Spouses, Children and Entrepreneurship*

João Galindo da Fonseca† Charles Berubé ‡

March 4, 2020

Abstract

We develop a model of endogenous entrepreneurship and marriage. Spouses influence entrepreneurship via three channels: they reduce benefits by working less the more profitable the business is, they reduce costs by working more in case of business failure, and children, associated with a spouse, increase the cost of failure. We use administrative matched owner-employer-employee-spouse data to estimate the specifications derived from our model. The model is informative on the sources of endogeneity and the IV strategy. We show that higher marriage rates induce less entry but larger firms on average. Through the lens of our model, marriage increases firm productivity.

(JEL-Codes : E24, E23, J12, J60)

Keywords : Labor Markets, Family Economics, Macroeconomics, Firm dynamics.

*We would like to thank Giovanni Gallipoli, Thomas Lemieux, David Green, Raphaël Godefroy, Joshua Lewis, 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

How does having a spouse affect the decision to start a firm? How does it affect the type of firms created by individuals? The answer to this question is crucial if we wish to understand how changes in household composition affect firm creation and productivity.

We know there have been secular changes to household composition. This is best portrayed by the decline in marriage rates, from 62% in 1990 to 55% in 2013 in the US and from 59.71% in 1984 to 45.58% in 2013 for Canada, and the rise of one person households, from 9% to 12% in the US and from 20.3% in 1981 to 27.6% in 2011 in Canada.[1] To evaluate how these changes in household composition affected productivity in the economy, we need to understand the interplay between firm formation, composition of new firms and marriage.[2] Yet, despite the large literature on entrepreneurship and firm formation (Lucas Jr (1978), Hurst and Lusardi (2004), Cagetti et al. (2006), Quadrini (2000), Beaudry et al. (2018) and Haltiwanger et al. (2013)), there has been no research on how marriage can affect both entry into entrepreneurship and firm size.

To study the role of spouses on the entrepreneurship of the main earner we propose a tractable model of endogenous entrepreneurship, endogenous marriage, endogenous spousal labour supply, heterogeneity in innate entrepreneurial ability and business project specific productivity. In the model, main earners choose ex-ante whether to marry based on a taste for marriage, their innate entrepreneurial ability and the overall productivity of the economy. Households can either be two persons or single person households.[3]

Main earners in the household draw business opportunities from an exogenous distribution. In equilibrium, for all business opportunities with productivity above their optimal threshold, the household decides the main earner should start a firm.

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

[2]Of course 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]To our knowledge we are the first to consider a model with endogenous marriage, endogenous entrepreneurship and endogenous spousal labor supply.

In two person households, main earner and spouse share their income, such that the consumption of each individual depends on both their income and of their partner. The household chooses whether spouses work or not (associated to a fixed cost) and if so how many hours they work.

This tractable framework delivers three channels via which being a couple influences the decision to start a firm. First, any increase in profits is partly offset by spouses working less hours or by the need for it to be shared with non-working spouses (spousal sharing effect). This channel makes entrepreneurship less attractive to married households which in turn become more selective on which business projects to implement.

Secondly, if the business fails, the spouse works more hours (spousal insurance effect). This channel is a natural application of the added worker effect, vastly studied in labor economics (See Hyslop (2001), Stephens (2002), Gallipoli and Turner (2009), Blundell et al. (2016), Wu and Krueger (2018)), to business risk. This insurance channel decreases the cost of failure for married households. This pushes a married household to become less selective in which business projects to implement.

Finally, being in a couple is often associated to having children (offspring effect). Children have three ways of affecting the decision to start a firm. First of all, children increase the cost of business failure. Secondly, children decrease the benefits to a more profitable business since some of these higher profits are shared with the children. These first two channels make children an additional incentive for married households to be more selective on which business projects to implement. Third, when both spouses are working, children decrease total consumption of the couple for a given total household income. Since consumption when both are working represents the opportunity cost to entrepreneurship, all else equal, children decrease the opportunity cost to entrepreneurship.[4] Hence, there is ambiguity in whether children induce lower or higher entry rate and average size.

Although the theoretical response of entry rates into entrepreneurship to marriage is ambiguous, the model delivers a sharp prediction that marriage rates affect

[4]See Galindo da Fonseca (2019) for a recent paper on the importance of opportunity costs for entrepreneurship.

average size of firms in the opposite direction it affects entry rates. This empirical implication comes from our mechanism of selection upon entry into entrepreneurship driving the differences between married and unmarried individuals of same ability. Instead of imposing it in our empirical estimation, we verify it holds in data.

Next, we proceed to an empirical analysis with the goal of investigating the relative strenghts of these channels and testing the model's empirical prediction. We make use of full universe confidential tax data with links between firms, employees and firm owners. The dataset is further merged to immigration landing files since 1980. Finally, the additional linkage between individuals and their spouses allows us to perform robustness tests otherwise impossible in other datasets. In particular, we verify our results are not being driven by firms jointly owned by spouses or those for which the spouse is employed by their partner.

We use the model to derive our instrument, the conditions under which it provides consistent estimates of the coefficients of interest and to derive implied restrictions in the data. This strategy fits with a literature that puts together structural modelling and instrumental variable estimation (Blundell et al. (1998), Beaudry et al. (2012), Beaudry et al. (2018), Tschopp (2015) and Green et al. (2017))). We focus on estimating first order implications of the theory as implied by its linear approximation. We use city-country of birth level variation to test the model. Our instrument for the marriage rate is the gender ratio (number of women divided by men) for a particular city-country of birth pair. The intuition is that men belonging to groups with higher number of women relative to men are more likely to find a potential partner (Angrist (2002)).

Through the lens of the model, we allow for the gender ratio to directly affect entrepreneurship. 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 this possible concern. In particular, the gender ratio affects intra household income sharing.[5] This in turn induces changes in female labor force participation and entrepreneurship of the mar-

[5]This is consistent with empirical work on sex ratios and intrahousehold resource allocation (Chiappori et al. (2002) and Angrist (2002)).

ried household. Hence, from the model, we derive a strict relationship between the intra household income sharing parameter and female labor force participation. We show that controlling for female labor force participation we control for variation in intra household income sharing. This shuts down the direct effect of the gender ratio on entrepreneurship. In absence of this direct effect, the exclusion restriction holds and we can use the gender ratio as an instrument for marriage rates.

We find that a 1 percentage point increase in the marriage rate is associated to a 0.2 percentage point decrease in the entry rate and a 1.13% increase in average size of firms. These findings are consistent with our prediction that the effect of marriage on average size has an opposite sign to the effect on entry. We interpret this as a general test of selection mechanisms commonly used in the firm dynamics literature. Although widely used, little empirical work has been done to test this underlying restriction present in selection-based models. Our results are consistent with the spousal sharing effect or the larger cost of failure associated to children dominating over the spousal insurance effect.

To distance our instrument from any endogenous migration choice in the part of immigrants, we verify our results are robust to using only immigrants having arrived at age 15 or younger. One concern is that our results are being driven by borrowing constraints, more easily overcome by two person households. To adress this issue, we verify our results are robust in both magnitude and significance to excluding high capital industries. We also verify our results hold for both men and women.

Finally, we use our estimates to perform a bounding exercise for the effect of changes in the marriage rate on firm productivity. Our results imply a decrease in marriage rates is associated to a fall in average firm productivity.

2 Model

In this section I go over the main model of the paper. The main objective is to derive the intuition for the different forces that come into play when an individual decides to start a firm as a function of having a spouse or not. The model also delivers the main empirical specifications to be estimated in the data. We start by describing individual choices conditional on marriage and individual entrepreneurial ability θ. After having done that, we describe the endogenous decision to marry as a function of entrepreneurial ability θ and an individual taste for marriage.

In the model we abstract from the possibility of the spouse helping out in the business started by their partner. This assumption is consistent with our results in the empirical section. In particular, in our data we find that the majority of new entries are started exclusively by only one member of the household and do not list the spouse as an employee. We also verify our main empirical results are not being driven by the presence of this subgroup of firms in which the spouse is participating in the business. We also abstract from the benefit of entrepreneurs receiving employer provided health insurance via their spouse. This choice is consistent with our data where such considerations are likely of second order due to Canada's public health care system. In this sense, the usage of Canadian data allows us to control for a channel that otherwise would be present in the case of the US. Finally, we do not have borrowing constraints in the model. This is motivated by our finding in the empirical section that the results are unchanged once we restrict our attention to low capital requirement industries.[6]

In the economy there are two types of households. Single individual households are composed of only one individual who is also the main earner. The second type of household is that of couples. These are composed of one main earner and one spouse. There is no savings and each household consumes their current income. We consider CRRA utility with 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

$$U^s(I) \equiv \frac{I^{1-\sigma}}{1-\sigma}. \tag{1}$$

For households composed of two individuals, there are two sources of income, one is the income of the main earner, I, and the other is the income of the spouse, wh. The income of the spouse is a function of how many hours the household chooses

[6]This is not to say that borrowing constraints are not important for the decision to start a firm. Rather, it says that the differences in entrepreneurship caused by marriage do not seem to be driven by borrowing constraint considerations.

for the spouse to work, h, and the wage paid to do so, w.

