

NBER WORKING PAPER SERIES

NATURE OR NURTURE? LEARNING AND THE GEOGRAPHY OF FEMALE LABOR  
FORCE PARTICIPATION

Alessandra Fogli  
Laura Veldkamp

Working Paper 14097  
<http://www.nber.org/papers/w14097>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
June 2008

We thank seminar participants at Chicago GSB, Wisconsin Madison, Minneapolis Federal Reserve, Princeton, European University in Florence, University of Southern California, New York University, Boston University, Bocconi, Pompeu Fabra, Ente Einaudi, Boston Federal Reserve and Harvard University and conference participants at 2008 AEA, SITE, the 2007 NBER Summer Institute, the SED conference, LAEF Households, Gender and Fertility conference, the NBER group on Macroeconomics across Time and Space, Midwest Macro Meetings, the NY/Philadelphia Workshop on Quantitative Macro, IZA/SOLE and Ammersee. We especially thank Stefania Marcassa for excellent research assistance and Stefania Albanesi, Roland Benabou, Raquel Bernal, Jason Faberman, Jeremy Greenwood, Luigi Guiso, Larry Jones, Patrick Kehoe, Narayana Kocherlakota, Ellen McGrattan and Fabrizio Perri for comments and suggestions. Laura Veldkamp thanks Princeton University for their hospitality and financial support through the Kenen fellowship. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2008 by Alessandra Fogli and Laura Veldkamp. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Nature or Nurture? Learning and the Geography of Female Labor Force Participation  
Alessandra Fogli and Laura Veldkamp  
NBER Working Paper No. 14097  
June 2008  
JEL No. E2,J16,N32,R1

**ABSTRACT**

One of the most dramatic economic transformations of the past century has been the entry of women into the labor force. While many theories explain why this change took place, we investigate the process of transition itself. We argue that local information transmission generates changes in participation that are geographically heterogeneous, locally correlated and smooth in the aggregate, just like those observed in our data. In our model, women learn about the effects of maternal employment on children by observing nearby employed women. When few women participate in the labor force, data is scarce and participation rises slowly. As information accumulates in some regions, the effects of maternal employment become less uncertain, and more women in that region participate. Learning accelerates, labor force participation rises faster, and regional participation rates diverge. Eventually, information diffuses throughout the economy, beliefs converge to the truth, participation flattens out and regions become more similar again. To investigate the empirical relevance of our theory, we use a new county-level data set to compare our calibrated model to the time-series and geographic patterns of participation.

Alessandra Fogli  
Federal Reserve Bank of Minneapolis,  
Research Department  
90 Hennepin Avenue, P.O. Box 291  
Minneapolis, MN 55480-0291  
afogli@umn.edu

Laura Veldkamp  
Stern School of Business  
New York University  
44 W Fourth Street, Suite 7-77  
New York, NY 10012  
and NBER  
lveldkam@stern.nyu.edu

Over the twentieth century, there has been a dramatic rise in female labor force participation in the United States. Many theories of this phenomenon have been proposed. Some of them emphasize the role played by market prices and technological factors; others focus on the role played by policies and institutions, and a few recent ones investigate the role of cultural factors. All of them, however, focus on aggregate shocks that explain why the transition took place and abstract from the local interactions that could explain how the transition took place.

We use new data and theory to argue that women's labor force participation decisions rely on information that is transmitted from one woman to another, located nearby. The local nature of information transmission smooths the effects of changes in the environment and generates geographically heterogeneous, but locally correlated reactions, like those observed in our data. Our theory focuses on learning and participation of women with children, because this sub-group is responsible for most of the rise in participation. A crucial factor in mothers' participation decisions is the effect of employment on their children. However, this effect is uncertain. The uncertainty makes risk-averse women less likely to participate. Learning resolves their uncertainty, causing participation to rise.

In our overlapping generations model, women learn from their neighbors about the relative importance of nature (innate ability) and nurture (the role of maternal employment) in determining children's outcomes (section 1). Women inherit their parents' beliefs and update them after observing the outcomes of neighboring women in the previous generation. Those outcomes reveal information about the effect of maternal employment only if those nearby mothers were employed. Section 2 shows that higher local participation generates more information, which reduces uncertainty about the effect of maternal employment and makes participation of nearby women more likely. Thus, local participation snowballs and a gradual, but geographically-concentrated rise in participation rates ensues.

Using county-level U.S. data from 1940-2000, section 3 documents how the growth rate of women's labor force varied over time and across counties.<sup>1</sup> The female labor force grew slowly

---

<sup>1</sup>To our knowledge this county-level data from the Integrated Public Use Microdata series has not been explored before in economics research.

during the post-war decades, accelerated during the 1970s and 1980s, and recently flattened out, generating an S-shaped time path. Furthermore, this growth was uneven across geographic regions: High participation rates emerged first in a few geographic centers and spread from there to nearby regions, over the course of several decades. This process gave rise to significant spatial correlation across the participation rates of US counties that is only marginally explained by common economic and demographic factors. This residual correlation slowly rose at the beginning of the period, peaked when aggregate labor force increased fastest and finally declined as aggregate labor force stagnated. Finally, survey evidence and natural experiments offer direct evidence of heterogeneity and changes in beliefs about maternal employment.

Sections 4 and 5 use moments of the labor force participation distribution across US counties in 1940 to calibrate and simulate a dynamic learning model and explore its quantitative properties. The results are consistent with the S-shaped evolution of aggregate labor force, and the rise and fall in the spatial correlation of county-level participation rates. The model generates S-shaped dynamics because initially, when uncertainty is high, very few women participate in the labor market; information about the role of nurture diffuses slowly and beliefs are nearly constant. As information accumulates and the effects of labor force participation become less uncertain, more women participate, learning accelerates and labor force participation rises more quickly. As uncertainty is resolved, beliefs converge to the truth, and participation flattens out.

The local nature of the learning process generates the rise and fall of spatial correlation in participation. Initially, female labor force participation is low everywhere and the minute differences are spatially uncorrelated. As women in some locations start working, their neighbors observe them and learn from them. This makes the neighbors more likely to work in the next generation, generating an increase in geographic heterogeneity and spatial correlation. Eventually as the truth about maternal employment is learned everywhere, heterogeneity and spatial correlation in local participation rates falls.

The model provides a simple framework for examining the transition dynamics and geography of a wide array of social and economic phenomena. Section 6 illustrates this potential by setting up a model of career choice that predicts patterns of wages and occupational sorting, issues also

addressed in recent work by Doepke and Zilibotti (2008). Section 7 concludes by describing further extensions of the model that could capture the effects of policy change, heterogeneity among social groups or the process of cultural change.

**Relationship to other theories** Many recent papers have explored the rise in female labor force participation. Among these theories, some focus on changes that affect the costs or benefits of employment for all women: changes in wages, less discrimination, the introduction of household appliances, the less physical nature of jobs, or the ability to control fertility.<sup>2</sup> In contrast, our theory focuses on why the participation of women with children rose so much faster than the aggregate participation rate. A complete understanding of the rise in participation requires both pieces, an explanation of what changed for all women and what made married mothers behave so differently.

Another group of theories shares our focus on changes that affect mothers specifically, but unlike our theory, rely on aggregate shocks. For example, the decline in child care costs, the invention of baby formula, or public news shocks are changes that spread quickly because there are no geographic barriers or distance-related frictions causing some regions to be unaffected.<sup>3</sup> Obviously, one can modify these theories to introduce geographic heterogeneity by adding income or preference heterogeneity.<sup>4</sup> What is harder to explain is why the participation transition happened at different times in different places. The rates of change in participation were vastly different across counties, resulting in a rise and then a fall in the cross-county dispersion of participation rates. This is not a pattern that a typical aggregate shock would generate.

One would think that any local coordination motive (e.g. social pressure) or thick market externality (e.g. child care markets) could generate local differences in the speed of transition. But such a coordination model typically predicts a simultaneous switch from a low-participation

---

<sup>2</sup>See Greenwood, Seshadri, and Yorukoglu (2005), Goldin and Katz (2002), and Goldin (1990), Jones, Manuelli, and McGrattan (2003) on nature of jobs.

<sup>3</sup>See Attanasio, Low, and Sanchez-Marcos (2008) and Del Boca and Vuri (2007) for child care costs, Albanesi and Olivetti (2007) for baby formula, Fernández and Fogli (2005), Antecol (2000) and Fernández, Fogli, and Olivetti (2004) on the role of cultural change, and Fernández (2007) for an aggregate information-based learning theory. Note that this work was done independently and was published as Minneapolis Federal Reserve Staff Working paper #386, prior to Fernández (2007).

<sup>4</sup>See Fuchs-Schundeln and Izem (2007) for a static theory of geographic heterogeneity in labor productivity between East and West Germany.

to a high-participation outcome, unless there is some friction preventing perfect coordination. Our local information externality generates locally correlated behavior, while the imperfect nature of the information is the friction that prevents perfect economy-wide coordination.

A third strand of related literature, on technology diffusion, does not focus on labor force participation, but does consider the geographic diffusion of information (see e.g. Munshi (2004)). One way to interpret our message is that ideas about how technology diffuses should be applied to female labor force participation. In this case, the technology being learned about is outsourcing the care of one's children. Of course, the spread of more traditional technologies like washing machines and dishwashers could also explain the geographic diffusion of participation. But, such technologies diffused throughout the country in the span of a decade or two. Part of the puzzle this paper wrestles with is isolating the information frictions that make learning about maternal employment so much slower than learning about consumer technologies.<sup>5</sup>

Facts about geographic heterogeneity do not prove that aggregate changes are irrelevant. Rather, they suggest such changes operate in conjunction with a mechanism that causes their effect to disseminate gradually across the country. We argue that this mechanism is the local transmission of information. Considering how beliefs react to changing circumstances and how these beliefs, in turn, affect participation decisions can help us understand and evaluate the effects of many other important changes to the benefits and costs of labor force participation.

## 1 The Model

In this section, we develop a theory in which the dramatic change in female labor force participation emerges solely as the result of local interactions. Because the bulk of the change came from married women with small children, we focus on their participation. We model local interactions that transmit information about the effect of maternal employment on children.

---

<sup>5</sup>In the macro learning literature, our model fills a gap between the literature on S-shaped learning dynamics and on endogenous information. The S-shaped learning dynamic is similar to the model of Amador and Weill (2006) where agents learn what their neighbors know while the idea that information is a by-product of economic activity appears in Veldkamp (2005). The idea that learning is slow because agents only observe outcomes of those near them is similar to work on government policy contagion by Buera, Monge-Naranjo, and Primiceri (2006).

The model makes two key assumptions. First, women were initially uncertain about the consequences of maternal employment on their children. The shift from agriculture to industry at the end of the 19th century changed the nature of work. In agriculture, women allocated time continuously between work and child-rearing. This was possible because home and work were in the same location. Industrialization required women who took jobs to outsource their child care. At that time, the effects of outsourcing were unknown. Women held beliefs about those effects which were very uncertain.<sup>6</sup>

The second key assumption is that learning happens only at the local level from a small number of observations, as in the Lucas (1972) island model. This allows learning to take place gradually, over the course of a century. In a richer model, this strong assumption could be relaxed. Appendix E sets up and simulates a model with multiple types where women need to observe others like themselves to learn their type-specific cost of maternal employment. For example, professionals do not learn from seeing hourly workers; urban mothers face different costs than rural ones. Instead of learning about what the cost of maternal employment is for the average woman, these women are learning about the difference between the average cost and the cost for their type. In this richer model, women can observe many more signals, as well as aggregate information like the true aggregate participation rate, and still learn slowly about the cost of maternal employment for their type. The results of the simple model below are nearly identical this richer model.

