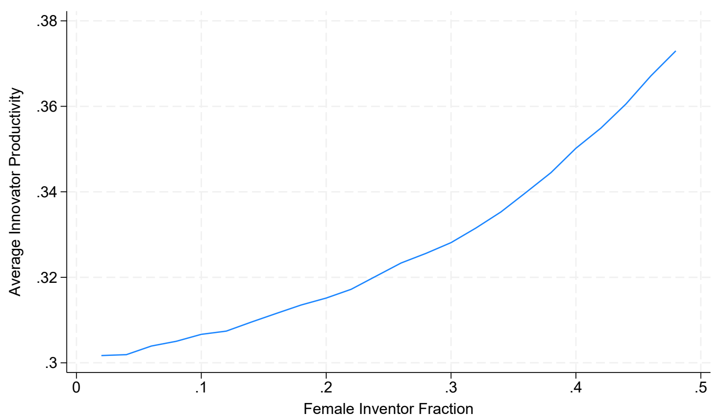
Notes: The panels of this figure report the distributions of the fraction of all-female, all-male, and mixed-gender teams under random assignment (histogram), their 5th and 95th percentile values (black lines), and means for these statistics in the data (red line). The distribution of the fraction of mix-gender teams under random assignments is obtained via bootstrap. The analysis is conducted separately on the phone apps and consumer packaged goods samples.

Figure C7: Change in Growth Rate and Innovator Productivity vs. Change in Female Inventor Fraction

A. Growth Rate



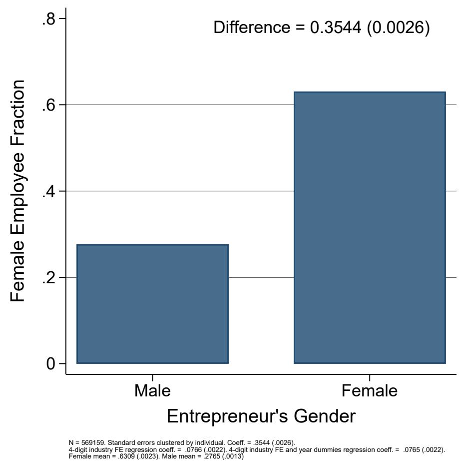
B. Average Innovator Productivity



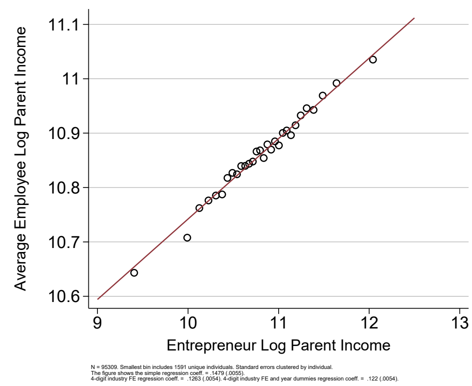
Notes: This figure reports the equilibrium change in growth rate (Panel A) and average innovator productivity (Panel B) against the equilibrium change in the fraction of women among innovators (x-axis), as we reduce access barriers in the model with Type 1 wedge.

Figure C8: Entrepreneur-Employee Homophily

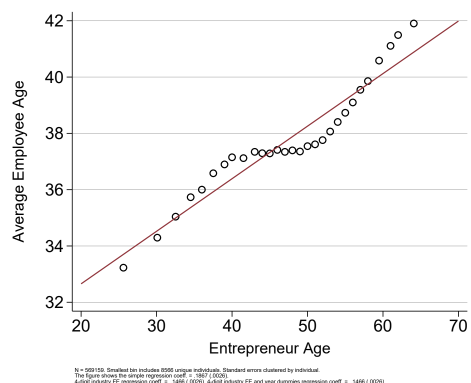
A. Gender



B. Parent Income



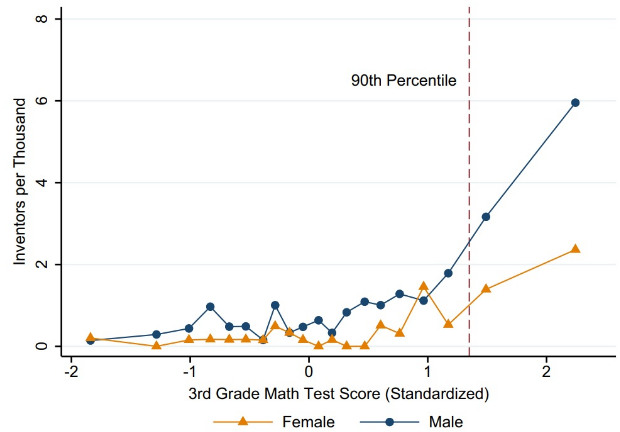
C. Age



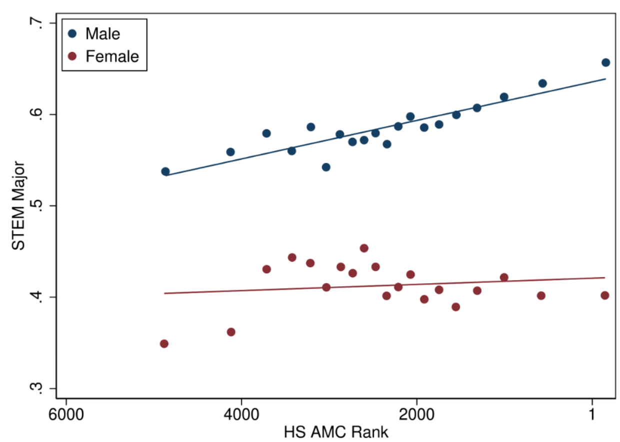
Notes: This figure depicts the relationship between entrepreneur and employee gender, parent income, and age, in the Finnish administrative data set.