Being in a couple is often associated to having children. We model children by a share χ of the married household income that is neither consumed by the spouse or the main earner. The choice of modelling children this way is to keep the model tractable while already allowing us to talk about the higher cost of business failure for couples with children.[7] Of the remaining $1-\chi$, let γ be the share of income that is left for the spouse to consume. Let each individual belong to a group g. Each group g defines the relevant marriage pool for an individual. Next, to allow for some reduced form intra household reallocation let γ be a function of the gender ratio in the individual's group g, 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)).[8] Individuals take ϑ as given. The household has a cost ϕh associated to a spouse working h hours. This disutility is paid by the entire household. 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

$$\mu\frac{[(1-\chi)\gamma(\vartheta_g)(I+wh)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))(I+wh)]^{1-\sigma}}{1-\sigma}-\phi h \quad (2)$$

where μ is the exogenous weight placed on the utility of the spouse and $\gamma(\vartheta_g)$ is such that $0\leq\gamma(\vartheta_g)\leq 1, \quad \forall\vartheta_g$. Note that μ is a preference parameter while γ is how the income is shared within the household. The married household solves the static problem of how many hours should the spouse work. As a result, conditional on the spouse working, the household solves

$$\max_{0\leq h\leq 1}\mu\frac{[(1-\chi)\gamma(\vartheta_g)(I+wh)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))(I+wh)]^{1-\sigma}}{1-\sigma}-\phi h \quad (3)$$

[7] We thank Tiago Cavalcanti for suggesting this modelling strategy.

[8] Note that the objective here is not to micro found changes in intra household income sharing or to study this channel. Rather, the goal is to allow for the possibility that the gender ratio affects intra household income sharing and to show we can still obtain identification.

This gives us

$$h^* = \frac{[\mu((1-\chi)\gamma(\vartheta_g))^{1-\sigma} + (1-\mu)((1-\chi)(1-\gamma(\vartheta_g)))^{1-\sigma}]^{\frac{1}{\sigma}}}{\phi^{\frac{1}{\sigma}} w^{\frac{\sigma-1}{\sigma}}} - \frac{I}{w} \tag{4}$$

when h is an interior solution and 0 or 1 when we get corner solutions. Once we replace this expression in the flow utility of the couple, we obtain as total flow utility for the household with the spouse working, $U^m_s(I)$:

$$\begin{aligned} U^m_s(I,\gamma(\vartheta)) \equiv & \frac{\phi^{\frac{\sigma-1}{\sigma}}}{w^{\frac{\sigma-1}{\sigma}}}[\mu((1-\chi)\gamma(\vartheta_g))^{1-\sigma} + (1-\mu)((1-\chi)(1-\gamma(\vartheta_g)))^{1-\sigma}]^{\frac{1-\sigma}{\sigma}} \\ & \cdot [\frac{\mu[(1-\chi)\gamma(\vartheta_g)]^{1-\sigma} + (1-\mu)[(1-\chi)(1-\gamma(\vartheta_g))]^{1-\sigma}}{1-\sigma}] \\ & - \frac{\phi^{\frac{\sigma-1}{\sigma}}}{w^{\frac{\sigma-1}{\sigma}}}[\mu((1-\chi)\gamma(\vartheta_g))^{1-\sigma} + (1-\mu)((1-\chi)(1-\gamma(\vartheta_g)))^{1-\sigma}]^{\frac{1}{\sigma}} + \phi\frac{I}{w} \end{aligned} \tag{5}$$

Now, consider as well that the household makes the decision on whether the spouse should work or not. Each spouse has a fixed cost of working $\kappa \sim M(\kappa)$. This decision is done ex-ante by the household and is irreversible. κ is paid only once upon the decision. This choice is done by the household with knowledge of their type g but not knowing which income the main earner is going to have. Define $U^m_{ns}(I,\gamma(\vartheta_g))$ as the flow utility of a married household composed of a main earner with income I and a non-working spouse,

$$U^m_{ns}(I,\gamma(\vartheta_g)) = \mu\frac{[(1-\chi)\gamma(\vartheta_g)I]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))I]^{1-\sigma}}{1-\sigma}. \tag{6}$$

Since the choice of the spouse working or not is done by the household, the household pays κ and the spouse works if

$$\kappa^*_g \equiv \int (U^m_s(I,\gamma(\vartheta_g)) - U^m_{ns}(I,\gamma(\vartheta_g)))f(I|g)dI > \kappa \tag{7}$$

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,

$$Prob(work) = M[\int (U_s^m(I, \gamma(\vartheta_g)) - U_{ns}^m(I, \gamma(\vartheta_g)))f(I|g)dI]. \tag{8}$$

Note that once we condition on income of the main earner, 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 married women. In what follows, I omit the notation $\gamma(\vartheta_g)$ to make it lighter. However, when we arrive to the empirical section, 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.

Let us consider the problem of a currently operating entrepreneur. Each main earner has innate entrepreneurial ability of θ. Let $\theta \in [1, T]$ and $\theta \sim G(\theta|g)$. Let the productivity of the firm be a function of an economy-wide productivity component y. In other words, y is the same for all entrepreneurs in a same economy while θ is an individual ability component that varies across entrepreneurs. Finally, let each business project an entrepreneur starts to be characterized by a productivity z. Let $z \in [0, \Upsilon]$. Throughout the life of the firm, productivity z is fixed. The individual entrepreneur takes wages w and firm productivity components z, θ and y as given and chooses how many individuals n to hire. Hence, they solve[9]

$$\pi(z) = \max_n y\theta e^z n^\alpha - wn \tag{9}$$

[9]The profit maximization problem is static because there are no search frictions.

which gives

$$n(z,w) = (\frac{\alpha y\theta e^z}{w})^{\frac{1}{1-\alpha}} \tag{10}$$

and

$$\pi(z) = (1-\alpha)(\frac{\alpha}{w})^{\frac{\alpha}{1-\alpha}} e^{\frac{z}{1-\alpha}} (y\theta)^{\frac{1}{1-\alpha}}. \tag{11}$$

Next we go over the value functions of the married and unmarried households. Note that for the married households these depend on the fixed cost, κ, of the spouse.

For the remainder of this section we omit the dependance of value functions on θ to make the notation lighter. 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 and B^u represent the value of failing a business without a spouse, then,

$$rJ^u(z) = \frac{\pi(z)^{1-\sigma}}{1-\sigma} + \lambda(B^u - J^u(z)). \tag{12}$$

For married households, the main earner gets a share $(1-\chi)(1-\gamma)$ of the income of the household and the spouse gets a share $(1-\chi)\gamma$. The total income of the household is composed of both the income of the entrepreneur $\pi(z)$ and of the spouse wh^*. 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 and $B^m(\kappa)$ represent the value function for a household composed of a failed entrepreneur with a spouse of working cost κ. Then for $\kappa < \kappa^*$, the spouse works and the value function of the household, $J^m(z,\kappa)$, is

$$\begin{aligned} rJ^m(z,\kappa) = \mu\frac{[(1-\chi)\gamma(\vartheta_g)(\pi(z)+wh^*)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))(\pi(z)+wh^*)]^{1-\sigma}}{1-\sigma} \\ - \phi h^* + \lambda(B^w(\kappa) - J^m(z,\kappa)). \end{aligned} \tag{13}$$

Note that the difference between the value of being an entrepreneur between the two groups depends on any differences in the value of having failed a business, B^m versus B^u, and differences in their flow utility. In particular, while the unmarried individual gets to keep all of the profits, the married household pools the income of the main earner and the spouse. Note that the income as an entrepreneur when unmarried is increasing in z.[10] On the other hand, for the married household as their entrepreneurial profits, $\pi(z)$, increase in z, hours worked by the spouse, h, weakly decrease. Hence, spousal income decreases, partially offsetting some of the increase in income due to higher profits. This 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. This is an important effect which we call the **spousal sharing effect**. It can be shown that there exists a $\hat{\phi}$, such that $\forall \phi < \hat{\phi}$ the pecuniary effect of lower total household income dominates over the saved cost of working. Proposition 1 below states this formally.

Proposition 1 *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.*

From hereafter we focus on the cases where $\phi < \hat{\phi}$. This is the first part of the **spousal sharing effect**. It compresses the benefits to entrepreneurship for married individuals relative to the unmarried. Note that conditional on already operating a firm, business project productivity z is fixed over time. Hence, the **spousal sharing effect** does not change the risk for an already operating entrepreneur. Instead, it compresses the value of becoming an entrepreneur, $J^m(z, \kappa)$.

When $\kappa > \kappa^*$, the spouse does not work. In this case, the value of a being a married household with an entrepreneur is given by

$$rJ^m(z,\kappa) = \mu \frac{[(1-\chi)\gamma(\vartheta_g)\pi(z)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))\pi(z)]^{1-\sigma}}{1-\sigma} + \lambda(B^m(\kappa) - J^m(z,\kappa)). \quad (14)$$

When this happens, the married household gets less flow utility than the unmarried

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

individual.[11] This is the second part of the **spousal sharing effect**. It decreases the incentives to entrepreneurship by decreasing the benefits of entrepreneurship. As shall be clear in what follows, the **spousal sharing effect** also increases the cost of failure in entrepreneurship when the spouse is not working.

Finally, note that having children, $\chi > 0$, decreases total household income. This decreases the benefits to entrepreneurship since a part of the profits is shared with the children. This effect of children on decreasing the benefits to entrepreneurship is present independent if the spouse is working or not. This is the first part of the **offspring effect**. As shall become clear children also increase the cost of business failure and change the opportunity cost to entrepreneurship.

