Spouses, Children, and Entrepreneurship*

João Galindo da Fonseca† Charles Berubé ‡

March 12, 2021

Abstract

We study the relationship between family and entrepreneurship decisions. We show that married individuals are less likely to start a business, but married entrepreneurs who still enter found larger and more productive firms. Family influences entrepreneurship decisions via three channels: (1) spouses work less hours the more profitable the business is, saving on the disutility of working for the household, (2) spouses work more in case of business failure, and (3) children increase the cost of failure. We use administrative matched owner-employer-employee-spouse Canadian data to estimate the specifications derived from our model.
(JEL-Codes : J23, J12, E24, E23)
Keywords : Labor Markets, Family Economics, Firm Dynamics, Entrepreneurship.

*We would like to thank Camille Hebert, David Green, Giovanni Gallipoli, Fabian Lange, Francesco Amodio, Thomas Lemieux, Nicole Fortin, Raphaël Godefroy, Leonie Baumann, Tiago Cavalcanti, Bernardo Guimarães and Rogério Santarrosa for feedback and discussions. Main corresponding author email : ja.galindo.da.fonseca@gmail.com
†Université de Montréal and CIREQ. E-mail: ja.galindo.da.fonseca@gmail.com
‡Innovation, Science and Economic Development Canada. E-mail: charles.berube@canada.ca

1 Introduction

We know there have been structural changes to household composition; a decline in marriage rates in the US from 62% in 1990 to 55% in 2020, and in Canada from 59.81% in 1990 to 44.9% in 2020, and a rise of one-person households in the US from 24.5% in 1989 to 28.4% in 2019, and in Canada from 22.8% in 1991 to 28.2% in 2016.[1] How does family affect the decision to start a company? How does it affect the type of firms created by individuals? To evaluate how changes in household composition affect job creation in the economy, we need to understand the interplay between firm formation, the composition of new firms, and marriage.[2] Despite the large literature on entrepreneurship and firm formation (e.g. Fonseca et al. (2001), Cagetti and De Nardi (2006, 2009), Hurst and Pugsley (2011), Poschke (2013, 2018), Beaudry et al. (2018) Sedláček (2020)), there is no research on how marriage affects both firm creation and firm size.

We find that a 1 percentage point increase in the marriage rate decreases the entry rate into entrepreneurship by 0.2 percentage points and increases average size of firms by 1.13% conditional on entry. Given the baseline entry rate in the data, this corresponds to a 20% decrease in the entry rate. We measure marriage as being in a couple, either legally married or cohabiting. Furthermore, we use our estimates to perform a bounding exercise to approximate the effect of changes in the marriage rate on firm productivity. Our results imply that a decrease in marriage rates is associated with a fall in average firm productivity.

To study the role of spouses on entrepreneurial outcomes, we propose a tractable model of endogenous entrepreneurship, marriage and labor supply. The model features heterogeneity in innate entrepreneurial ability and project-specific productivity. Individuals choose ex-ante whether to marry based on a taste for marriage, their innate entrepreneurial ability and the economy's overall productivity.

Main earners in the household draw business opportunities from an exogenous distribution. In equilibrium, main earners open a firm if the business opportunity productivity is above the household's optimal threshold. The main earner and spouse share their income, such that individual-level consumption depends on income of both members of the household.[3] The household chooses whether spouses work or not (associated with a fixed cost) and, if so, how many hours.

This framework delivers three channels via which being a couple influences the decision to start a firm.

[1]These numbers for the US are based on CPS data. Individuals who stated they had a common-law marriage were coded as "married".

[2]This presumes that firm formation is important for productivity and aggregate activity. Haltiwanger (2011) and Clementi and Palazzo (2016) study the contribution of firm creation to productivity.

[3]In the benchmark model, household income sharing is constant. However, as a robustness, we show our results are unchanged if we allow income sharing to depend on the gender ratio.

First, any increase in profits allows the spouse to work less hours, saving on working costs for the household (spousal sharing effect). This channel increases the marginal benefit of a more productive business. It makes entrepreneurship more attractive to married households who, as a result, become less selective on which business projects to implement. Second, if the business fails, the spouse works more hours (spousal insurance effect). Insurance from the spouse decreases the cost of failure, making married households less selective in which business projects to implement. Finally, a family might also include children. Children increase the cost of failure (offspring effect), pushing married households to be more selective on which business projects to implement.[4] As a result, the response of entry into entrepreneurship and firm size to marriage is ambiguous. To test for the relative strenght of these three channels we bring the model to the data.

We use full universe confidential Canadian tax data with information on firms and households from 2001 to 2013. We observe the entire universe of individual tax filers, linked to their employer firm, to any firm they might own and to their spouse. We observe all individuals including those without a spouse. For each individual we observe gender, marital status, age, and income. Individuals and firms are linked over time, allowing us to observe firm entry. For each firm we observe industry and number of employees. The use of Canadian data allows us to abstract from employer provided health insurance cosiderations present in the US. Second, to our knowledge, this is the first dataset to contain links between employees and employers, owners and firms, and individuals and their spouses. The linkage between spouses allows us to eliminate alternative mechanisms such as firms jointly owned by spouses or an entrepreneur employing their spouse. By verifying our results are not driven by these type of companies we can focus on a smaller number of mechanisms.

The model shows that both aggregate productivity and entrepreneurial ability impact marriage rates and entrepreneurship. To address this endogeneity, we use an instrument derived from our model: the gender ratio. We exploit variation in gender ratio changes across country of birth groups and cities over time. The intuition is that individuals in a group-city with a higher ratio of people from the other gender relative to theirs are more likely to find a potential partner.[5] Our instrument strongly predicts marriages rates. To obtain information on country of birth, the dataset is merged to immigration landing files since 1980.

Because of the gender gap in entrepreneurship (Tracy (2011), Brush et al. (2014) and Hebert (2020)),

[4]We abstract from any motive to save associated with children. See Abbott (2020) for a recent paper on the topic.
[5]Angrist (2002).

changes in the gender ratio could affect aggregate entry rates even without changes in marriage rates. To address this issue, for all our specifications, dependent variables are constructed using only women or only men. As a result, the entry rates and average sizes we consider are women specific and men specific rates, hence, not mechanically affected by the quantity of individuals from the other gender. Since we use specifications in first differences, all city and country of origin time-invariant effects are controlled for.

We use the model to derive our instrument, the conditions under which it provides consistent estimates, and implied restrictions in the data. In this sense, we follow the literature that puts together structural modelling and instrumental variable estimation using macro-labor models (Beaudry et al. (2012), Beaudry et al. (2018), Tschopp (2015) and Green et al. (2017)). Our identification strategy relates to the discussion in Kahn and Whited (2018) on the sufficient conditions for causation versus identification.

We find that a rise in the marriage rate is associated with a decrease in the entry rate into entrepreneurship and a increase in the average size of firms conditional on entry. These findings are consistent with our prediction that the effect of marriage on average size has an opposite sign to the effect on entry. Our results are consistent with the higher cost of failure among families with children being strong enough to dominate over the spousal insurance effect and the spousal sharing effect for all families.

Our findings hold qualitatively for both women and men but magnitudes are larger for women. A rise in marriage rates decreases entry rates into entrepreneurship of women by 3 times as much as that of men and increases average size by 6 times as much as that of men. Such a pattern speaks to the growing literature on the gender gap in entrepreneurship (e.g. Gottlieb et al. (2016), Scott and Shu (2017), Howell and Nanda (2019), Zandberg (2020), Ewens and Townsend (2020), Hebert (2020)) by showing that the response of entrepreneurship to marriage is different among genders.

Because our results hold qualitatively for both genders, they are inconsistent with explanations based on gender differences in risk aversion. In particular, if women are more risk averse, a couple: a man and a woman, would be more risk averse than a single man but less risk averse than a single woman. If this channel drove our results, we would expect marriage to decrease entry rates into entrepreneurship among men but to increase them among women. Instead, we find that for both women and men, higher marriage rates decrease entry rates into entrepreneurship.

Our results hold across several robustness checks. First, we might be worried that our results are driven by financial constraints, more easily overcome by two-person households. All else equal, such a mechanism would imply that married individuals are more likely to enter entrepreneurship. We find the exact opposite.

To further address this issue, we verify that our estimates are robust in sign, magnitude and significance to excluding high capital industries. Second, cultural norms might simultaneously determine gender ratios and entrepreneurship. To control for this possibility, we verify our findings are unchanged if we include dummies for each country of birth and year pair. Third, to rule out the possibility that our results are being driven by a correlation between city-specific shocks and the gender ratio, we check our findings are unchanged if we include dummies for each city and year pair. Magnitudes, sign, and significance are unchanged for these different specifications.

Fourth, to further distance our instrument from any endogenous migration choice of immigrants, we verify our results are robust to using only immigrants that arrived at age 15 or younger.

Finally, as further robustness, we consider an extension of the benchmark model that allows for the gender ratio to affect entrepreneurship directly. Given our objective of using the gender ratio as an instrument, this corresponds to a violation of the exclusion restriction. Then, using the structure of the model, we derive the correct specification to address the issue. In particular, we allow for the gender ratio to affect intra-household income sharing[6] , inducing changes in labor force participation of the spouse and, as a result, entrepreneurship of the main earner. We show that controlling for the labor force participation of the spouse controls for variation in intra-household income sharing, shutting down the direct effect of the gender ratio on entrepreneurship. In the absence of this direct effect, the exclusion restriction holds and we can use the gender ratio as an instrument for marriage rates.

Our paper connects the literature on the gender gap in entrepreneurship (Gottlieb et al. (2016), Scott and Shu (2017), Howell and Nanda (2019), Zandberg (2020), Ewens and Townsend (2020), Hebert (2020)) with that on spousal insurance (Hyslop (2001), Gallipoli and Turner (2009), Blundell et al. (2016), Wu and Krueger (2018) and Bellou and Cardia (2020)). The larger response of entrepreneurial outcomes to marriage for women is consistent with women receiving less spousal insurance from their spouses than what men receive from theirs. This happens when working hours are less responsive to spousal income for men than for women. The result highlights how assymmetries between genders in the labor market map to gender differences in entrepreneurship rates.

[6]This is consistent with work on the relation between gender ratios and intrahousehold resource allocation (Chiappori et al. (2002)).

2 Dataset

We use the Canadian Employer-Employee Dynamic Database (CEEDD) to shed light on an issue never before studied. The dataset links employees to firms and firms to their corresponding owners and individuals to their spouses across space and time from 2001 to 2013. This linkage is achieved by using individual tax information (T1 files, individual tax returns), with linked employer-employee information (T4 files) and firm ownership and structure information (T2 files). According to Canadian law, each employer must file a T4 file for each of her employees. The equivalent in the US is the W-2, Wage and Tax Statement. In this file, the employer identifies herself, the employee, and reports the employee's labor earnings. T2 forms, the Canadian Corporate Income Tax forms, contain the schedule 50 in which corporations must list any owner with at least 10% of ownership, allowing us to link each firm to individual entrepreneurs. The equivalent in the US to the schedule 50 of the T2 form is the schedule G of the 1120 form. Our dataset is an improvement relative to employer-employee firm population data from the US, which does not allow the researcher to identify the firm's owners or an individual's spouse. To our knowledge, there is no other dataset with employer-employee, owner-firm, individual-spouse linkages.

The linkage between individuals and their spouses allows us to directly observe whether individuals start their firm with their spouse and whether the spouse works for the firm. Because we can control for these channels empirically, we omit them from our model. We do this by verifying our main results are unchanged once we exclude companies started by both spouses and those for which individuals hire their spouses. Finally, given that the data is from Canada, we can abstract from any employer-provided health insurance considerations present in the US.

Because the data contains information on both firms and individuals, it is possible to disentangle the characteristics of the business owner and the firm. We concentrate on firms that contribute to job creation by hiring employees. This is done by focusing on employers instead of self-employed individuals.

Business owners are identified as individuals present in the schedule 50 files that have employees. Wage workers are identified as individuals who are not business owners and report a positive employment income on their T4. We use the information in the T1 files to obtain information on characteristics such as gender, age, and marital status. Marriage is defined as including both legal marriage and cohabitation, which we observe in the data. Finally, the dataset is also linked to immigration landing files dating back to 1980. This allows us to observe the country of birth of all individuals that arrived at 1980 or later in Canada.

The linkage between each firm and its corresponding owner is only available for privately owned incorporated firms. Incorporated firms with employees are the appropriate measure of firms to consider to study the interplay between entrepreneurship and the aggregate economy. First, they have two key characteristics that correspond closely to how economists typically think about firms: limited liability and separate legal identity. Chandler (1977) and Harris (2000) highlight how the incorporated business structure was created with the explicit goal of fostering investment in large, long gestation, innovative and risky activities. Furthermore, incorporated firms tend to be larger, more productive, and are more likely to contribute to aggregate employment (Glover and Short (2010)). Third, there is little transition that takes place from unincorporated to incorporated status (Levine and Rubinstein (2017)). The puzzle, that entrepreneurs earn less than they would have as wage workers, is no longer true once we consider incorporated business owners (Levine and Rubinstein (2017)). In this sense, incorporated businesses are closer to firms in traditional macro models. Our dataset represents an important improvement in that aspect; it allows us to focus on firms contributing to aggregate output and employment.

For the remainder of the paper, the empirical definition of an entrepreneur is an owner and founder of a privately owned incorporated firm with employees. We consider only the founders of firms to restrict ourselves to entrepreneurs, like those in our model, that start new ventures. Although interesting, the study of shares purchase in an already existing firm is left for future research. In our dataset, the average size of startups, firms at most 1 year old, is 6 employees. If we do not restrict on age, the average size is 13 employees. The US economy equivalent is an average of 6 employees for startups and an average of 23 employees for the entire firm population.

Grekou and Gueye (2021) document both women and men in Canada starting a firm are on average 45 years old and mostly already married (77.1% among women and 72.1% among men). Motivated by this pattern, we model marriage decisions and the birth of children as preceding the decision to open a firm.

Table 1: Startups Fraction by province/year in Canada

	N.L	P.E.I	N.S	N.B	QC	ON	MN	Sas.	Alb.	B.C
2002	17.6%	16%	14%	14.3%	12.2%	14.9%	14.9%	14.2%	16.9%	16.3%
2003	19.3%	15.7%	14.3%	14.7%	12%	15.2%	14.3%	14.1%	16.7%	16.4%
2004	20.3%	17%	15.1%	16.1%	12.9%	16.7%	15.8%	15.2%	18.4%	18.4%
2005	17.5%	15.9%	13.2%	15.2%	12.1%	16.1%	15%	14.8%	18.8%	18.1%
2006	16%	15.2%	12.9%	13.6%	11.7%	15.4%	14.3%	14.7%	19.1%	17.8%
2007	18.2%	16.4%	13%	14.1%	11.9%	17%	14.6%	16.2%	19.5%	18.1%
2008	17.7%	16.7%	12.7%	14.4%	11.7%	16%	15%	17.5%	17.8%	16.5%
2009	16.2%	16%	12.5%	13.5%	10.9%	14.9%	14%	15.8%	15.2%	14.5%
2010	18.4%	15.2%	12.5%	13.3%	10.9%	14.9%	13.9%	14.8%	15.4%	14.6%
2011	16.2%	14.6%	11.7%	12.6%	10.8%	14.4%	13.5%	15.1%	16.3.%	14.6%
2012	15.9%	16%	11.2%	12.2%	10.5%	13.9%	13.6%	15.5%	16.7%	14%
2013	14.5%	14.7%	10.9%	12%	10.3%	13.3%	13.1%	14.6%	15.4%	13.7%

Notes: Startup rates by province and time in Canada. N.L stands for Newfoundland and Labrador. P.E.I stands for Prince Edwards Island. N.B stands for New Brunswick. QC stands for Quebec. ON stands for Ontario. MN stands for Manitoba. Sas. stands for Saskatchewan. Alb. stands for Alberta. B.C stands for British Columbia.

To estimate the effect of marriage on entrepreneurship, we need enough variation in entrepreneurship. Our estimation is done using variation across cities, country of birth groups, and time. To get a sense of the variation in the data, Table 1 presents the share of startups across different Canadian provinces over time.

We see there is considerable variation across space and time in Canada in the share of startups. For the year 2010, the share goes from as high as 18.4% in Newfoundland and Labrador (N.L.) to as low as 10.9% in Quebec (QC). We also see a sizeable variation across time; the share goes from as high as 18.1% to as low as 13.7% for British Columbia (B.C). Furthermore, across all provinces we see a downward trend in the share of startups. Relative to the US, startup shares are higher. This indicates that young firms and the process of firm creation are equally or even more important to the Canadian economy.