Figure C9: Probability of Becoming an Innovator by Gender and Latent Ability

Panel A – Bell et al. (2019): Gender Gaps in Patent Rates
by 3rd Grade Math Test Scores



Panel B – Calaway (2025): Gender Gap in STEM
for Exceptional Math Students



Notes: This figure reports evidence from Bell et al. (2019a) on Panel A and Calaway (2025) in Panel B, using math test scores as (noisy) proxies for latent innovation abilities. The panels show that gender gaps in becoming patent inventors and STEM majors persist even among exceptionally talented students.

Table C1: Innovator-Consumer Geographic Homophily for Phone Applications

	User State Share	
	(1)	(2)
Founder State	0.088***	0.038***
	(0.007)	(0.009)
Fixed Effects	None	Category-by-State
Sample Size	46,512	46,176

Notes: The sample used in this table includes all phone applications for VC-backed startups. The analysis is at the app by state level. The outcome variable is the fraction of time usage of a given app from a given state, which has a sample mean of 0.02. The independent variable is an indicator for whether the company is located in the state. With category-by-state fixed effects, the estimate is identified off whether apps in the same category have different shares in a given state based on where the startups are located. The sample size varies because singleton observations are dropped in the second column. Standard errors are clustered at the startup level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: Innovator-Consumer Homophily for Consumer Packaged Goods, Alternative Consumer Gender Measure

	Share of Sales to Women	
	(1)	(2)
Female Founder Fraction	0.0367**	
	(0.0159)	
Female Patent Inventor Fraction		0.0361***
		(0.0128)
Product Module F.E.	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 1,094,229$

Notes: This table is identical to Columns (1) and (2) of Table 3 in the main text, except that we measure consumer gender by weighting sales by the fraction of female members in the household, rather than the gender of the household head. The coefficients are very similar to the main text, showing that our results are robust to alternative measures of consumer gender. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Innovator-Consumer Homophily by Age Group, Alternative Consumer Age Measure

	Share of Sales to Age Group	
	(1)	(2)
Founder of Same Age Group	0.0433*	0.0434***
	(0.0244)	(0.0155)
Fixed Effects	None	Module-by-Age Group
Sample size	$N = 4,058$	

Notes: This table presents an alternative analysis of age-based homophily. We create five age group bins based on quintiles of the consumer age distribution in Nielsen (the cutoffs are 42.5, 52.5, 60, and 65). We then compute the share of sales to each age group for each product. Finally, we create a binary indicator, "Founder of Same Age Group," equal to one if the average age of the founding team at founding matches the given age group. The baseline probability of sales is mechanically 20% across all groups. The results show that having a founder in the same age group increases the probability of sales by about 21.5% relative to baseline. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C4: Gender Homophily – Revised Analysis for CPG Sample with Patents

	Share of Sales to Women							
	All UPCs		New UPCs		New Brands		New Manufacturers	
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0248*	0.0373***	0.0243	0.0338***	0.0315**	0.0220*	0.0763	0.0384
	(0.0149)	(0.0108)	(0.0149)	(0.0103)	(0.0148)	(0.0119)	(0.0486)	(0.0264)
Constant	0.269***	0.537***	0.271***	0.540***	0.266***	0.539***	0.259***	0.540***
	(0.00242)	(0.00265)	(0.00268)	(0.00238)	(0.00251)	(0.00198)	(0.00909)	(0.00504)
Fixed Effects	Module							
Unique firms	1,178	1,178	1,112	1,112	909	909	397	397
Sample Size	295,292	295,292	222,794	222,794	51,080	51,080	15,917	15,917

Notes: This table presents an alternative analysis of homophily, in the sample of firms that obtain a patent, within the consumer packaged goods sample. While the first two columns keep all barcodes, as in the main text, in the other columns we construct three new subsamples of products: 1) under new UPCs, 2) under new brands, and 3) by new manufacturers, entering the data in 2005 or later. The estimates remain broadly similar in these subsamples and suggesting that the results in the baseline sample are not driven by established products. The coefficients for the new manufacturers' subsample are large but imprecisely estimated. Keeping all barcodes, as in the first two rows, has the advantage of keeping patents affecting process innovations, rather than new products alone. The outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: Gender Homophily – Subsample with High Revenue Per Patent

	Share of Sales to Women							
	All UPCs		New UPCs		New Brands		New Manufacturers	
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0260*	0.0391***	0.0264*	0.0365***	0.0326**	0.0251**	0.108*	0.0578**
	(0.0155)	(0.0112)	(0.0156)	(0.0105)	(0.0164)	(0.0127)	(0.0556)	(0.0291)
Constant	0.269***	0.537***	0.271***	0.539***	0.266***	0.538***	0.260***	0.539***
	(0.00251)	(0.00274)	(0.00279)	(0.00244)	(0.00267)	(0.00213)	(0.0110)	(0.00572)
Fixed Effects	Module							
Unique firms	592	592	584	584	512	512	102	102
Sample Size	284,819	284,819	215,154	215,154	48,298	48,298	12,717	12,717

Notes: This table presents an alternative analysis of homophily, in the sample of firms that obtain a patent, within the consumer packaged goods sample. The analysis is the same as in Table C4, except that we further reduce the sample to the subset of firms with an average revenue in the Nielsen data per patent above the median of the distribution. Indeed, some firms may have activities outside the domain of consumer packaged goods and may obtain patents for these activities rather than for the set of products we observe. Focusing on firms with a high Nielsen revenue per patent may mitigate this potential issue and identify more reliably a set of innovative firms within consumer packaged goods. The table shows that the results are similar to the full sample. The outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers. Standard errors are clustered at the firm level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table C6: Innovator-Consumer Homophily across Industries in Finland, Excluding Agriculture