**Preferences and Constraints** Time is discrete and infinite ( $t = 1, 2, \dots$ ). We consider an overlapping generation economy made up of a large finite number of agents living for two periods. Each agent is nurtured in the first period and consumes and has one child in the second period of her life. Preferences of an individual in family  $i$  born at time  $t - 1$  depend on their consumption  $c_{it}$  and the potential wage of their child  $w_{i,t+1}$ .

$$U = \frac{c_{it}^{1-\gamma}}{1-\gamma} + \beta \frac{w_{i,t+1}^{1-\gamma}}{1-\gamma} \quad \gamma > 1 \quad (1)$$

---

<sup>6</sup>This is consistent with the decline in the labor market participation rate of married women observed during the turn of the century by Goldin (1995), and with the findings of Mammen and Paxson (2000) who document a U-shaped relationship between women's labor force rates and development in a cross section of countries.

This utility function captures the idea that parents care about their child’s earning potential, but not about the choices they make.<sup>7</sup>

The budget constraint of the individual from family  $i$  born at time  $t - 1$  is

$$c_{it} = n_{it}w_{it} + \omega_{it} \tag{2}$$

where  $\omega_{it}$  is an endowment which could represent a spouse’s income and  $n_{it} \in \{0, 1\}$  is the discrete labor force participation choice. If the agent works in the labor force,  $n_{it} = 1$ .

The key feature of the model is that an individual’s earning potential is determined by a combination of endowed ability and nurturing, that cannot be perfectly disentangled. Endowed ability is an unobserved normal random variable  $a_{i,t} \sim N(\mu_a, \sigma_a^2)$ . If a mother stays home with her child, the child’s full natural ability is achieved. If the mother joins the labor force, some unknown amount  $\theta$  of the child’s ability will be lost. Wages depend exponentially on ability:

$$w_{i,t} = \exp(a_{i,t} - n_{i,t-1}\theta) \tag{3}$$

Of course, a child also benefits from higher household income when its mother joins the labor force. While this benefit is not explicitly modeled,  $\theta$  represents the cost to the child of maternal employment, net of the gain from higher income. The net effect could be positive for child welfare. When we model beliefs, women will not rule out the possibility that employment has a net positive effect on their child’s development. Furthermore, appendix D explores a model where all women initially believe that maternal employment is beneficial and shows that uncertainty alone can deter participation.

**Information Sets** The constant  $\theta$  determines the importance of nurture and is not known when making labor supply decisions. Women have two sources of information about  $\theta$ : beliefs passed

---

<sup>7</sup>Using utility over the future potential wage, rather than recursive utility shuts down an experimentation motive where mothers participate in order to create information that their decedents can observe. Such a motive makes the problem both intractable and unrealistic. Most parents do not gamble with their children’s future just to observe what happens.



down through their family and the wage outcomes of themselves and their neighbors. Agents do not learn from aggregate outcomes.

Young agents inherit their prior beliefs about  $\theta$  from their parents' beliefs. In the first generation, initial beliefs are identical for all families  $\theta_{i,0} \sim N(\mu_0, \sigma_0^2)$ ,  $\forall i$ . Each subsequent generation updates these beliefs and passes down their updated beliefs to their child. To update beliefs at the beginning of time  $t$ , agents use both potential earnings and parental employment decisions for themselves and for  $J - 1$  peers. We refer to  $w$  as the potential wage because it is observed, regardless of whether the agent chooses to work.<sup>8</sup> Ability  $a$  is never observed so that  $\theta$  can never be perfectly inferred from observed wages. But, these potential wages are only informative about the effect of maternal employment on wages if a mother actually worked. Note from equation (3) that if  $n_{i,t-1} = 0$ , then  $w_{i,t}$  only reflects innate ability and contains no information about  $\theta$ . Since the content of the signals in the first period depends on the previous period's participation rate, the model requires a set of initial participation decisions  $n_{i,0}$  for each woman  $i$ .

The set of family indices for the outcomes observed by agent  $i$  is  $\mathbb{J}_i$ . Spatial location matters in the model because it determines the composition of the signals in this information set. Each agent  $i$  has a location on a two-dimensional map with indices  $(x_i, y_i)$ . Signals are drawn uniformly from the set of agents within a distance  $d$  in each direction:  $\mathbb{J}_i \sim \text{unif}\{[x_i - d, x_i + d] \times [y_i - d, y_i + d]\}^{J-1}$ .

Agents use the information in observed potential wages to update their prior, according to Bayes' law. Bayesian updating with  $J$  signals is equivalent to the following two-step procedure: First, run a regression of children's potential wages on parents' labor choices:

$$W - \mu_a = N\theta + \varepsilon_i$$

where  $W$  and  $N$  are the  $J \times 1$  vectors  $\{\log w_{j,t}\}_{j \in \mathbb{J}_i}$  and  $\{n_{i,t-1}\}_{j \in \mathbb{J}_i}$ . Let  $\bar{n}_{i,t}$  be the sum of the labor decisions for the set of families that  $(i, t)$  observes:  $\bar{n}_{i,t} = \sum_{j \in \mathbb{J}_i} n_{j,t}$ . The resulting estimated coefficient  $\hat{\theta}$  is normally distributed with mean  $\hat{\mu}_{i,t} = \sum_{j \in \mathbb{J}_i} (\log w_{j,t} - \mu_a) n_{j,t} / \bar{n}_{i,t}$  and variance

---

<sup>8</sup>This assumption could be relaxed. If  $w_{i,t}$  were only observed once agent  $(i, t)$  decided to work, then an informative signal about  $\theta$  would only be observed if both  $n_{i,t} = 1$  and  $n_{i,t-1} = 1$ . Since this condition is satisfied less frequently, such a model would make fewer signals observed and make learning slower.

$\hat{\sigma}_{i,t}^2 = \sigma_a^2 / \bar{n}_{i,t}$ . Second, form the posterior mean as a linear combination of the estimated coefficient  $\hat{\theta}$  and the prior beliefs  $\mu_t$ , where each component's weight is its relative precision:

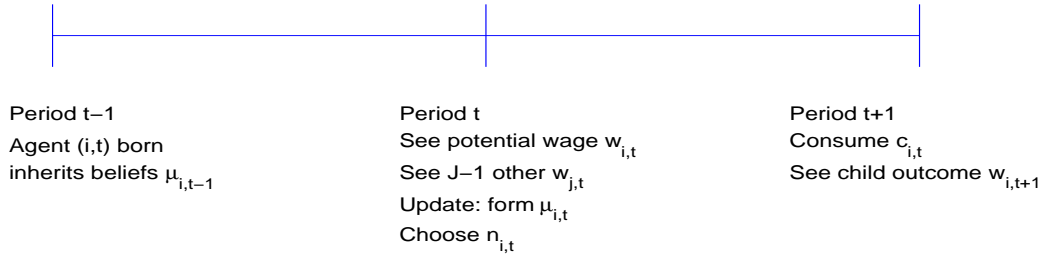
$$\mu_{i,t} = \frac{\hat{\sigma}_{i,t}^2}{\sigma_{i,t-1}^2 + \hat{\sigma}_{i,t}^2} \mu_{i,t-1} + \frac{\sigma_{i,t-1}^2}{\sigma_{i,t}^2 + \hat{\sigma}_{i,t}^2} \hat{\mu}_{i,t} \quad (4)$$

Posterior beliefs about the value of nurturing are normally distributed  $\theta \sim N(\mu_{i,t}, \sigma_{i,t}^2)$ . The posterior precision (inverse of the variance) is the sum of the prior precision and the signal precision.<sup>9</sup>

Thus posterior variance is

$$\sigma_{i,t}^2 = (\sigma_{i,t-1}^{-2} + \hat{\sigma}_{i,t}^{-2})^{-1}. \quad (5)$$

The timing of information revelation and decision-making is as follows.



**Equilibrium** An equilibrium is a sequence of wages, distributions that characterize beliefs about  $\theta$ , work and consumption choices, for each individual  $i$  in each generation  $t$  such that the following four conditions are satisfied: First, taking beliefs and wages as given, consumption and labor decisions maximize expected utility (1) subject to the budget constraint (2). The expectation is conditioned on beliefs  $\mu_{i,t}, \sigma_{i,t}$ . Second, wages of agents born in period  $t-1$  are consistent with the labor choice of the parents, as in (3). Third, priors  $\mu_{i,t-1}, \sigma_{i,t-1}$  are equal to the posterior beliefs of the parent, born at  $t-1$ . Priors are updated using observed wage outcomes  $\mathbb{J}_{i,t}$ , according to Bayes' law (4). Fourth, distributions of elements  $\mathbb{J}_{i,t}$  are consistent with distribution of optimal

<sup>9</sup>The fact that another woman's mother chose to work is potentially an additional signal. But the information content of this signal is very low because the outside observer does not know whether this person worked because they were highly able, very poor, less uncertain or had low expectations for the value of theta. Since these observations contain much more noise than wage signals, and the binary nature of the working decision makes updating much more complicated, we approximate beliefs by ignoring this small effect. We solve an extended model where women use this extra information in the appendix. Over the 70-year simulation, the extra information increases participation by 2.4%.

labor choices  $n_{i,(t-1)}$  and each agent's spatial location.

## 2 Analytical Results

In this section we establish some cross sectional and dynamic predictions of our theory that distinguish it from other theories. We begin by solving for the optimal participation decision. Substituting the budget constraint (2) and the law of motion for wages (3) into expected utility (1) produces the following optimization problem for agent  $i$  born at date  $t - 1$ :

$$\max_{n_{it} \in \{0,1\}} \frac{(n_{it}w_{it} + \omega_{it})^{1-\gamma}}{1-\gamma} + \beta E_{a_{i,t+1},\theta} \left[ \frac{\exp((a_{i,t+1} - n_{i,t}\theta)(1-\gamma))}{1-\gamma} \right]. \quad (6)$$

Taking the expectation over the unknown ability  $a$  and the importance of nurture  $\theta$  delivers expected utilities from each choice. If a woman stays out of the labor force, her expected utility is

$$EUO_{it} = \frac{(\omega_{it})^{1-\gamma}}{1-\gamma} + \frac{\beta}{1-\gamma} \exp\left(\mu_a(1-\gamma) + \frac{1}{2}\sigma_a^2(1-\gamma)^2\right). \quad (7)$$

If she participates in the labor force, her expected utility is

$$EUW_{it} = \frac{(w_{it} + \omega_{it})^{1-\gamma}}{1-\gamma} + \frac{\beta}{1-\gamma} \exp\left((\mu_a - \mu_{i,t})(1-\gamma) + \frac{1}{2}(\sigma_a^2 + \sigma_{i,t}^2)(1-\gamma)^2\right). \quad (8)$$

The optimal policy is to join the labor force when the expected utility from employment is greater than the expected utility from staying home ( $EUW_{it} > EUO_{it}$ ). Define  $\mathcal{N}_{it} \equiv EUW_{it} - EUO_{it}$  to be the expected net benefit of labor force participation, conditional on information  $(\mu_{i,t}, \sigma_{i,t})$ .