3 Model

In this section, we go over the main model of the paper. The main objective is to derive the intuition for the different economic mechanisms, formally derive our empirical specifications, and obtain identification. We start by describing individual choices conditional on marriage and entrepreneurial ability, θ. Afterwards, we describe the endogenous decision to marry as a function of ability θ and a taste for marriage. Consistent with our empirical results, we abstract from the possibility of the spouse helping out in the business started by their partner. We also abstract from employer-provided health insurance considerations, given Canada's public health care system. Finally, given new entrants in Canada tend to be middle aged, we consider the decision to marry is done prior to the decision to open a firm.

3.1 Outcomes for two types of Households : Married and Unmarried

In the economy, there are two types of households. Unmarried households are composed of only one individual: the main earner. The second type of household is that of couples composed of a main earner and a spouse (hereafter, married households). There are no savings, households consume their income. We consider CRRA utility with a coefficient of risk aversion σ. There are no search frictions. All individuals searching for a job find one instantaneously. The flow utility derived from income I for the single person household is given by $U^s(I) \equiv I^{1-\sigma}/(1-\sigma)$.

For married households, there are two sources of income, the income of the main earner, I, and the income of the spouse, wh. The income of the spouse is a function of hours chosen by the household for the spouse to work, h, and their wage, w.

We model children by a quantity χ of the married household income that is neither consumed by the spouse or the main earner. This modeling choice keeps the model tractable while already allowing us to talk about the higher cost of failure for households with children. Of the remaining, let γ be the share of income that the spouse consumes and $(1-\gamma)$ the share of income that the main earner consumes. We consider there is a fraction ν of couples that have kids, $\chi = \chi_1$, and a fraction $1-\nu$ that have no kids, $\chi = \chi_2 = 0$.

Let each individual belong to a group g, that defines the relevant marriage pool for an individual. The household has a cost ϕh associated with the spouse working h hours. This disutility is paid for by the entire household. Finally, being in a couple is associated with an additional utility of φ. The parameter φ captures both the love component of being in a couple and the utility from having children. In particular for couples with children $\varphi = \varphi_1$ and for couples without children $\varphi = \varphi_2$. Then, the flow utility of a married household composed of a main earner that makes I income and a spouse that earns wh income is given by

$$\mu \frac{[\gamma(I + wh - \chi)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\gamma)(I + wh - \chi)]^{1-\sigma}}{1-\sigma} + \varphi - \phi h \tag{1}$$

where μ is the exogenous weight placed on the utility of the spouse. The standard collective, cooperative, household model where households choose the share of income each member consumes to maximize total household utility,[7] is captured by the particular case where $\gamma = (\frac{\mu^{\frac{1}{\sigma}}}{(1-\mu)^{\frac{1}{\sigma}}+\mu^{\frac{1}{\sigma}}})$.[8] In Section J of the Appendix, we relax this assumption of constant γ and allow for it to be a function of the gender ratio. Such an extension allows us to capture some aspects of non-cooperative households in reduced form.

The married household solves the static problem of how many hours should the spouse work. Proposition 1 states that conditional on working, hours worked by the spouse decreases with main earner income, I, and increases with the wage of the spouse, w.

Proposition 1 *Let h^* denote the optimal choice of how many hours the spouse should work conditional on working. Then,*

$$\frac{\partial h^*}{\partial I} \leq 0 \quad \textit{and} \quad \frac{\partial h^*}{\partial w}|_{\Delta I=0} \geq 0. \tag{2}$$

Now, consider as well that the household must decide ex-ante whether the spouse works or not. This decision is made before knowing the income I of the husband and is irreversible. The choice to have the spouse work is associated with a one time cost of κ for the household. This modelling choice makes the model tractable without changing the main intuition. Proposition 2 states how the decision to work or not for the spouse is characterized by a threshold rule for their cost of working κ.[9]

Proposition 2 *Let κ^* be the value of κ that makes the household indifferent between choosing the spouse to work or not. Then, the spouse works if $\kappa \leq \kappa^*$.*

Each main earner has innate entrepreneurial ability of θ, where $1 \leq \theta < \infty$ and $\theta \sim G(\theta|g)$. Let the productivity of the firm be a function of an economy-wide productivity component y. The productivity component y is the same for all entrepreneurs in the same economy, while θ varies across entrepreneurs. Finally, let each business project started by an entrepreneur be characterized by a productivity z, where $0 \leq z < \infty$. Firm production is given by $y\theta e^z n^\alpha$ where $0 < \alpha < 1$ and n is number of employees. Throughout the life of the firm, z is fixed. The entrepreneur takes wage w and firm productivity components z, θ, and y as given and chooses how many individuals n to hire. Proposition 3 below states that flow profits, $\pi(z)$, and a firm's number of employees, n, are increasing in y, z, and θ.

[7] See Browning et al. (2006)

[8] See Section I of the Appendix for proof.

[9] See the Proof of Proposition 2 in Appendix Section E for the expressions for the probability the spouse works $Prob(work)$ and of the threshold κ^*.

Proposition 3 *Let $\pi(z)$ denote flow profits of the firm and n denote the optimal choice of how many individual to hire then*

$$\frac{\partial \pi(z)}{\partial y} > 0, \frac{\partial \pi(z)}{\partial \theta} > 0, \frac{\partial \pi(z)}{\partial z} > 0, \frac{\partial n(z)}{\partial y} > 0, \frac{\partial n(z)}{\partial \theta} > 0, \frac{\partial n(z)}{\partial z} > 0. \tag{3}$$

For the remainder of this section, we omit the dependence of value functions on θ to make the notation lighter. Households discount the future according to parameter r. With probability λ, the firm fails, and the main earner is forced to shut down the firm. Let $J^u(z)$ represent the value of being an entrepreneur without a spouse. Let B^u represent the value of failing a business without a spouse, then,

$$rJ^u(z) = \frac{\pi(z)^{1-\sigma}}{1-\sigma} + \lambda(B^u - J^u(z)). \tag{4}$$

For married households, the main earner gets a share $(1-\gamma)$ of total household income, and the spouse gets a share γ. Total household income is composed of the income of the entrepreneur, $\pi(z)$, and that of the spouse wh^*, net of children expenditures χ. Let $J^m(z,\kappa)$ represent the value function for a household composed of an entrepreneur running a project of quality z and a spouse that has working cost κ. Let $B^m(\kappa)$ represent the value function for a household composed of a failed entrepreneur with a spouse of working cost κ. Then, for $\kappa \leq \kappa^*$, the spouse works and the value function of the household, $J^m(z,\kappa)$, is

$$\begin{aligned} rJ^m(z,\kappa) = \mu\frac{[\gamma(\pi(z)+wh^*-\chi)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\gamma)(\pi(z)+wh^*-\chi)]^{1-\sigma}}{1-\sigma} \\ + \varphi - \phi h^* + \lambda(B^w(\kappa) - J^m(z,\kappa)). \end{aligned} \tag{5}$$

The difference in the value of being an entrepreneur between the two households depends on their flow utility differences. For single person households the only source of income is that of the main earner. The married household pools the income of the main earner and the income of the spouse. The household income of an unmarried entrepreneur is increasing in z.[10] On the other hand, for the married household as profits, $\pi(z)$, increase in z, hours worked by the spouse, h, weakly decrease. The decrease in hours worked by the spouse increases household flow utility by saving on the costs of working for the spouse, but decreases total household income. We call this the spousal sharing effect. Proposition 9 in the Appendix shows there exists

[10]Recall that profits, $\pi(z)$, are increasing in z.

a $\hat{\phi}$, such that $\forall \phi > \hat{\phi}$, the saved cost of working dominates over the effect of lower household income. From hereafter, we focus on the cases where $\phi > \hat{\phi}$. The spousal sharing effect increases the benefits to entrepreneurship for married households relative to unmarried households, by allowing the household to decrease working hours of the spouse. An equivalent interpretation is that larger profits by the firm allows the spouse to increase their hours of leisure. Since z is fixed over the life of the firm, the spousal sharing effect does not change the risk for an operating entrepreneur. Instead, it increases the value of becoming an entrepreneur, $J^m(z, \kappa)$. This channel is being driven by the fact that the household internalizes the cost of the spouse working. The motivation behind this assumption is that both members of the household value each other's utility.

When $\kappa > \kappa^*$, the spouse does not work.[11] Note that when the spouse does not work, it is not obvious if the married household gets less or more flow utility than the unmarried household. Crucially this depends on the size of the extra utility from being in a couple, φ. However, it is important to note that φ does not affect the decision to start a firm. For each household, what matters is the difference in value functions for entrepreneurship and wage work, which for the married household both contain φ.

Once the business fails, the main earner transitions to employment but is forced to pay a cost c and is not allowed to enter entrepreneurship. Hence, the main earner's flow income is $w - c$. The main earner exits this state of failed entrepreneur with probability p. Let W^u denote the value function for a one person household where the main earner is employed and did not recently fail a business. Then, the value for the one person household with a main earner that failed entrepreneurship, B^u, is

$$rB^u = \frac{(w - c)^{1-\sigma}}{1 - \sigma} + p(W^u - B^u). \tag{6}$$

For married households, the main earner consumes a share $(1 - \gamma)$ of their income $w - c$ and of the income of their spouse, wh^* net of child expenditures χ. The remaining γ is consumed by the spouse. Let $W^m(\kappa)$ represent the value function for married household where the main earner is working and the spouse has a cost of working κ. Then, for $\kappa \leq \kappa^*$, the spouse works, and the value function for a married household with

[11] See Section A of the Appendix for the explicit expression for the value function in this case.

a failed entrepreneur, $B^m(\kappa)$, is given by

$$rB^m(\kappa) = \mu\frac{[\gamma(w-c+wh^*-\chi)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\gamma)(w-c+wh^*-\chi)]^{1-\sigma}}{1-\sigma} + \varphi - \phi h^* + p(W^m(\kappa)-B^m(\kappa)). \quad (7)$$

The difference in the value of failing entrepreneurship between the two groups depends on differences in flow utility. While for unmarried individuals, income falls by the cost of bankruptcy, c, married households benefit from the income of their spouse, conditional on the spouse working. When the business of the main earner fails, spouses weakly increase working hours, h, partially offsetting the fall in household income. We call this effect the spousal insurance effect. It compresses the costs of business failure for married households. Proposition 4 shows that, all else equal, the spousal insurance effect implies the flow utility of married households after business failure is larger than the flow utility of single person households after business failure.

Proposition 4 *Let $u_M^B(\chi=0)$ denote the flow utility of a married household, with no children ($\chi=0$), for which the main earner failed their business. Let u_S^B denote the flow utility of a one person household for which the main earner failed their business. Then, there exists a value $\hat{c}$ such that for all $c>\hat{c}$:*

$$u_M^B(\chi=0) > u_S^B. \quad (8)$$

Proposition 4 makes clear that in the absence of children, the flow utility upon business failure of married households is higher than the equivalent for single person households. The presence of children makes the relative ranking of these flow utilities less clear. Intuitively, having children, $\chi>0$, increase further the utility cost of business failure. Proposition 5 shows that the drop in flow utility of married households upon business failure is increasing in child expenditures χ.

Proposition 5 *Let u_M^B denote the flow utility of a married household for which the main earner failed their business. Then,*

$$\frac{\partial u_M^B}{\partial c} < 0 \quad \textit{and} \quad \frac{\partial^2 u_M^B}{\partial c \partial \chi} < 0. \quad (9)$$

Intuitively, children increase the marginal utility of consumption upon failure, raising the cost of business failure, regardless of whether the spouse is working or not. This is the offspring effect.

The last value functions are those for households when the main earner is working. Recall $W^m(\kappa)$ and W^u are the value functions associated with married and unmarried households with the main earner working. For both types of households, at rate ψ, a business project z is drawn from the exogenous distribution $F(z)$. Households choose optimally which projects to implement by comparing the value of opening a firm ($J^m(z,\kappa)$ if married and $J^u(z)$ if unmarried) to the value of being employed ($W^m(\kappa)$ if married and W^u if unmarried). Let $\underline{z}^u$ represent the firm productivity that makes the unmarried household indifferent between opening a firm and continuing to work. Then, $J^u(\underline{z}^u) = W^u$. It follows that the value of an unmarried household composed of a working main earner is

$$rW^u = \frac{w^{1-\sigma}}{1-\sigma} + \psi \int_{\underline{z}^u} (J^u(z) - W^u) dF(z). \tag{10}$$

Let $\underline{z}^m(\kappa)$ represent the firm productivity that makes the married household indifferent between allocating the main earner to entrepreneurship or to employment. Then, $J^m(\underline{z}^m(\kappa), \kappa) = W^m(\kappa)$.

For married households, the spouse consumes a share γ of the main earner income, w, and of their own income, wh^* net of child expenditures χ. The remaining $(1-\gamma)$ is consumed by the main earner. Let $W^m(\kappa)$ be the value function of the married household with main earner working and spouse with cost of working κ. Then, for $\kappa \leq \kappa^*$, the spouse works and the value function of a married household with main earner working is

$$\begin{aligned} rW^m(\kappa) = \mu \frac{[\gamma(w + wh^* - \chi)]^{1-\sigma}}{1-\sigma} + (1-\mu) \frac{[(1-\gamma)(w + wh^* - \chi)]^{1-\sigma}}{1-\sigma} \\ + \varphi - \phi h^* + \psi \int_{\underline{z}^m(\kappa)} (J^m(z,\kappa) - W^m(\kappa)) dF(z). \end{aligned} \tag{11}$$

If ($\kappa > \kappa^*$), then the spouse does not work. Readers interested in the explicit expression for the value function in this case are referred to Section A of the Appendix.

In total, we have three channels via which family affects entrepreneurship: (1) the spousal sharing effect, (2) the spousal insurance effect and (3) the offspring effect. All three affect entrepreneurial outcomes by changing the selection threshold of each group ($\underline{z}^m$ and $\underline{z}^u$). Proposition 6 highlights how these thresholds determine the relative entry into entrepreneurship and firm size of married versus unmarried households.

Proposition 6 *Let ER^m and ER^u be the entry rate into entrepreneurship of married and unmarried individuals. Let $E[s]^m$ and $E[s]^u$ be the average size of firms owned by married and unmarried individuals.*

Then, conditional on θ

$$\underline{z}^m \geq \underline{z}^u \Leftrightarrow (ER^m \leq ER^u \quad and \quad E[s]^m \geq E[s]^m) \tag{12}$$

$$\underline{z}^m \leq \underline{z}^u \Leftrightarrow (ER^m \geq ER^u \quad and \quad E[s]^m \leq E[s]^m) \tag{13}$$

The spousal sharing effect increases the benefits of entrepreneurship for married households. Higher profits allow households to decrease working hours of the spouse, saving on the disutility of working for the household. All else equal, this pushes married households to be less selective on which business projects to implement (higher $\underline{z}^m$) relative to unmarried individuals ($\underline{z}^u$). The result is a higher entry rate, lower average productivity, and lower average size among firms created by married individuals.

The spousal insurance effect decreases the cost of business failure for married households. The drop in income of the main earner upon failing their business is compensated by a rise in working hours by the spouse. All else equal, this pushes married households to be less selective on which business projects to implement (lower $\underline{z}^m$) relative to unmarried individuals ($\underline{z}^u$). The result is a higher entry rate, lower average productivity, and smaller average size among firms created by married households.

The offspring effect increases the cost of business failure. It follows that children induce married households to be more selective on which business projects to implement (higher $\underline{z}^m$) relative to unmarried individuals ($\underline{z}^u$). As a consequence, all else equal, children lower the entry rate and raise average productivity and size for firms created by married households.

The model implies an ambiguous response of entry rates and average firm size to changes in marriage rates; however, it predicts that, conditional on ability, the effect of marriage on entry must be the opposite of its effect on firm size. Intuitively, once we control for entrepreneurial ability θ, we are identifying changes in average size due to changes in business project selection upon entry. It turns out this restriction holds in the data. On the other hand, without conditioning in ability there is no reason to believe that the effect of marriage has opposite signs regarding entry and firm size.