Once the business fails, the individual works as a wage worker but is forced to pay a cost c and is not allowed to enter entrepreneurship. Hence, the flow income of the main earner in this case is $w - c$. The individual exits bankruptcy back to wage work with probability p. Let W^u denote the value of being an unmarried wage worker. Then, the value of being bankrupt and unmarried, B^u, is

$$rB^u = \frac{(w-c)^{1-\sigma}}{1-\sigma} + p(W^u - B^u). \tag{16}$$

For married households, the main earner gets a share $(1-\chi)(1-\gamma)$ of total household income and the spouse gets a share $(1-\chi)\gamma$. The total income of the household in this case is composed of the income of the main earner, wage minus cost of bankruptcy, $w - c$, and the income of the spouse, $w^s h^*$. Let $W^m(\kappa)$ represent the value function for a household composed of a worker and a spouse with cost of working κ. Then, for $\kappa < \kappa^*$, the spouse works and the value function for a

[11] To see this note that

$$\begin{aligned}\mu\frac{[(1-\chi)(\gamma(\vartheta_g))\pi(z)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))\pi(z)]^{1-\sigma}}{1-\sigma} \\ < \mu\frac{[(1-\chi)\pi(z)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\chi)\pi(z)]^{1-\sigma}}{1-\sigma} = \frac{[(1-\chi)\pi(z)]^{1-\sigma}}{1-\sigma} < \frac{\pi(z)^{1-\sigma}}{1-\sigma}\end{aligned} \tag{15}$$

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

$$rB^m(\kappa) = \mu\frac{[(1-\chi)\gamma(\vartheta_g)(w-c+wh^*)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))(w-c+wh^*)]^{1-\sigma}}{1-\sigma} - \phi h^* + p(W^m(\kappa)-B^m(\kappa)) \quad (17)$$

The difference in the value of being bankrupt between the two groups depends on the difference in continuation values, W^m versus W^u and in differences in flow utility. In particular, while for unmarried individuals income falls by the cost of bankruptcy, c, married housholds benefit from the income of the spouse, conditional on the spouse working. When the business of the main earner fails, spouses weakly increase working hours h. The result is an increase in spousal earning which partially offsets the decrease in income suffered by married failed entrepreneurs. We call this effect the **spousal insurance effect**. This effect compresses the costs of business failure for married households.

If the spouse does not work ($\kappa > \kappa^*$), then the value of being a married household with a failed entrepreneur is given by

$$rB^m(\kappa) = \mu\frac{[(1-\chi)\gamma(\vartheta_g)(w-c)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))(w-c)]^{1-\sigma}}{1-\sigma} + p(W^m(\kappa)-B^m(\kappa)). \quad (18)$$

In this case the married household flow utility is less than that of the unmarried individual.[12] This increases the costs to entrepreneurship for married households with a spouse not working. Note that this comes from having to share income with a not working spouse, even when income is low like in a situation of business failure. This is another part of the **spousal sharing effect**.

[12]To see this note that

$$\mu\frac{[(1-\chi)(\gamma(\vartheta_g))(w-c)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))(w-c)]^{1-\sigma}}{1-\sigma} < \mu\frac{[(1-\chi)(w-c)]^{1-\sigma}}{1-\sigma}+(1-\mu)\frac{[(1-\chi)(w-c)]^{1-\sigma}}{1-\sigma} = \frac{[(1-\chi)(w-c)]^{1-\sigma}}{1-\sigma} < \frac{(w-c)^{1-\sigma}}{1-\sigma}. \quad (19)$$

Finally, note that having children, $\chi > 0$, decreases even more the income of the household with a failed entrepreneur. Children increase the cost of business failure. Furthermore, this effect of children is present regardless of whether the spouse is working or not. This is the second part of the **offspring effect**. Note that this second part goes in the same direction as the first part by making entrepreneurship less desirable for married households relative to unmarried households due to the presence of children.

The three parts of the **spousal sharing effect** and both parts of the **offspring effect** described up until now make entrepreneurship less desirable for married households relative to the unmarried. On the other hand, the **spousal insurance effect** makes entrepreneurship more desirable among married households by compressing the costs of failure.

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

$$J^u(\underline{z}^u) = W^u. \tag{20}$$

It follows that the value of an unmarried household composed of a working main earner is given by

$$rW^u = \frac{w^{1-\sigma}}{1-\sigma} + \psi \int_{\underline{z}^u} (J^u(z) - W^u) dF(z). \tag{21}$$

Let $\underline{z}^m(\kappa)$ represent the firm productivity that makes the married household indifferent between allocating the main earner to entrepreneurship or to wage work.

Then,

$$J^m(\underline{z}^m(\kappa), \kappa) = W^m(\kappa). \tag{22}$$

For married households, the spouse gets a share $(1-\chi)\gamma$ of total household income and the main earner gets a share $(1-\chi)(1-\gamma)$. When the main earner is working, total household income is given by the income of the main earner, their wage, w, and the income of the spouse, wh^*. 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 $\kappa < \kappa^*$, the spouse works and the value function of a married household with main earner working is

$$rW^m(\kappa) = \mu\frac{[(1-\chi)\gamma(\vartheta_g)(w+wh^*)]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))(w+wh^*)]^{1-\sigma}}{1-\sigma} - \phi h^* + \psi\int_{\underline{z}^m(\kappa)}(J^m(z,\kappa) - W^m(\kappa))dF(z) \tag{23}$$

It is important to note that throughout we have omitted the dependancy of value functions on θ to make the notation lighter. But of course, given that dependency, $\underline{z}^m$ and $\underline{z}^u$ both depend on entrepreneurial ability θ.

If the spouse does not work ($\kappa > \kappa^*$), then the value of a married household with the main earner working, $W^m(\kappa)$, is given by

$$rW^m(\kappa) = \mu\frac{[(1-\chi)\gamma(\vartheta_g)w]^{1-\sigma}}{1-\sigma} + (1-\mu)\frac{[(1-\chi)(1-\gamma(\vartheta_g))w]^{1-\sigma}}{1-\sigma} + \psi\int_{\underline{z}^m(\kappa)}(J^m(z,\kappa) - W^m(\kappa))dF(z). \tag{24}$$

Finally, note that having children, $\chi > 0$, decreases the total income of the married household with the main earner working. This is true independent if the spouse is working or not. This is the third part of the **offspring effect**. By decreasing the total income of the household when the main earner is working, children decrease the opportunity cost to entrepreneurship. We conclude that the **offspring effect** has an ambiguous effect on the incentives to start a firm. Despite the simplified modelling of children, we already obtain quite a lot of channels via which children affect the entrepreneurship decision of married households. In the model, children induce a higher cost of failing the business, a lower benefit to a productive business and a lower opportunity cost to entrepreneurship.

In total, we have three channels via which spouses affect entrepreneurship : **spousal sharing effect**, **spousal insurance effect** and **offspring effect**.

Next, to understand the relevance of the selection channel for productivity, Proposition 2 states the condition for average productivity to be increasing in the measure of married individuals MR.

Proposition 2 *Average firm productivity, $E[z]$, is increasing in the measure of married individuals, MR, if $\underline{z}^m > \underline{z}^u$ 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.

2.1 Outcomes and Spouses

We are interested in how having a spouse affects the entry into entrepreneurship and the size of firms. In our model, these objects are both captured by the selection thresholds, $\underline{z}^m$ for the married, and $\underline{z}^u$ for the unmarried. 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

$$\psi(1 - F(\underline{z}^m)) \tag{25}$$

and

$$\psi(1 - F(\underline{z}^u)). \tag{26}$$

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

Given the expression in equation (73), the average size of firms among married households (for fixed κ and θ) is given by

$$E[s]^m = \int_{\underline{z}^m} n(z,w)\frac{\Gamma_m(z)}{\Gamma_m}dz = \int_{\underline{z}^m} n(z,w)\frac{f(z)}{1-F(\underline{z}^m)}dz$$
$$= \int_{\underline{z}^m} (\frac{\alpha y \theta e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{(1-F(\underline{z}^m))}dz \quad (27)$$

where

$$\Gamma_m = \int \Gamma_m(z)dz. \quad (28)$$

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

$$E[s]^u = \int_{\underline{z}^u} n(z,w)\frac{\Gamma_u(z)}{\Gamma_u}dz = \int_{\underline{z}^u} n(z,w)\frac{f(z)}{1-F(\underline{z}^u)}dz$$
$$= \int_{\underline{z}^u} (\frac{\alpha y \theta e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{(1-F(\underline{z}^u))}dz \quad (29)$$

where

$$\Gamma_u = \int \Gamma_u(z)dz. \quad (30)$$

From these expressions we see that

$$\frac{\partial E[s]^m}{\partial \underline{z}^m} > 0 \quad (31)$$

and

$$\frac{\partial E[s]^u}{\partial \underline{z}^u} > 0. \quad (32)$$

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:

$$\underline{z}^m \lesseqgtr \underline{z}^u. \quad (33)$$

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

The **spousal sharing effect** decreases the benefits to entrepreneurship for married households. Firstly, it means that a higher profit for the main earner in a married couple is partially offset by a spouse working less hours. Secondly, for couples with non-working spouses, firm failure becomes more costly because the little income left still needs to be shared with the non working spouse. All else equal, this pushes married households to be more selective on which business projects to implement (higher $\underline{z}^m$) relative to unmarried individuals ($\underline{z}^u$). The result is a lower entry rate, higher average productivity and higher average size among firms created by married individuals.

The **spousal insurance effect** decreases the cost of business failure for married households. 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 lower average size among firms created by married households.

Finally, the **offspring effect** has an ambiguous effect on entrepreneurship. Firstly, children increase the cost of business failure. The reason is that the lower household income during business failure is decreased further by the need to set aside some of that income for the children. Secondly, children decrease the benefit to a successful business. Any increase in firm profits gives rise to a smaller increase in household consumption due to the requirement to share a part of that income with the children. These two channels make married individuals more selective on which business projects to implement (higher $\underline{z}^m$) relative to the unmarried individuals ($\underline{z}^u$). The result is a lower entry rate and higher average productivity and size among firms created by married households. On the other hand, children decrease total household income when the main earner is working. Since this represents the opportunity cost to entrepreneurship, this pushes married households to be less selective on which business projects to implement (higher $\underline{z}^m$) relative to unmarried households ($\underline{z}^u$). Hence, this third part of the **offspring effect** induces higher entry rate and lower average productivity and size for firms created by married households. Taken together, these three channels imply the **offspring effect** has an ambiguous effect on entry rates into entrepreneurship, average productivity and firm size for married relative to unmarried households.