### 2.1 Comparative statics: The Role of Beliefs, Wages and Wealth

**Beliefs** The key variable whose evolution drives the increase in labor force participation is beliefs, and particularly uncertainty. We begin by establishing two intuitive properties of labor force participation (both derived formally in appendix A). First, a higher expected value of nurture reduces the probability that a woman will participate in the labor force, holding all else equal.

The logic of this result appears in equation (8). Increasing the expected value of nurture decreases the net expected utility of labor force participation:  $\partial \mathcal{N}_{i,t} / \partial \mu_{i,t} = -\beta$ , times an exponential term, which is always non-negative. Since  $-\beta < 0$ , a higher  $\mu_{i,t}$  reduces the utility gain from labor force participation and therefore reduces the probability that a woman will participate.

Second, greater uncertainty about the value of nurture reduces the probability that a woman will participate in the labor force, holding all else equal. More uncertainty about the cost of maternal employment on children makes labor force participation more risky. Participation falls because agents are risk-averse. Over time as information accumulates and uncertainty falls, the net benefit of participating rises:  $\partial \mathcal{N}_{i,t} / \partial \sigma_{i,t} = (1 - \gamma)\beta$ , times a non-negative (exponential) term. Higher risk aversion makes  $(1 - \gamma)$  more negative and amplifies this effect.

Thus, there are two ways our model could produce an increase in participation. First, women could have started with biased, pessimistic beliefs (low  $\mu_0$ ) and participation rates would rise as women learned that participation is not as bad as they thought. This is the driving force in Fernández (2007). Instead, our calibration will give women unbiased beliefs about  $\theta$ . Our women will work more over time because they start out uncertain (high  $\sigma_0$ ) and learning reduces their uncertainty. It is possible that some force in the economy caused women around the world to be systematically deceived about the effect maternal employment has on their children. But the economic transition from agricultural work to the modern age, and the new requirement that employed women outsource their children's care, undoubtedly created uncertainty.

**Wages** Wages in our model have standard role: Women work more if wages are higher. While other theories give wages and human capital a more central role (Olivetti (2006), Goldin and Katz (1999), Jones, Manuelli, and McGrattan (2003)), our baseline model holds the distribution of wages fixed. We explore the effects of a changing wage process in our technical appendix.

**Wealth** Greater initial wealth  $\omega_{i,t}$  reduces the probability that a woman will participate in the labor force. Poorer women join the labor force before richer ones because poorer women have a higher marginal value of wage income.

## 2.2 Dynamic Properties

One might think that the initial state after industrialization would be no women participating and no information being produced and that this would be an absorbing state. The following result shows that zero participation is a state that can persist for many periods but is exited each period with a small probability (proof in appendix A.2).

**Result 1** *In any period where the labor force participation rate is zero ( $\sum_j n_{j,t-1} = 0$ ), there is a positive probability that at least one woman will work in the following period ( $\sum_j n_{j,t} \geq 1$ ).*

All it takes to escape a zero-participation state is for one extremely able woman to be born. She generates information that makes the women around her less uncertain about the effects of maternal employment. That information encourages these women to work. They, in turn, generate more information for women around them. Gradually, the information and participation disseminate.

Condition (8) also suggests circumstances in which such a woman is likely to emerge. One example is a low endowment  $\omega_{jt}$ , which raises the marginal value of labor income. Depressions or wars, which reduce endowments by eliminating husbands' incomes, can hasten the transition. Learning amplifies those kinds of shocks and causes them to persist long after their direct effects have disappeared. Shocks that cause more women to participate persist through their effects on the information that gets transmitted from generation to generation.

**S-shaped Evolution of Participation Rates** One of the hallmarks of information diffusion models is that learning is slow at first, speeds up, and then slows down again as beliefs converge to the truth. The concave portion of this S-shaped pattern can be explained by any theory. Because the participation rate is bounded above by one, any shock to participation must eventually taper off. But many shocks to labor force participation would be strongest when they first hit. The interesting feature of this model is its prediction that participation will first rise slowly and then speed up.

The information gleaned from observing others' labor market outcomes can be described as a signal with mean  $\hat{\mu}_{i,t} = \sum_{j \in \mathbb{J}_i} (\log w_{j,t} - \mu_a) n_{j,t} / \bar{n}_{i,t}$  and variance  $\hat{\sigma}_{i,t}^2 = \sigma_a^2 / \bar{n}_{i,t}$ . Let  $\rho$  be the

fraction of women who participate in the labor force. Then, the expected precision of this signal is  $E[\hat{\sigma}_{i,t}^{-2}] = \rho N \sigma_a^2$ . A higher signal precision increases the expected magnitude of changes in beliefs. This conditional variance of  $t$  beliefs is the difference between prior variance and posterior variance:  $var(\mu_{i,t}|\mu_{i,t-1}) = \sigma_{i,t-1}^2 - \sigma_{i,t}^2$ . Substituting in for posterior variance using equation (5),

$$var(\mu_{i,t}|\mu_{i,t-1}) = \sigma_{i,t-1}^2 - \frac{1}{\sigma_{i,t-1}^{-2} + \hat{\sigma}_{i,t}^{-2}}. \quad (9)$$

Since  $\partial var(\mu_{i,t}|\mu_{i,t-1})/\partial \hat{\sigma}_{i,t}^{-2} > 0$ , the expected size of revisions is increasing in the precision of the observed signals and therefore in the fraction of women who work. This is the first force: As beliefs change more rapidly, so does labor force participation, early in the century.

The concave part of the S-shaped increase in participation comes later, from convergence of beliefs to the truth. Over time, new information reduces posterior variance:  $\sigma_{i,t}^2 < \sigma_{i,t-1}^2$  (equation 5). As posterior variance falls, beliefs change less:  $\partial var(\mu_{i,t}|\mu_{i,t-1})/\partial \sigma_{i,t-1}^2 > 0$ .

**Endogenous Pessimism** At the start of the transition, there is another force that suppresses participation: Women become more pessimistic about the benefits of maternal employment, on average ( $\int_i \mu_{i,t} di$  rises). Women who have pessimistic beliefs ( $\mu_{i,t-1} > \theta$ ) do not participate and thus generate less information for their children than women with optimistic beliefs ( $\mu_{i,t-1} < \theta$ ). Since new information  $\hat{\mu}_{i,t}$  is unbiased, on average, it moves beliefs toward the true  $\theta$  (equation 4). Since the children of pessimistic women observe less new information, their posterior beliefs remain closer to their prior beliefs. The children of optimistic women revise their beliefs more, which brings them closer to the truth. Since pessimism is persistent and optimism is undone by learning, the average belief is pessimistic, until information disseminates fully.

### 2.3 Geographic Properties

The model produces two effects relating to geography: dispersion and spatial correlation in participation rates. Differences in participation rates come from differences in beliefs. Each child's potential wage is a random realization. Differences in these realizations create differences in beliefs

across women. These differences are amplified when women who get information suggesting that maternal employment is not very costly join the labor force and generate more information for the women around them. Locations with high mean beliefs generate more information, which lowers the variance of their beliefs. Both high means and lower variance (less uncertainty) promote higher labor force participation rates. More participation feeds back by creating more information, which further reduces the uncertainty and risk associated with maternal employment. Local information diffusion creates a learning feedback mechanism that amplifies the effect of small differences in signal realizations.

We formalize this local information effect in the following result. Suppose that a woman has location  $(x_i, y_i)$ . Define her region to be the set of agents whose outcomes are in her information set with positive probability:  $[x_i - d, x_i + d] \times [y_i - d, y_i + d]$ .

**Result 2** *A woman with an average prior belief who observes average signal draws in a region with a high participation rate at time  $t$  is more likely to participate at time  $t + 1$ , all else equal.*

Information diffusion makes cross-region dispersion in participation rates rise and then fall. All women have identical initial prior beliefs by assumption. Dispersion in beliefs is zero. In the limit as  $t \rightarrow \infty$ , beliefs converge to the truth and their dispersion converges back to zero. In between, beliefs among women differ and therefore have positive dispersion. The rise and fall in belief dispersion is what will create a rise and fall in the dispersion of participation rates.

### 3 Empirical Evidence: Time Series and Geographic

To examine the transition in female labor force participation predicted by our model, we calibrate and simulate it. Before turning to those results, this section describes the data and the measures we use to compare the model to the data. It also presents direct evidence that changing beliefs played a role in the transition.

### 3.1 Time Series Evidence

We study the labor force participation behavior of white women over the period 1940-2005 using data from the US decennial Census and from the Census Bureau’s American Community Survey. Figure 1 reports the labor force participation rate in each decade for women between 25 and 34 years old.<sup>10</sup> This implies that the data for each decade comes from a distinct cohort of women. The increase is quite large: The fraction of women in the labor force rose from one-third in 1940 to nearly 75% in 2005.

However, this increase in the aggregate rate hides large differences among subgroups of women. The increase comes mainly from the change in working behavior of married women with children. Women without children or unmarried women have always worked in large numbers: In 1940, their participation rate was already around 60%. On the other hand, the participation rate of married women with children at that time was only 10% and dramatically increased, reaching 62% in 2005. Therefore, to understand the large aggregate rise over the period we need to understand what kept married women with children out of the labor market at the beginning of the period and why their behavior has changed so dramatically.<sup>11</sup>

Another interesting feature of the phenomenon that emerges from Figure 1 shows that the increase took place at different rates over the period: steady but slow in the first part of the sample, it significantly accelerated during the 1970s and 1980s and has recently flattened out, generating an-S shaped path.

### 3.2 Geographic Evidence

The geographic predictions of our model are a distinctive feature: The rise of women’s labor force participation started in few locations and gradually spread to nearby areas, as information diffused. This section explores the geographic patterns of female labor force participation, using county-level

---

<sup>10</sup>We exclude women living in institutions. We also exclude individuals living on a farm or employed in agricultural occupations since agricultural occupations may make working compatible with child-rearing. We also exclude residents of Alaska and Hawaii. All observations are weighted using the relevant person weights.

<sup>11</sup>There were also changes in the composition of the population over the period: the fraction of married women with children (the group with the lowest participation rate), first increased and then decreased between 1940 and 2005. However, the reduction in the percentage of married women with children, from 53% in 1940 to 45% in 2005, was too small to account for the observed rise in the aggregate.