The prediction that average firm size and entry into entrepreneurship move in opposite direction is a feature in models where both are simultaneously determined by one factor. The selection in this factor affects entry and firm size in opposite directions. In our model this common factor is business project productivity. By controlling for ability, this is the only factor left affecting entry and firm size. Alternatively, in a model where entry and firm size are determined by separate factors or more than one factor affects both, this opposite relationship need not hold. In this sense, this testable prediction helps us rule out models in which, after controlling for ability, there are still multiple factors affecting entry and firm size.

Finally, Proposition 10 in Section C of the Appendix shows average firm productivity increases with marriage rates if married households are more selective on projects ($\underline{z}^m > \underline{z}^u$) and decreasing otherwise.

To better understand the sources of endogeneity when bringing the model to the data, we now consider the endogenous decision to marry.

3.2 Endogenous decision to marry

Up until now, we haven taken the marital status of an individual as given. In this section, we formalize the decision of individuals to marry or not. The objective is not to provide a detailed theory of the formation and dissolution of marriage. Rather, we use this formalization to inform us how entrepreneurial ability θ and economy-wide productivity y affect both marriage and entrepreneurial outcomes. This is an important point given our desire to estimate the effect of marriage on entrepreneurship.

We consider a tractable form of endogenous marriage formation. Ex-ante, individuals choose to marry based on the expected value of being married and unmarried. In particular, individuals make this choice under the veil of ignorance before knowing whether they enter or not entrepreneurship. Assume individuals cannot direct their search towards spouses of a particular cost of working, κ. They weigh each state by the equilibrium measure of individuals of the same entrepreneurial ability as themselves in each state. Furthermore, recall that each individual belongs to a group g. Each group g has an idiosyncratic utility value of $v_g \sim H(v)$ associated with being married. Once married, the couple draws a cost for the spouse to work, κ. The household pays this cost if and only if $\kappa \leq \kappa_g^*$. Next, a fraction ν of couples have children.

Individuals search for a partner to marry. The probability of meeting someone depends on the gender ratio ϑ_g, of their particular group in the their economy.[12] Let $q(\vartheta_g)$ be the probability an individual finds a potential partner, with $q'(\vartheta_g) > 0$. Upon meeting the potential partner, the individual makes the decision to marry or not based on the expected value of being married, $V^M(\theta, g) + v_g$ versus unmarried, $V^U(\theta, g)$. Then, the probability of marriage $Pr(M = 1)$ is given by Proposition 7.

[12] In our empirical estimation we consider the ratio of women relative to men when looking at outcomes for men and the ratio of men relative to women when looking at outcomes for women.

Proposition 7

$$Pr(M=1) = q(\vartheta)(1 - H(V^U(\theta, g) - V^M(\theta, g))). \tag{14}$$

Readers interested in the explicit expressions for $V^M(\theta, g)$ and $V^U(\theta, g)$ are referered to the proof of Proposition 7. Taking a first order Taylor approximation for the value functions gives us $Pr(M = 1)$ as a linear function of individual entrepreneurial ability, θ, economy specific productivity, y, the gender ratio, ϑ_g, and the probability of spouses working, $Prob(work)_g$.

Corollary 7.1 *A first order linear approximation of the probability of marriage $Pr(M = 1)$ gives*

$$Pr(M=1) = C_0 + C_1\theta + C_2 y + C_3\vartheta_g + C_4 Prob(work)_g. \tag{15}$$

A higher entrepreneurial ability, θ, increases the incentive to start a firm for an individual, making the effect of marriage on entrepreneurship more important for these individuals. Intuitively, entrepreneurial ability, θ, affects both the decision to start a firm conditional on marital status as well as the decision to marry. Similarly, when economy-wide productivity, y, is high, individuals are more prone to start a firm. This makes the differences in the value of entrepreneurship between married and unmarried all the more important. As a result, a change in y affects both the decision to start a firm and the decision to marry. It follows that a naive regression of entry or firm size on marriage suffers from endogeneity due to both y and θ. To overcome this endogeneity problem, we need an instrument that captures variation in $Pr(M = 1)$ uncorrelated to the decision to start a firm. From equation (15) we see that a natural candidate is the gender ratio, ϑ_g.

4 Empirical Strategy

This section goes over the main empirical strategy and specifications used to bring the model to the data. Our main objective is to verify which of the three effects: the spousal sharing effect, the spousal insurance effect or the offspring effect is the strongest. We also want to test the prediction that, once we control for ability, the effect of marriage on the entry rate into entrepreneurship must have the opposite sign to it's effect on the average size of firms. This restriction comes directly from our selection mechanism in which the entry decision of entrepreneurs is driving firm heterogeneity.

Our instrument is at the city c, country of birth group g and year t level. For the sake of transparency when reporting empirical results, we aggregate all variables to the same level as our instrument. We include Canadian born as a group but our results are robust to using only foreign born. Consider each city as a local economy described by our model in the previous section. Our approach is to linearize the model, deriving formally our empirical specifications. Define $ER_{c,g,t}$ as the entry rate into entrepreneurship: the share of individuals in city c, country of birth g that were not entrepreneurs in year $t-1$ that opened a firm at year t. Let $log(SY_{c,g,t})$ be the log of average firm size for a particular city c, country of birth g, year t. Proposition 8 shows how $ER_{c,g,t}$ and $log(SY_{c,g,t})$ are a function of the marriage rate among individuals of that same city c, group g, year t: $MR_{c,g,t}$.

Proposition 8 *Suppose there exists a large number of economies c all of which are characterized by the model described in the previous subsection. Let $G_c(\theta|g)$ be the distribution of innate ability θ for group g, in economy c. Let $Prob(work)_{c,g,t}$ be the probability that spouses from group g in local economy c, year t work. Finally, let the entry rate into entrepreneurship ER have measurement error, $\varepsilon_{c,g,t,1}$, characterized by its first differences, $\Delta\varepsilon_{c,g,t,1}$, being i.i.d and $E[\Delta\varepsilon_{c,g,t,1}] = 0$ and the average size of firms SY have measurement error, $\varepsilon_{c,g,t,2}$, characterized by its first differences, $\Delta\varepsilon_{c,g,t,2}$, being i.i.d and $E[\Delta\varepsilon_{c,g,t,2}] = 0$. Then, the entry rate into entrepreneurship ER and the average size of firms SY in each of these c economies for a group g can be written as*

$$ER_{c,g,t} = \beta_{0,1} + \beta_{1,1}MR_{c,g,t} + \beta_{2,1}Prob(work)_{c,g,t} + \beta_{3,1}y_{c,t} + \beta_{4,1}\int \theta dG_c(\theta|g) + \varepsilon_{c,g,t,1}. \quad (16)$$

and

$$log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}MR_{c,g,t} + \beta_{2,2}Prob(work)_{c,g,t} + \beta_{3,2}y_{c,t} + \beta_{4,2}\int \theta dG_c(\theta|g) + \varepsilon_{c,g,t,2}, \quad (17)$$

where $\int \theta dG_c(\theta|g)$ is the average entrepreneurial ability in city c among individuals of group g. Proposition 8 conveys that any instrument for marriage rates $MR_{c,g,t}$ must be independant of variation in $y_{c,t}$ and $\int \theta dG_c(\theta|g)$. Note that $Prob(Work)_{c,g,t}$ is the labor force participation of spouses. This leaves only $y_{c,t}$,

$\int \theta dG_c(\theta|g)$ and $\varepsilon_{c,g,t}$ as unobserved error terms. Next, we take first differences to obtain

$$\Delta ER_{c,g,t} = \beta_{1,1}\Delta MR_{c,g,t} + \beta_{2,1}\Delta Prob(Work)_{c,g,t} + \beta_{3,1}\Delta y_{c,t} + \Delta\varepsilon_{c,g,t,1}. \tag{18}$$

and

$$\Delta log(SY)_{c,g,t} = \beta_{1,2}\Delta MR_{c,g,t} + \zeta_{3,1}\Delta Prob(Work)_{c,g,t} + \beta_{3,2}\Delta y_{c,t} + \Delta\varepsilon_{c,g,t,2}. \tag{19}$$

Hence, after taking first differences across c and g the term $\int \theta dG_c(\theta|g)$ disappears. By using first differences we control for entrepreneurial ability, θ. Given our equations above, all we need is an instrument for the change in marriage rates, $\Delta MR_{c,g,t}$, which is uncorrelated to changes in economy specific productivity shocks, $\Delta y_{c,t}$. To obtain an expression for marriage rates $\Delta MR_{c,g,t}$ we can aggregate our linear equation for the probability of marriage $Pr(M = 1)$, in Corollary 15, to the city, country of origin and year level (c, g, t). This is done in Corollary 8.1.

Corollary 8.1 *Suppose there exists a large number of economies c all of which are characterized by the model described in this section. Let $G_c(\theta|g)$ be the distribution of entrepreneurial ability θ conditional on being from group g in economy c. Then, the aggregate marriage rate $MR_{c,g,t}$ for group g in economy c at time t can be written as*

$$MR_{c,g,t} = C_0 + C_1 \int \theta dG_c(\theta|g) + C_2 y_{c,t} + C_3\vartheta_{c,g,t} + C_4 log(Prob(work))_{c,g,t}. \tag{20}$$

Corollary 8.1 states formally that the marriage rate for a particular group g of an economy c and year t is a function of the average entrepreneurial ability, $\int \theta dG_c(\theta|g)$, the economy-specific shock, $y_{c,t}$, and the gender ratio for that group g in that economy c, $\vartheta_{c,g,t}$.

Taking first differences, we get

$$\Delta MR_{c,g,t} = C_2\Delta y_{c,t} + C_3\Delta\vartheta_{c,g,t} + C_4\Delta Prob(work)_{c,g,t}. \tag{21}$$

The expression for $\Delta MR_{c,g,t}$ highlights how the gender ratio $\Delta\vartheta_{c,g,t}$ directly affects $\Delta MR_{c,g,t}$. Furthermore, from equations 18 and 19, we know $\Delta\vartheta_{c,g,t}$ has no impact on our outcomes of interest, $\Delta ER_{c,g,t}$ and $\Delta SY_{c,g,t}$, other than via $\Delta MR_{c,g,t}$. These two conditions make $\Delta\vartheta_{c,g,t}$ the ideal candidate as an instrument for $\Delta MR_{c,g,t}$. We adopt this approach, such that our instrument $IV_{c,g,t}$ for $\Delta MR_{c,g,t}$ is defined as

$$IV_{c,g,t} = \Delta\vartheta_{c,g,t}. \tag{22}$$

The key restriction that allows for this empirical strategy is that the gender ratio $\Delta\vartheta_{c,g,t}$ has no effect on ER and SY other than via changes in marriage rates. In Section J of the Appendix, we relax this assumption by allowing within household income sharing to depend on the gender ratio, and show how identification is still possible if we include the appropriate controls. To summarize our two main specifications are

$$\Delta ER_{c,g,t} = \beta_{1,1}\Delta MR_{c,g,t} + \beta_{2,1}\Delta Prob(Work)_{c,g,t} + \beta_{3,1}\Delta y_{c,t} + \Delta\varepsilon_{c,g,t,1}. \tag{23}$$

and

$$\Delta log(SY)_{c,g,t} = \beta_{1,2}\Delta MR_{c,g,t} + \zeta_{3,1}\Delta Prob(Work)_{c,g,t} + \beta_{3,2}\Delta y_{c,t} + \Delta\varepsilon_{c,g,t,2}. \tag{24}$$

where $\Delta MR_{c,g,t}$ is instrumented by $\Delta\vartheta_{c,g,t}$. Recall that we have the prediction from the model that

- $\beta_{1,1} > 0$ if and only if $\beta_{1,2} < 0$,
- $\beta_{1,1} < 0$ if and only if $\beta_{1,2} > 0$.

5 Measurement and Empirical Results

5.1 Measurement

To test the model in the data, we need to assign the status of primary earner to one individual in each couple. The assignment must be based on a characteristic exogenous to individual choices to avoid it drives selection into entrepreneurship. Recall that in the model, an important difference between primary earner and spouse is the opportunity to become an entrepreneur. As discussed by Tracy (2011), Brush et al. (2014) and Hebert (2020), roughly 85% to 90% of startups are started by men. Consistent with this empirical pattern, we assign primary earner status to the male in the couple.[13]

An alternative would have been to assign primary earner status in each couple randomly. However, the instrument must be changed accordingly for each gender. The instrument for marriage of women is the ratio

[13] We find that for our results, it is irrelevant which gender we assign the status of primary earner. We verify the results are robust to considering women as primary earners.

of men to women and the instrument for marriage of men is the ratio of women to men. Hence, it is not possible to include both genders in one regression. As a result, the best way to verify our results are robust to the assumption that only one household member can open a firm, is to verify our results hold for women. This is the approach we take. We find that our results continue to hold.

Given the gender gap in entrepreneurship, changes in the gender ratio could affect aggregate entry rates even without changes in marriage rates. To address this concern, for all our specifications, dependent variables are constructed using only women or only men. As a result, the entry rates and average sizes we consider are women specific and men specific rates, hence, not mechanically affected by the quantity of individuals from the other gender.

Our measure of the local labor market is that of economic regions. These are equivalent to commuting zones in the US. There is a total of 76 economic regions in Canada. We focus on individuals between 25 to 65 years old.

Below we present summary statistics for the gender ratio (number of women relative to men). Table 2 presents summary statistics for the gender ratio across different economic regions and countries of birth for the year 2005. Each observation represents an economic region-country of birth pair. We see there is a large variation in the gender ratio in our data.

Table 2: Summary Statistics for Gender Ratio across Economic Regions and Immigrant groups for 2005

Mean	Standard Deviation	# of Observations
0.881	0.8813	7737

Notes: Summary Statistics for Gender Ratio across Economic Regions and Country of Birth groups for 2005.

Next, we want to see how much variation we have across different country of birth groups within an economic region. We plot summary statistics for the gender ratio across different countries of birth within the economic region of Toronto in 2005 (first row) and Winnipeg in 2005 (second row). Table 3 shows there is quite a lot of variation across the country of birth groups. In Toronto, the average gender ratio is 101 women per 100 men, and the standard deviation is 55 women per 100 men. The number of observations in the first row is the number of country of birth groups in Toronto in 2005. In Winnipeg, the average gender ratio is 92.7 women per 100 men, and the standard deviation is 92.5 women per 100 men. The number of observations in the second row is the number of country of birth groups in Winnipeg in 2005.

Table 3: Summary Statistics for Gender Ratio across Immigrant groups for 2005 in Toronto and Winnipeg

Economic Region	Mean	Standard Deviation	# of Observations
Toronto	1.01	0.55	217
Winnipeg	0.927	0.925	169

Notes: Summary Statistics for Gender Ratio across Country of Birth groups for 2005 in Toronto and for 2005 in Winnipeg. Each observation is a country of birth group in Toronto in 2005 (first row) or a country of birth group in Winnipeg in 2005 (second row).

Next, we plot summary statistics for the gender ratio across different economic regions for a given country of birth at 2005. Table 4 below plots the summary statistics for the gender ratio across economic regions for the Argentinean-born population (first row), South Korean-born population (second row), and Russian-born population (third row).[14] In the first row we see that the average gender ratio is 103 women per 100 men for the Argentinean-born population, and the standard deviation is 78 women per 100 men. In the second row, we see that for the South Korean-born population, the average gender ratio is 105 women per 100 men and the standard deviation is 77 women per 100 men. The third row shows that, for the Russian-born population, the average gender ratio is 152 women per 100 men and the standard deviation is 159 women per 100 men. Such a high number of women per men among Russian-born individuals is consistent with the high number of women per men currently in Russia[15] The number of observations in each row represents the number of economic regions in 2005 for which there were people of that country of birth. Hence, each observation represents an economic region in 2005.

Table 4: Summary Statistics for Gender Ratio across Economic Regions for given Country of Birth in 2005

Country of Birth	Mean	Standard Deviation	# of Observations
Argentina	1.03	0.785	56
South Korea	1.05	0.767	60
Russia	1.52	1.595	61

Notes: Summary Statistics for Gender Ratio across Economic Regions in 2005 for Argentina born population (first row), for South Korea born population (second row), for Russia born population (third row). Number of observations in each row represent the number of economic regions in 2005 that had the given country of birth.

[14]These are just some examples, in our analysis we use all countries of birth.