	Share of Industry Sales to Women		Industry Income Elasticity		Average Consumer Age, Sales-weighted	
	(1)	(2)	(4)	(5)	(8)	(9)
Female Patent Inventor	0.0488***					
	(0.0037)					
Female Entrepreneur		0.0401***				
		(0.0005)				
Patent Inventor's Log Parent Income			0.0269**			
			(0.0125)			
Entrepreneur's Log Parent Income				0.0119***		
				(0.0027)		
Patent Inventor Age					0.0537***	
					(0.0051)	
Entrepreneur Age						0.0253***
						(0.0013)
Country	Finland	Finland	Finland	Finland	Finland	Finland
Mean	0.5688	0.6369	1.279	1.300	0.5688	51.81
N industries	330	451	245	417	330	451
N individuals	9,592	274,785	3800	83,316	9,592	274,785

Notes: All regressions are run at the level of an individual innovator, with outcomes measured at the industry level. Standard errors are clustered at the individual level. This table is identical to Table 4 in the main text, except that we exclude agriculture. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C7: Innovator-Consumer Homophily across Industries, by Income Group Expenditure Shares

	Above 100k vs. below 30k			Above 60k vs. below 60k		
	(1)	(2)	(3)	(4)	(5)	(6)
Patent Inventor's Log Parent Income		0.0082**			0.0053*	
		(0.0034)			(0.0027)	
Entrepreneur's Log Parent Income	0.0101**		0.0371***	0.0076**		0.0247***
	(0.00428)		(0.0008)	(0.00299)		(0.0006)
Country	U.S	Finland	Finland	U.S	Finland	Finland
Mean	0.7053	0.6997	0.6809	0.6328	0.6241	0.6122
N industries	17	253	441	17	253	441
N individuals	275	3,812	99,189	275	3,812	99,189

Notes: This table reports industry-level homophily estimates for the United States and Finland. To compute the share of sales to households earning above \$100k ("high-income") or below \$30k ("low-income") depending on an entrepreneur's own family income background in the U.S., we proceed as follows. We take the regression coefficient in column (1) and use average parent income across the U.S. income distribution, equal to \$14,859 for entrepreneurs from the bottom 20%, and \$269,356 for the top 20% in 2021, according to the U.S. Census Bureau Historical Income Tables (see here, Table H-3). We obtain that the share of sales to high-income households increases by 4.15% relative to the baseline rate when an entrepreneur comes from a family from the top income quintile instead of the bottom ($= 0.0101 \times \log(269,356/14,859)/0.7053$). We can use this between-industry homophily estimate to extrapolate and estimate the overall homophily coefficient. Specifically, we assume that the relative magnitudes of gender homophily and income homophily are the same within and between industries; therefore by rescaling our overall gender homophily estimate (= 21.8) by the ratio of between-industry income and gender homophily estimates (= 4.15/3.9), we obtain that the overall income homophily coefficient for the United States is 23.8%. For Finland, we take the regression coefficient in column (3) and use average income across the Finnish disposable income distribution, equal to \$16,581 for the bottom 20% and \$84,547 for the top 20%, according to official statistics in 2013. Thus, the share of sales to high-income households increases by 3.85% relative to baseline when an entrepreneur comes from the top income quintile instead of the bottom ($= 0.0371 \times \log(84,547/16,581)/0.6809$). *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C8: Industry-Level Regressions for Innovator-Consumer Homophily, with Industry-Level Independent Variables

	Share of Industry Sales to Women			Industry Income Elasticity		Average Consumer Age, Sales-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Patent Inventor Fraction	0.677***	0.1032***						
	(0.133)	(0.0163)						
Female Entrepreneur Fraction			0.120***					
			(0.021)					
Patent Inventor's Log Parent Income				0.0812**				
				(0.0393)				
Entrepreneur's Log Parent Income					0.1377**			
					(0.0581)			
Patent Inventor Age						0.121	0.0486*	
						(0.105)	(0.0278)	
Entrepreneur Age								0.1097***
								(0.0357)
Country	U.S.	Finland	Finland	Finland	Finland	U.S.	Finland	Finland
Mean	0.593	0.5843	0.592	1.1478	1.1267	48.97	49.62	49.4101
N industries	325	476	476	253	441	323	342	476

Notes: This table reports industry-level homophily estimates for the United States and Finland. The independent variables are observed at the individual level but averaged to the industry level. This table is thus identical to Table 4 in the main text, except that the regressions are implemented at the industry level instead of the individual level. Since most of the variation in innovator covariates occurs within industries, where the outcome does not vary, the point estimates in this table tend to be larger than in Table 4. Standard errors are clustered at the industry level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table C9: Homophily Regression Coefficients – U.S.-Based Inventors

	Share of Industry Sales to Women	
	(1)	(2)
Female Patent Inventor	0.0233***	0.0218***
	(0.000293)	(0.000304)
Weighting	Unweighted	Log patents
N industries	325	325
N individuals	993,610	993,610

Notes: The table displays the estimates of the industry-level homophily regression, including patents with only U.S.-based inventors instead of all U.S. patents.

Table C10: Homophily, Dissimilarity, and Consumption Shares

Panel A – Nielsen CPG

	Female HH			Female Wt			Share of startup rev	Fem. Innov. Frac.	#Firms
	β	Dissimilarity	Fem. Rev. Frac.	β	Dissimilarity	Fem. Rev. Frac.			
Food	0.048 (0.030)	0.21	0.24	0.036 (0.018)	0.08	0.50	0.84	0.14	85
Health and Household	0.147 (0.032)	0.31	0.27	0.132 (0.017)	0.12	0.52	0.13	0.08	52
General Merchandise	-0.007 (0.043)	0.33	0.25	-0.012 (0.042)	0.13	0.50	0.03	0.21	45

Panel B – Phone Apps

	β	Dissimilarity	Fem. Usage Frac.	Share of Startup Usage	Female Innov. Frac.	#Firms
Lifestyle, Health and Leisure	0.097 (0.043)	0.31	0.54	0.45	0.12	795
Technology and Tools	0.112 (0.066)	0.18	0.56	0.42	0.066	532
Information and Commerce	0.053 (0.091)	0.25	0.55	0.13	0.067	352

Notes: These tables present gender homophily estimates in subsets of the product space, for consumer packaged goods (Panel A) and phone applications (Panel B). For each subset, we report the gender homophily coefficient (β), the consumption dissimilarity index between male and female consumers across all goods in the subset ("dissimilarity"), the fraction of revenue or usage from female consumers across both startup and other goods in the subset ("Fem. Rev. Frac." and "Fem. Usage Frac."), the share of startup revenue that the subset takes up, the female innovator fraction ("Fem. Innov. Frac.") across all firms selling in that subset (not weighted by number of founders and firms can be in multiple areas), and number of firms. For consumer packaged goods, the outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers.