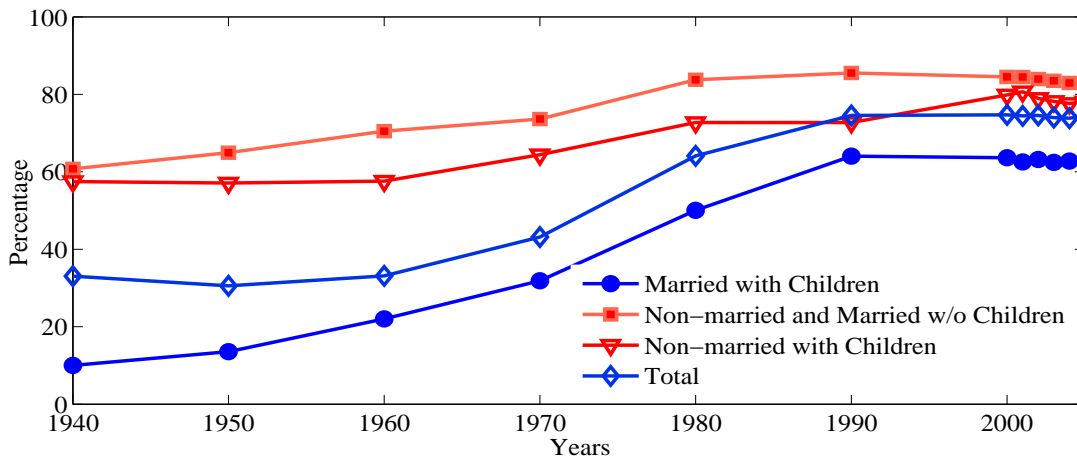


Figure 1: LABOR FORCE PARTICIPATION AMONG SUB-GROUPS OF WOMEN.  
 Details of the data are in appendix B.

U.S. data. The data source is “Historical, Demographic, Economic, and Social Data: The United States, 1790-2000” produced by the Inter-university Consortium for Political and Social Research. We start our analysis in 1940 because the wage data we need for our calibration begin only in 1940. There are 3107 U.S. counties in 1940. After eliminating counties with incomplete information over our entire sample period and excluding Hawaii and Alaska, 3074 counties remain. Our participation series is the number of working-age females in the civilian labor force, divided by the total working-age female population. See appendix B for data details.

Figure 2 maps the labor force participation rate for each U.S. county every twenty years. Darker colors indicate higher levels of female labor force participation. There are three salient features of the data. First, the levels of labor force participation are not uniform: while the average 1940 participation rate was 18.5%, there were counties with participation rates as low as 4.6% and as high as 50%. Second, the changes in participation rates are not uniform. While some areas increased their participation rate dramatically between 1940 and 1960 (for example, the Lake Tahoe region), others stayed stagnant until the 1980’s and witnessed a surge in participation between 1980 and 2000 (for example, southern Minnesota). Third, there is spatial clustering: counties where the female participation rate is over 40% tend to be geographically close to other such counties. These counties are concentrated in the foothills of the southern Appalachians (Piedmont region), in the

North East, Florida, Great Lakes and West coast. Central regions display much lower participation.

To quantify the spatial features of the data and compare those features to the model, we use two statistics, cross-county dispersion and spatial correlation. For each county  $i$  and time  $t$ , we first estimate  $LFP_{it} = \beta_{1t} + \beta_{2t}\text{controls}_{it} + \epsilon_{it}$ . As control variables, we use the county’s demographic characteristics, industrial composition and occupational data.<sup>12</sup>

For dispersion, we compute the standard deviation of the residuals across counties. This is a measure of geographic heterogeneity not attributable to observable economic features. For spatial correlation, we estimate correlation in the residuals of all counties  $i$  and  $j$  within a distance  $d$ :

$$I = \left( \frac{N}{\sum_i \sum_j \iota_{i,j,d}} \right) \frac{\sum_i \sum_j \iota_{i,j,d} \epsilon_i \epsilon_j}{\sum_j \epsilon_j^2}. \quad (10)$$

where  $N$  is the number of counties and  $\iota_{i,j,d} = 1$  if counties  $i$  and  $j$  are within distance  $d$ , meaning that  $(x_j, y_j) \in [x_i - d, x_i + d] \times [y_i - d, y_i + d]$ . This spatial correlation measure is also known as Moran’s I (Moran 1950). It is a measure of local geographic similarity commonly used in fields such as geography, sociology and epidemiology to measure spatial effects.<sup>13</sup> We report both dispersion and correlation, for each decade, and compare them to the model simulation results in section 5.

### 3.3 Direct evidence about changes in beliefs

**Survey responses** Our empirical measure of beliefs is survey responses from 1930-2005. The precise wording of the survey question varies between four different surveys (see appendix B.2 for details and sources). But all of these surveys ask men and women whether they believe that a married woman – some are specific to a woman with children, or preschool-aged children – should participate in the labor force.

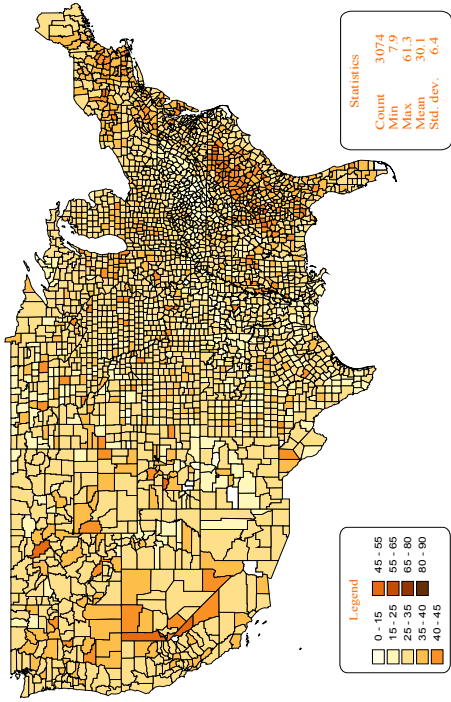
Figure 3 displays the fraction of survey respondents supportive of female employment. It rises over time, in an S-shape pattern that mimics the participation rate. Of course, this does not prove

---

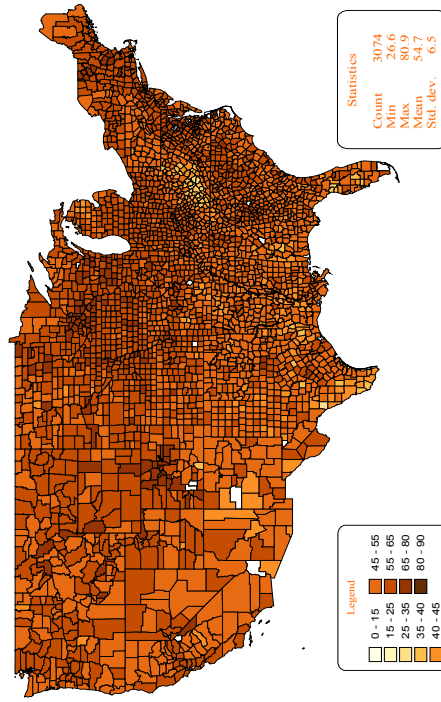
<sup>12</sup>Several different data sets were used in the construction of the panel data of the control variables. Details are in table 2 of the appendix. Table 3 presents the summary statistics for each decade.

<sup>13</sup>While these other literatures frequently try to identify a causal relationship that drives spatial correlation, we make no such attempt here. In both the model and the data, issues like Manski (1993) reflection problems arise. We compare the contaminated moment in the model to the equivalent contaminated moments in the data.

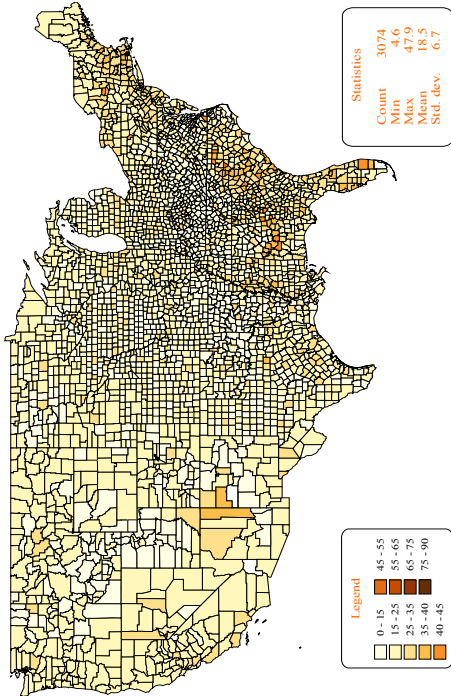
Female LFP 1960



Female LFP 2000



Female LFP 1940



Female LFP 1980

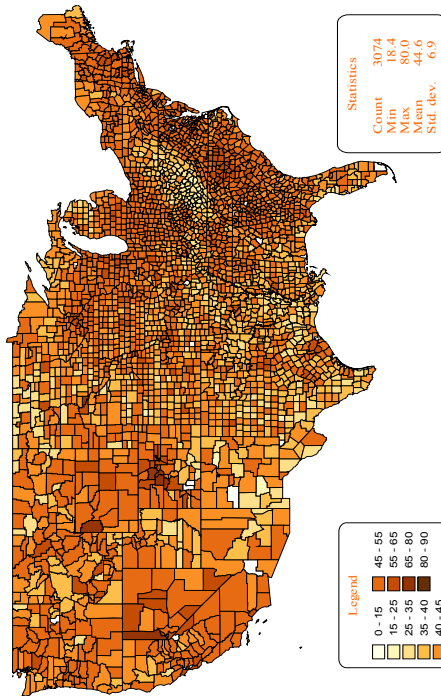


Figure 2: FEMALE LABOR FORCE PARTICIPATION RATE BY U.S. COUNTY.

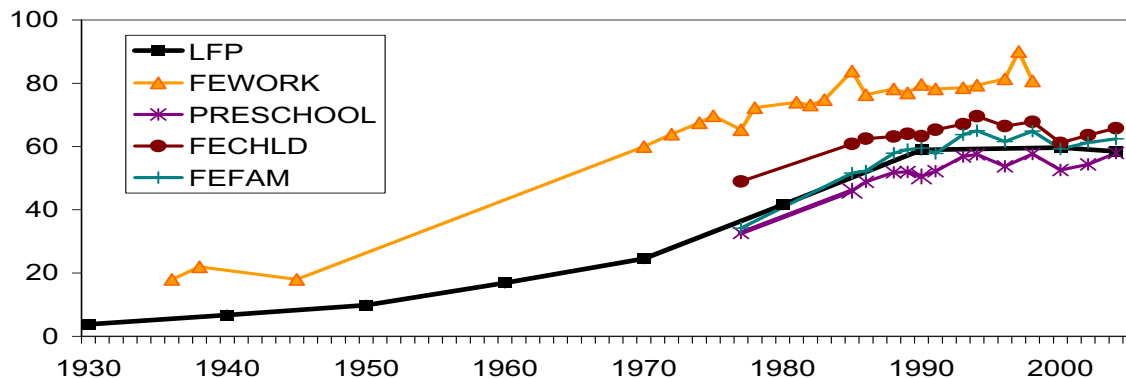


Figure 3: THE FRACTION OF SURVEY RESPONDENTS SUPPORTIVE OF FEMALE EMPLOYMENT. Details of the survey questions are in appendix B.2.

that changes in beliefs caused participation to rise. It could be that people report more support for participation when they see participation rise. However, Farre and Vella (2007) show that women who have more positive responses are more likely to work and more likely to have daughters that work. Causal or not, this is direct evidence that beliefs did change in the way the model predicts.