[15]This is often explained by higher mortality rates among men, due to alcoholism. See the recent article by the Pew Research Center https://www.pewresearch.org/fact-tank/2015/08/14/why-the-former-ussr-has-far-fewer-men-than-women/

One concern is that most businesses are started jointly by spouses, which is absent in our model. As explained previously, we verify that our results remain unchanged once we exclude these joint ventures. Second, we might be worried if spouses substitute from working for another firm to working for the firm started by their partner. In this case, we might wrongly interpret a positive impact of marriage on entrepreneurship as due to insurance when instead it is due to spouses working for their partner. In the data, we find that only 16.26% of new entries into entrepreneurship are accompanied by the spouse being reported as an employee. Furthermore, we also verify our main specifications are robust to excluding the subset of businesses for which the spouse works for the new firm started by their partner.

Our identification strategy relies on a certain degree of homophily in marriage in Canada. We might be worried that due to the high level of immigration from all around the world, most couples are composed of two individuals born in different countries. A recent brief by Statistics Canada for 2011[16] shows that of all couples in Canada: 66.9% are between two Canadian-born individuals, 18.2% are between two immigrants born in the same country, 3.7% are between two immigrants of different countries of origin and 11.2% between one Canadian born and one immigrant. It follows that 85% of couples were between individuals born in the same country. Among immigrant only couples, 83% of couples were between individuals born in the same country. This highlights how, despite the high immigration rates to Canada, homophily is still relatively high in the population. This degree of homophily is also observed in other dimensions. The same brief reports that 87.6% of couples have one or more common mother tongues. Similarly, 90.2% of couples are composed of either two individuals that share the same religious affiliation or two individuals with no religious affiliation.

Fourth, one possible concern is that the gender ratio variation in our data comes from small populations of individuals of a particular country in a given economic region. In the robustness section, we show our results are robust to restricting the sample to cells with a minimum of men and women.

5.2 Results

In this section, we present the main results of our specifications. We include the change in female labor force participation of married women for each triplet (c, g, t) as a control. Given our choice of using men as main earners, this represents the change in labor force participation of spouses. Results are robust to not including it. We include additional controls: year dummies, changes in the average income of married

[16] NHS in Brief, Catalogue no. 99-010-X2011003: Mixed Unions in Canada

men for each triplet (c, g, t), changes in the share of total employment in the oil, gas and mining sector, $\Delta Share^{oil}_{c,t}$, changes in the share of total employment in the manufacturing sector, $\Delta Share^{manuf}_{c,t}$, changes in the share of total employment in the service sector, $\Delta Share^{serv}_{c,t}$, and changes in the share of population in different age groups for triplet (c, g, t). Results are robust to not including any of these controls and to clustering at the economic region c level. Each observation in our regressions represent a city-country of birth group-year (c, g, t) triplet.

Table 5: 1st Stage Regressions

Outcome in 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta ER_{c,g,t}$	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$	$\Delta SY_{c,g,t}$	$\Delta SY_{c,g,t}$
$IV_{c,g,t}$	0.0371	0.0371	0.0372	0.0302	0.0302	0.0302
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Dummies for cities	-	Yes	-	-	Yes	-
Dummies for birth country	-	-	Yes	-	-	Yes
Observations	70551	70551	70551	32305	32305	32305

Notes: First Stage regression results. The endogenous variable is the change in the marriage rate of men in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument, $IV_{c,g,t}$ is the change in the gender ratio at economic region c, country of birth g and year t. The gender ratio is defined as the total amount of women of that city c, group g, year t divided by the total amount of men of that city c, group g, year t. Standard errors are clustered at the economic region c and country of birth g. Results are robust to clustering at the economic region c level.

Table 5 reports first stage results. Our model predicts that larger changes in the gender ratio ($IV_{c,g,t}$) should increase changes in the marriage rate ($\Delta MR_{c,g,t}$). Column 1 of Table 5 indicates that one additional woman per man increases the marriage rate of men by 3.71 percentage points. Columns 2 and 3 show the results are robust to including dummies for cities and country of birth. Columns 4, 5 and 6 report similar results when restricting the sample to the one used in the average firm size regressions. Intuitively, the average size regressions only include city, country of birth and years cells that contain active entrepreneurs.

Column 1 of Table 6 presents the OLS results for our entry rate specification. We see that marriage rates are significantly and positively correlated to entry into entrepreneurship if we do not instrument for it. Column 2 shows that once we instrument for marriage rates, the coefficient flips sign and increases in magnitude. A 1 percentage point increase in the marriage rate is associated with a 0.2 percentage point drop in the entry rate into entrepreneurship. This corresponds to a 20% drop in the entry rate. Column 3 indicates the results are robust to including city dummies. The corresponding first stage results are those in Column 1 and 2 of Table 5.

Table 6: Main specifications: Entry Rate Regressions

	OLS	IV	IV	IV
$\Delta MR_{c,g,t}$	0.0258	-0.219	-0.219	-0.22
	(0.004)	(0.035)	(0.035)	(0.035)
Significance IV for $\Delta MR_{c,t}$	-	Yes	Yes	Yes
Dummies for cities	-	-	Yes	-
Dummies for country of birth	-	-	-	Yes
Observations	70551	70551	70551	70551

Notes: Regressions of changes in the entry rate into entrepreneurship for men in economic region c, country of birth g and year t, $\Delta ER_{c,g,t}$, on the change in the marriage rate of men in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. Column 1 reports OLS results. Columns 2, 3 and 4 reports results when using our instrument. Our instrument is the change in the gender ratio. The gender ratio is defined as the total amount of women divided by total amount of men for that city c, group g, year t. Specifications include year dummies, changes in female labor force among married women, in average income among married men, in the share of population at each triplet (c, g, t) within 3 age groups, in the employment shares in oil and gas, manufacturing and services sectors. Standard errors are clustered at economic region c and country of birth g.

One concern with our results is that cultural determinants can simultaneously influence the change in gender ratio and entrepreneurship. To address this issue, we use variation in changes in the gender ratio of the same immigrant group across two different locations. This is achieved by including country of birth dummies. Column 4 shows that our results are robust to using this variation. Results are also unchanged if we consider dummies for each country of birth and year pair. Finally, to rule out that results are driven by the gender ratio correlating to city-specific shocks, we verify our results remain unchanged in magnitude, significance, and sign if we include dummies for each city and year pair.[17]

Next, we turn to the results on the average number of employees of firms. Column 1 of Table 7 indicates that a rise in marriage rates, $\Delta MR_{c,g,t}$, predicts a small insignificant drop in average size of firms if not instrumented for. Column 2 of Table 7 shows results once we instrument for marriage rates by our instrument, the gender ratio. Our estimate implies that a 10 percentage point increase in the marriage rate increases average size by 11.3%.

[17] These results for the entry regressions, including dummies for each country of origin and year pair or each city and year pair, are available upon request from the authors.

Table 7: Main specifications: Average Number of Employees Regressions

	OLS	IV	IV	IV
$\Delta MR_{c,g,t}$	-0.038	1.127	1.12	1.13
	(0.037)	(0.349)	(0.35)	(0.351)
Significance IV for $\Delta MR_{c,g,t}$	-	Yes	Yes	Yes
Dummies for cities	-	-	Yes	-
Dummies for country of birth	-	-	-	Yes
Observations	32305	32305	32305	32305

Notes: Regressions of changes in the log of average number of employees of firms owned by men in economic region c, country of birth g and year t, $\Delta log(SY)_{c,g,t}$, on the change in the marriage rate of men in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. Column 1 reports OLS results. Columns 2, 3 and 4 reports results when using our instrument. Our instrument is the change in the gender ratio. The gender ratio is defined as the total amount of women divided by total amount of men for that city c, group g, year t. Specifications include year dummies, changes in female labor force among married women, in average income among married men, in the share of population at each triplet (c, g, t) within 3 age groups, in the employment shares in oil and gas, manufacturing and services sectors. Standard errors are clustered at the economic region c and country of birth g.

Column 3 and 4 indicate our finding is robust to the inclusion of city dummies or country of birth dummies. The corresponding first stage results are those in Column 4, 5 and 6 of Table 5. Our estimate is unchanged in sign, magnitude and significance if we consider dummies for each country of origin and year pair. Finally, to address the concern that the results are explained by a correlation between the gender ratio and city-specific shocks, we verify our results are robust to including dummies for each city and year pair.[18]

When we include city dummies we use variation in gender ratio changes across country of birth groups within the same city. When we include country of birth dummies we use variation in gender ratio changes for the same immigrant group across different cities. These are two different sources of variation for which there is no reason to expect similar results. Nevertheless, despite this different source of variation, we get a similar sign, significance and magnitude regardless of which of these two sources of variation we use.

Next, we verify that our results are present for women. Changes in marriage rates of women are instrumented by changes in the ratio of men to women. We construct our key variables $\Delta ER_{c,g,t}$, $\Delta SY_{c,g,t}$ and $\Delta MR_{c,g,t}$ for women.

[18]These results are available upon request from the authors.

Table 8: 1st Stage Regressions - Outcomes for Women

Outcome in 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$IV_{c,g,t}$	0.006	0.009
	(0.001)	(0.001)
Observations	65919	25801

Notes: First Stage regression results. The endogenous variable is change in the marriage rate of women in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument, $IV_{c,g,t}$ is the change in the relevant gender ratio for women at economic region c, country of birth g and year t. The relevant gender ratio is defined as the total amount of men of that city c, group g, year t divided by the total amount of women of that city c, group g, year t. Standard errors are clustered at the economic region c and country of birth g. Results are robust to clustering at just the economic region c level.

Table 8 presents first stage results. Column 1 shows that one additional man per woman increases the marriage rate of women by 0.6% percentage points. Column 2 reports similar results when restricting the sample to the one used in the average firm size regressions.

Table 9 shows that a 1 percentage point increase in the marriage rate of women is associated with a 0.685% percentage point drop in entry rates into entrepreneurship and a 6.27% increase in the average size of firms. These estimates indicate that a rise in marriage rates decreases entry rates into entrepreneurship of women by 3 times as much as that of men and increases average size by 6 times as much as that of men. Such a pattern relates to the growing literature on differences in entrepreneurship among genders and the entrepreneurship gender gap (e.g. Scott and Shu (2017), Howell and Nanda (2019), Ewens and Townsend (2020), and Hebert (2020)). In particular, we contribute to the discussion by showing there is a gender gap in the responsiveness of entrepreneurship to lower marriage rates.

Table 9: Main specifications: Regressions for women

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
	IV	IV
$\Delta MR_{c,g,t}$	-0.685	6.268
	(0.232)	(3.649)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
Observations	65919	25801

Notes: Regressions of changes in entry rate into entrepreneurship for women, $\Delta ER_{c,g,t}$ and changes in log of average firm size for women, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$ for women. The instrument for $\Delta MR_{c,g,t}$ for women is the number of men relative to women, $\vartheta^{-1}_{c,g,t}$. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level. First stage results can be found in the Appendix.

The results indicate that differences between men and women do not explain the qualitative impact of marriage on entry into entrepreneurship and firm average size. In particular, one potential interpretation of our results for men is that men are less risk averse than women. Then, a couple, a man and a women, are more risk averse than a single man, entering less and creating larger firms. If this narrative were true, married women should have higher entry rates into entrepreneurship and lower average size than single women. We find the exact opposite. It follows that differences in risk aversion between genders do not explain our results.

5.3 Discussion on Alternative Mechanisms

The theoretical model is purposely tractable to keep the intuition clear and concise. However, there might be other economic mechanisms affecting entrepreneurship not present in the model.

Firstly, spouses might help individuals overcome borrowing constraints via wealth sharing or faster wealth accumulation. There are two ways that borrowing constraints affect outcomes for an individual considering to start a firm. The first is that individuals are constrained in the scale of the firm they create. If married individuals are wealthier, we expect higher marriage rates to be associated with a larger average size. This is an alternative narrative for our finding that average size increases with the marriage rate. The second way borrowing constraints might matter is in the presence of startup costs. However, both these channels would predict higher marriage rates increase the entry rate into entrepreneurship. We find exactly the opposite. To further verify this borrowing constraint channel is not driving our findings, we verify our results are robust to excluding businesses started in high capital demanding sectors (See Table 12 of Appendix).

Another potential mechanism ignored by the model is the possibility of joint entrepreneurship by both the main earner and spouse. In particular, married individuals can be more likely to start a firm because they can start a firm jointly with their spouses. Of course, this is inherently hard to measure. Nonetheless, to the extent that a couple running together a firm are both listed as owners of the business or have the spouse listed as an employee of the firm, we can verify our results are robust to excluding these types of businesses. Our results continue to hold. See Tables 13 and 14 of the Appendix for results taking out firms started jointly by the couple and taking out firms for which the spouse is listed as an employee. Such robustness tests would be impossible without our link between business owner, firm, employee, and spouse.

Another concern is that in non-cooperative households, changes in the gender ratio could affect how income is shared within the household. In our model, this would imply the shares of income kept by the spouse, γ, and by the main earner, $1 - \gamma$, would be a function of the gender ratio ϑ. Given that we use the gender ratio as an instrument, this would violate the exclusion restriction. In Section J of the Appendix, we consider an extension of the model where we allow the income sharing parameter γ to be a function of the gender ratio ϑ. To address this issue, we use this model extension to derive the correct specification in this case. In particular, if the gender ratio affects intra-household income sharing, it induces changes in labor force participation of spouses. This allows us to derive a strict relationship between the intra-household income sharing parameter and labor force participation of spouses. Once we control for labor force participation of spouses and the average income of married main earners, we control for variation in intra-household income sharing, allowing us to recover the exclusion restriction. Since our main specifications already include labor force participation of spouses and the average income of married main earners as controls, our results are robust to this potential critique.

5.4 Robustness Checks

One concern with our strategy is that there might be economic regions with a small number of individuals from a particular country. To address this concern, we verify our results are robust to restricting our analysis to economic region, country of birth, and year triplets with a minimum of women and men. Our results are unchanged. Magnitudes actually increase. See Columns 1 and 2 of Table 11 of the Appendix.

A second concern with our identification strategy arises if city-specific shocks $\Delta y_{c,t}$ are gender-biased. Since our results are robust to including city-year pair dummies, they are robust to this concern if the gender-bias is the same across groups. As an example, this means that if the productivity shock Calgary receives is male biased, this bias is not different between Colombian-born and Chilean-born individuals. In this case, Colombian-born men and Chilean-born men would be more likely to move to Calgary than Colombian-born women and Chilean-born women. But because this bias is the same across all groups, city-year dummies control for this mechanism.

Of course, it is possible that $\Delta y_{c,t}$ are gender-biased differently across groups. This would be the case if the productivity shock Calgary received is male biased for both Colombian-born and Chilean-born individuals but this bias is stronger for the Colombian-born. In this case, men would be more likely to move to Calgary than women, but this difference in likelihood to move to Calgary between genders would be larger for the Colombian-born. As a result, our instrument would be correlated to the error term. This is not a concern for individuals that arrived at an early age in Canada. As long as the choice of an immigrant, of where to immigrate in Canada, is uncorrelated to the gender of their child, we can use these early age arrival immigrants to address this concern. We verify our results are robust to using a gender ratio constructed using only individuals that arrived at age 15 or younger in Canada. This robustness also confirms that our findings are unchanged when we exclude individuals that arrived already married in Canada. See Columns 3 and 4 of Table 11 in the Appendix.

5.5 Brief Conclusion

Our findings are consistent with marriage making individuals more selective on which business projects to implement. These results hold for both women and men. Through our model, we interpret that the spousal sharing effect and spousal insurance effect are dominated by the larger cost of failure for families with children. Our estimates confirm the prediction of our model that the group (married or unmarried) with the highest entry rate into entrepreneurship is also the one with the smaller average firm size.

Second, the stronger effect for women is consistent with women receiving less spousal insurance from their male spouses than what their male spouses receive from them. This arises when working hours are less responsive to spousal income for men than for women.

The rise in the selection of business projects due to marriage implies that marriages rates are connected to average firm productivity. In the next section, we formally look at this implication.