Table C11: Innovator-Consumer Homophily for Consumer Packaged Goods, Weighted Regressions (Logarithm of Sales)

	Share of Sales to Women		Average Consumer Age, Sales-weighted
	(1)	(2)	(3)
Female Founder Fraction	0.0408*		
	(0.0212)		
Female Patent Inventor Fraction		0.0272*	
		(0.0153)	
Founder Age			0.127**
			(0.0532)
Product Module F.E.	Yes	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 1,094,229$	Startups, $N = 4,058$

Notes: This table is identical to Table 3 in the main text, except that all products are now weighted by the logarithm of sales. In columns (1) and (2), the outcome variable is the fraction of sales to households with a female head. The sample means are 0.256 in column (1) and 0.265 in column (2). In column (3), the outcome variable is the average age of consumers, where the average is obtained using sales weights. The level of observation is a product. The coefficients are very similar to the main text, showing that the results are not driven by "marginal" product innovations. Standard errors are clustered at the firm level. $^*p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table C12: Innovator-Consumer Homophily for Phone Applications, Weighted Regressions (Logarithm of Time Use)

	Female User Share			Founder State User Share	
	(1)	(2)	(3)	(4)	(5)
Female Founder Fraction	0.0997**	0.0829**	0.0772*		
	(0.0369)	(0.0357)	(0.0433)		
Female VC Fraction			0.152**		
			(0.0761)		
Founder State				0.0875***	0.0365***
				(0.0062)	(0.0079)
Fixed Effects	None	Subcategory	None	None	Subcategory
Sample Size	3211	3211	1391	43680	43392

Notes: The sample used in this table includes all phone applications for VC-backed startups. This table is identical to Table 2 in the main text, except that all applications are now weighted by the logarithm of time use. The specification thus gives more weight to applications that are more widely used. In columns (1) to (3), the outcome variable is the fraction of time usage of an app accounted for by female users. In columns (4) and (5), the outcome variable is the fraction of time usage of the app by users located in the same U.S. state as the founder of the app. The coefficients are very similar to the main text, showing that the results are not driven by "marginal" innovations. The level of observation is an app. Standard errors are clustered at the startup level. $^*p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table C13: Innovator-Consumer Homophily for Phone Applications, Raw Usage (Logarithm of Time Use)

	Female Usage Share	
	(1)	(2)
Female Founder Fraction	0.0903	0.0267
	(0.715)	(0.741)
Fixed Effects	None	Subcategory
Sample Size	$N = 3,211$	

Notes: The sample used in this table includes all phone applications for VC-backed startups. This table is identical to Table 2 in the main text, except that the outcome is log(1+female usage), rather than usage shares. The level of observation is an app. Standard errors are clustered at the startup level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C14: Innovator-Consumer Homophily across Industries, Weighted Regressions

	Share of Industry Sales to Women			Industry Income Elasticity			Average Consumer Age, Sales-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Patent Inventor	0.0209***	0.0478***							
	(0.000203)	(0.0042)							
Female Entrepreneur			0.0306***						
			(0.0003)						
Patent Inventor's Log Parent Income				0.0300**					
				(0.0145)					
Entrepreneur's Log Parent Income					0.0385**	0.1418***			
					(0.013013)	(0.0034)			
Patent Inventor Age							0.00313***	0.0579***	
							(0.000814)	(0.0057)	
Entrepreneur Age									0.0160***
									(0.0010)
Country	U.S.	Finland	Finland	Finland	U.S.	Finland	U.S.	Finland	Finland
N industries	325	342	476	253	17	441	323	342	476
N individuals	2,219,193	9,643	344,698	3,812	275	99,189	48,156	9,643	344,698

Notes: All regressions are run at the level of an individual innovator, with outcomes measured at the industry level. Regressions are weighted as follows: columns studying patent inventors (col. (1), (2), (4), (7) and (8)) use $\log(1+\text{patents}_i)$ as weights; columns studying entrepreneurship (cl. (3), (5), (6) and (9)) use $\log(1+\text{income}_i)$ as weights. The coefficients are similar to those reported in the main text, indicating that homophily between industries is not driven by marginal innovators. Standard errors are clustered at the individual level. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C15: Price versus founding team composition

	Log Price	
	(1)	(2)
Female Founder Fraction	0.281**	0.243*
	(0.129)	(0.144)
Constant	1.497***	1.503***
	(0.0336)	(0.0233)
Fixed Effects	Group	Module
Sample Size	4044	4054

Notes: This table documents the relationship between founder gender and log unit prices within the consumer packaged goods sample. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C16: Gender Homophily with Price Controls

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0485**	0.0379**	0.0530***	0.0416***	0.0513**	0.0401***
	(0.0209)	(0.0154)	(0.0202)	(0.0155)	(0.0204)	(0.0154)
Unit Price	-0.000770	-0.000862***				
	(0.000597)	(0.000315)				
Log Price			-0.0233***	-0.0204***	-0.0266	-0.0189
			(0.00825)	(0.00459)	(0.0224)	(0.0122)
Log Price^2					0.00931	0.00814**
					(0.00865)	(0.00406)
Log Price^3					-0.00310	-0.00302***
					(0.00226)	(0.000891)
Log Price^4					0.000253	0.000180
					(0.000474)	(0.000218)
Constant	0.257***	0.541***	0.285***	0.564***	0.279***	0.556***
	(0.00656)	(0.00410)	(0.0125)	(0.00697)	(0.0172)	(0.00902)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	4054	4054	4054	4054	4054	4054