**Ancestry Evidence** An empirical literature identifies variation in preferences and beliefs that are influenced by one’s society as an important factor in explaining the large differences in women’s labor force participation. Fernández and Fogli (2005) study the working behavior of second generation American women to isolate the effect of preferences and beliefs from that of markets and institutions (see also Antecol (2000), Fortin (2005) and Alesina and Giuliano (2007)). They show that the geographic heritage of these women, as captured by the aggregate labor force of the country of origin of their parents, is significant in explaining their labor force participation behavior and find these results to be even stronger when the women live in an ethnically dense neighborhood. These results suggest that preferences/beliefs matter for women’s participation and that these beliefs are influenced by the parents and local society.

## 4 Calibration

To explore the quantitative predictions of our theory, we calibrate the economy to reproduce some key aggregate statistics in the 1940’s and then compare its evolution over time and across regions

with the data. Because we have census data every 10 years, we consider a period in the model to be 10 years. There are 3025 counties because this is the closest square number to the actual number of U.S. counties (3074). 100 women live in each county. We focus on the dynamics generated by local interactions alone and abstract from changes due to wages, wealth and technology, by holding the costs and benefits of maternal employment fixed over time. Table 1 summarizes our calibrated parameters.

We construct initial 1930 participation to have a geographic pattern that resembles the U.S. data. This enables us to start with reasonable initial dispersion and spatial correlation. Initial participation rates affect subsequent local participation because they determine the probability of observing an informative signal. Appendix C offers additional detail about the calibration targets and initial conditions.

mean log ability	$\mu_a$	-0.90	women's 1940 earnings distribution
std log ability	$\sigma_a$	0.57	women's 1940 earnings distribution
mean log endowment	$\mu_\omega$	-0.28	average endowment = 1
std log endowment	$\sigma_\omega$	0.75	men's 1940 earnings distribution
true value of nurture	$\theta$	0.04	children's test scores (Bernal and Keane 2006)
radius of interaction	d	2	40 miles
outcomes observed	J	4	growth of <i>LFP</i> in 1940's
prior mean $\theta$	$\mu_0$	0.04	unbiased beliefs
prior std $\theta$	$\sigma_0$	0.76	average 1940 <i>LFP</i> level
utility of leisure	L	0.3	1940 <i>LFP</i> of women without kids
risk aversion	$\gamma$	3	commonly used

Table 1: Parameter values for the simulated model and the calibration targets.

**Wages and endowments** The ability and endowment distributions in our model match the empirical distributions of annual labor income of full-time employed, married women with children under age 5 and their husbands. We match the moments for 1940, the earliest year for which we have wage data. Since we interpret women's endowment  $\omega$  as being husbands' earnings, and earnings are usually described as log-normal, we assume  $\ln(\omega) \sim N(\mu_\omega, \sigma_\omega^2)$ . We normalize the average endowment (not in logs) to 1 and use  $\sigma_\omega$  to match the dispersion of 1940 annual log earnings of husbands with children under 5. For the mean  $\mu_a$  and standard deviation  $\sigma_a$  of women's ability,

we match the censored distribution of working women's earnings in the first period of the model to the censored earnings distribution in the 1940 data. Our estimates imply that full-time employed women earn 81% of their husbands' annual earnings, on average.<sup>14</sup>

**True value of nurture** Our theory is based on the premise that the effect of mothers' employment on children is uncertain. This is realistic because only in the last 10 years have researchers begun to agree on the effects of maternal employment in early childhood. Harvey (1999) summarizes studies on the effects of early maternal employment on children's development that started in the early 60s and flourished in the 1980s when the children of the women interviewed in the National Longitudinal Survey of Youth reached adulthood. She concludes that working more hours is associated with slightly less cognitive development and academic achievement, before age 7. More recent work confirms this finding (Hill, Waldfogel, Brooks-Gunn, and Han 2005). Combining Bernal and Keane (2006)'s estimates of the reduction in children's test scores from full-time maternal employment of married women with estimates of the effect of these test scores on educational attainment and on expected wages (Goldin and Katz 1999), delivers a loss of 4% of lifetime income from maternal employment ( $\theta = 0.04$ ).

**Information parameters** Without direct observable counterparts for our information variables, we need to infer them from participation data. Initial beliefs are assumed to be the same for all women and unbiased, implying  $\mu_0 = \theta$ . The alternative, a theory driven by initially biased beliefs, is difficult to rationalize. The same bias would have to be present in every country; otherwise, female labor force participation would start out high and decrease in some countries.

Initial uncertainty  $\sigma_0$  is chosen to match women's 1940 average labor force participation rate in the U.S.. Of course, 1940 participation decisions depend not only on initial uncertainty, but also on the number of signals that women use to update those beliefs  $J$ . We choose  $J$  to match the aggregate growth in labor force participation between 1940 and 1950.

---

<sup>14</sup>A wage gap where women earn 81% of their husbands' income is higher than most estimates. This is due to two factors. First, we do not require husbands to be full-time workers because we want to capture the reality that women's endowments can be high or low. Second, poor women are more likely to be employed. By comparing only husbands of employed women to their wives, we are selecting poorer husbands.

The distance of social interaction  $d$  is difficult to calibrate because the model results are not very sensitive to it. We use a value that is equivalent to 40 miles because spatial correlation drops off quickly beyond that distance. To map this physical distance into the model, we use our county location data to ask: For an average county, how many other counties have centroids within 40 miles of its centroid? The answer is approximately eight. Therefore, we set  $d$  in the model to the length of 2 counties, so that eight neighboring county centroids are located within that radius.

**Preference parameters** Risk aversion  $\gamma$  is 3, a commonly used value. We also add one new parameter, a value for leisure  $L$ , to give women without children some reason not to participate. For women who remain out of the labor force, expected utility is now  $EUO$  (as in equation 7), plus  $L$ . We calibrate  $L$  such that a woman who knows for sure that  $\theta = 0$ , (because she has no child who could be harmed by her employment) participates with a 60% probability, just like women without children in 1940. The exogenous  $L$  parameter explains why some women without children do not work. Our theory explains the difference between women with and without children.

**Alternative parameter values and model timing** Appendix D shows that moderate differences in calibrated parameters do not overturn our results. The exact value of the true  $\theta$ , even a zero or negative value, has only a modest effect on the participation rate that the model converges to at the end. The radius of social interaction  $d$  can be doubled or halved, with no perceptible differences in the results. Replacing some of the initial uncertainty with pessimism (lowering  $\sigma$ , lowering  $\mu_0$ ) slows learning initially. Even optimism can be offset with initial uncertainty. Increasing the number of signals  $J$  speeds the transition but does not change the participation level that the model converges to.

The appendix also explores more significant changes to the model. One extension allows for women with many types (different  $\theta$ 's); the same dynamic emerges, even when women observe more outcomes and aggregate information. Another extension changes the model timing: Women spend 25 years growing up and 10 years having children under age 5.

## 5 Simulation Results

This section compares the model’s predictions for labor force participation rates to the data – first the time series and then the geography. Finally, it examines wage and wealth predictions.

### 5.1 Time Series Results

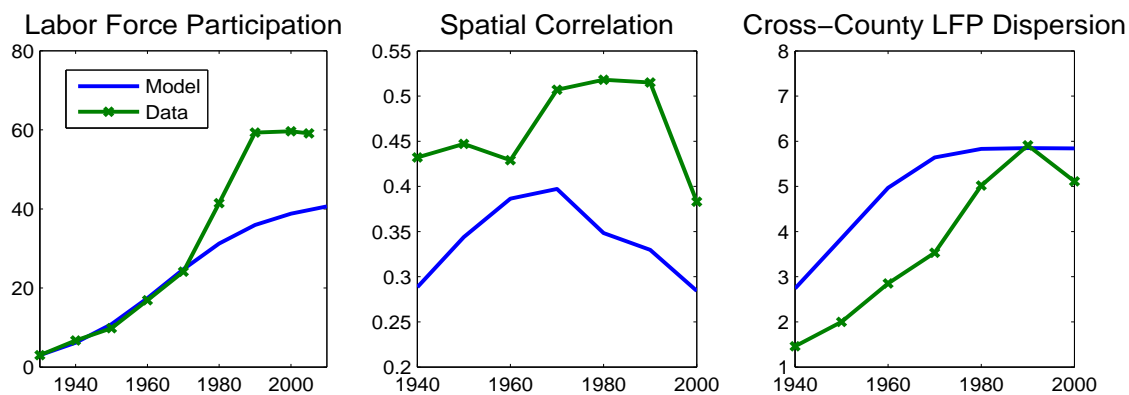


Figure 4: Aggregate level, cross-county heterogeneity and spatial correlation of female labor force participation: data and calibrated model. See section 3 for the construction of dispersion and spatial correlation measures.

By itself, learning can generate a large increase in labor force participation (figure 4). By 2010, our model predicts a 41% participation rate. While this falls short of the 62% rate observed in the 2005 data, the model is missing features like increasing wages, a decline in the social stigma associated with female employment and changes in household durable technologies. One indicator of the size of these effects is the increase in the participation rate of women without children. While 56% of these women participated in 1940, 85% participated in 2005, a 29% increase. If the changes that affected all women were added to the learning effects specific to mothers of small children that this model captures, the results would more than account for the full increase. Yet, the results suggest that up to 2/3rds of the increase in participation could be due to learning.

Participation rises slowly at first, just like in the data. But, the model does not match the sudden take off in the 1970’s. Participation growth is governed by three key parameters: First, the number of signals observed  $J$  matters because more signals means faster learning and faster



participation growth. Second, the amount of noise in each signal  $\sigma_a$  matters because noisy signals slow down learning. Third, the initial degree of uncertainty matters because more uncertain agents weight new information more and thus their beliefs change quickly. This also speeds the transition.

## 5.2 Geographic Results

The most novel results of our model are the geographic ones. While models have not attempted to match these facts, they provide clues about how the female labor force transition took place. The right two panels of figure 4 plot our two geographic measures, dispersion and spatial correlation, for the model and the data.

The “LFP dispersion” measure captures the heterogeneity of participation rates across counties. In both the model and the data, the level of dispersion is similar and is humped-shaped; it rises then falls. For the model, the fall is tiny in year 2000 and only becomes noticeable later. The pronounced drop in dispersion in the 2000 data is partly due to missing county data for that year. See appendix B for a discussion.

Dispersion rises because of the information externality: Regions that initially have high participation generate more informative signals that cause regional participation to rise more quickly. Regions with low participation have slower participation growth; with few women working, not enough information is being generated to cause other women to join the labor force. Later in the century, dispersion falls. This happens because beliefs are converging to the truth. Since differences in beliefs generate dispersion, resolving those differences reduces dispersion.

The second measure is spatial correlation, as defined in (10). This measures how similar a location is to nearby locations and captures the strength of the information externalities. Spatial correlation also rises, then falls. The increase comes from the the information externality. Initially, this effect is weak because when few women work, information is scarce. In the long run, this effect diminishes because once most information has diffused throughout the economy, the remaining cross-country differences are due to ability and endowments, which are spatially uncorrelated. The middle of the transition is when correlation is strongest. Nothing in the calibration procedure ensures that dispersion or spatial correlation looks like the data, after 1930. Therefore, these

patterns are supportive of the model’s mechanism.