6 Implications for Average Firm Productivity

The results in the previous section make clear that higher marriage rates induce lower entry rates but larger firms on average. Through the lens of our model, the results indicate that higher marriage rates induce less firm creation but increase average firm size and productivity. In this section, we use our results to discipline a back of the enveloppe bounding exercise of the implied change in average firm productivity, $E[z]$. To do so, note that the optimal size for a firm, $n(z, w)$ is given by

$$n(z, w) = (\frac{\alpha y \theta}{w})^{\frac{1}{1-\alpha}} e^{\frac{z}{1-\alpha}}. \tag{25}$$

This in turn implies

$$E[n] = (\frac{\alpha}{w})^{\frac{1}{1-\alpha}} E[z] \tag{26}$$

where $E[z]$ is average firm productivity[19] and $E[n]$ is average number of employees. It follows that

$$\Delta log(E[z]) = \Delta log(E[n]) - \Delta log(\frac{\alpha}{w})^{\frac{1}{1-\alpha}}. \tag{27}$$

In the short and medium run we can argue $\Delta w \approx 0 \Rightarrow \Delta log(\frac{\alpha}{w})^{\frac{1}{1-\alpha}} \approx 0$. Hence,

$$\Delta log(E[z]) \approx \Delta log(E[n]). \tag{28}$$

Given our estimates of the previous section it follows that a 1 percentage point increase in marriage rates is associated to 1.13% percent increase in average productivity. Of course, in the long run we expect wages, w, to adjust. If the main source of variation in $(\frac{\alpha}{w})^{\frac{1}{1-\alpha}}$ in the long run is the rise in wages, w, then our estimate of the impact of marriage on average firm productivity is a lower bound to the true long run effect. More generally, our estimate of the response of average size to marriage rates allows us to calculate the implied changes in average firm productivity given any estimate of changes in wages.

[19] See Equation (106) of the Appendix.

7 Conclusion

We explore a topic never before studied, the importance of family in the individual decision to start a firm, and the outcomes of that firm. Through a tractable model we show that family affects the decision through three opposing channels. First, spouses increase the benefit to a more productive business by working less hours, allowing households to save on the costs of working. Second, spouses provide insurance by working more hours in case of business failure. Finally, families might have children, increasing the cost of failure.

Using the model, we derive empirical specifications, the source of endogeneity, and an instrumental variable that allows us to bring the model to the data. We show that marriage is associated with lower entry into entrepreneurship and a higher average number of employees. This is consistent with the spousal sharing effect and spousal insurance effect being dominated by the higher cost of failure associated to having children. Furthermore, the results are consistent with our selection mechanism in which the effect of marriage on entry into entrepreneurship and the average size of firms has opposite signs. We also verify that our results are robust to a series of alternative mechanisms. Through the lens of our model, our results indicate that a fall in aggregate marriage rates induces a decline in average firm productivity.

References

Abbott, Brant, "Incomplete Markets and Parental Investments in Children," Technical Report, Working Paper 2020.

Angrist, Josh, "How do sex ratios affect marriage and labor markets? Evidence from America's second generation," *The Quarterly Journal of Economics*, 2002, *117* (3), 997–1038.

Beaudry, Paul, David A Green, and Benjamin M Sand, "In search of labor demand," *American Economic Review*, 2018, *108* (9), 2714–57.

_ , _ , and Benjamin Sand, "Does industrial composition matter for wages? a test of search and bargaining theory," *Econometrica*, 2012, *80* (3), 1063–1104.

Bellou, Andriana and Emanuela Cardia, "The Great Depression and the rise of female employment: A new hypothesis," 2020.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten, "Consumption inequality and family labor supply," *American Economic Review*, 2016, *106* (2), 387–435.

Browning, Martin, Pierre-Andre Chiappori, and Valerie Lechene, "Collective and unitary models: A clarification," *Review of Economics of the Household*, 2006, *4* (1), 5–14.

Brush, CG, Patricia G Greene, Lakshmi Balachandra, and Amy E Davis, "Diana Report. Women Entrepreneurs 2014: Bridging the gender gap in venture capital. Arthur M," *Blank Center for Entrepreneurship, Babson College*, 2014.

Cagetti, Marco and Mariacristina De Nardi, "Entrepreneurship, frictions, and wealth," *Journal of Political Economy*, 2006, *114* (5), 835–870.

_ **and** _ , "Estate taxation, entrepreneurship, and wealth," *American Economic Review*, 2009, *99* (1), 85–111.

Chandler, Alfred Dupont, *The visible hand: The Managerial Revolution in American Business*, Cambridge, MA : Harvard University Press, 1977.

Chiappori, Pierre-Andre, Bernard Fortin, and Guy Lacroix, "Marriage market, divorce legislation, and household labor supply," *Journal of Political Economy*, 2002, *110* (1), 37–72.

Clementi, Gian Luca and Berardino Palazzo, "Entry, exit, firm dynamics, and aggregate fluctuations," *American Economic Journal: Macroeconomics*, 2016, *8* (3), 1–41.

Ewens, Michael and Richard R Townsend, "Are early stage investors biased against women?," *Journal of Financial Economics*, 2020, *135* (3), 653–677.

Fonseca, Raquel, Paloma Lopez-Garcia, and Christopher A Pissarides, "Entrepreneurship, start-up costs and employment," *European Economic Review*, 2001, *45* (4), 692–705.

Gallipoli, Giovanni and Laura Turner, "Household responses to individual shocks: Disability and labor supply," Technical Report 2009.

Glover, Andy and Jacob Short, "Bankruptcy, Incorporation, and the Nature of Entrepreneurial Risk," 2010.

Gottlieb, Joshua D, Richard R Townsend, and Ting Xu, "Experimenting with entrepreneurship: the effect of job-protected leave," Technical Report, National Bureau of Economic Research 2016.

Green, David A, Rene Morissette, Benjamin M Sand et al., "Economy Wide Spillovers From Booms: Long Distance Commuting and the Spread of Wage Effects," Technical Report 2017.

Grekou, D. and B. Gueye, "Who were the men and women entering business ownership in Canada?," *Analytical Studies Branch Research Paper Series*, 2021.

Haltiwanger, John, "Job Creation and firm dynamics in the US," *Unpublished manuscript, May*, 2011.

Harris, Ron, *Industrializing English law: entrepreneurship and business organization, 1720-1844*, Cambridge University Press, 2000.

Hebert, Camille, "Gender stereotypes and entrepreneur financing," *Working Paper*, 2020.

Howell, Sabrina T and Ramana Nanda, "Networking frictions in venture capital, and the gender gap in entrepreneurship," Technical Report, National Bureau of Economic Research 2019.

Hurst, Erik and Benjamin Wild Pugsley, "What Do Small Businesses Do?," *Brookings Papers on Economic Activity*, 2011, *2011* (2), 73–118.

Hyslop, Dean R, "Rising US earnings inequality and family labor supply: The covariance structure of intrafamily earnings," *American Economic Review*, 2001, *91* (4), 755–777.

Kahn, R and Toni M Whited, "Identification is not causality, and vice versa," *Review of Corporate Finance Studies*, 2018, *7* (1), 1–21.

Levine, Ross and Yona Rubinstein, "Smart and illicit: who becomes an entrepreneur and do they earn more?," *The Quarterly Journal of Economics*, 2017, *132* (2), 963–1018.

Poschke, Markus, "Who becomes an entrepreneur? Labor market prospects and occupational choice," *Journal of Economic Dynamics and Control*, 2013, *37* (3), 693–710.

__ , "The firm size distribution across countries and skill-biased change in entrepreneurial technology," *American Economic Journal: Macroeconomics*, 2018, *10* (3), 1–41.

Scott, Erin L and Pian Shu, "Gender gap in high-growth ventures: Evidence from a university venture mentoring program," *American Economic Review*, 2017, *107* (5), 308–11.

Sedláček, Petr, "Lost generations of firms and aggregate labor market dynamics," *Journal of Monetary Economics*, 2020, *111*, 16–31.

Tracy, Spencer L, "Accelerating job creation in America: The promise of high-impact companies," *US Small Business Administration, Office of Advocacy*, 2011.

Tschopp, Jeanne, "The Wage Response to Shocks: The Role of Inter-Occupational Labour Adjustment," *Labour Economics*, 2015, *37*, 28–37.

Wu, Chunzan and Dirk Krueger, "How Much Consumption Insurance in Bewley Models with Endogenous Family Labor Supply?," Technical Report, National Bureau of Economic Research 2018.

Zandberg, Jonathan, "Family Comes First: Reproductive Rights and the Gender Gap in Entrepreneurship," *Journal of Financial Economics, forthcoming*, 2020.

A Value function for households where the spouse does not work

In this section I go over the value functions of the household when $\kappa > \kappa^*$. When that is the case, the spouse does not work. The value function for married households with spouse not working and main earner currently running a firm is given by

$$rJ^m(z,\kappa) = \mu\frac{[\gamma(\pi(z)-\chi)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\gamma)(\pi(z)-\chi)]^{1-\sigma}}{1-\sigma} + \varphi + \lambda(B^m(\kappa) - J^m(z,\kappa)). \quad (29)$$

The value function for married households with spouse not working and main earner having failed a business is given by

$$rB^m(\kappa) = \mu\frac{[\gamma(w-c-\chi)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\gamma)(w-c-\chi)]^{1-\sigma}}{1-\sigma} + \varphi + p(W^m(\kappa) - B^m(\kappa)). \quad (30)$$

The value function for married households with spouse not working and main earner currently working is given by

$$rW^m(\kappa) = \mu\frac{[\gamma(w-\chi)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\gamma)(w-\chi)]^{1-\sigma}}{1-\sigma} + \varphi + \psi\int_{\underline{z}^m(\kappa)}(J^m(z,\kappa) - W^m(\kappa))dF(z). \quad (31)$$

B Proposition concerning threshold $\hat{\phi}$ for spousal sharing effect

Proposition 9 *There exists a $\hat{\phi}$ such that $\forall\phi < \hat{\phi}$, the derivative of the flow utility of an unmarried household running a firm with respect to the business productivity z is larger than the derivative of the flow utility of a married household composed of an entrepreneur and a working spouse with respect to business productivity z.*

C Proposition concerning relationship between $E[z]$ and measure of individuals married MR

Proposition 10 *Average firm productivity, $E[z]$, is increasing in the measure of married individuals, MR, if $\underline{z}^m > \underline{z}^u$ and decreasing otherwise.*

D Value function for married main earner

In this subsection of the Appendix we describe the value functions for married main earners. Note that these are not the same as the value function of the married household. First, define as in the body of the text $J_h^m(z, \kappa, \theta)$ as the value function of the main earner of ability θ, married to a spouse with fixed cost of working κ and running a firm of business project productivity z. Let $W_h^m(\kappa, \theta)$ be the value function of the main earner of ability θ married to a spouse of fixed cost κ and working. Let $B_h^m(\kappa, \theta)$ be the value function of the main earner of ability θ married to a spouse of fixed cost κ and having just failed a business. Then, it follows that

$$rW_h^m(\kappa, \theta) = \frac{(\gamma(w + wh^* - \chi))^{1-\sigma}}{1 - \sigma} + \mu\varphi - \mu\phi h^* \\ + \psi \int_{\underline{z}^M} (J_h^m(z, \kappa, \theta) - W_h^m(z, \kappa, \theta))dF(z) \quad (32)$$

$$rJ_h^m(z, \kappa, \theta) = \frac{(\gamma(\pi(z) + wh^* - \chi))^{1-\sigma}}{1 - \sigma} + \mu\varphi - \mu\phi h^* \\ + \lambda(B_h^m(\kappa, \theta) - J_h^m(z, \kappa, \theta)) \quad (33)$$

and

$$rB_h^m(\kappa, \theta) = \frac{(\gamma(w - c + wh^* - \chi))^{1-\sigma}}{1 - \sigma} + \mu\varphi - \mu\phi h^* + p(W_h^m(\kappa, \theta) - B_h^m(\kappa, \theta)). \quad (34)$$

Recall that μ is the weight of the main earner in the household. Hence, μ is the share of spouse's cost of working ϕh paid by the main earner and the share of the utility associated to being in a couple received by the main earner.

E Proofs

Proof of Proposition 1.

Conditional on the spouse working, the household solves

$$\max_{0\leq h\leq 1}\mu\frac{[\gamma(I+wh-\chi)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\gamma)(I+wh-\chi)]^{1-\sigma}}{1-\sigma}+\varphi-\phi h \tag{35}$$

This gives us

$$h^{*}=\frac{(\mu\gamma^{1-\sigma}+(1-\mu)(1-\gamma)^{1-\sigma})^{\frac{1}{\sigma}}}{\phi^{\frac{1}{\sigma}}w^{\frac{\sigma-1}{\sigma}}}+\frac{\chi}{w}-\frac{I}{w} \tag{36}$$

when h is an interior solution and 0 or 1 when we get corner solutions.

Proof of Proposition 2.

Define $U_s^M(I)$ as the total flow utility for the household with spouse working and $U_{ns}^m(I,\gamma)$ as the flow utility of a married household composed of a main earner with income I and a non-working spouse. Since the choice of the spouse working or not is done by the household, the household pays κ and the spouse works if

$$\kappa_g^{*}\equiv\int(U_s^m(I,\gamma)-U_{ns}^m(I,\gamma))f(I|g)dI>\kappa \tag{37}$$

where $f(I|g)$ is the endogenous distribution of main earner income for group g. Define $Prob(work)$ as the probability of working for a spouse, then,

$$Prob(work)=M[\int(U_s^m(I,\gamma)-U_{ns}^m(I,\gamma))f(I|g)dI]. \tag{38}$$

Proof of Proposition 3.

Because there are no search frictions, the profit maximization problem is state. Hence, entrepreneurs solve

$$\pi(z)=\max_{n}y\theta e^{z}n^{\alpha}-wn \tag{39}$$

which gives

$$n(z)=(\frac{\alpha y\theta e^{z}}{w})^{\frac{1}{1-\alpha}} \tag{40}$$

and

$$\pi(z) = (1-\alpha)(\frac{\alpha}{w})^{\frac{\alpha}{1-\alpha}} e^{\frac{z}{1-\alpha}} (y\theta)^{\frac{1}{1-\alpha}}. \tag{41}$$

Taking derivaties gives

$$\frac{\partial \pi(z)}{\partial y} > 0, \frac{\partial \pi(z)}{\partial z} > 0, \frac{\partial \pi(z)}{\partial \theta} > 0, \frac{\partial n(z)}{\partial y} > 0, \frac{\partial n(z)}{\partial z} > 0, \frac{\partial n(z)}{\partial \theta} > 0. \tag{42}$$

Proof of Proposition 4.

Using the expression for optimal hours of the spouse h^* and setting $\chi = 0$, we obtain

$$u_M^B(\chi = 0) = \frac{D^{1-\sigma}}{1-\sigma}(\mu\gamma^{1-\sigma} + (1-\mu)(1-\gamma)^{1-\sigma}) + \phi - \phi D + \frac{\phi(w-c)}{w}, \tag{43}$$

and

$$u_S^B = \frac{(w-c)^{1-\sigma}}{1-\sigma}, \tag{44}$$

where

$$D \equiv \frac{(\mu\gamma^{1-\sigma} + (1-\mu)(1-\gamma)^{1-\sigma})^{\frac{1}{\sigma}}}{\phi^{\frac{1}{\sigma}} w^{\frac{\sigma-1}{\sigma}}}. \tag{45}$$

Next, note that

$$u_M^B > u_S^B \Rightarrow \frac{D^{1-\sigma}}{1-\sigma}(\mu\gamma^{1-\sigma} + (1-\mu)(1-\gamma)^{1-\sigma}) + \phi - \phi D > \frac{(w-c)^{1-\sigma}}{1-\sigma} - \frac{\phi(w-c)}{w}. \tag{46}$$

Let

$$RHS \equiv \frac{(w-c)^{1-\sigma}}{1-\sigma} - \frac{\phi(w-c)}{w} \quad \text{and} \quad c^{**} = w - w^{\frac{1}{\sigma}}. \tag{47}$$

Then,

$$\frac{\partial RHS}{\partial c} = -(w-c)^{-\sigma} + \frac{\phi}{w} < 0, \quad \text{for all} \quad c > c^{**}, \text{and} \quad \frac{\partial^2 RHS}{\partial c^2} < 0, \quad \text{for all } c. \tag{48}$$

From the above we see that for all $c > c^{**}$, RHS becomes decreasing in c. Given the left hand side of the inequality in 46 does not depend on c, it follows that there exists a $\hat{c}$ such that for all $c > \hat{c}$, $u_M^B(\chi = 0) > u_S^B$.

Proof of Proposition 5.

Take derivative of u_M^B with respect to c to get

$$\frac{\partial u_M^B}{\partial c} = -\mu\gamma(\gamma(w - c + wh - \chi))^{-\sigma} - (1 - \mu)(1 - \gamma)((1 - \gamma)(w - c + wh - \chi))^{-\sigma} < 0 \tag{49}$$

Note take a derivative of the above expression with respect to χ to get

$$\frac{\partial^2 u_M^B}{\partial c \partial \chi} = -\mu\gamma^2\sigma(\gamma(w-c+wh-\chi))^{-\sigma-1} - (1-\mu)(1-\gamma)^2\sigma((1-\gamma)(w-c+wh-\chi))^{-\sigma-1} < 0 \tag{50}$$

Proof of Proposition 6.