Notes: This table presents homophily estimates with flexible price controls, using the consumer packaged goods sample. The outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C17: Nielsen – Adding Basic Controls

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0485**	0.0376**	0.0540***	0.0437***	0.0615	0.0802***
	(0.0212)	(0.0157)	(0.0195)	(0.0154)	(0.0398)	(0.0233)
Constant	0.247***	0.532***	0.286***	0.564***	0.289***	0.571***
	(0.00367)	(0.00297)	(0.0123)	(0.00689)	(0.0193)	(0.00989)
Controls						
Module FEs	x	x	x	x	x	x
Log(Price), Size, Multipack, Organic, etc.			x	x	x	x
Indicators for other characteristics					x	x
Fixed Effects	Module					
Sample Size	4015	4015	4015	4015	4015	4015

Notes: This table documents homophily after controlling for specific sets of product characteristics, to assess the role of horizontal product differentiation. The second set of controls ("Log(Price), Size...") includes log unit price, multipack indicator, product size, USDA organic indicator, and male/female targeted variables based on Nielsen Homescan data and the brand text approach in Moshary et al. (2023). We then include indicators for different values of the following variables in the Nielsen extra product attributes file: flavor code, form code, formula code, container code, salt content code, style code, type code, product code, variety code, strength code, scent code, and target skin condition code. The outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C18: Nielsen – Adding Controls with LASSO

Panel A – Startup Sample

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0455**	0.0399***	0.0510**	0.0434***	0.0391	0.0684**
	(0.0213)	(0.0144)	(0.0214)	(0.0146)	(0.0429)	(0.0273)
Available Controls (all with LASSO)						
Module FEs	x	x	x	x	x	x
Log(Price), Size, Organic, etc.			x	x	x	x
Indicators for other characteristics with LASSO					x	x
Sample Size	4015	4015	4015	4015	4015	4015

Panel B – Broader Sample (50% of all firms)

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Startup	-0.0409***	-0.00818**	-0.0370***	-0.00756**	-0.0289***	-0.00917*
	(0.00567)	(0.00380)	(0.00568)	(0.00381)	(0.00789)	(0.00510)
Startup ×Female Founder Fraction	0.0573***	0.0339***	0.0580***	0.0332***	0.0573**	0.0421***
	(0.0173)	(0.0121)	(0.0173)	(0.0120)	(0.0233)	(0.0141)
Available Controls						
Module FEs	x	x	x	x	x	x
Log(Price), Size, Organic, etc.			x	x	x	x
Indicators for other characteristics					x	x
Sample Size	250367	250367	250361	250361	250361	250361

Notes: This table documents homophily after controlling for specific sets of product characteristics, under a penalized regression (LASSO) approach. We provide results from our core startup sample and from adding a random sample of 50% of non-startup manufacturers in order to better populate the characteristics space. The second set of controls ("Log(Price), Size...") includes log unit price, multipack indicator, product size, USDA organic indicator, and male/female targeted variables based on Nielsen Homescan data and the brand text approach in Moshary et al. (2023). We then include indicators for different values of the following variables in the Nielsen extra product attributes file: flavor code, form code, formula code, container code, salt content code, style code, type code, product code, variety code, strength code, scent code, and target skin condition code. The outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers. In Panel B, "Startup" represents the difference between all-male founded startups and the overall module average and the interaction term represents the difference between female startups and male startups. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C19: Gender homophily within consumer subgroups, consumer packaged goods

Panel A – Female Weighted

	Share of Sales to Women					
	Education		Income		Age	
	College	Non-College	>= 60k	< 60k	45+	<45
Female Founder Fraction	0.0373**	0.0173	0.0434***	0.0166	0.00641	0.0557***
	(0.0146)	(0.0252)	(0.0133)	(0.0205)	(0.0191)	(0.0166)
Constant	0.526***	0.553***	0.502***	0.593***	0.545***	0.530***
	(0.00248)	(0.00368)	(0.00228)	(0.00406)	(0.00349)	(0.00267)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	3867	3469	3797	3558	3870	3369

Panel B – Single Households

	Share of Sales to Women					
	Education		Income		Age	
	College	Non-College	>= 60k	< 60k	45+	<45
Female Founder Fraction	0.0787**	0.0317	0.0555	0.0547	0.0232	0.103**
	(0.0370)	(0.0721)	(0.0506)	(0.0354)	(0.0522)	(0.0422)
Constant	0.673***	0.718***	0.638***	0.725***	0.686***	0.720***
	(0.00577)	(0.00801)	(0.00724)	(0.00626)	(0.00725)	(0.00670)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	2666	2388	2217	2772	2844	2014

Notes: This table investigates heterogeneity in gender homophily depending on characteristics of the customers, in the consumer packaged goods sample. Panel A uses weights within the households to compute female consumer fraction. Panel B focuses households with one member. The column labels note the demographic variable and the subgroup of consumers within which we compute the female consumer fraction (e.g., within the set of college-educated households). The number of data points varies across subgroups because some products do not have any purchasing by households from that group. Standard errors are clustered at the firm level. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table C20: Gender homophily within consumer subgroups, phone applications

	Female Usage Share					
	Education		Income		Age	
	College	Non-College	>= 75k	< 75k	45+	<45
Female Founder Fraction	0.144***	0.0320	0.128***	0.0453	0.0437	0.0981***
	(0.0420)	(0.0374)	(0.0402)	(0.0406)	(0.0425)	(0.0365)
Constant	0.504***	0.562***	0.482***	0.568***	0.546***	0.526***
	(0.0112)	(0.0122)	(0.00947)	(0.0129)	(0.0114)	(0.0118)
Fixed Effects	Subcategory	Subcategory	Subcategory	Subcategory	Subcategory	Subcategory
Sample Size	2,364	2,812	2,141	2,921	2,443	2,770