### 5.3 Selection Effects on Wealth and Wages

The speed at which women switch from staying at home to joining the labor force depends not just on their location, but also on their socioeconomic status. There are two components to this status: a woman’s own wage and her endowment. Figure 5 shows the mean endowment and wage of a woman, relative to her husband, for employed women. In both the model and the data, wages are censored; they are only measured for the subset of women who participate. The model’s unconditional distribution of endowments and abilities is constant. What is changing is the selection of women who work. In other words, this is primarily a selection effect.



Figure 5: AVERAGE ENDOWMENT AND RELATIVE WAGE FOR WORKING WOMEN, BELIEF DISPERSION FOR ALL WOMEN.

Average relative wage is the woman’s wage divided by her husband’s wage ( $w_{it}/\omega_{it}$ ), averaged over all employed women. Belief dispersion is  $\text{std}(\mu_{i,t})$ , taken in each period  $t$  over all women  $i$ .

Employed women’s endowments are low at the start of the sample since many women joined the labor force because they are poor and desperate for income. As women learn and employment poses less of a risk, less poor women also join. The average endowment of working women rises. This prediction distinguishes our theory from others. For example, since women with larger endowments can afford new appliances and child care first, technology-based explanations predict that richer women join first.

The finding that women’s relative wages declined in the early part of the sample is supported by O’Neill (1984) who documents a widening of the male-female wage gap in the mid-50’s to 70’s.

She attributes it to the same selection effects that operate in our model: Not only are husbands of employed women becoming richer, less skilled women are also entering the labor force. One reason that women worked at the start of the sample was that they were very highly skilled. Those women earned high wages. As learning made employment more attractive, less skilled women joined as well, lowering the average wage women earn.

**Belief heterogeneity and measurement error** One question that remains is: Why do endowments fall and relative wages rise at the end of the sample? This small effect is not simulation error; extending the simulation a few more decades reveals this is a persistent trend. Instead, the answer lies in the heterogeneity of beliefs.

The cross-sectional dispersion of both means and variances of beliefs about  $\theta$  rises initially as women in different locations see varying amounts of information and signal realizations. But as information accumulates, beliefs converge to the truth, uncertainty converges to zero and belief dispersion falls. This matters for the relationship between aggregate variables in the model because belief dispersion is a source of unmeasured heterogeneity that affects participation decisions. In other words, it acts like noise in an estimation and makes variables look less related.

Starting in 1990, differences in women's participation decisions are driven less by differences in beliefs, which are starting to converge, and more by differences in endowments and abilities. Thus, the endowment and wage selection effects become stronger again. This finding offers a warning about interpreting a wide range of statistics concerning female labor force participation. If there are significant changes in belief heterogeneity that affect participation over the 20th century, many estimated relationships between participation and other economic determinants of labor force participation will be biased.

**Rising unconditional wages** While learning offers one explanation for wage changes, there are obviously other factors external to the model that have contributed to this trend. But feeding the time series of wages in to the model has a negligible effect. Wage-based theories rely on mechanisms that raise labor supply elasticity to make wages matter. Our model has no such

mechanism. Learning makes elasticity even lower because heterogeneity in beliefs makes fewer women marginal workers. The next section illustrates how rising wages and learning can interact.

## 6 Extending the model: wages and career choice

The increase in female labor force participation is not the only phenomenon that might be influenced by local information and that has rich geographic patterns. Many types of social change could be modeled using this type of framework. To give a sense of how this model framework might be used to address a broader set of issues, we illustrate one direction in which the model could be extended. This extension examines women’s career choice.

**Model** The timing, the number of agents and preferences are the same as in the standard model. What differs is that a woman has an additional career option. If she chooses a high-intensity career, she gets a known multiple  $\tilde{w} > 1$  of her baseline wage, but may further compromise her ability to nurture her child. Agents learn about two unknown parameters: the value of nurture  $\theta$  and the toll on a child of high-intensity maternal employment  $\tilde{\theta}$ .

The budget constraint of the individual from family  $i$  born at time  $t - 1$  is

$$c_{it} = (n_{it} + h_{it}\tilde{w})w_{it} + \omega_{it} \tag{11}$$

where  $h_{it} \in \{0, 1\}$  indicates the choice of a high-intensity career;  $\omega_{it}$  is an endowment which could represent a spouse’s income, and  $n_{it} \in \{0, 1\}$  is the choice to join a low-intensity career. If the agent works in the labor force,  $n_{it} = 1$ . If she works in a high intensity career,  $h_{it} = 1$ . A woman can only have one career:  $n_{it} + h_{it} \leq 1$ .

As before, endowed ability is  $a_{i,t} \sim N(\mu_a, \sigma_a^2)$ . If a mother stays home with her child, the child’s full natural ability is achieved. If the mother chooses a low-intensity career, some unknown amount  $\theta$  of the child’s ability will be lost; for a high-intensity career, the loss is  $\tilde{\theta} > \theta$ :

$$w_{i,t} = \exp(a_{i,t} - n_{i,t-1}\theta - h_{i,t-1}\tilde{\theta}). \tag{12}$$

The constants  $\theta$  and  $\tilde{\theta}$  are not known when making labor supply decisions. Initial beliefs are  $\theta \sim N(\mu_0, \sigma_0^2)$  and  $\tilde{\theta} \sim N(\tilde{\mu}_0, \tilde{\sigma}_0^2)$ , where  $(\mu_0 - \theta)$  and  $(\tilde{\mu}_0 - \tilde{\theta})$  are independent. The high-intensity career is initially thought to be more costly ( $\tilde{\mu}_0 > \mu_0$ ) and more risky ( $\tilde{\sigma}_0^2 > \sigma_0^2$ ). Each generation updates beliefs using Bayes law (4 and 5) and by observing wages and nurturing decisions for themselves and for the same set  $\mathbb{J}_i$  of peers as in the original model. Ability  $a$  is never observed so that neither  $\theta$ , nor  $\tilde{\theta}$  can be perfectly inferred from the wage. An important feature of (12) is that a wage  $w_{i,t}$  is only informative about  $\tilde{\theta}$  if the mother had an intense career ( $h_{i,t-1} = 1$ ).

**Discussion of model results** Since the high-intensity career is initially more uncertain, few women participate in it initially. Thus, for moderate levels of the wage premium, the early participants are primarily in low-intensity (regular) careers. Since high-intensity participation takes off later, the composition of careers changes over time. The growth in the fraction of employed women participating in high-intensity careers increases the average wage of working women. This could be one component of the explanation for a rise in female wages and its geographic patterns.

Women who work in the high-intensity sector early on are the highest-ability women. Because we assumed a multiplicative wage premium, high-ability women earn more additional income from high-intensity careers. Thus as women learn faster in some regions than in others, the degree of occupation sorting will diverge and then converge again as information diffuses and beliefs converge to the truth. Appendix H details the basic results, analytically and numerically.

## 7 Conclusion

Many changes have contributed to the increase in female labor force participation over the last century. We do not argue that beliefs were the only relevant change. Rather, the model abstracts from other changes to focus on how the transition from low to high participation can be regulated by learning in a way that matches the time-series and geographic data. Including local information transmission as part of the story of female labor force participation in the 20th century helps to explain its gradual dynamic and geographic evolution.

While this paper used the evolution of geographic heterogeneity of the labor force transition

to evaluate the strength of the information transmission channel, another empirical exercise could examine the effect of policy changes. Consider a policy designed to encourage maternal employment, but whose true effect on the cost and benefit of employment is uncertain. Upon policy passage, uncertainty would dampen the effect on participation. However, women in regions with higher participation would observe more outcomes that were informative about the policy's effect and would increase their participation faster than regions with initially low participation. Exploring this prediction could lend additional support to the model.

One direction to extend the theoretical framework is to think more broadly about how social change arises. One important feature of social behavior that this model does not capture is the desire to fit in or coordinate with others. Using an objective function like that in beauty contest games (Morris and Shin 2002), coupled with the geographic nature of information transmission, could provide a rich set of testable implications. Specifically, it could predict geographic patterns, like the spread from urban to rural areas, in the types of cultural changes investigated by Greenwood and Guner (2005), Guiso, Sapienza, and Zingales (2006) and Bisin and Verdier (2001). Such work could help differentiate exogenous changes in preferences from information-driven changes in coordination outcomes.

Another direction one could take this model is to interpret the concept of distance more broadly. Arguably, socioeconomic, ethnic, religious or educational differences create stronger social barriers between people than physical distance does. If that is the case, the learning dynamics that arise within each social group may be quite distinct. If the initial conditions in these social groups differ, changes in labor force participation, career choice, or social norms may arise earlier in one group than in another. This model provides a vehicle for thinking about the diffusion of new behaviors, with uncertain consequences, among communities of people.

## References

- ALBANESI, S., AND C. OLIVETTI (2007): “Gender Roles and Technological Progress,” NBER Working Paper 13179.
- ALESINA, A., AND P. GIULIANO (2007): “The Power of the Family,” NBER Working Paper 13051.
- AMADOR, M., AND P.-O. WEILL (2006): “Learning from Private and Public Observations of Others’ Actions,” Working Paper.
- ANTECOL, H. (2000): “An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates,” *Labour Economics*, 7, 409–426.
- ATTANASIO, O., H. LOW, AND V. SANCHEZ-MARCOS (2008): “Explaining Changes in Female Labour Supply in a Life-Cycle Model,” *American Economic Review*, forthcoming.
- BERNAL, R., AND M. KEANE (2006): “Child Care Choices and Childrens Cognitive Achievement: The Case of Single Mothers,” Northwestern University, Working Paper.
- BISIN, A., AND T. VERDIER (2001): “The Economics of Cultural Transmission and the Evolution of Preferences,” *Journal of Economic Theory*, 97(2), 298–319.
- BUERA, F., A. MONGE-NARANJO, AND G. PRIMICERI (2006): “Learning the Wealth of Nations,” Northwestern University working paper.
- COVER, T., AND J. THOMAS (1991): *Elements of Information Theory*. John Wiley and Sons, New York, New York, first edn.
- DEL BOCA, D., AND D. VURI (2007): “The Mismatch between labor supply and child care,” *Journal of Population Economics*, 4.
- DOEPKE, M., AND F. ZILIBOTTI (2008): “Occupational Choice and the Spirit of Capitalism,” *Quarterly Journal of Economics*, forthcoming.
- DUXBURY, L., AND C. HIGGINS (2003): *2001 National Work-Life Conflict Study: Report I*. Health Canada.
- FARRE, L., AND F. VELLA (2007): “The Intergenerational Transmission of Gender Role Attitudes and its Implications for Female Labor Force Participation,” Georgetown Working Paper.
- FERNÁNDEZ, R. (2007): “Culture as Learning: The Evolution of Female Labor Force Participation over a Century,” Working paper.
- FERNÁNDEZ, R., AND A. FOGLI (2005): “An Empirical Investigation of Beliefs, Work and Fertility,” NBER Working Paper 11268.
- FERNÁNDEZ, R., A. FOGLI, AND C. OLIVETTI (2004): “Mothers and Sons: Preference Formation and Female Labor Force Dynamics,” *Quarterly Journal of Economics*, 119(4), 1249–1299.
- FORTIN, N. (2005): “Gender Role Attitudes and the Labor Market Outcomes of Women Across OECD Countries,” *Oxford Review of Economic Policy*, 21, 416–438.