In what follows, we discuss how these thresholds map into differences in firm outcomes and the channels that generate differences in these thresholds. The discussion that follows should be understood as conditional on a value for θ and κ. Once, we bring model to the data we will take into account the dependency of our value functions on θ and κ.

The entry rate into entrepreneurship for married and unmarried individuals is determined respectively by

$$ER^m = \psi(1 - F(\underline{z}^m)) \quad \text{for married households} \tag{51}$$

and

$$ER^u = \psi(1 - F(\underline{z}^u)) \quad \text{for unmarried.} \tag{52}$$

From the expressions above we see that a higher threshold decreases the entry rate of the corresponding group since

$$\frac{\partial ER^m}{\partial z} = -\psi f(\underline{z}^m) < 0 \tag{53}$$

and

$$\frac{\partial ER^u}{\partial z} = -\psi f(\underline{z}^u < 0 \tag{54}$$

Let $\Gamma_i(z)$ be the fraction of individuals $i \in \{m, u\}$ running a business project of productivity e^z. The average size of firms among married households (for fixed κ and θ), $E[s]^m$, is given by

$$E[s]^m = \int_{\underline{z}^m} n(z,w)\frac{\Gamma_m(z)}{\Gamma_m}dz = \int_{\underline{z}^m} n(z,w)\frac{f(z)}{1-F(\underline{z}^m)}dz$$
$$= \int_{\underline{z}^m} (\frac{\alpha y \theta e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{(1-F(\underline{z}^m))}dz \quad (55)$$

where

$$\Gamma_m = \int \Gamma_m(z)dz. \quad (56)$$

Similarly, average size of firms among unmarried individuals, $E[s]^u$, is given by

$$E[s]^u = \int_{\underline{z}^u} n(z,w)\frac{\Gamma_u(z)}{\Gamma_u}dz = \int_{\underline{z}^u} n(z,w)\frac{f(z)}{1-F(\underline{z}^u)}dz$$
$$= \int_{\underline{z}^u} (\frac{\alpha y \theta e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{(1-F(\underline{z}^u))}dz \quad (57)$$

where

$$\Gamma_u = \int \Gamma_u(z)dz. \quad (58)$$

For the expression for $\Gamma^m(z)$ and $\Gamma^u(z)$ please see Proof of Proposition 10.

From these expressions we see that

$$\frac{\partial E[s]^m}{\partial \underline{z}^m} > 0 \quad (59)$$

and

$$\frac{\partial E[s]^u}{\partial \underline{z}^u} > 0. \quad (60)$$

Hence, we conclude that average size of firms for a group (married versus unmarried) is increasing in the thresholds chosen by that group. To summarize, if we want to know whether married individuals enter more or less and create smaller or larger firms, we must uncover the relationship between their productivity thresholds:

$$\underline{z}^m \lesseqgtr \underline{z}^u. \quad (61)$$

The direction of this inequality depends on the relative strenghts of the channels we discussed: **spousal sharing effect**, **spousal insurance effect** and **offspring effect**.

Proof of Proposition 7.

In this subsection I go over the proof of Proposition 7 as well as the expressions for $V^M(\theta, g)$ and $V^U(\theta, g)$.

The cost κ of the spouse working is shared between spouse and main earner according to their weight in the household (μ and $1-\mu$ respectively). Let $W_h^m(\kappa, \theta, \chi)$ represent the value for a main earner of ability θ of working and being married to a spouse of type κ and children expenditures χ. Let $J_h^m(z, \kappa, \theta, \chi)$ represent the value for a main earner of ability θ of being married to a spouse of type κ, children expenditures χ, and running a firm with business project productivity of z. Let $B_h^m(\kappa, \theta, \chi)$ be the value for a main earner of ability θ of being married to a spouse of type κ, children expenditures χ, and having failed a firm. In Appendix D we describe in further detail value functions $J_h^m(z, \kappa, \theta, \chi)$, $B_h^m(\kappa, \theta, \chi)$ and $W_h^m(\kappa, \theta, \chi)$. Note that the value functions J_h^M, W_h^M and B_h^m are the value functions for the married main earner and the values functions J^m, W^m and B^m are those for the married household. Let e be the measure of main earners in a household with value function W, b be the measure of main earners in households with value function B.

For the first part of the proof let us find the solution for the probability of marriage in the model. The ex-ante value function of being married, $V^M(\theta, g)$ for an individual of ability θ from group g is

$$\begin{aligned}V^M(\theta, g) + v_g \equiv \nu \int &(e_m(\kappa, \theta, \chi_1) W_h^m(\kappa, \theta, \chi_1) + b_m(\kappa, \theta, \chi_1) B_h^m(\kappa, \theta, \chi_1) \\ &+ \int_{\underline{z}^m(\kappa,\theta,\chi_1)} J_h^m(z, \kappa, \theta, \chi_1)\Gamma_M(z, \kappa, \theta, \chi_1) dz) dM(\kappa) + (1-\nu)\int (e_m(\kappa, \theta, 0) W_h^m(\kappa, \theta, 0) \\ &+ b_m(\kappa, \theta, 0) B_h^m(\kappa, \theta, 0) + \int_{\underline{z}^m(\kappa,\theta,0)} J_h^m(z, \kappa, \theta, 0)\Gamma_M(z, \kappa, \theta, 0) dz) dM(\kappa) - (1-\mu)\int^{\kappa_g^*} \kappa dM(\kappa) + v_g.\end{aligned} \tag{62}$$

The ex-ante value of being unmarried, $V^U(\theta, g)$, for an individual of ability θ from group g is

$$V^U(\theta, g) \equiv e_u(\theta) W^u(\theta) + b_u(\theta) B^u(\theta) + \int_{\underline{z}^u(\theta)} J^u(z, \theta)\Gamma_U(z, \theta) dz. \tag{63}$$

It follows that an individual of ability θ from group g wants to marry if

$$V^M(\theta,g)+v_g \equiv \nu\int(e_m(\kappa,\theta,\chi_1)W_h^m(\kappa,\theta,\chi_1)+b_m(\kappa,\theta,\chi_1)B_h^m(\kappa,\theta,\chi_1)$$
$$+\int_{\underline{z}^m(\kappa,\theta,\chi_1)} J_h^m(z,\kappa,\theta,\chi_1)\Gamma_M(z,\kappa,\theta,\chi_1)dz)dM(\kappa)+(1-\nu)\int(e_m(\kappa,\theta,0)W_h^m(\kappa,\theta,0)$$
$$+b_m(\kappa,\theta,0)B_h^m(\kappa,\theta,0)+\int_{\underline{z}^m(\kappa,\theta,0)} J_h^m(z,\kappa,\theta,0)\Gamma_M(z,\kappa,\theta,0)dz)dM(\kappa)-(1-\mu)\int^{\kappa_g^*}\kappa dM(\kappa)+v_g$$
$$> e_u(\theta)W^u(\theta)+b_u(\theta)B^u(\theta)+\int_{\underline{z}^u(\theta)} J^u(z,\theta)\Gamma_U(z,\theta)dz. \quad (64)$$

Now allow for the possibility that individuals need to find a partner to marry. This probability of meeting someone depends on the the gender ratio ϑ_g, of their particular group.[20] Hence, let $Pr(M=1)$ be the probability an individual marries, then,

$$Pr(M=1)=q(\vartheta_g)\cdot Pr(V^M(\theta,g)+v_g>V^U(\theta,g))$$
$$\equiv q(\vartheta_g)Pr(v_g>V^U(\theta,g)-V^M(\theta,g))$$
$$=q(\vartheta_g)(1-H(V^U(\theta,g)-V^M(\theta,g))). \quad (65)$$

where $q'()>0$.

Proof of Corollary 7.1.

For the first part of this proof we show how to obtain an equation defining thresholds $\underline{z}^M$ and $\underline{z}^U$. With a slight abuse of notation, define $u_S^x(\kappa)$ as the flow utility an individual receives when married or unmarried ($x\in\{u,m\}$) with spouse of cost of working κ if married[21] and in State S, where $S\in\{W,B,J)$ and $u_J^x(z,\kappa)$ represents the value of being $x\in\{u,m\}$ when running a firm of productivity z.

Now note the fact that, in equilibrium, $\underline{z}^M(\kappa)$ and $\underline{z}^U$ are defined by

$$J^U(\underline{z}^U)=W^U \quad (66)$$

[20] In our empirical estimation we consider the ratio of women relative to men when looking at outcomes for men and the ratio of men relative to women when looking at outcomes for women.

[21] This flow utility is independant of any κ for the unmarried. This is the sense in which it is a certain abuse of notation.

$$J^W(\underline{z}^W(\kappa),\kappa) = W^M(\kappa). \tag{67}$$

Define value function $V^x(\kappa)$ for $x = \{u, m\}$ where κ only has meaning for $x = m$. We can rewrite value functions $W^x(\kappa), J^x(z,\kappa), B^x(\kappa)$ as

$$J^x(z,\kappa) = \frac{u_J^x(z,\kappa) + \lambda B^x(\kappa)}{r+\lambda} \quad \forall x \in \{u,m\}. \tag{68}$$

$$B^x(\kappa) = \frac{u_B^x(\kappa) + pW^x(\kappa)}{r+p} \quad \forall x \in \{u,m\}. \tag{69}$$

$$W^x(\kappa) = \frac{u_W^x(\kappa) + \psi \int_{\underline{z}^x(\kappa)} J^x(z,\kappa)dF(z)}{r + \psi(1 - F(\underline{z}^x(\kappa)))} \tag{70}$$

Using the expression for $J^x(z,\kappa)$ and using $J^x(\underline{z}^x(\kappa),\kappa) = W^x(\kappa)$ we get

$$0 = J^x(\underline{z}^x(\kappa),\kappa) - W^x(\kappa) = \frac{u_J^x(\underline{z}^x,\kappa)}{r+\lambda} + \frac{\lambda u_B^x(\kappa)}{(r+\lambda)(r+p)} - \frac{rW^x(\kappa)(r+p+\lambda)}{(r+p)(r+\lambda)} \tag{71}$$

Now after some algebra we find

$$\begin{aligned} rW^x(\kappa) = &\frac{(r+p)(r+\lambda)u_W^x(\kappa)}{(r+\lambda)(r+p) + r(r+\lambda+p)\psi(1-F(\underline{z}^x(\kappa)))} \\ &+ \psi \int_{\underline{z}^x(\kappa)} \frac{u_J^x(z,\kappa)(r+p)dF(z)}{(r+\lambda)(r+p) + r(r+p+\lambda)\psi(1-F(\underline{z}^x(\kappa)))} \\ &+ \frac{\psi(1-F(\underline{z}^x(\kappa)))\lambda u_B^x(\kappa)}{r(r+p)(r+\lambda) + r(r+\lambda+p)\psi(1-F(\underline{z}^x(\kappa)))}. \end{aligned} \tag{72}$$

If we replace this expression for rW^x in Equation 71 we find equations defining each treshold $\underline{z}^x$ as a function of parameters and y, θ, ϑ. Using these optimal expression for each threshold we can linearize $\underline{z}^M(\kappa)$, $\underline{z}^U$, y, θ and γ_g around the point $(z^*, z^*, 1, 1, 1)$ which gives

$$\underline{z}^M(\kappa,\chi_1) = \zeta^M_{1,0,\chi_1} + \zeta^M_{1,1,\chi_1} y + \zeta^M_{1,2,\chi_1}\theta \quad \text{if} \quad \kappa \leq \kappa^* \tag{73}$$

$$\underline{z}^M(\kappa,\chi_1) = \zeta^M_{2,0,\chi_1} + \zeta^M_{2,1,\chi_1} y + \zeta^M_{2,2,\chi_1}\theta \quad \text{if} \quad \kappa > \kappa^* \tag{74}$$

$$\underline{z}^M(\kappa,\chi_2) = \zeta^M_{1,0,\chi_2} + \zeta^M_{1,1,\chi_2} y + \zeta^M_{1,2,\chi_2}\theta \quad \text{if} \quad \kappa \leq \kappa^* \tag{75}$$

$$\underline{z}^M(\kappa,\chi_2) = \zeta^M_{2,0,\chi_2} + \zeta^M_{2,1,\chi_2} y + \zeta^M_{2,2,\chi_2}\theta \quad \text{if} \quad \kappa > \kappa^* \tag{76}$$

$$\underline{z}^U = \zeta_0^U + \zeta_1^U y + \zeta_2^U \theta. \tag{77}$$

Note that besides the importance of whether $\kappa \lesseqgtr \kappa^*$, $\underline{z}^M(\kappa)$ does not depend on the level of κ.

Now integrate over all κ to get

$$\int \underline{z}^M(\kappa)dM(\kappa) = Prob(work)_g(\nu(\zeta^M_{1,0,\chi_1} + \zeta^M_{1,1,\chi_1} y + \zeta^M_{1,2,\chi_1}\theta + \zeta^M_{1,3,\chi_1}\gamma_g)$$
$$+ (1-\nu)(\zeta^M_{1,0,\chi_2} + \zeta^M_{1,1,\chi_2} y + \zeta^M_{1,2,\chi_2}\theta + \zeta^M_{1,3,\chi_2}\gamma_g)) + (1 - Prob(work)_g)(\nu(\zeta^M_{2,0,\chi_1} + \zeta^M_{2,1,\chi_1} y + \zeta^M_{2,2,\chi_1}\theta)$$
$$+ (1-\nu)(\zeta^M_{2,0,\chi_2} + \zeta^M_{2,1,\chi_2} y + \zeta^M_{2,2,\chi_2}\theta)) \tag{78}$$

and

$$\underline{z}^U = \zeta_0^U + \zeta_1^U y + \zeta_2^U \theta. \tag{79}$$

Next, linearize equation 78 with $\underline{z}^M(\kappa)$, $\underline{z}^U$, y, θ and $Prob(work)$ around the point $(z^*, z^*, 1, 1, 1, 1)$ to get

$$\int \underline{z}^M(\kappa)dM(\kappa) = \zeta^M_{3,0} + \zeta^M_{3,1} y + \zeta^M_{3,2}\theta + \zeta^M_{3,3}\gamma_g + \zeta^M_{3,4} Prob(work)_g \tag{80}$$

and

$$\underline{z}^U = \zeta_0^U + \zeta_1^U y + \zeta_2^U \theta. \tag{81}$$

Next, we linearize $Pr(M=1)$ for $\underline{z}^M(\kappa), \underline{z}^U, y, \theta, \vartheta$ and κ^* around the point $(z^*, z^*, 1, 1, 1, \overline{\kappa})$ to get

$$Pr(M=1) = D_0 + D_1 \int z^M dM(\kappa) + D_2 \underline{z}^U + D_3\theta + D_4\vartheta_g$$
$$+ D_5 Prob(work)_g + D_6 y + D_7 \kappa_g^* \tag{82}$$

Secondly, note that

$$Prob(work)_g = M(\kappa_g^*) \tag{83}$$

Next, linearize the above expression for $Prob(work)$ and κ^* around $(M(\overline{\kappa}), \overline{\kappa})$ giving us

$$\kappa^* = \alpha_0 + \alpha_1 Prob(work)_g. \tag{84}$$

Finally, we replace κ^*, $\int \underline{z}^M dM(\kappa)$ and $\underline{z}^U$ by their expressions given by Equations (80) and (77) to arrive at

$$Pr(M = 1) = C_0 + C_1\theta + C_2 y + C_3 log(\vartheta_g) + C_4 Prob(work)_g. \tag{85}$$

Proof of Corollary 8.1.

Let there be a large number of economies c characterized by the model described in the paper. Let $G_c(\theta|g)$ be the distribution of entrepreneurial ability θ conditional on being from group g in economy c. Then if we aggregate Equation (15) at the economy-group wide level we obtain

$$MR_{c,g,t} = C_0 + C_1 \int \theta dG_c(\theta|g) + C_2 y_{c,t} + C_3 \vartheta_{c,g,t} + C_4 Prob(work)_{c,g,t}. \tag{86}$$

Proof of Proposition 8.