In the next sections, I go over the data and the empirical strategy used to test the strength of these different effects. Note that although the model implies an ambiguous response of entry rates and average size of firms to changes in marriage rates, it imposes the restriction that any variable that changes the entry rate into entrepreneurship must change the average size of firms in the opposite direction. This comes from the model restriction that we are identifying changes in average size due to changes in the selection of business projects upon entry. As shall become clear, it turns out this restriction holds in the data.

Before proceeding to the data analysis we consider the endogenous decision to marry. This is important since it will make clear the source of endogeneity that needs to be overcome to estimate the effect of marriage on entrepreneurship.

2.2 Endogenous decision to marry

Up until now, we have considered the marital status of an individual as exogenous. In this section, we formalize the decision of individuals to marry or not. The objective is not to provide a full 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 empirically estimate the effect of marriage on entrepreneurship.

With this objective in mind, we consider a simple 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 knowledge of whether they enter or not entrepreneurship. Furthermore, assume individuals cannot direct search towards spouses of a particular cost of working, κ. They weigh each state by the equilibrium measure of individuals of same entrepreneurial ability as themselves in each state. Furthermore, recall individual belongs to a group g. Each group g has an idyossincratic utility value of v_g associated to being married. Let $v_g \sim H(v)$. Let $W_h^m(\kappa, \theta)$ represent the value for a main earner of ability θ of working and being married to a spouse of type κ. Let $J_h^m(z, \kappa, \theta)$ represent the value for a main earner of ability θ of being married to a spouse of type κ and running a firm with business project productivity of z. Let $B_h^m(\kappa,\theta)$ be the value for a main earner of ability θ of being married to a spouse of type κ and having failed a firm. In Online Appendix A we describe in further detail value functions $J_h^m(z,\kappa,\theta)$, $B_h^m(\kappa,\theta)$ and $W_h^m(\kappa,\theta)$. 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. Once married the couple draws a cost for the spouse to work, κ. This cost is paid by the household if and only if $\kappa < \kappa_g^*$. This cost κ is shared between spouse and main earner according to their weight in the household (μ and $1-\mu$ respectively). Hence, the ex-ante value of marriage, $V^M(\theta,g)$, for an individual of ability θ from group g is

$$V^M(\theta,g) + v_g \equiv \int (e_m(\kappa,\theta)W_h^m(\kappa,\theta) + b_m(\kappa,\theta)B_h^m(\kappa,\theta)$$
$$+ \int_{\underline{z}^m(\kappa,\theta)} J_h^m(z,\kappa,\theta)\Gamma_M(z,\kappa,\theta)dz)dM(\kappa) - (1-\mu)\int^{\kappa_g^*} \kappa dM(\kappa) + v_g. \quad (34)$$

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

$$V^U(\theta,g) \equiv e_u(\theta)W^u(\theta) + b_u(\theta)B^u(\theta) + \int_{\underline{z}^u(\theta)} J^u(z,\theta)\Gamma_U(z,\theta)dz. \quad (35)$$

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

$$\int (e_m(\kappa,\theta)W_h^m(\kappa,\theta) + b_m(\kappa,\theta)B_h^m(\kappa,\theta)$$
$$+ \int_{\underline{z}^m(\kappa,\theta)} J_h^m(z,\kappa,\theta)\Gamma_M(z,\kappa,\theta)dz)dM(\kappa) - (1-\mu)\int^{\kappa_g^*} \kappa dM(\kappa) + v_g >$$
$$e_u(\theta)W^u(\theta) + b_u(\theta)B^u(\theta) + \int_{\underline{z}^u(\theta)} J^u(z,\theta)\Gamma_U(z,\theta)dz. \quad (36)$$

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, ratio of women relative to men, ϑ_g, of their particular group. Hence, let $Pr(M=1)$ be the probability a man marries, then,

$$
\begin{aligned}
Pr(M=1) = q(\vartheta_g) \cdot Pr(V^M(\theta, g) - (1-\mu)\int^{\kappa_g^*} \kappa dM(\kappa) + \upsilon_g > V^U(\theta, g)) \\
\equiv q(\vartheta_g) Pr(\upsilon_g > V^U(\theta, g) - V^M(\theta, g) + (1-\mu)\int^{\kappa_g^*} \kappa dM(\kappa)) \\
= q(\vartheta_g)(1 - H(V^U(\theta, g) - V^M(\theta, g) + (1-\mu)\int^{\kappa_g^*} \kappa dM(\kappa))). \quad (37)
\end{aligned}
$$

where $q'() > 0$. Taking a first order linearization of the set of value functions for θ, y and κ_g^* gives us the probability of marriage as a linear function of individual entrepreneurial ability, θ, economy specific productivity, y, the gender ratio, ϑ_g, and the probability of spouses working, $Prob(work)_g$. The proposition below states this formally.

Proposition 3 *The probability to marry $Pr(M = 1)$ is characterized by*

$$Pr(M=1) = C_0 + C_1\theta + C_2 y + C_3\vartheta_g + C_4 Prob(work)_g. \quad (38)$$

A higher entrepreneurial ability increases the individual's incentive to start a firm which in turn makes the effect of marriage on the value of entrepreneurship more important for these individuals. This intuition makes clear that entrepreneurial ability, θ, affects both the decision to start a firm conditional on marital status as well as the decision to marry. Similarly, in productive economies (with higher y), individuals are more prone to start a firm. This in turn makes the differences in the value of entrepreneurship between married and unmarried all the more important for this individual. As a result, a change in y affects both the decision to start a firm conditional on marriage and the decision to marry.

It follows that a naive regression of entry into entrepreneurship on marriage suffers from endogeneity due to both y and θ. When we bring the model to the data we implement a strategy to overcome this endogeneity problem. Crucially, we need an instrument that captures variation in $Pr(M = 1)$ uncorrelated to the decision to start a firm. From equation (38) above we see that natural candidate is the gender ratio, ϑ_g. To do so, we will need to control for any direct effect of the gender ratio,

ϑ_g, on the decision to start a firm.

When we bring the model to the data we will consider variation across local economies and countries of birth. In particular, we consider each local economy-time period pair to be described by the model layed out in this section. Each group g will be a country of birth. In line with this logic, Proposition 4 below derives the marriage rate for a particular group g of an economy c and year t as a function of the average entrepreneurial ability, $\int \theta dG_c(\theta|g)(\theta)$, the economy-specific shock, $y_{c,t}$, and the gender ratio for that group g in that economy c, $\vartheta_{c,g,t}$.

Proposition 4 *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*

$$MR_{c,g,t} = C_0 + C_1 \int \theta dG_c(\theta|g) + C_2 y_{c,t} + C_3 \vartheta_{c,g,t} + C_4 log(Prob(work))_{c,g,t}. \tag{39}$$

2.3 Empirical Analysis

In this section I go over the main empirical strategies used to disentangle the relative strengths of each channel via which spouses affect the individual's decision to start a firm. The strategy uses variation across cities and countries of birth. Intuitively, consider each city as a local economy described by our model in the previous section.

We want to test the model prediction that the effect of marriage on the entry rate into entrepreneurship must have the opposite sign of the effect of marriage on the average size of firms. This restriction comes directly from our selection mechanism, in which firm heterogeneity is being driven by the entry decision of entrepreneurs. To our knowledge this is one of the first papers to empirically test this restriction of firm selection mechanisms.

Our main objective is to verify which effect from the theory (**spousal sharing effect**, **spousal insurance effect** and **offspring effect**) is the strongest. To verify this we use variation across different cities and immigrant groups in Canada.[13] Let c denote city, t year and g denote individuals born in country g, then the entry into entrepreneurship $ER_{c,g,t}$ and the log of average size of firms $log(SY_{c,g,t})$ can be written as a function of the marriage rate for that group g in that city c, $MR_{c,g,t}$. This is formally stated in Proposition 5 below.

Proposition 5 *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 main earner income 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. 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*

$$ER_{c,g,t} = \beta_{0,1} + \beta_{1,1}MR_{c,g,t} + \beta_{2,1}\gamma_{c,g,t} + \beta_{3,1}y_{c,t} + \beta_{4,1}\int \theta dG(\theta|g) + \beta_{5,1}Prob(work)_{c,g,t}. \quad (40)$$

and

$$log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}MR_{c,g,t} + \beta_{2,2}\gamma_{c,g,t} + \beta_{3,2}y_{c,t} + \beta_{4,2}\int \theta dG(\theta|g) + \beta_{5,2}Prob(work)_{c,g,t}. \quad (41)$$

where

$$\gamma_{c,g,t} = \zeta_0 + \zeta_1\vartheta_{c,g,t} \quad (42)$$

This proposition 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

[13]We include native Canadians in our groups but our results are robust to using only foreign born.