- FUCHS-SCHUNDELN, N., AND R. IZEM (2007): “Explaining the Low Labor Productivity in East Germany - A Spatial Analysis,” Harvard University Working Paper.
- GOLDIN, C. (1990): *Understanding the Gender Gap*. Oxford University Press.
- (1995): “The U-shaped Female Labor Force Function in Economic Development and Economic History,” in *Investment in Human Capital*, ed. by T. P. Schultz. University of Chicago Press.
- GOLDIN, C., AND L. KATZ (1999): “The Returns to Skill in the United States across the Twentieth Century,” NBER Working Paper # 7126.
- (2002): “The Power of the Pill: Oral Contraceptives and Women’s Career and Marriage Decisions,” *Journal of Political Economy*, 100, 730–770.
- GREENWOOD, J., AND N. GUNER (2005): “Social Change,” *Economie d’avant gard*, research Report 9, University of Rochester.
- GREENWOOD, J., A. SESHADRI, AND M. YORUKOGLU (2005): “Engines of Liberation,” *Review of Economic Studies*, 72(1), 109–133.
- GUIISO, L., P. SAPIENZA, AND L. ZINGALES (2006): “Does Culture Affect Economic Outcomes?,” *Journal of Economic Perspectives*, 20(2), 23–48.
- HARVEY, E. (1999): “Short-Term and Long-Term Effects of Early Parental Employment on Children of the National Longitudinal Survey of Youth,” *Developmental Psychology*, 35(2), 445–459.
- HILL, J., J. WALDFOGEL, J. BROOKS-GUNN, AND W. HAN (2005): “Maternal Employment and Child Development: A Fresh Look Using Newer Methods,” *Developmental Psychology*, 41(6), 833–850.
- JONES, L., R. MANUELLI, AND E. MCGRATTAN (2003): “Why Are Married Women Working So Much?,” Research Department Staff Report 317, Federal Reserve Bank of Minneapolis.
- LUCAS, R. (1972): “Expectations and the Neutrality of Money,” *Journal of Economic Theory*, 4(2), 103–124.
- MAMMEN, K., AND C. PAXSON (2000): “Women’s Work and Economic Development,” *Journal of Economic Perspectives*, 14(4), 141–164.
- MANSKI, C. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *The Review of Economic Studies*, 60(3), 531–542.
- MORAN, P. (1950): “Notes on continuous stochastic phenomena,” *Biometrika*, 37, 17–23.
- MORRIS, S., AND H. SHIN (2002): “The Social Value of Public Information,” *American Economic Review*, 92, 1521–1534.
- MUNSHI, K. (2004): “Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution,” *Journal of Development Economics*, 73, 185–213.



- OLIVETTI, C. (2006): “Changes in Women’s Hours of Market Work: The Effect of Changing Returns to Experience,” *Review of Economic Dynamics*, 9, 557–587.
- O’NEILL, J. (1984): “The Trend in the Male-Female Wage Gap in the United States,” *Journal of Labor Economics*, 3(1), S91–S116.
- VELDKAMP, L. (2005): “Slow Boom, Sudden Crash,” *Journal of Economic Theory*, 124(2), 230–257.

## A Proofs of analytical results

### A.1 Derivation of comparative statics

Step 1: Define a cutoff wage  $\bar{w}$  such that all women who observe  $w_{i,t} > \bar{w}$  choose to join the labor force. A woman joins the labor force when  $EUW_{it} - EUO_{it} > 0$ . Note that  $\partial \mathcal{N}_{i,t} / \partial w_{it} = (n_{it} w_{it} + \omega_{it})^{-\gamma} > 0$ . Since  $\mathcal{N}_{i,t}$  is monotonically increasing in the wage  $w$ , there is a unique  $\bar{w}$  for each set of parameters, such that at  $w = \bar{w}$ ,  $\mathcal{N}_{i,t} = 0$ .

Step 2: Describe the probability of labor force participation. Let  $\Phi$  denote the cumulative density function for the unconditional distribution of wages in the population. This is a log-normal c.d.f. Since the lognormal is unbounded and has positive probability on every outcomes, its c.d.f. is therefore strictly increasing in its argument. Then, the probability that a woman participates is  $1 - \Phi(\bar{w})$ , which is then strictly decreasing in  $\bar{w}$ .

Step 3: The effect of mean beliefs on labor force participation. Taking the partial derivative of the net utility gain from labor force participation yields  $\partial \mathcal{N}_{i,t} / \partial \mu_{i,t} = -\beta$ . By the implicit function theorem,  $\partial \bar{w} / \partial \mu_{i,t} > 0$ . Thus,  $\partial(1 - \Phi(\bar{w})) / \partial \mu_{i,t} = (\partial(1 - \Phi(\bar{w})) / \partial \bar{w})(\partial \bar{w} / \partial \mu_{i,t}) < 0$ .

Step 4: Calculate the effect of uncertainty on labor force participation. The benefit to participating is falling in uncertainty:  $\partial \mathcal{N}_{i,t} / \partial \sigma_{i,t} = (1 - \gamma)\beta \exp((\mu_a - \mu_{i,t})(1 - \gamma) + \frac{1}{2}(\sigma_a^2 + \sigma_{i,t}^2)(1 - \gamma)^2)$ . Since  $\gamma > 1$ ,  $\beta > 0$  by assumption, and the exponential term must be non-negative, this means that  $\partial \mathcal{N}_{i,t} / \partial \sigma_{i,t}^2 < 0$ . As before, the implicit function theorem tells us that  $\partial \bar{w} / \partial \sigma_{i,t}^2 > 0$ . Thus,  $\partial(1 - \Phi(\bar{w})) / \partial \sigma_{i,t}^2 = (\partial(1 - \Phi(\bar{w})) / \partial \bar{w})(\partial \bar{w} / \partial \sigma_{i,t}^2) < 0$ .

### A.2 Proof of result 1: Zero participation is not a steady state

Proof: For any arbitrary beliefs  $\mu_{jt}$ ,  $\sigma_{jt}$  and endowment  $\omega_{jt}$ , there is some finite level of ability  $a^*$  and an associated wage  $w^* = \exp(a^*)$ , such that  $EUW_{it} > EUO_{it} > 0$ ,  $\forall a_{jt} \geq a^*$ . The fact that  $a_{jt}$  is normally distributed means that  $Prob(a_{jt} \geq a^*) > 0$  for all finite  $a^*$ . Since woman  $j$  enters the labor force whenever  $EUW_{it} > EUO_{it} > 0$ , and this happens with positive probability,  $n_{jt} = 1$  with positive probability. Since this is true for all women  $j$ , it is also true that  $\sum_j n_{jt} \geq 1$  with positive probability.

### A.3 Proof of result 2: Geographic correlation

Let  $\alpha$  be the fraction of women who participate in family  $i$ 's region. The region a woman lives in does not affect her endowments or ability. Therefore,  $\mathcal{N}_{i,t+1}$  can be rewritten, using (7) and (8) as

$$\mathcal{N}_{i,t+1} = A - B \exp\{(\gamma - 1)\mu_{i,t+1} + \frac{1}{2}(\gamma - 1)^2 \sigma_{i,t+1}^2\}$$

for positive constants  $A$  and  $B$ . Since woman born at time  $t$  participates if  $\mathcal{N}_{i,t+1} > 0$ , it suffices to show that  $\partial \mathcal{N}_{i,t+1} / \partial \alpha > 0$ , for a woman with average prior beliefs and an average signal.

The number of informative signals that a woman in family  $i$ , with an average signal draw, would see is  $\bar{n}_{it} = \alpha J$ . Since beliefs and signals are unbiased by construction, then a woman with average prior beliefs has  $\mu_{it} = \theta$  and a woman with an average signal has  $\hat{\mu}_{it} = \theta$ . By equation (4), her posterior belief is  $\mu_{i,t+1} = \theta$ , for any fraction  $\alpha$ . Her posterior precision does depend on  $\alpha$ : According to equation 5, the definition of  $\hat{\sigma}_{i,t+1}^2$ , and the equation for  $\bar{n}_{it}$  above,  $\sigma_{i,t+1}^{-2} = \sigma_{i,t}^{-2} + \alpha J / \sigma_a$ . Since  $J$  and  $\sigma_a$  are both positive, posterior precision is increasing in  $\alpha$ . Thus, posterior variance  $\sigma_{i,t+1}^2$  is decreasing in  $\alpha$ , and  $\mathcal{N}_{i,t+1}$  is increasing in  $\alpha$ .

## B Data: Sources and Definitions

### B.1 County-level data

County-level data come from “Historical, Demographic, Economic, and Social Data: The United States, 1790-2000” produced by the Inter-university Consortium for Political and Social Research (series 2896). This data set is a consistency-checked and augmented version of the the Integrated Public Use Microdata series, produced by the Minnesota Population Center. Table 2 lists the demographic, industrial and occupation control variables and their data sources. Table 3 documents their summary statistics, by decade.

The matrix of distances between county centroids is the “ground distance circle” that comes from CDA Transportation Network.

**Missing observations** One data issue we were concerned with was potential bias in our estimates from excluding counties with missing data. We also did not control for wages because that data was so scarce. As can be seen in table 3, we are missing the sectoral composition for some counties in 1940 and in 2000. We are also missing 7 observations on education in 1950. We re-calculate the residuals from the regression  $LFP_{it} = \beta_{1t} + \beta_{2t}\text{controls}_{it} + \epsilon_{it}$ , excluding sectoral composition and wages and found no discernable difference between the properties of these residuals and those from an unbalanced panel, with one exception. In 2000, many counties are missing entries for the industrial sector. When we balanced the panel by excluding industrial sector data for all years and recovered the additional counties for 2000, the spatial correlation measure rose from 0.38 to 0.45. It is possible that spatial correlation rose because of spatial correlation in industrial sector composition that is now attributed to information. However, the correlation did not rise (in the first two significant digits) in the previous decades when industrial sectors were excluded. This suggests that most of the variation in sectors is also captured by occupational and demographic variables and that the change in correlation is due to the sparser data available in 2000. Therefore, we use the higher estimate, on the full sample of data for the 2000 spatial correlation estimate in figure 4.

### B.2 Survey data

The survey data from GSS begin only in 1972. However, the increasing speed of female entry in the labor force (start of the S) precedes that date. To establish the contemporaneous S-shaped evolution of beliefs, it is vital to have more historical data. We have one measure of beliefs that is collected infrequently, since the 1930’s. This data are from IPOLL databank, maintained by the Roper Center for Public Opinion Research. Unfortunately, the phrasing of the questions differs slightly over time. We describe below the questions and the replies.

**August 1936** The Gallup Poll asked: “Should a married woman earn money if she has a husband capable of supporting her?” 18% said yes, 82% no. No uncertain or no response entries were allowed.

**October 1938** The Gallup Poll asked: “Do you approve of a married woman earning money in business or industry if she has a husband capable of supporting her?” 22% approve, 78% disapprove.

**November 1945** The Gallup Poll (AIPO) asked: “Do you approve or disapprove of a married woman holding a job in business and industry if her husband is able to support her?” 62% disapprove, 18% approve. The rest of the replies are miscellaneous open answers (e.g., if she has a good job, if she has no children, etc.).