For the first part of this proof we need to obtain an equation defining tresholds $\underline{z}^M(\theta, \kappa)$ and $\underline{z}^U$. This just comes directly from the Proof of Corollary 7.1. Next, recall the expression for the entry rate in the economy is given by

$$\begin{aligned} ER_{c,g,t} = MR_{c,g,t} \int\int (\nu\psi(1-F(\underline{z}^M(\theta,\kappa,\chi_1)_{i,c,t})) + (1-\nu)\psi(1-F(\underline{z}^M(\theta,\kappa,\chi_2)_{i,c,t}))dG_c(\theta|g)dM(\kappa) \\ + (1 - MR_{c,g,t}) \int \psi(1 - F(\underline{z}^U(\theta)_{i,c,t})))dG_c(\theta|g) \end{aligned} \tag{87}$$

where i denotes individual. The expression for log average size of firms is given by

$$log(SY)_{c,g,t} = log(MR_{c,g,t} \int\int \nu \int_{\underline{z}^M(\theta,\kappa,\chi_1)_{i,c,t}} (\frac{\alpha y_{c,t}\theta_i e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{1-F(\underline{z}^M(\theta,\kappa,\chi_1)_{i,c,t})}$$
$$+ (1-\nu)\int_{\underline{z}^M(\theta,\kappa,\chi_2)_{i,c,t}} (\frac{\alpha y_{c,t}\theta_i e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{1-F(\underline{z}^M(\theta,\kappa,\chi_2)_{i,c,t})} dzdG_c(\theta|g)dM(\kappa)$$
$$+ (1-MR_{c,g,t})\int\int_{\underline{z}^U(\theta)_{i,c,t}} (\frac{\alpha y_{c,t}\theta_i e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{1-F(\underline{z}^U(\theta)_{i,c,t})} dzdG_c(\theta|g)). \quad (88)$$

Next, linearize both these expression for $(\underline{z}^M_{i,c,t}, \underline{z}^U_{i,c,t}, y_{c,t}, \theta_i)$ around $(z^*, z^*, 1, 1)$ and linearize $MR_{c,t}$ around $\hat{M}$. Then replace $\underline{z}^M_{i,c,t}$ and $\underline{z}^U_{i,c,t}$ by their expressions given by Equations (80) and (81). Next, linearize for (y, θ and $Prob(work)$) around the point $(1, 1, 1)$ to get :

$$ER_{c,g,t} = \beta_{0,1} + \beta_{1,1}MR_{c,g,t} + \beta_{2,1}Prob(work)_{c,g,t} + \beta_{3,1}y_{c,t} + \beta_{4,1}\int \theta dG(\theta|g)$$
$$+ \varepsilon_{c,g,t,1}. \quad (89)$$

and

$$log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}MR_{c,g,t} + \beta_{2,2}Prob(work)_{c,g,t} + \beta_{3,2}y_{c,t} + \beta_{4,2}\int \theta dG(\theta|g)$$
$$+ \varepsilon_{c,g,t,2}. \quad (90)$$

Proof of Proposition 9.

Start by replacing h^* by its optimal expression in the flow utility of a married household running a firm. Next, take a derivative with respect to z to obtain

$$\frac{\phi}{w}\frac{\partial \pi(z)}{\partial z} \quad (91)$$

Now take the derivative of the flow utility of an unmarried household running a firm with respect to z to obtain

$$\pi(z)^{-\sigma}\frac{\partial \pi(z)}{\partial z} \tag{92}$$

Hence, the pecuniary effect of lower total household income dominates over the saved cost of working if

$$\frac{\phi}{w} > \frac{1}{\pi(z)^{\sigma}} \tag{93}$$

In other words, if the above condition holds, the flow utility of an married household with spouse working increases faster with z than the flow utility of a unmarried household. Now note that the minimum possible profit of a firm $\min_{z,\theta} \pi(z)$ is the one evaluated at minimum values of z and θ

$$\min_{z,\theta} \pi(z) = (1-\alpha)(\frac{\alpha}{w})^{\frac{\alpha}{1-\alpha}} y^{\frac{1}{1-\alpha}}. \tag{94}$$

Then note that

$$\frac{1}{(\min_{z,\theta} \pi(z))^{\sigma}} > \frac{1}{\pi(z)^{\sigma}}. \tag{95}$$

Then it follows that if

$$\frac{\phi}{w} > \frac{1}{(\min_{z,\theta} \pi(z))^{\sigma}} \Rightarrow \frac{\phi}{w} > \frac{1}{\pi(z)^{\sigma}} \quad \text{for all } z, \theta, y. \tag{96}$$

Hence all we need is

$$\phi > \hat{\phi} \equiv \frac{w}{(\min_{z,\theta} \pi(z))^{\sigma}} \tag{97}$$

Proof of Proposition 10.

In this section, we solve for the expression for average firm productivity in the economy.

Firstly, we solve for the equilibrium measure of entrepreneurs conditional on a share of married individuals. Let MR be the measure of married individuals in the economy. Let $\eta_m(\kappa, \theta, \chi)$ be the share of individuals married to a spouse with cost of working κ, entrepreneurial ability θ, and children expenditures χ that are entrepreneurs. Let $e_m(\kappa, \theta, \chi)$ be the share of individuals married to a spouse with cost of working κ, with entrepreneurial ability θ, and children expenditures χ that are workers. Let $b_m(\kappa, \theta, \chi)$ be the share of individuals married to a spouse with cost of working κ, entrepreneurial ability θ, and children expenditures χ that are failed entrepreneurs. Similarly define $\eta_u(\theta)$, $e_u(\theta)$ and $b_u(\theta)$ as the same corresponding shares for unmarried individuals. With a slight abuse of notation, we can write the share of entrepreneurs for both groups as characterized by

$$\dot{\eta}_i(\kappa,\theta,\chi) = \psi(1 - F(\underline{z}^i(\kappa,\theta,\chi)))e_i(\kappa,\theta,\chi) - \lambda\eta_i(\kappa,\theta,\chi), \quad \forall i \in \{m,u\} \tag{98}$$

where $m =$ married, $u =$ unmarried. Similarly, the share of individuals bankrupt in both groups is characterized by

$$\dot{b}_i(\kappa,\theta,\chi) = \lambda\eta_i(\kappa,\theta,\chi) - pb_i(\kappa,\theta,\chi), \quad \forall i \in \{m,u\}. \tag{99}$$

Finally, the share of workers in both groups is characterized by

$$\eta_i(\kappa,\theta,\chi) + b_i(\kappa,\theta,\chi) + e_i(\kappa,\theta,\chi) = 1, \quad \forall i \in \{m,u\}. \tag{100}$$

Setting $\dot{\eta}(\kappa,\theta,\chi) = 0$ and $\dot{b}_i(\kappa,\theta,\chi) = 0$ and solving for these shares gives us

$$e_i(\kappa,\theta,\chi) = \frac{\lambda p}{\lambda p + \psi(1 - F(\underline{z}^i(\kappa,\theta,\chi)))(p+\lambda)}, \quad \forall i \in \{m,u\}. \tag{101}$$

$$b_i(\kappa,\theta,\chi) = \frac{\psi(1 - F(\underline{z}^i(\kappa,\theta,\chi)))\lambda}{\lambda p + \psi(1 - F(\underline{z}^i(\kappa,\theta,\chi)))(p+\lambda)}, \quad \forall i \in \{m,u\}. \tag{102}$$

$$\eta_i(\kappa,\theta,\chi) = \frac{\psi(1 - F(\underline{z}^i(\kappa,\theta,\chi)))p}{\lambda p + \psi(1 - F(\underline{z}^i(\kappa,\theta,\chi)))(p+\lambda)}, \quad \forall i \in \{m,u\}. \tag{103}$$

Futhermore, the fraction of individuals $i \in \{m,u\}$ of productivity e^z, with entrepreneurial ability θ and with cost of working κ for the spouse, and children expenditures χ, $\Gamma_i(z,\kappa,\theta,\chi)$, is characterized by

$$\dot{\Gamma}_i(z,\kappa,\theta,\chi) = \psi f(z)e_i(\kappa,\theta,\chi) - \lambda\Gamma_i(z,\kappa,\theta,\chi), \quad \forall z \geq \underline{z}^i(\kappa,\theta,\chi) \quad i \in \{m,u\} \tag{104}$$

Using the expression for $e_i(\kappa, \theta, \chi)$ and setting $\dot{\Gamma}_i(z, \kappa, \theta, \chi) = 0$ we get

$$\Gamma_i(z, \kappa, \theta, \chi) = \frac{\psi f(z) p}{\lambda p + \psi(1 - F(\underline{z}^i(\kappa, \theta, \chi)))(p + \lambda)}. \tag{105}$$

Let $G(\theta|g)$ be the measure of individuals of ability θ conditional on belonging to group g. Let η_g be the measure of individuals of group g in the economy. Then, we can define average firm productivity $E[z]$ in this economy by

$$\begin{aligned} E[z] &= MR \cdot E[z]^m + (1 - MR) \cdot E[z]^u = \\ MR \cdot \sum \eta_g \int\int (\nu \int_{\underline{z}^m(\kappa,\theta,\chi_1)} &\frac{(y\theta)^{\frac{1}{1-\alpha}} e^{\frac{z}{1-\alpha}} f(z)}{1 - F(\underline{z}^m(\kappa, \theta, \chi_1))} dz + (1-\nu) \int_{\underline{z}^m(\kappa,\theta,\chi_2)} \frac{(y\theta)^{\frac{1}{1-\alpha}} e^{\frac{z}{1-\alpha}} f(z)}{1 - F(\underline{z}^m(\kappa, \theta, \chi_2))} dz) dM(\kappa) G(\theta|g) d\theta \\ &+ (1 - MR) \cdot \sum \eta_g \int\int\int_{\underline{z}^u(\kappa,\theta)} \frac{(y\theta)^{\frac{1}{1-\alpha}} e^{\frac{z}{1-\alpha}} f(z)}{1 - F(\underline{z}^u(\kappa, \theta))} dz dM(\kappa) G(\theta|g) d\theta. \end{aligned} \tag{106}$$

where $E[z]^m$ is the average productivity among firms created by married individuals and $E[z]^u$ is the average productivity among firms created by unmarried individuals. From this expression we can see that $E[z]$ is increasing in the measure of married individuals, MR, if $\underline{z}^m(\chi_1)$ and $\underline{z}^m(\chi_2) > \underline{z}^u$ and decreasing otherwise. It follows that relative selection of both groups ($\underline{z}^m$ versus $\underline{z}^u$) is crucial for our understanding of how changes in the marriage rate affect the firm productivity distribution.

F Robustness Regressions

Table 10: Main specifications : Robustness Regressions

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta log(SY)_{c,g,t}$	$\Delta ER_{c,g,t}$	$\Delta log(SY)_{c,g,t}$
$\Delta MR_{c,g,t}$	-0.468 (0.079)	3.01 (1.05)	-0.149 (0.048)	3.64 (1.45)
IV using immigrants arrived age ≤ 15	No	No	Yes	Yes
Significance IV for $\Delta MR_{c,t}$	Yes	Yes	Yes	Yes
Cells with at least 10 men and 10 women	Yes	Yes	Yes	Yes
Observations	30906	23287	10506	9284

Notes: Regressions of changes in entry rate into entrepreneurship, $\Delta ER_{c,g,t}$ and changes in log of average firm size, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. Columns 1 and 2 report the results when excluding all triples (c, g, t) for which there is less than 10 men and 10 women. Columns 3 and 4 reports the results when constructing the gender ratio using only immigrants that arrived to Canada at the age of 15 or earlier. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

Table 11: Main specifications : Robustness Regressions - 1st Stage

Dependant Variable at 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta IV_{c,g,t}$	0.033 (0.002)	0.040 (0.002)	0.022 (0.001)	0.020 (0.001)
IV constructed with immigrants arrived age ≤ 15	No	No	Yes	Yes
Significance IV for $\Delta MR_{c,t}$	Yes	Yes	Yes	Yes
Only cells with at least 10 men and 10 women	Yes	Yes	Yes	Yes
Observations	30906	23287	10506	9284

Notes: First Stage regression of our endogenous variable (for the entry rate regressions), change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$ in our instrument, the change in the gender ratio at economic region c, country of birth g and year t $IV_{c,g,t}$. Second Stage are regressions of changes in entry rate into entrepreneurship, $\Delta ER_{c,g,t}$ and changes in log of average firm size, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. Columns 1 and 2 report the results when excluding all triples (c, g, t) for which there is less than 10 men and 10 women. Columns 3 and 4 reports the results when constructing the gender ratio using only immigrants that arrived to Canada at the age of 15 or earlier. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

G Alternative mechanisms

Table 12 reports results for when we exclude high capital requirement industries. We exclude firms started in Oil, Mining, Gas, Utilities, Manufacturing, Wholesale Trade, Retail Trade, Transportation and Warehousing, Information and Cultural Industries and Real Estate. We set entry into entrepreneurship into high capital industries equal to zero (coded as a non-entry for the purpose of these regressions). Hence, the number of observations for our entry into entrepreneurship regression remains unchanged relative to our benchmark specification. For the average size regressions on the other hand we all businesses started in high capital requirement industries are no longer included. First Stage results can be found at Table 15.

Table 12: Regressions excluding high capital industries

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta MR_{c,g,t}$	-0.167 (0.027)	1.098 (0.385)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
Observations	70551	27881

Notes: Regressions of changes in entry rate into entrepreneurship excluding entry into high capital requirement industries, $\Delta ER_{c,g,t}$ and changes in log of average firm size excluding firms in high capital requirement industries, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument for $\Delta MR_{c,g,t}$ is the change in the gender ratio $\Delta\vartheta^{-1}_{c,g,t}$. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

Table 13 reports results for when we exclude businesses owned by both in couple. We set entry into entrepreneurship equal to zero (coded as non-entry for the purpose of these regressions) if both individuals in the couple are listed as owners. Note that this implies the number of observations for our entry into entrepreneurship regression remains unchanged relative to our benchmark specification. On the other hand, for the average size regressions, since we exclude all firms owned by both individuals of the couple, the number of observations is smaller relative to our benchmark specification. First Stage results can be found at Table 16.

Table 13: Regressions excluding businesses owned by both in couple

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta MR_{c,g,t}$	-0.19	1.32
	(0.031)	(0.464)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
Observations	70551	27702

Notes: Regressions of changes in entry rate into entrepreneurship excluding entry into entrepreneurship of firms owned by both in couple, $\Delta ER_{c,g,t}$ and changes in log of average firm size excluding firms owned by both in couple, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument for $\Delta MR_{c,g,t}$ is the change in the gender ratio, $\Delta\vartheta^{-1}_{c,g,t}$. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

Table 14 reports results for when we exclude businesses for which the spouse is listed as an employee. We set entry into entrepreneurship equal to zero (coded as non-entry for the purpose of these regressions) if the spouse is listed as an employee of the firm. The result is that the number of observations for our entry into entrepreneurship regression remains unchanged relative to our benchmark specification. On the other hand, for the average size regressions since we exclude all firms for which the spouse is listed as employee of the firm, the number of observations is smaller relative to our benchmark specification. First Stage results can be found at Table 17.