$\gamma_{c,g,t}$, 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 married women, $Prob(work)_{c,g,t}$, and the income sharing parameter, $\gamma_{c,g,t}$. Proposition 6 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 6 *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. Finally, let $Prob(work)_{c,g,t}$ have measurement error, $\varepsilon_{c,g,t}$, characterized by its first differences, $\Delta\varepsilon_{c,g,t}$, being i.i.d and $E[\Delta\varepsilon_{c,g,t}] = 0$. Then the probability of working for married women in economy c in group g, time t, $Prob(work)_{c,g,t}$, can be approximated by*

$$Prob(work)_{c,g,t} = \beta_{0,4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4}\int Id\mu_{c,t}(I|g) + \varepsilon_{c,g,t} \tag{43}$$

where $\varepsilon_{c,g,t}$ is measurement error.

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 (40) and (41) we get

$$\begin{aligned} ER_{c,g,t} = \beta_{0,1} + \beta_{1,1}MR_{c,g,t} + \zeta_{2,1}Prob(Work)_{c,g,t} + \zeta_{2,2}\int Id\mu_{c,t}(I|g) + \beta_{3,1}y_{c,t} \\ + \beta_{4,1}\int \theta dG(\theta|g) + \zeta_{2,3}\varepsilon_{c,g,t}. \end{aligned} \tag{44}$$

and

$$\begin{aligned} log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}MR_{c,g,t} + \zeta_{3,1}Prob(Work_{c,g,t}) + \zeta_{3,2}\int Id\mu_{c,t}(I|g) + \beta_{3,2}y_{c,t} \\ + \beta_{4,2}\int \theta dG(\theta|g) + \zeta_{3,3}\varepsilon_{c,g,t}. \end{aligned} \tag{45}$$

Note that $Prob(Work)_{c,g,t}$ is the female labor force participation and $\int Id\mu_{c,t}(I|g)$ is average income of married men 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

$$\Delta ER_{c,g,t} = \beta_{1,1}\Delta MR_{c,g,t} + \zeta_{2,1}\Delta Prob(Work)_{c,g,t} + \zeta_{2,2}\Delta \int Id\mu_{c,t}(I|g) \\ + \beta_{3,1}\Delta y_{c,t} + \zeta_{2,3}\Delta\varepsilon_{c,g,t}. \quad (46)$$

and

$$\Delta log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2}\Delta MR_{c,g,t} + \zeta_{3,1}\Delta Prob(Work)_{c,g,t} + \zeta_{3,2}\Delta \int Id\mu_{c,t}(I|g) \\ + \beta_{3,2}\Delta y_{c,t} + \zeta_{3,3}\Delta\varepsilon_{c,g,t}. \quad (47)$$

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 (39) and taking first differences we get

$$\Delta MR_{c,g,t} = C_2\Delta y_{c,t} + C_3\Delta\vartheta_{c,g,t} + C_4\Delta Prob(work)_{c,g,t}. \quad (48)$$

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

$$IV_{c,g,t} = \Delta\vartheta_{c,g,t}. \quad (49)$$

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 female labor force participation, $\Delta Prob(Work)_{c,g,t}$, and average income of married men,

$\Delta \int Id\mu_{c,t}(I|g)$. To summarize our two main specifications are

$$\Delta ER_{c,g,t} = \beta_{1,1}\Delta MR_{c,g,t} + \zeta_{2,1}\Delta Prob(Work)_{c,g,t} + \zeta_{2,2}\Delta \int Id\mu_{c,t}(I|g) \\ + \beta_{3,1}\Delta y_{c,t} + \zeta_{2,3}\Delta\varepsilon_{c,g,t}. \quad (50)$$

and

$$\Delta log(SY)_{c,g,t} = \beta_{0,2}+\beta_{1,2}\Delta MR_{c,g,t}+\zeta_{3,1}\Delta Prob(Work)_{c,g,t}+\zeta_{3,2}\Delta \int Id\mu_{c,t}(I|g) \\ + \beta_{3,2}\Delta y_{c,t} + \zeta_{3,3}\Delta\varepsilon_{c,g,t}. \quad (51)$$

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 $\beta_{1,2} > 0$.

3 Data and Empirical Results

In this section we go over the data we use and our empirical results.

3.1 Data and Measurement

The data used for the empirical analysis is the Canadian Employer-Employee Dynamics Database (CEEDD). It contains the entire universe of Canadian tax filers, and privately owned incorporated firms. The dataset links employees to firms and firms to their corresponding owners across space and time. This is achieved by linking individual tax information (T1 files, individual tax returns), with linked employer-employee information (T4 files)[14] and firm ownership and structure information (T2 files). T2 forms are the Canadian Corporate Income Tax forms. In-

[14]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, identifies the employee and reports the labour earnings of the employee.

side the T2 files we find the schedule 50 in which each corporation must list all owners with at least 10% of ownership. This allows 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 1120 form (Corporate Income Tax Form in the US). The data is annual and is available from 2001 to 2013. This constitutes an advantage relative to employer-employee firm population data from the US, which does not allow the researcher to identify the owners of the firm.

Furthermore, relative to other employer-employee linked datasets it is unique since it contains as well the link between individuals and their spouses. This allows us to directly observe whether individuals start their firm with their spouse and whether the spouse works for the firm. These are important channels that we omit from our model exactly because we are able to control for them empirically. In particular, we do this by considering our main specifications excluding joint ventures started by both spouses and businesses for which individuals hire their spouses.

The data is annual with information on both firms and individuals. Using this database, 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. We focus on outcomes for men between 25 and 65 years old to focus on individuals with high labor market attachment.[15]

Business owners are identified as individuals present in the schedule 50 files from the T2 that have employees. Wage workers are identified as those who are not entrepreneurs and report a positive employment income on their T4. We use the information in the T1 files to control for characteristics such as gender, age and to identify marital status. We identify as married all couples and not just individuals legally married.[16] Finally, the dataset is also linked to immigrant 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 have two key charac-

[15] In the robustness section we show our results are robust to considering women.

[16] This is possible in data because we are able to identify cohabitation.

teristics which correspond closely to how economists typically think about firms : limited liability and separate legal identity. Furthermore, there is a growing literature showing that incorporated firms tend to be larger and that they are more likely to contribute to aggregate employment. [17] There is also evidence that there is little transition from unincorporated to incorporated status.[18] These facts highlight how incorporated firms with employees are the most appropriate measure of firms to consider if we are interested in the interplay between entrepreneurship and the aggregate economy. Another reason to focus on incorporated firms with employees is Canadian corporte law. In Canada there are significant tax advantages for incorporating as a higher earner. So to exclude from my analysis high-earning workers that incorporate exclusively due to tax purposes, I focus on incorporated firms with employees. This dataset represents an important improvement in that aspect, by allowing us to focus on firms that contribute 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 founders of firms to restrict ourselves to entrepreneurs, as those in our model, that start new ventures. Although interesting in its own right, the study of the choice to buy shares in an already existing firm is left for future research.

Our measure of local labor market is that of economic regions. These are equivalent to commuting zones in the US. There is a total of 76 such regions in Canada.

We define a startup as a firm that is at most 1 year old. In our dataset the average size of startups is 6 employees while that of all firms is 13 employees. The equivalent for the US economy is an average of 6 employees for startups and an average of 23 employees for the entire firm population during the period of $2001 - 2013$.[19] Hence, while the average size of startups is similar, older firms tend to be smaller in Canada.

Below we present summary statistics for the gender ratio (number of women relative to men). Table 1 presents summary statistics for the gender ratio across dif-

[17]Glover and Short (2010) document that incorporated entrepreneurs operate larger businesses, accumulate more wealth, and are on average more productive than unincorporated entrepreneurs.

[18]Levine and Rubinstein (2017) show that there is little transition from unincorporated to incorporated status.

[19]The numbers for the US were calculated by the author using BDS data.

ferent economic regions and countries of birth for the year 2005. Each observation represents an economic region/country of birth pair. We see there is large variation in the gender ratio in our data.

Table 1: 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 country of birth groups within the economic region of Toronto in 2005 (first row) and of Winnipeg in 2005 (second row). Table 2 shows there is quite a lot of variation across country of birth groups. In the case of Toronto, the average gender ratio is of 101 women per 100 men and the standard deviation is of 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 the case of Winnipeg, the average gender ratio is 92.7 women per 100 men and the standard deviation is of 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 2: 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 in 2005. Table 3 below plots the summary statistics for the gender ratio across economic regions for the Argetina born population (first row), South Korea born population (second row), Russia born population (third row) and Uganda born population (fourth row).[20] In the first row we see that, for the Argentina born population, the average gender ratio is 103 women per 100 men and the standard deviation is 78 women per 100 men. In the second row we see that, for the South Korea 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 Russia born population, the average gender ratio is 152 women per 100 men and the standard deviation is 159 women per 100 men. Finally, in the fourth row, the Table shows that for the Uganda born population the average gender ratio is 83 women per 100 men and the standard deviation is 75 women per 100 men. 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 3: 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
Uganda	0.83	0.746	44

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), for Uganda born population (fourth row). Number of observations in each row represent the number of economic regions in 2005 that had the given country of birth.

[20] These are just some examples since it would be too much to present information for all countries of birth. Since we have the full universe of Canada, in our data we have information for all countries of birth and all economic regions.

Like previously mentioned one concern is that most businesses are started jointly by spouses something that is absent in our model. Looking in our data we find that only 29% of new firms started are started by both spouses. Furthermore, as explained previously in our main specifications we verify our results remain unchanged once we proceed to excluding these joint ventures.

Another possible concern is 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 employee of the business. 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.

One possible concern with our identification strategy is the potential lack of homophily in marriage in Canada. In particular, 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[21] shows that of all couples in Canada : 66.9% are between two Canadian born, 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 immigrantion 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.

Finally, we might be worried that some of the variation in the gender ratio 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

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

restricting the sample to cells with a minimum of men and women.