Notes: This table investigates heterogeneity in gender homophily depending on characteristics of the customers, in the sample of phone applications. The column labels note the demographic variable and the subgroup of consumers within which we compute the female consumer fraction (e.g., within the set of college-educated households). The number of data points varies across subgroups because some products do not have any purchasing by households from that group. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C21: Innovator-Consumer Homophily across Industries by Sector

	Patent Inventors				Entrepreneurs			
	(1) Manuf.	(2) Manuf.	(3) Other	(4) Other	(5) Manuf.	(6) Manuf.	(7) Other	(8) Other
	A. Outcome: Female Consumption Share							
Female	0.0332***	0.0035*	0.0540***	0.0076***	0.0680***	0.0127***	0.0254***	0.0071***
	(0.0039)	(0.0019)	(0.0045)	(0.0023)	(0.0017)	(0.0015)	(0.0003)	(0.0002)
Constant	0.5444***	0.5465***	0.6284***	0.6332***	0.5505***	0.5640***	0.6328***	0.6394***
	(0.0011)	(0.0007)	(0.0017)	(0.0008)	(0.0008)	(0.0006)	(0.0002)	(0.0001)
2-digit industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
N individuals	6945	6945	3082	3082	28147	28147	322613	322613
N industries	21	21	41	41	22	22	42	42
	B. Outcome: Income Elasticity							
Log Parent Income	0.0289**	0.0168*	0.0302	0.0030	0.0193**	0.0087*	0.1487***	0.0062***
	(0.0132)	(0.0099)	(0.0286)	(0.0059)	(0.0089)	(0.0050)	(0.0035)	(0.0012)
Constant	0.9496***	1.0839***	0.9634***	1.2643***	0.9186***	1.0324***	-0.4779***	1.0526***
	(0.1471)	(0.1101)	(0.3186)	(0.0669)	(0.0961)	(0.0541)	(0.0384)	(0.0134)
2-digit industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
N individuals	2769	2769	1194	1194	6675	6675	93829	93829
N industries	21	21	39	39	20	20	42	42

Notes: This table presents homophily estimates for different sectors in Finland. We present the results separately for patent inventors (Columns 1-4) and entrepreneurs (Columns 5-8) in manufacturing and non-manufacturing sectors. Each even column includes fixed effects for 2-digit industries. For gender homophily (Panel A), we observe consistently positive and statistically significant coefficients, indicating that female patent inventors and entrepreneurs cater more to female consumers within sectors and industries. The fixed effects estimates are considerably smaller but statistically significant, meaning that while gender homophily is prevalent also within 2-digit industries, variation across industries accounts for a significant fraction of it. For income homophily (Panel B), all coefficients are positive, with six out of ten being statistically significant at the 10 percent or lower risk level. The estimate for patent inventors in the non-manufacturing sector in Column 3 has low precision due to limited sample size (this sample includes only patent inventors in non-manufacturing sector whose parental income is observed). However, the point estimate of 0.0302 is almost equivalent to the baseline estimate of 0.0304 in the full sample in the main text. Standard errors are clustered at the individual level.
*$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C22: Innovator-Consumer Homophily across Industries by Innovator Type

	Patent Inventors			Entrepreneurs		
	(1)	(2)	(3)	(4)	(5)	(6)
		A. Outcome: Female Consumption Share				
Female	0.0447***	0.0510***	0.0456***	0.0312***	0.0262***	0.0305***
	(0.0083)	(0.0109)	(0.0044)	(0.0004)	(0.0010)	(0.0006)
Female ×College	0.0049			-0.0039***		
	(0.0092)			(0.0010)		
Female ×Age > 40		0.0084			-0.0006	
		(0.0068)			(0.0007)	
Female ×High-income			-0.0075			0.0106***
			(0.0120)			(0.0013)
Sample Size	9643	9643	3812	344698	344698	99189
		B. Outcome: Income Elasticity				
Log Parent Income	0.0213	0.0289**	0.0278**	0.1396***	0.1596***	0.1363***
	(0.0248)	(0.0129)	(0.0130)	(0.0037)	(0.0043)	(0.0034)
Log Parent Income ×College	0.0057			0.0160*		
	(0.0284)			(0.0089)		
Log Parent Income ×Age > 40		0.0214			0.0532***	
		(0.0351)			(0.0090)	
Log Parent Income ×Female			0.0142			-0.0452***
			(0.0463)			(0.0067)
Sample Size	3812	3812	3812	99189	99189	99189

Notes: This table estimates heterogeneity in homophily by innovator type in Finland. We present the results separately for patent inventors (Columns 1-3) and entrepreneurs (Columns 4-6). Panel A investigates heterogneity in gender homophily by whether the patent inventor/entrepreneur holds a college/university degree (Columns 1 and 4), is more than 40 years old (Columns 2 and 5), and has parental income above the median (Columns 3 and 6). Panel B investigates heterogneity in income homophily by the same characteristics, except in Columns 3 and 6 the interaction is with a binary indicator for female. Standard errors are clustered at the individual level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C23: Gender Homophily by Founding Team Composition

	Share of Sales to Women					
	All Startups			Multi-founder Startups		
	CPG (HH) (1)	CPG (Wt) (2)	Phone Apps (3)	CPG (HH) (4)	CPG (Wt) (5)	Phone Apps (6)
1 Female Founder	0.0283**	0.0260***	0.0870***	0.00515	0.0115	0.0788***
	(0.0142)	(0.00946)	(0.0184)	(0.0271)	(0.0176)	(0.0217)
2+ Female Founders	0.0168	-0.0113	-0.0155	0.0367	-0.0137	-0.0231
	(0.0553)	(0.0369)	(0.0516)	(0.0701)	(0.0454)	(0.0504)
Constant	0.251***	0.533***	0.528***	0.242***	0.530***	0.533***
	(0.00453)	(0.00303)	(0.0112)	(0.00824)	(0.00533)	(0.0122)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	4054	4054	3211	1595	1595	2057