Table 14: Regressions excluding businesses where spouse listed as employee

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta MR_{c,g,t}$	-0.180 (0.031)	1.74 (0.56)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
Observations	70551	26945

Notes: Regressions of changes in entry rate into entrepreneurship excluding entry of firms for which the spouse is listed as an employee, $\Delta ER_{c,g,t}$ and changes in log of average firm size excluding firms in which the spouse is listed as an employee, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument for $\Delta MR_{c,g,t}$ is the change in the gender ratio, $\Delta \vartheta^{-1}_{c,g,t}$. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

Table 15: Regressions excluding high capital industries - 1st Stage

Dependant Variable at 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta IV_{c,g,t}$	0.037 (0.001)	0.03 (0.001)
Observations	70551	27881

Notes: First Stage regression of our endogenous variable (for the entry rate regressions), change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$ in our instrument, the change in the gender ratio at economic region c, country of birth g and year t $IV_{c,g,t}$. Second Stage are regressions of changes in entry rate into entrepreneurship excluding entry into high capital requirement industries, $\Delta ER_{c,g,t}$ and changes in log of average firm size excluding firms in high capital requirement industries, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument for $\Delta MR_{c,g,t}$ is the change in the gender ratio, $\Delta \vartheta^{-1}_{c,g,t}$. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

Table 16: Regressions excluding businesses owned by both in couple - 1st Stage

Dependant Variable at 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta IV_{c,g,t}$	0.0371	0.032
	(0.001)	(0.002)
Observations	70551	27702

Notes: First Stage regression of our endogenous variable (for the entry rate regressions), change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$ in our instrument, the change in the gender ratio at economic region c, country of birth g and year t $IV_{c,g,t}$. Second Stage are regressions of changes in entry rate into entrepreneurship excluding entry by firms owned by both in couple, $\Delta ER_{c,g,t}$ and changes in log of average firm size excluding firms owned by both in couple, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument for $\Delta MR_{c,g,t}$ is the change in the gender ratio, $\Delta \vartheta^{-1}_{c,g,t}$. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

Table 17: Regressions excluding businesses where spouse listed as employee - 1st Stage

Dependant Variable in the 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta IV_{c,g,t}$	0.037	0.035
	(0.001)	(0.002)
Observations	70551	26945

Notes: First Stage regression of our endogenous variable (for the entry rate regressions), change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$ in our instrument, the change in the gender ratio at economic region c, country of birth g and year t $IV_{c,g,t}$. Second Stage are regressions of changes in entry rate into entrepreneurship excluding entry by firms where spouse is listed as an employee, $\Delta ER_{c,g,t}$ and changes in log of average firm size excluding firms where spouse is listed as an employee, $\Delta SY_{c,g,t}$ in economic region c, country of birth g and year t, on the change in the marriage rate in economic region c, country of birth g, year t, $\Delta MR_{c,g,t}$. The instrument for $\Delta MR_{c,g,t}$ is the change in the gender ratio, $\Delta \vartheta^{-1}_{c,g,t}$. All specifications include year dummies to capture national trends, changes in female labor force among married women, changes in average income among married men, changes in the share of population at each triplet (c, g, t) within different age ranges (for 3 different age ranges), changes in the employment shares of oil and gas sector, changes in the employment shares of the manufacturing sector and changes in employment shares of the services sector. Standard errors are clustered at the economic region c and country of birth g level.

H Ranking of flow utility when spouse not working

In this Section we show that whether or not married households have larger flow utility relative to unmarried households when the spouse is not working depends on the size of the utility associated to being in a couple φ.

To see this compare two households where the main earner is in the same state S, where S is either entrepreneurship, working or having failed a firm. Let I_S be the income associated to each of these states. Then,

$$\mu\frac{[\gamma(I_S-\chi)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\gamma)(I_S-\chi)]^{1-\sigma}}{1-\sigma}+\varphi$$
$$<\mu\frac{(I_S-\chi)^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{(I_S-\chi)^{1-\sigma}}{1-\sigma}+\varphi=\frac{(I_S-\chi)^{1-\sigma}}{1-\sigma}+\varphi\lesseqgtr\frac{I_S^{1-\sigma}}{1-\sigma}\Rightarrow\text{depends on size of }\varphi. \tag{107}$$

I Proof of solution for standard collective household model - Internet Appendix

In this Section I show that the restriction that the income sharing parameter γ is exogenous is not without loss of generality relative to standard collective household models. In particular, it suffices to consider $\gamma\equiv\frac{c_f}{I+wh}=(\frac{\mu^{\frac{1}{\sigma}}}{(1-\mu)^{\frac{1}{\sigma}}+\mu^{\frac{1}{\sigma}}})$ to get the standard collective household solution.

In our context, a standard collective household model would imply the following household problem

$$\max_{c_m,c_f,h}\mu\frac{c_f^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{c_m^{1-\sigma}}{1-\sigma}+\varphi-\phi h\quad\text{subject to}\quad c_m+c_f=wh+I-\chi \tag{108}$$

where c_m is the main earner's consumption and c_f is the consumption of the spouse. The solution to the problem above gives $\gamma\equiv\frac{c_f}{I+wh-\chi}=(\frac{\mu^{\frac{1}{\sigma}}}{(1-\mu)^{\frac{1}{\sigma}}+\mu^{\frac{1}{\sigma}}})$.We would like to thank an anonymous referee for drawing our attention to this point.

J Internet Appendix : Relaxing the Exclusion Restriction

In this Section of the Appendix, I show how we can relax the exclusion restriction of our instrument and still obtain our main results. In particular, let the parameter how income is shared within the household, γ, to be a function of the gender ratio ϑ_g. This is consistent with recent empirical literature showing how changes in household behaviour are consistent with the gender ratio affecting intra-household income sharing (Chiappori et al. (2002) and Angrist (2002)).[22]

Note that this extension of the baseline model differs from standard collective, cooperative, household models (Browning et al. (2006)). The motivation for this choice is the intuition that if there is an increase in number of gender g_1 individuals relative to gender g_2, individual of gender g_1 can threaten to leave the relationship to extract a higher share of household income. In this sense, our model has a structure that combines non-cooperative household behaviour components to a standard cooperative household setting. We do not consider a full fledged non-cooperative household to keep the model tractable enough to bring to the data. In particular, the main objective of allowing for the income sharing parameter γ to be a function of the gender ratio is to show how the exclusion restriction for our instrument guides the set of controls we need to include in our regressions.

The probability of working for a spouse can be written as

$$Prob(work) = M[\int (U_s^m(I, \gamma(\vartheta_g)) - U_{ns}^m(I, \gamma(\vartheta_g)))f(I|g)dI]. \tag{109}$$

where the only difference relative to the expression in the main text is that γ is now a function of ϑ_g. Note that once we condition on the main earner's income, I, ϑ_g only impacts the probability of working of the spouse via $\gamma(\vartheta_g)$. Later, when we proceed to our empirical analysis, this is useful by allowing us to control for variation in γ due to ϑ via $Prob(Work)$. Empirically, $Prob(Work)$ is the labor force participation of spouses. This dependancy will be important as it will be informative for our identification strategy and the conditions for it's validity.

Furthermore, note that eventhough the household is making a joint decision, variation in $\gamma(\vartheta)$ still affects total household utility. The intuition is that a larger share of total household income going to the spouse, larger γ, affects the household differently depending on μ. If there is a large weight on spousal utility (high μ) and most of the income is being taken by the main earner (low γ), we expect that a larger γ increases household utility. If, on the other hand, little weight is put on spousal utility (low μ) and most of the income is being taken by the spouse (high γ) then an increase in γ is likely to be detrimental for total household utility.

[22] We leave for future work a micro foundation of this dependancy of income sharing on the gender ratio in the context of our model.

Next, we can derive an equivalent to Proposition 8 in the main text to obtain linear expressions for the entry rate into entrepreneurship $ER_{c,g,t}$ and the average size of firms $log(SY)_{c,g,t}$

Proposition 11 *Suppose there exists a large number of economies c all of which are characterized by the model described in the previous section. Let $G_c(\theta|g)$ be the distribution of innate ability θ for group g, in economy c. Let $\gamma_{c,g,t}$ be the share of the income of the main earner that the spouse consumes in economy c for group g at time t. Let $Prob(work)_{c,g,t}$ be the probability that spouses from group g in local economy c, year t work. Finally, let the entry rate into entrepreneurship ER have measurement error, $\varepsilon_{c,g,t,1}$, characterized by its first differences, $\Delta\varepsilon_{c,g,t,1}$, being i.i.d and $E[\Delta\varepsilon_{c,g,t,1}] = 0$ and the average size of firms SY have measurement error, $\varepsilon_{c,g,t,2}$, characterized by its first differences, $\Delta\varepsilon_{c,g,t,2}$, being i.i.d and $E[\Delta\varepsilon_{c,g,t,2}] = 0$. Then, the entry rate into entrepreneurship ER and the average size of firms SY in each of these c economies for a group g can be written as*

$$ER_{c,g,t} = \beta_{0,1} + \beta_{1,1}MR_{c,g,t} + \beta_{2,1}\gamma_{c,g,t} + \beta_{3,1}y_{c,t} + \beta_{4,1}\int \theta dG(\theta|g) + \beta_{5,1}Prob(work)_{c,g,t} + \varepsilon_{c,g,t,1}. \quad (110)$$

and

$$log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}MR_{c,g,t} + \beta_{2,2}\gamma_{c,g,t} + \beta_{3,2}y_{c,t} + \beta_{4,2}\int \theta dG(\theta|g) + \beta_{5,2}Prob(work)_{c,g,t} + \varepsilon_{c,g,t,2}. \quad (111)$$

where

$$\gamma_{c,g,t} = \zeta_0 + \zeta_1\vartheta_{c,g,t} \quad (112)$$

Proposition 11 makes clear that any instrument for marriage rates $MR_{c,g,t}$ must be independant of variation in $\gamma_{c,g,t}$, $y_{c,t}$ and $\int \theta dG(\theta|g)$. Importantly, note that the only channel via which the gender ratio, $\vartheta_{c,g,t}$, has an effect on $ER_{c,g,t}$ and $log(SY)_{c,g,t}$ is via the income sharing parameter, $\gamma_{c,g,t}$. If we are able to control for $\gamma_{c,g,t}$, we can control for the effect of $\vartheta_{c,g,t}$ on entrepreneurial outcomes, and use $\vartheta_{c,g,t}$ as an instrument for marriage rates.

Now recall that our model implies a tight relationship between the probability of working for spouses,

$Prob(work)_{c,g,t}$, and the income sharing parameter, $\gamma_{c,g,t}$. Proposition 12 below makes clear that we can use $Prob(work)_{c,g,t}$ and average income of married men, $\int Id\mu_{c,t}(I|g)$, as a proxies to control for $\gamma_{c,g,t}$.

Proposition 12 *Suppose there exists a large number of economies c all of which are characterized by the model described in the previous section. Let $\mu_{c,t}(I|g)$ be the income distribution of married main earners in group g in economy c at time t. Then the probability of working for spouses in economy c in group g, time t, $Prob(work)_{c,g,t}$, can be approximated by*

$$Prob(work)_{c,g,t} = \beta_{0,4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4}\int Id\mu_{c,t}(I|g) \tag{113}$$

From the expression above we can write $\gamma_{c,g,t}$ as a function of $Prob(Work)_{c,g,t}$ and $\int log(I)d\mu_{c,t}(I|g)$. If we replace this expression for $\gamma_{c,g,t}$ in Equations (89) and (111) we get

$$ER_{c,g,t} = \beta_{0,1} + \beta_{1,1}MR_{c,g,t} + \zeta_{2,1}Prob(Work)_{c,g,t} + \zeta_{2,2}\int Id\mu_{c,t}(I|g) + \beta_{3,1}y_{c,t} + \beta_{4,1}\int \theta dG(\theta|g) + \varepsilon_{c,g,t,1}. \tag{114}$$

and

$$log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}MR_{c,g,t} + \zeta_{3,1}Prob(Work_{c,g,t}) + \zeta_{3,2}\int Id\mu_{c,t}(I|g) + \beta_{3,2}y_{c,t} + \beta_{4,2}\int \theta dG(\theta|g) + \varepsilon_{c,g,t,2}. \tag{115}$$

Note that $Prob(Work)_{c,g,t}$ is the labor force participation of spouses and $\int Id\mu_{c,t}(I|g)$ is average income of married main earners for group g, economy c and time t. We can measure and control for these variables, which leaves only $y_{c,t}$, $\int \theta dG(\theta|g)$ and $\varepsilon_{c,g,t}$ as unobserved error terms. Next, we take first differences to obtain

$$\Delta ER_{c,g,t} = \beta_{1,1}\Delta MR_{c,g,t} + \zeta_{2,1}\Delta Prob(Work)_{c,g,t} + \zeta_{2,2}\Delta\int Id\mu_{c,t}(I|g) + \beta_{3,1}\Delta y_{c,t} + \Delta\varepsilon_{c,g,t,1}. \tag{116}$$

and

$$\Delta log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}\Delta MR_{c,g,t} + \zeta_{3,1}\Delta Prob(Work)_{c,g,t} + \zeta_{3,2}\Delta \int Id\mu_{c,t}(I|g) + \beta_{3,2}\Delta y_{c,t} + \Delta\varepsilon_{c,g,t,2}. \quad (117)$$

Hence, after taking first differences the term $\int \theta dG(\theta|g)$ disappears. Given our equations above, all we need is an instrument for the change in marriage rates, $\Delta MR_{c,g,t}$, which is uncorrelated to changes in economy specific productivity shocks, $\Delta y_{c,t}$. Looking back at our expression for marriage rates in equation (20) and taking first differences we get

$$\Delta MR_{c,g,t} = C_2\Delta y_{c,t} + C_3\Delta\vartheta_{c,g,t} + C_4\Delta Prob(work)_{c,g,t}. \quad (118)$$

The expression above for $\Delta MR_{c,g,t}$ makes explicit that a candidate for changes in marriage rates independant of $\Delta y_{c,t}$ are changes in the gender ratio, $\Delta\vartheta_{c,g,t}$. We adopt this approach, such that our instrument $IV_{c,g,t}$ for $\Delta MR_{c,g,t}$ is defined as

$$IV_{c,g,t} = \Delta\vartheta_{c,g,t}. \quad (119)$$

The key restriction that allows for this empirical strategy is that we control for the impact of $\Delta\vartheta_{c,g,t}$ on entrepreneurial outcomes by controlling for labor force participation of spouses, $\Delta Prob(Work)_{c,g,t}$, and average income of married main earners, $\Delta \int Id\mu_{c,t}(I|g)$. To summarize our two main specifications are

$$\Delta ER_{c,g,t} = \beta_{1,1}\Delta MR_{c,g,t} + \zeta_{2,1}\Delta Prob(Work)_{c,g,t} + \zeta_{2,2}\Delta \int Id\mu_{c,t}(I|g) + \beta_{3,1}\Delta y_{c,t} + \zeta_{2,3}\Delta\varepsilon_{c,g,t}. \quad (120)$$

and

$$\Delta log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}\Delta MR_{c,g,t} + \zeta_{3,1}\Delta Prob(Work)_{c,g,t} + \zeta_{3,2}\Delta \int Id\mu_{c,t}(I|g) + \beta_{3,2}\Delta y_{c,t} + \zeta_{3,3}\Delta\varepsilon_{c,g,t}. \quad (121)$$

where $\Delta MR_{c,g,t}$ is instrumented by $\Delta\vartheta_{c,g,t}$. Recall that we have the prediction from the model that

- $\beta_{1,1} > 0$ if and only if $\beta_{1,2} < 0$,
- $\beta_{1,1} < 0$ if and only if $\beta_{1,2} > 0$.

Hence, we conclude that if we control for changes in labor force participation of spouses and changes in average income of married main earners, our exclusion restriction continues to hold. Since our final empirical specifications already include these controls, we conclude that allowing for an effect of the gender ratio in the income sharing parameter γ does not change our results and their interpretation.

Proof of Proposition 11.

Follow the steps for the proof of Corollaries 7.1, 8.1 and Proposition 8 except that linearizing $\underline{z}^M(\kappa), \underline{z}^U, y, \theta$ and γ instead of just $\underline{z}^M(\kappa), \underline{z}^U, y, \theta$ like before. To obtain the expression

$$\gamma_{c,g,t} = \zeta_0 + \zeta_1 \vartheta_{c,g,t} \tag{122}$$

linearize $\gamma = \gamma(\vartheta)$ for (γ, ϑ) around $(1, 1)$. To get an expression of $Pr(M = 1)$ and $MR_{c,g,t}$ that do not depend on $\vartheta_{c,g,t}$ just replace $\gamma_{c,g,t}$ in the expression for $Pr(M = 1)$ by $\zeta_0 + \zeta_1 \vartheta_{c,g,t}$.

Proof of Proposition 12.

To start consider equation 109. Then, log-linearize the terms (I, γ) inside $U_s^m(I, \gamma_g)$ and $U_{ns}^m(I, \gamma_g)$ around $(I^*, \frac{1}{2})$ to get

$$Prob(work|I)_{c,g,t} = \beta_{0.4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4} I. \tag{123}$$

Now if we integrate over I we obtain

$$Prob(work)_{c,g,t} = \beta_{0.4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4} \int I d\mu_{c,t}(I|g). \tag{124}$$

where $\mu(I|g)$ is the distribution of main earner income for individuals in group g.

Finally, with the assumption that $log(Prob(work)_{c,g,t})$ is observed with measurement error $\epsilon_{c,g,t}$ we obtain the desired result

$$Prob(work)_{c,g,t} = \beta_{0.4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4} \int I d\mu_{c,t}(I|g) + \epsilon_{c,g,t}. \tag{125}$$