3.1.1 Results

In this section we present the main results of our specifications. We also include as additional controls : year dummies, 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 each economic region, country of birth and year triplet. Results are robust to not including this extra set of controls.

Column 1 of Table 4 presents the OLS results for our entry rate specification. We see that marriage rates are significant and positive if we don't instrument for it. Column 2 shows that once we instrument for marriage rates, the coefficient flips sign and increases in magnitude. In particular, a 1 percentage point increase in the marriage rate is associated to a 0.2 percentage point drop in the entry rate into entrepreneurship. Given the 1% baseline entry rate in the data, this corresponds to a 20% drop in the entry rate into entrepreneurship. Finally, row 2 indicates that the instrument is significant at the first stage. Column 3 indicates the results are robust to including city dummies.

One concern with our results is that cultural determinants can be simulatenously determining the change in gender ratio and in entrepreneurship. To deal with this we can use the variation in changes in the gender ratio of a same immigrant group across two different locations. To do that we include country of origin dummies. Column 4 shows our results are robust to using this variation. First stage results for Table 4 can be found in Table 6 of the Appendix.

Next, we turn to the results on average number of employees of firms. Column 1 of Table 5 indicates marriage, $\Delta MR_{c,g,t}$, has a small negative insignificant effect on average size of firms if not instrumented for. Column 2 of Table 5 shows results for average size of firms once we instrument for marriage rates by our instrument, the gender ratio. Our estimate implies that a 1 percentage point increase in the marriage rate increases the average size by 1.13%. In other words, a 10 percentage point increase in the marriage rate increases average size of firms by 11.3%. Finally row 2 indicates that the instrument is significant at the first stage. Column 3 and 4 indicate that the results for average size are robust to the inclusion of city dummies or country of birth dummies. First stage results for Table 5 can be found in Table 7 of the Online Appendix.[22]

Note that when we include city dummies we are using variation in changes of the gender ratio across country of birth groups within a same city. In turn, when we include country of birth dummies we are using variation in changes of the gender ratio for a same immigrant group across different cities. Hence, these are two different sources of variation for which there is no reason ex-ante to expect similar results. Yet, despite this different source of varation, we get similar results in sign, significance and magnitude regardless of which of these two sources of variation we use. Furthermore, results for both entry and average size are robust in sign, mangnitude and significance to including dummies for each country of origin and year pair.[23] Results are also robust to clustering at the economic region c level.

These results are consistent with the prediction of our model that the group (married vs unmarried) with the highest entry rate into entrepreneurship is also the one with the lower average firm size. Futhermore, note that this is true for both the OLS and IV specifications. Our results are consistent with the notion that the **spousal insurance effect** is dominated by the **spousal sharing effect** or by the negative effects on entrepreneurship of the **offspring effect**.

The results are in line with married households being more selective in which business projects to implement. This implies that a decrease in marriage rates induces a fall in average productivity of the economy. In the next section, we look at this implication more formally.

[22] The number of observations for the average size regressions are smaller because for a country of birth, economic region and year triplet to be included we need there to be entrepreneurs in that triplet. For the entry rate regression we just need there to be individuals in that country of birth, economic region and year triplet.

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

Table 4: Main specifications : Entry Rate Regressions

	OLS	IV	IV	IV
$\Delta MR_{c,g,t}$	0.0258 (0.004)	-0.219 (0.035)	-0.219 (0.035)	-0.22 (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 in economic region c, country of birth g and year t, $\Delta ER_{c,g,t}$, on the change in the marriage rate 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.

Table 5: Main specifications : Average Number of Employees Regressions

	OLS	IV	IV	IV
$\Delta MR_{c,g,t}$	-0.038 (0.037)	1.127 (0.349)	1.12 (0.35)	1.13 (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 in economic region c, country of birth g and year t, $\Delta log(SY)_{c,g,t}$, on the change in the marriage rate 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.

4 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 productivity, $E[z]$. To do so, note that for all firms in the model we have

$$n(z, w) = (\frac{\alpha y \theta}{w})^{\frac{1}{1-\alpha}} e^{\frac{z}{1-\alpha}}. \tag{52}$$

This in turn implies

$$E[n] = (\frac{\alpha}{w})^{\frac{1}{1-\alpha}} E[z] \tag{53}$$

where $E[z]$ is as defined in Equation 74 and $E[n]$ is average number of employees. It follows that

$$\Delta log(E[z]) = \Delta log(E[n]) - \Delta log(\frac{\alpha}{w})^{\frac{1}{1-\alpha}}. \tag{54}$$

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,

$$\Delta log(E[z]) \approx \Delta log(E[n]). \tag{55}$$

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 for us to calculate the implied changes in average firm productivity given any estimate of changes in wages.

5 Robustness Checks

One concern with our strategy is that there might be economic regions with very low absolute numbers of individuals from a particular country. This would mean that a small arrival of individuals from that group produces large fluctuations in the gender ratio. To address this concern we verify our results are robust to restricting the use of economic region, country of birth and year triplets with at least 10 women and 10 men. Our results are unchanged in both sign and significance. In fact, the magnitudes actually increase. Results can be found in Columns 1 and 2 of Table 9 of the Online Appendix.

A second concern with our identification strategy is if there are city specific shocks $\Delta y_{c,t}$ that are gender biased differentially across different immigrant groups. As such, men from particular immigrant groups could be more prone to move to particular cities relative to men of other immigrant groups. The result is that our instrument would be correlated to the error term. However, 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 to in Canada is uncorrelated to the gender of their child, we can use these early age arrival immigrants to address this concern. Consistent with this argument, we verify our results are robust to using a gender ratio constructed using only individuals that arrived at age 15 or younger in Canada. Our results are unchanged in both sign and significance. For entry rates the magnitude is unchanged while for average size the effect is now stronger. Results can be found in Columns 3 and 4 of Table 9 in the Online Appendix.

6 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 borrowing constraints affects outcomes for an individual considering starting a firm. The first is that individuals are constrained in the scale of the firm they create. If married individuals on average are wealthier we expect higher marriage rates to be associated to higher 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 if there are startup costs to opening a firm. However, both these channels go in the opposite direction to our finding that higher marriage rates decrease entry rates into entrepreneurship.

To further verify our findings are not being driven by this borrowing constraint channel we check that our results are robust to excluding businesses started in high capital demanding sectors (See Table 10 of the Online Appendix).[24] This robustness check alleviates the concern that our results are being driven by borrowing constraint considerations.

Another potential mechanism ignored by the model is the possibility of joint entrepreneurship by both main earner and spouse. In particular, married individuals can be more likely to start a firm because they have the option of starting with their spouse which also contributes to running the business. Of course this is inherently hard to measure. But to the extent that a couple running together a firm are both listed as owners of the business or has the spouse listed as employee of the firm, we can verify our results are robust to excluding these types of businesses. Indeed, we verify that our results continue to hold, in both magnitude and significance once we exclude firms started jointly by both main earner and spouse and firms for which the spouse is listed as an employee. See Table 11 of the Online Appendix for results taking out firms started jointly by the couple. See Table 12 of the Online Appendix for results taking out firms for which the spouse is listed as employee of the firm. Such robustness tests would be impossible without our link between business owner, firm, employee and spouse.

Finally, it is tempting to think that our results are driven by lower risk aversion among men relative to women. This channel can generate a higher risk aversion among a couple composed of a man and a woman relative to a man. Such a narrative is consistent with married individuals being more selective on which business

[24] We exclude firms started in : Oil, Mining, Gas, Utilities, Manufacturing, Wholesale Trade, Retail Trade, Transportation and Wharehousing, Information and Cultural Industries and Real Estate.

projects to implement but for reasons not explicit in the model. However, if this narrative were true we expect the results for women to be the opposite of what we find for men. We should expect married women to start more and smaller firms relative to unmarried women. We verified our results are of similar sign and significance for women (Table 13 of the Online Appendix). In fact, the magnitudes are even stronger for women. This evidence for women is inconsistent with the narrative of our results being driven by gender specific risk aversion.

7 Conclusion

In this paper we explore for the first time the importance of spouses and children for an individdual's decision to start a firm and the outcomes of that firm. Through a tractable model we show that spouses affect the decision through different opposing channels. At one hand, spouses provide insurance against business risk. On the other hand, spouses decrease the marginal benefit to entrepreneurship by cutting on working hours when the business is more productive. Finally, having a spouse is also associated to having children which increase the cost of failure and decrease the benefits to entrepreneurship.

Using the model, we derive empirical specifications, the source of endogeneity and a instrumental variable that allow us to bring the model to the data. We show that marriage is associated to lower entry into entrepreneurship and higher average number of employees. This is consistent with the importance of the negative effect of having to share the benefits of a successfull business with a spouse or with the higher cost of failure/lower benefits associated to having children. Furthermore, the results are consistent with our selection mechanism in which the effect of marriage on entry into entrepreneurship and average size of firms has opposite signs. We also verify 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

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.

Blundell, Richard, Alan Duncan, and Costas Meghir, "Estimating labor supply responses using tax reforms," *Econometrica*, 1998, pp. 827–861.

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

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

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.

da Fonseca, Joao Galindo, "Unemployment, Entrepreneurship and Firm Outcomes," 2019.

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.

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.

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

_ , Ron S Jarmin, and Javier Miranda, "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Hurst, Erik and Annamaria Lusardi, "Liquidity constraints, household wealth, and entrepreneurship," *Journal of Political Economy*, 2004, *112* (2), 319–347.

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.

Jr, Melvin Stephens, "Worker displacement and the added worker effect," *Journal of Labor Economics*, 2002, *20* (3), 504–537.