Notes: This table investigates heterogeneity in homophily depending on team composition, using the consumer packaged goods on phone apps datasets. To examine whether one woman in the team is sufficient, we regress the market share of female customers ("HH" refers to measurement based on female head of household and "Wt" refers to measurement based on overall gender composition of the household) on dummies for the number of female founders, specifically whether the firm has one female founder or more than one. Columns (1) to (3) consider all startups, while Columns (4) to (6) focus on the subset of startups with more than one founder. For phone applications, Columns (3) and (6) show that it is sufficient to have at least one female founder for the application to be more tailored towards female customers. The other columns show that, for consumer packaged goods, the patterns appear to be driven by single founders. Indeed, in columns (4) and (5) the point estimates become much smaller in magnitude (although we cannot rule out substantial homophily given the size of the standard error). Standard errors are clustered at the firm level. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table C24: Homophily by Team Composition, Finland

	Patent Inventors		Entrepreneurs	
Outcome:	Female Share	Income Elasticity	Female Share	Income Elasticity
1 Female Inventor	0.03423***		0.0346***	
	(0.0090)		(0.0105)	
2+ Female Inventors	0.0458***		0.0488***	
	(0.0114)		(0.0114)	
1 Low-Income Inventor		-0.0157		-0.0221
		(0.0307)		(0.0338)
2+ Low-Income Inventors		-0.0579*		-0.0560*
		(0.0313)		(0.0313)
Constant	0.5396***	1.2504***	0.5366***	1.2670***
	(0.0029)	(0.0164)	(0.0035)	(0.0134)
N firms	2172	1060	1442	867

Notes: This table examines whether homophily varies with team composition in Finland. Inventor teams are defined based on firm-year cells. We find that, in line with the the U.S. results for founder teams, inventor teams with at least one female member cater more female-intensive consumer markets, on average. We find similar results for entrepreneurs. Moreover, unlike for U.S. founder teams, the degree of homophily is not tied to having at least one female inventor or entrepreneur in the team, as having additional female patent inventors or entrepreneurs in the team further strengthens it. We recover similar pattern for entrepreneurs from low-income backgrounds, with homophily increasing as we move from teams with a single low-income member to teams with multiple low-income members. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C25: Consumption Dissimilarity Indices

	Consumer Expenditure Survey	Consumer Packaged Goods	Phone Applications
	(1)	(2)	(3)
Male vs. female, single-gender households	0.14	0.320	0.242
Male vs. female, reference person gender	0.067	0.247	
Top vs. bottom income quintiles	0.34	0.335	0.354
Top vs. bottom income deciles	0.38	0.472	0.441

Notes: The table displays the estimates of dissimilarity indices between male and female consumers, as well as across the income distribution, for three datasets. For the consumer expenditure survey and the Nielsen dataset covering consumer packaged goods, we compute the dissimilarity index between the two consumer groups using expenditure shares and the most detailed products available in each dataset. For gender, we conduct the analysis either by focusing on single-gender households or by using the reference person gender for all households. For phone applications, we compute the dissimilarity index by using time shares, rather than expenditures. Since the use of phone applications is directly observed at the individual level, we do not draw a distinction between single-gender households and reference person gender in Column (3). Column (1) only takes into account differences in expenditure shares that arise between product categories, while Columns (2) and (3) account for the differences within categories. The between-category estimate in Column (1) ignores some of the relevant variation, while the within-category dissimilarity indices in Columns (2) and (3) may not be representative of other categories. For the analysis by gender, we take 0.24 as our baseline dissimilarity index, which we view as conservative since the dissimilarity index is much higher in the first row of Column (2). For income groups, the dissimilarity indices are very similar for the top and bottom income quintiles; we take 0.34 as our baseline value.

Table C26: Probability of Becoming a Top Innovator by Gender, Model vs. Data

Panel A – Model Predictions

	Probability of Top 10%		Probability of Top 5%	
	Men	Women	Men	Women
Wedge 1 (“Untapped Marie Curies”)	10.4%	7.4%	5.2%	3.7%
Wedge 2 (“Untapped Marginal Female Inventors”)	5.74%	39.0%	2.87%	19.5%
Wedge 3 (“Stifled Marie Curies”)	10.4%	7.4%	5.2%	3.7%

Panel B – Empirical Patterns

	Probability of Top 10%		Probability of Top 5%	
	Men	Women	Men	Women
Phone Apps	10.5%	5.7%	5.3%	3.4%
Nielsen CPG	11.3%	8.7%	6.7%	2.2%
Citation Weighted Patents	10.3%	8.1%	5.1%	4.0%

Notes: Using models with the three types of wedges described in the text, Panel A presents the fractions of male and female innovators above the top 10% or top 5% productivity threshold, using model parameter η_i for models with Types 1 or 2 wedges and $(1 - \hat{\tau}_{gi})\eta_i$ for the model with the Type 3 wedge. Panel B reports empirical patterns, examining the probability of being in the top 10% or top 5% of innovators for phone applications (ranking by time usage of apps), consumer packaged goods (ranking by sales), and patents (ranking by citations). All ranks are computed within starting cohorts (founding year for phone apps, five-year groups for consumer packaged goods, and year of first filing for patent inventors).

Table C27: Productivity of New/Exiting Inventors with Changes in Access vs. Market Size

		Avg. Productivity	
		Removing Access Barriers (1)	Equalizing Market Size (2)
Male Inventors, Male-Focused Sector	Enter	7.34	–
	Exit	–	7.578
Male Inventors, Female-Focused Sector	Enter	–	4.12
	Exit	5.43	–
Female Innovators, Male-Focused Sector	Enter	32.79	–
	Exit	–	7.56
Female Innovators, Female-Focused Sector	Enter	34.06	4.12
	Exit	5.43	–

Notes: This table presents additional statistics related to two of the counterfactuals discussed in the main text. For each counterfactual (removing access barriers in Column (1) and equalizing market size in Column (2)), we identify the individuals who enter and exit innovation in each sector between the status quo and the counterfactual scenario, and compute the average research productivity (parameter η_i) within each group. The model uses the Type 1 wedge.