Jr, Robert E Lucas, "On the size distribution of business firms," *The Bell Journal of Economics*, 1978, pp. 508–523.

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.

Quadrini, Vincenzo, "Entrepreneurship, saving, and social mobility," *Review of Economic Dynamics*, 2000, *3* (1), 1–40.

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.

A For Online Publication : Value function for married main earner

In this subsection of the Appendix I 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

$$\begin{aligned} rW_h^m(\kappa, \theta) = u((1-\chi)\gamma(\vartheta_g)(w + wh^*)) - \mu\phi h^* \\ + \psi \int_{\underline{z}^M} (J_h^m(z, \kappa, \theta) - W_h^m(z, \kappa, \theta)) dF(z) \end{aligned} \tag{56}$$

$$\begin{aligned} rJ_h^m(z, \kappa, \theta) = u((1-\chi)\gamma(\vartheta_g)(\pi(z) + wh^*)) - \mu\phi h^* \\ + \lambda(B_h^m(\kappa, \theta) - J_h^m(z, \kappa, \theta)) \end{aligned} \tag{57}$$

and

$$rB_h^m(\kappa, \theta) = u((1-\chi)\gamma(\vartheta_g)(w - c + wh^*)) - \mu\phi h^* + p(W_h^m(\kappa, \theta) - B_h^m(\kappa, \theta)) \tag{58}$$

where

$$u(x) = \frac{x^{1-\sigma}}{1-\sigma}, \quad \forall x. \tag{59}$$

Recall that μ is the weight of the main earner in the household. Hence, μ is also the share of cost of spouse working and the fixed cost of the spouse paid by the married main earner.

B For Online Publication : Proofs

Proof of Proposition 1.

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

$$\frac{\phi}{w}\frac{\partial \pi(z)}{\partial z} \tag{60}$$

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

$$\pi(z)^{-\sigma}\frac{\partial \pi(z)}{\partial z} \tag{61}$$

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

$$\frac{1}{\pi(z)^{\sigma}} > \frac{\phi}{w} \tag{62}$$

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

$$\max_{z,\theta} \pi(z) = (1-\alpha)(\frac{\alpha}{w})^{\frac{\alpha}{1-\alpha}} e^{\frac{\Upsilon}{1-\alpha}} (yT)^{\frac{1}{1-\alpha}} \tag{63}$$

Then note that

$$\frac{1}{\pi(z)^{\sigma}} > \frac{1}{(\max_{z,\theta} \pi(z))^{\sigma}}. \tag{64}$$

Then it follows that if

$$\frac{1}{(\max_{z,\theta} \pi(z))^{\sigma}} > \frac{\phi}{w} \Rightarrow \frac{w}{\max_{z,\theta} \pi(z)^{\sigma}} \equiv \hat{\phi} > \phi \tag{65}$$

we have our desired result.

Proof of Proposition 2. 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 $\eta_m(\kappa, \theta)$ be the share of individuals married to a spouse with cost of working κ and entrepreneurial ability θ that are entrepreneurs. Let $e_m(\kappa, \theta)$ be the share of individuals married to a spouse with cost of working κ, with entrepreneurial ability θ that are workers. Let $b_m(\kappa, \theta)$ be the share of individuals married to a spouse with cost of working κ with entrepreneurial ability θ 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

$$\dot{\eta}_i(\kappa, \theta) = \psi(1 - F(\underline{z}^i(\kappa, \theta)))e_i(\kappa, \theta) - \lambda\eta_i(\kappa, \theta), \quad \forall i \in \{m, u\} \tag{66}$$

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

$$\dot{b}_i(\kappa, \theta) = \lambda\eta_i(\kappa, \theta) - pb_i(\kappa, \theta), \quad \forall i \in \{m, u\}. \tag{67}$$

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

$$\eta_i(\kappa, \theta) + b_i(\kappa, \theta) + e_i(\kappa, \theta) = 1, \quad \forall i \in \{m, u\}. \tag{68}$$

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

$$e_i(\kappa, \theta) = \frac{\lambda p}{\lambda p + \psi(1 - F(\underline{z}^i(\kappa, \theta)))(p + \lambda)}, \quad \forall i \in \{m, u\}. \tag{69}$$

$$b_i(\kappa, \theta) = \frac{\psi(1 - F(\underline{z}^i(\kappa, \theta)))\lambda}{\lambda p + \psi(1 - F(\underline{z}^i(\kappa, \theta)))(p + \lambda)}, \quad \forall i \in \{m, u\}. \tag{70}$$

$$\eta_i(\kappa,\theta) = \frac{\psi(1-F(\underline{z}^i(\kappa,\theta)))p}{\lambda p + \psi(1-F(\underline{z}^i(\kappa,\theta)))(p+\lambda)}, \quad \forall i \in \{m,u\}. \tag{71}$$

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, $\Gamma_i(z,\kappa,\theta)$, is characterized by

$$\dot{\Gamma}_i(z,\kappa,\theta) = \psi f(z) e_i(\kappa,\theta) - \lambda \Gamma_i(z,\kappa,\theta), \quad \forall z \geq \underline{z}^i(\kappa,\theta) \quad i \in \{m,u\} \tag{72}$$

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

$$\Gamma_i(z,\kappa,\theta) = \frac{\psi f(z) p}{\lambda p + \psi(1-F(\underline{z}^i(\kappa,\theta)))(p+\lambda)}. \tag{73}$$

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

$$\begin{aligned}
E[z] &= MR \cdot E[z]^m + (1-MR) \cdot E[z]^u = \\
&\qquad MR \cdot \sum \eta_g \int \int \int_{\underline{z}^m(\kappa,\theta)} \frac{(y\theta)^{\frac{1}{1-\alpha}} e^{\frac{z}{1-\alpha}} f(z)}{1-F(\underline{z}^m(\kappa,\theta))} dz dM(\kappa) G(\theta|g) d\theta \\
&+ (1-MR) \cdot \sum \eta_g \int \int \int_{\underline{z}^u(\kappa,\theta)} \frac{(y\theta)^{\frac{1}{1-\alpha}} e^{\frac{z}{1-\alpha}} f(z)}{1-F(\underline{z}^u(\kappa,\theta))} dz dM(\kappa) G(\theta|g) d\theta.
\end{aligned} \tag{74}$$

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 $\underline{z}^m > \underline{z}^u$ 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.

Proof of Proposition 3.

For the first part of this proof I show how to obtain an equation defining thresholds $\underline{z}^M$ and $\underline{z}^U$. With a slight abuse of notation, define $u^x_S(\kappa)$ as the flow utility an individual receives when married or unmarried ($x \in \{u, m\}$) with spouse of cost of working κ if married[25] and in State S, where $S \in \{W, B, J)$ and $u^x_J(z, \kappa)$ 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

$$J^U(\underline{z}^U) = W^U \tag{75}$$

$$J^W(\underline{z}^W(\kappa), \kappa) = W^M(\kappa). \tag{76}$$

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

$$J^x(z, \kappa) = \frac{u^x_J(z, \kappa) + \lambda B^x(\kappa)}{r + \lambda} \quad \forall x \in \{u, m\}. \tag{77}$$

$$B^x(\kappa) = \frac{u^x_B(\kappa) + pW^x(\kappa)}{r + p} \quad \forall x \in \{u, m\}. \tag{78}$$

$$W^x(\kappa) = \frac{u^x_W(\kappa) + \psi \int_{\underline{z}^x(\kappa)} J^x(z, \kappa) dF(z)}{r + \psi(1 - F(\underline{z}^x(\kappa)))} \tag{79}$$

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

$$0 = J^x(\underline{z}^x(\kappa), \kappa) - W^x(\kappa) = \frac{u^x_J(\underline{z}^x, \kappa)}{r + \lambda} + \frac{\lambda u^x_B(\kappa)}{(r + \lambda)(r + p)} - \frac{rW^x(\kappa)(r + p + \lambda)}{(r + p)(r + \lambda)} \tag{80}$$

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

Now after some algebra we find

$$
\begin{aligned}
rW^x(\kappa) = &\frac{(r+p)(r+\lambda)u_W^x(\kappa)}{(r+\lambda)(r+p)+r(r+\lambda+p)\psi(1-F(\underline{z}^x(\kappa)))} \\
&+ \psi \int_{\underline{z}^x(\kappa)} \frac{u_J^x(z,\kappa)(r+p)dF(z)}{(r+\lambda)(r+p)+r(r+p+\lambda)\psi(1-F(\underline{z}^x(\kappa)))} \\
&+ \frac{\psi(1-F(\underline{z}^x(\kappa)))\lambda u_B^x(\kappa)}{r(r+p)(r+\lambda)+r(r+\lambda+p)\psi(1-F(\underline{z}^x(\kappa)))}. \quad (81)
\end{aligned}
$$

If we replace this expression for rW^x in Equation 80 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

$$\underline{z}^M(\kappa) = \zeta_{1,0}^M + \zeta_{1,1}^M y + \zeta_{1,2}^M \theta + \zeta_{1,3}^M \gamma_g \quad \text{if} \quad \kappa \leq \kappa^* \quad (82)$$

$$\underline{z}^M(\kappa) = \zeta_{2,0}^M + \zeta_{2,1}^M y + \zeta_{2,2}^M \theta + \zeta_{2,3}^M \gamma_g \quad \text{if} \quad \kappa > \kappa^* \quad (83)$$

$$\underline{z}^U = \zeta_0^U + \zeta_1^U y + \zeta_2^U \theta. \quad (84)$$

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

$$
\begin{aligned}
\int \underline{z}^M(\kappa) dM(\kappa) = &Prob(work)_g(\zeta_{1,0}^M + \zeta_{1,1}^M y + \zeta_{1,2}^M \theta + \zeta_{1,3}^M \gamma_g) \\
&+ (1 - Prob(work)_g)(\zeta_{2,0}^M + \zeta_{2,1}^M y + \zeta_{2,2}^M \theta + \zeta_{2,3}^M \gamma_g) \quad (85)
\end{aligned}
$$

and

$$\underline{z}^U = \zeta_0^U + \zeta_1^U y + \zeta_2^U \theta. \quad (86)$$

Next, linearize equation 85 with $\underline{z}^M(\kappa)$, $\underline{z}^U$, y, θ, γ and $Prob(work)$ around the point $(z^*, z^*, 1, 1, 1, 1)$ to get

$$\int \underline{z}^M(\kappa)dM(\kappa) = \zeta^M_{3,0} + \zeta^M_{3,1}y + \zeta^M_{3,2}\theta + \zeta^M_{3,3}\gamma_g + \zeta^M_{3,4}Prob(work)_g \tag{87}$$

and

$$\underline{z}^U = \zeta^U_0 + \zeta^U_1 y + \zeta^U_2\theta. \tag{88}$$

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

$$Pr(M = 1) = D_0 + D_1 \int z^M dM(\kappa) + D_2\underline{z}^U + D_3\theta + D_4\vartheta_g + D_5 Prob(work)_g + D_6 y + D_7\kappa^*_g \tag{89}$$

Secondly, note that

$$Prob(work)_g = M(\kappa^*_g) \tag{90}$$

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

$$\kappa^* = \alpha_0 + \alpha_1 Prob(work)_g. \tag{91}$$

Finally, we replace κ^*, $\int \underline{z}^M dM(\kappa)$ and $\underline{z}^U$ by their expressions given by Equations (87) and (84) to arrive at

$$Pr(M = 1) = B_0 + B_1\theta + B_2 y + B_3\gamma_g + B_4 log(\vartheta_g) + B_5 Prob(work)_g. \tag{92}$$

Finally linearize (γ_g, ϑ_g) (for $\gamma(\vartheta_g)$) around $(1, 1)$ to get

$$\gamma_g = \zeta_0 + \zeta_1\vartheta_g. \tag{93}$$

Once we replace this expression for γ_g at the Equation (92) we obtain the desired result

$$Pr(M = 1) = C_0 + C_1\theta + C_2 y + C_3\vartheta_g + C_5 Prob(work)_g. \tag{94}$$

Proof of Proposition 4.

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 (38) at the economy-group wide level we obtain

$$MR_{c,g,t} = C_0 + C_1 \int \theta dG_c(\theta|g) + C_2 y_{c,t} + C_3 \vartheta_{c,g,t} + C_4 Prob(work)_{c,g,t}. \quad (95)$$

Proof of Proposition 5.

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 Proposition 3. Next, recall the expression for the entry rate in the economy is given by

$$ER_{c,g,t} = MR_{c,g,t} \int\int \psi(1 - F(\underline{z}^M(\theta,\kappa)_{i,c,t}))dG_c(\theta|g)dM(\kappa) + (1 - MR_{c,g,t}) \int \psi(1 - F(\underline{z}^U(\theta)_{i,c,t}))dG_c(\theta|g) \quad (96)$$

and the one for log average size of firms is given by

$$log(SY)_{c,g,t} = log(MR_{c,g,t} \int\int\int_{\underline{z}^M(\theta,\kappa)_{i,c,t}} (\frac{\alpha y_{c,t}\theta_i e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{1 - F(\underline{z}^M(\theta,\kappa)_{i,c,t})} dz dG_c(\theta|g)dM(\kappa) + (1 - MR_{c,g,t}) \int\int_{\underline{z}^U(\theta)_{i,c,t}} (\frac{\alpha y_{c,t}\theta_i e^z}{w})^{\frac{1}{1-\alpha}} \frac{f(z)}{1 - F(\underline{z}^U(\theta)_{i,c,t})} dz dG_c(\theta|g)). \quad (97)$$

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 (82), (83) and (84). Next, linearize for (y, θ, γ and $Prob(work)$) around the point $(1, 1, 1, 1)$ to get :

$$ER_{c,g,t} = \beta_{0,1} + \beta_{1,1} MR_{c,g,t} + \beta_{2,1}\gamma_{c,g,t} + \beta_{3,1} y_{c,t} + \beta_{4,1} \int \theta dG(\theta|g) + \beta_{5,1} Prob(work)_{c,g,t}. \quad (98)$$

and

$$log(SY)_{c,g,t} = \beta_{0,2} + \beta_{1,2} MR_{c,g,t} + \beta_{2,2}\gamma_{c,g,t} + \beta_{3,2} y_{c,t} + \beta_{4,2}\int \theta dG(\theta|g) + \beta_{5,2} Prob(work)_{c,g,t}. \quad (99)$$

Finally linearize (γ_g, ϑ_g) around $(1, 1)$ to get

$$\gamma_{c,g,t} = \zeta_0 + \zeta_1 \vartheta_{c,g,t}. \quad (100)$$

Proof of Proposition 6.

To start consider equation 8. Then, log-linearize the terms (I, γ) inside $U_s^m(I, \gamma_g)$ and $U_{ns}^m(I, \gamma)$ around $(I^*, \frac{1}{2})$ to get

$$Prob(work|I)_{c,g,t} = \beta_{0.4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4} I. \quad (101)$$

Now if we integrate over I we obtain

$$Prob(work)_{c,g,t} = \beta_{0.4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4}\int I d\mu_{c,t}(I|g). \quad (102)$$

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

$$Prob(work)_{c,g,t} = \beta_{0.4} + \beta_{1,4}\gamma_{c,g,t} + \beta_{2,4}\int I d\mu_{c,t}(I|g) + \epsilon_{c,g,t}. \quad (103)$$

C For Online Publication : First Stage Regressions

In this section we present the results for the first stage regression of our IV regressions. Our endogenous variable is the change in the marriage rate in economic region c year t, $\Delta MR_{c,t}$. Our instrument is the sum of the share of individuals from group g in a economic region c in the first year of the sample multiplied by the change in marriage rate of individual at group g at the national level in year t.

Table 6: 1st Stage Regression - Entry Rate Regressions

$IV_{c,g,t}$		0.0371	0.0371	0.0372
		(0.001)	(0.001)	(0.001)
Dummies for cities	-	-	Yes	-
Dummies for country of birth	-	-	-	Yes
Observations		70551	70551	70551

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 $IV_{c,g,t}$. 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 just the economic region c level.

Table 7: 1st Stage Regression - Average Size Regressions

$IV_{c,g,t}$		0.0302	0.0302	0.0302
		(0.002)	(0.002)	(0.002)
Dummies for cities	-	-	Yes	-
Dummies for country of birth	-	-	-	Yes
Observations		32305	32305	32305

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 $IV_{c,g,t}$.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 just the economic region c level.

D For Online Publication : Robustness Regressions

Table 8: Main specifications : Robustness Regressions

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta log(SY)_{c,g,t}$	$\Delta ER_{c,g,t}$	$\Delta log(SY)_{c,g,t}$
	IV	IV	IV	IV
$\Delta MR_{c,g,t}$	-0.468	3.01	-0.149	3.64
	(0.079)	(1.05)	(0.048)	(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 9: 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	0.040	0.022	0.020
	(0.002)	(0.002)	(0.001)	(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.

E For Online Publication : Alternative mechanisms

Table 10 reports results for when we exclude high capital requirement industries. 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 14.

Table 10: Regressions excluding high capital industries

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
	IV	IV
$\Delta MR_{c,g,t}$	-0.167	1.098
	(0.027)	(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 11 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 15.

Table 11: Regressions excluding businesses owned by both in couple

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
	IV	IV
$\Delta MR_{c,g,t}$	-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 12 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 16.

Table 12: Regressions excluding businesses where spouse listed as employee

Dependant Variable	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
	IV	IV
$\Delta MR_{c,g,t}$	-0.180	1.74
	(0.031)	(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 12 reports results for when we consider outcomes for women rather than men. We construct changes in entry rates $\Delta ER_{c,g,t}$, changes in log average number of employees $\Delta SY_{c,g,t}$ and changes in the marriage rate $\Delta MR_{c,g,t}$ for women. Our instrument in this case is the change in the ratio of number of men relative to women $\Delta\vartheta^{-1}_{c,g,t}$. First Stage results can be found at Table 17.

Table 13: 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.

Table 14: Regressions excluding high capital industries - 1st Stage

Dependant Variable at 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
	IV	IV
$\Delta IV_{c,g,t}$	0.037	0.03
	(0.001)	(0.001)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
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 15: Regressions excluding businesses owned by both in couple - 1st Stage

Dependant Variable at 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
	IV	IV
$\Delta IV_{c,g,t}$	0.0371	0.032
	(0.001)	(0.002)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
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 16: Regressions excluding businesses where spouse listed as employee - 1st Stage

Dependant Variable in the 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
	IV	IV
$\Delta IV_{c,g,t}$	0.037	0.035
	(0.001)	(0.002)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
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.

Table 17: Main specifications : Regressions for women - 1st Stage

Dependant Variable at 2nd Stage	$\Delta ER_{c,g,t}$	$\Delta SY_{c,g,t}$
$\Delta IV_{c,g,t}$	0.006	0.009
	(0.001)	(0.001)
Significance IV for $\Delta MR_{c,t}$	Yes	Yes
Observations	65919	25801

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 for women, $\Delta ER_{c,g,t}$ and changes in 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 change in number of men relative to women, $\Delta\vartheta^{-1}_{c,g,t}$. 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.