Table C28: Features of the Baseline and Counterfactual Economies

	Baseline			Full Equal Access Counterfactual			Top 1% Equal Access Counterfactual		
	All	Sector 1	Sector 2	All	Sector 1	Sector 2	All	Sector 1	Sector 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fraction of innovators									
All	0.059	0.028	0.031	0.081	0.040	0.041	0.053	0.032	0.021
Men	0.102	0.054	0.048	0.080	0.058	0.022	0.087	0.059	0.028
Women	0.016	0.002	0.014	0.082	0.022	0.060	0.019	0.005	0.015
Average productivity of innovators									
All	0.329	0.470	0.202	0.404	0.426	0.384	0.523	0.488	0.574
Men	0.345	0.474	0.201	0.424	0.446	0.366	0.396	0.439	0.305
Women	0.227	0.374	0.205	0.385	0.371	0.390	1.100	1.106	1.098

Notes: The table displays the fraction of innovators and their average productivity in different scenarios. Columns (1) to (3) report the results for the baseline equilibrium. Columns (4) to (6) focus on the "full equal access" counterfactual, setting $\tau_i = 1$ for all women. Columns (7) to (9) describe the "top 1% equal access" scenario, with $\tau_i = 1$ for all women in the top 1% of the innovation productivity distribution.

Table C29: Baseline Parameters for the Analysis by Income Quintiles

Panel A: Parameters calibrated outside of the model			
Model Parameter	Parameter Definition	Source	Value
$\lvert\alpha_{Q1} - \alpha_{Q5}\rvert$	Expenditure dissimilarity index by income quintiles	Consumer Expenditure Survey, Nielsen data, phone applications data (cf. Table C25, third row)	0.34
ε	Elasticity of substitution between varieties	DellaVigna and Gentzkow (2019)	1.9
ρ	Discount rate, annual	Kaplan et al. (2018)	0.051
λ	Pareto parameter of innovators' productivity	Bell et al. (2019a)	1.26
$\frac{1-\delta}{\delta}$	Ratio of disposable income (after taxes and transfers) between top and bottom income quintiles	Congressional Budget Office	6.46

Panel B: Jointly estimated model parameters		
Model Parameter	Parameter Definition	Value
τ	Barrier to enter innovation careers	0.121
ϕ	Sectoral assignment	0.596
$\bar{\eta}$	Pareto scale parameter of of innovators' productivity	0.011

Panel C: Targeted moments and model fit		
Targeted Moment [Source]	Data	Model
Share of patent inventors in bottom vs. top parent income quintiles [Bell et al. (2019b)]	0.116	0.116
Income homophily regression coefficients [Table C7, column (1) and table notes]	0.238	0.238
Annual growth rate of labor productivity, 1990-2020 [Saint Louis Fed]	0.02	0.02

Notes: This table presents the baseline parameters of the growth model for the analysis by income quintiles. In Panel A, the model parameters are set directly to match the value observed in data or taken from the literature. In Panel B, the three parameters are estimated jointly to match the moments from the model with moments observed in the data, displayed in Panel C.

Table C30: Model Predictions: Growth Impacts of Observed Changes in Women's Representation in Innovation Careers from 1985 to 2014

	1985	2014
Female inventor share (data)	0.050	0.128
τ	0.033	0.105
Growth	1.88pp	2.00pp
Cost-of-Living Inequality	22.20pp	18.70pp

Notes: This table presents results from the backwards looking counterfactual. 2014 quantities are based on the core calibration. 1985 equilibrium numbers are computed under a value of τ that generates a female innovator fraction of 0.05.

Table C31: Clean Patents and Inventor Gender

	Clean Patent	
	(1)	(2)
Female Inventor Fraction	0.326***	
	(0.015)	
Any Female Inventor		0.212***
		(0.007)
N	1403	1401

Notes: This table reports the propensity to create "clean patents" by inventor gender, following Aghion et al. (2016) to classify clean vs. other energy patents. The level of observation is a patent, in the sample of energy patents. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C32: Models with Endogenous vs. Exogenous Gender Wage Gap

Panel A: Growth and Cost-of-Living Counterfactuals

	Model with Exogenous Gender Wage Gap	Model with Endogenous Gender Wage Gap
Δ Cost-of-Living Inequality, Baseline vs. Equal Access	-18.20pp	-16.31pp
Δ Labor Productivity Growth, Baseline vs. Equal Access	+1.44pp	+1.38pp

Panel B: Calibration Parameters

	Model with Exogenous Gender Wage Gap	Model with Endogenous Gender Wage Gap
ϕ	0.7321	0.4056
τ	0.1050	0.1163
η	0.0114	0.0109

Panel C: Male vs. Female Earnings Ratios

	Model with Exogenous Gender Wage Gap	Model with Endogenous Gender Wage Gap
Male-to-Female Earnings Ratio, Total Earnings		
Baseline	1.205	1.568
Equal Access Counterfactual	1.205	1.212
Share of Women in Innovation Sector		
Baseline	12.5 %	12.5 %
Equal Access Counterfactual	50 %	50 %
Earnings Ratios, Innovators vs. Production Workers		
Baseline	N/A	4.93
Equal Access Counterfactual	N/A	4.63
Male-to-Female Earnings Ratio for Production Workers		
Baseline	1.205	1.205
Equal Access Counterfactual	1.205	1.205
Male-to-Female Earnings Ratio for Innovators		
Baseline	N/A	1.235
Equal Access Counterfactual	N/A	1.220

Notes: Equilibrium market size ratio and counterfactual growth and inequality results under a model that allows for endogenous market size. All results are medians over 100 simulations. For the model with endogenous wage gap, we apply a wage penalty (83%) to women who pursue entrepreneurship and production work.