**June 1970** The Gallup Poll asked: “Do you approve of a married woman earning money in business or industry if she has a husband capable of supporting her?” 60% approve, 36% disapprove, 4% do not know.

From 1977 on, data come from <http://webapp.icpsr.umich.edu/GSS/>. The question is: *Do you agree with the following statement: A preschool child is likely to suffer if his or her mother works.* (Strongly agree=1, agree=2, disagree=3, strongly disagree=4, don't know=8, no answer=9, na=0). The only modification we make is to treat “don't know” and “na” replies as missing observations. There are 14 observations, one in 1977, and then at least every two years from 1995-2004. There are between 890 and 2,344 responses per year, totalling 19,005 observations. The average reply ranges from 2.2 in 1977 to 2.6 in 2004.

Merging the two data series: From the Roper data, there are 3 observations available before 1967 and then regular observations starting in 1970. For each of the pre-1977 observations, we compute the growth rate from one data point to the next. Then, we apply these same growth rates to project our preschool data back from 1977 to the earlier observations. We believe that using one series to infer another is a reasonably accurate procedure because for years in which both survey questions are asked, the correlation in the replies is 0.75.

## C Calibration

Throughout, we look at women 25-54, with their own child younger than 5 living in the household. We use whites not living in an institution or on a farm, and not working in agriculture. The time-series data we have from the census is much more detailed than the county-level data. That additional detail allows us to distinguish which women are married with children under five. Since it is these women our theory is oriented toward, it makes sense to compare the model results to this restricted sample of women. But in the county data, we only have participation rates for all women. Therefore, we adjust all the county data by a decade-specific scaling factor that is the ratio of the participation rate of all women in the census data to the participation rate of married mothers with children under five in the census. This re-scaling ensures that the average participation rate across counties is the aggregate participation rate in each decade.

**Abilities** The distribution of women's abilities is constructed so that their wages in the model match the distribution of women's wages in the 1940 census data.  $\sigma_a = .57$  is the standard deviation of log ability and  $\mu_a = \ln(\text{earnings gap}) - (\sigma_a^2)/2$  is the mean of log ability. These parameters target the initial ratio between average earnings of working women and average earnings of all husbands (0.8 in the data) and target the standard deviation of log earnings of employed women in the data (0.53).

**Selection effects in the model** The distribution of observed wages in the data needs to be matched with the distribution of wages for employed women in the model. Employed women are not a representative sample. They are disproportionately high-skill women. The calibration deals with this issue by matching the truncated distribution of wages in the data to the same truncated sample in the model. In other words, we use the model to back out how much selection bias there is.

**Endowment distribution** Data come from the census. We use husbands' wages in 1940 (first available year). From this, we construct two pools of matched data: One is only married women; the other is their husbands.

The log endowment is normal. For these two sets of wage data, we take the log of wages over previous year. For husbands,  $\text{mean}(\log\_incwage\_husb) = 7.04$  and  $\text{std}(\log\_incwage\_husb) = 0.73$ . Therefore, we set  $\sigma_\omega = 0.73$ . We choose the mean log endowment  $\mu_\omega = -(\sigma_\omega^2)/2$  such that mean endowment is normalized to 1.

**True value of nurture** To calibrate the  $\theta$  parameter, we use micro evidence on the effect of maternal employment on the future earnings of children. Our evidence on the effect of maternal employment comes from the National Longitudinal Survey of Youth (NLSY), in particular the Peabody Picture Vocabulary Test (PPVT) at age 4 and the Peabody Individual Achievement Test (PIAT) for math and reading recognition scores measured at age 5 and 6. One year of full time maternal employment plus informal day care reduces test scores by roughly 3.4% (Bernal and Keane 2006). If a mother works from one year after birth until age six, these five years of employment translate in to a score reduction of 17%.

The childhood test scores are significantly correlated with educational attainment at 18. A 1% increase in the math at age 6 is associated with .019 years of additional schooling. A 1% increase in the reading test score at age 6 is associated with .025 additional school years. Therefore, five years of maternal employment translates into between 0.32 ( $17 \cdot .019$ ) and 0.42 ( $17 \cdot .025$ ) fewer years of school.

The final step is to multiply the change in educational attainment by the returns to a college education. We use the returns to a year of college from 1940 to 1995 from Goldin and Katz (1999). Their estimates are the composition-adjusted log weekly wage for full-time/full-year, non agricultural, white males. Those estimates are 0.1, 0.077, 0.091, 0.099, 0.089, 0.124, and 0.129 for the years 1939, 1949, 1959, 1969, 1979, 1989, and 1995. The average return to a year of college is 10%. Since maternal employment reduces education by 0.32-0.42 years, the expected loss in terms of foregone yearly log earnings is about 4%, or  $\theta = 0.04$ .

**Number of signals**  $J$  is calibrated to get the aggregate labor force participation to rise from 6% in 1940 to 10% in 1950.

**Initial Participation in 1930 (heterogeneous across regions)** We want to preserve some of the spatial information in our data. However, the model is on a square grid. Mapping irregular-sized US counties onto this grid is a challenge. To do this, we used regions, which are larger than counties. Regions are constructed by taking the 48 contiguous states, computing the county centroid with the highest and the lowest longitude (call the difference between the maximum and minimum *lodist*), and dividing the US map into  $n$  vertical strips, each with width  $lodist/n$ . Then, for each strip, we compute the maximum and minimum latitude, and divide the strip into  $n$  boxes of equal height. We choose  $n = 10$  because it is the largest possible number that does not result in there being boxes containing no county centroids.

In the model, we divide the evenly-spaced agents into 100 regions of equal size and population. For each of these 100 regions, we assign the participation rate of the corresponding box on the U.S. map and assign agents randomly to participate or not. Each participates with a probability given by the regional participation rate. After calibrating initial participation, this regional aggregation structure is never used again and we compute statistics at the more local, county level.

## D Robustness Checks

Increasing the number of signals makes agents learn faster and makes participation rise faster, while reducing the number of signals has the opposite effect (panel A). Eliminating the cost of maternal employment increases the ending participation rate moderately, while doubling the cost lowers it (panel B). Making agents more uncertain and more optimistic initially about the costs of labor force participation has very little effect on the model. The reverse, lowering uncertainty but raising the estimated cost of maternal employment has a net positive effect on participation, in the first few decades (panel C). Doubling or halving the distance of social interaction has no perceptible effect on participation (panel D).

## E Model with Multiple Types of Women

This extension of the model introduces multiple types of women with different  $\theta$ 's. The idea is that women need to observe other women like themselves to determine what the cost of maternal employment is for their type. Professionals do not learn from seeing hourly workers. A female doctor who is on call all night does not learn about her  $\theta$  from seeing the children of 9-5 workers, and urban mothers face different challenges and costs from rural ones. In this richer model, women can observe many more signals as well as aggregate information and still learn slowly about the  $\theta$  for their type.

The model is the same as the benchmark except for the following changes. Suppose there are  $\Omega$  different types of women, indexed by  $\omega$ . A woman of type  $\omega$  has a cost of maternal employment  $\theta_\omega \sim N(\bar{\theta}, \sigma_\theta^2)$ , where the  $\theta$ 's are i.i.d. across types. A woman's type  $\omega$  is publicly observable.

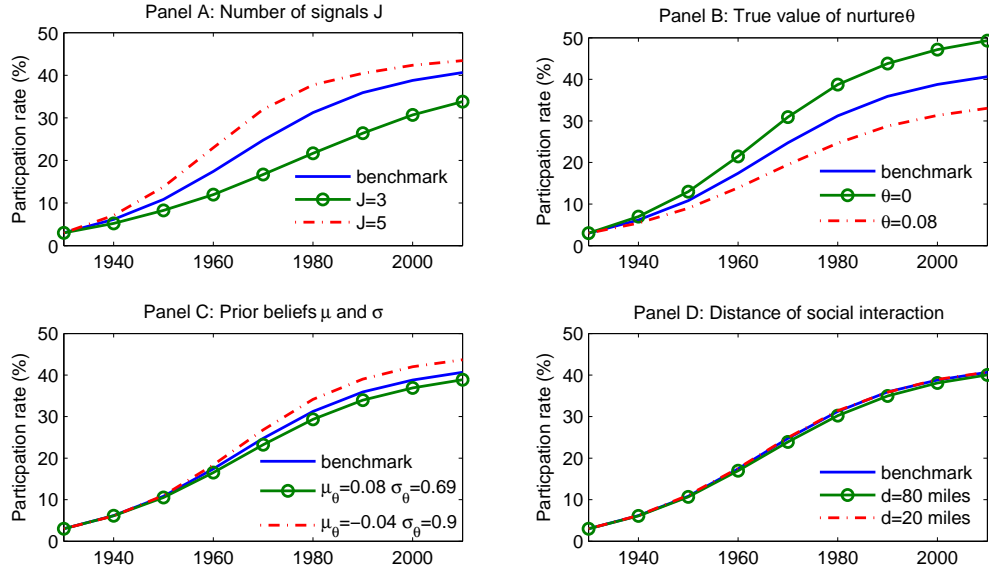


Figure 6: Robustness exercises.

Note that women now know the true cost of maternal employment for the average woman,  $\bar{\theta}$ . Therefore, new research, magazine articles, or aggregate statistics contain no new information. Instead of learning about what the cost of maternal employment is for the average woman, this woman is now learning about how the cost of maternal employment for her type of woman differs from that average.

**Simulation results** We use the same calibration as the benchmark model, except that there are now 5 types of women, with  $\theta_\omega$ 's equally spaced between 0.3 and 0.5. Each woman observed 20 signals and knew that the true mean of  $\theta$  across all types was 0.4. The results in figure 7 are almost indistinguishable from those of the benchmark model (figure 4).

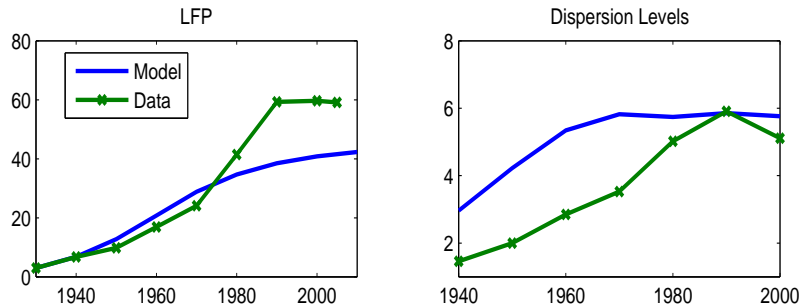


Figure 7: LABOR FORCE PARTICIPATION WITH MULTIPLE TYPES OF WOMEN WHO OBSERVE AGGREGATE INFORMATION.

## F Model with Learning from Others' Choices

To keep the model simple and tractable, we assumed that women do not draw any inference from the labor decisions of other women. They use the knowledge of whether  $J$  of their peers were nurtured in order to estimate the cost of maternal employment. But they do not take advantage of the fact that the mother's

employment decision reveals something about the mother’s beliefs, which is additional information about the true value of nurture  $\theta$ .

This section shows that our simplifying assumption is innocuous. Seeing other women’s labor force decisions does not significantly speed up learning for five reasons: 1) Participation is a binary choice. The binary nature of the signal eliminates much of its information. 2) Early on, most women do not work and other women expect that the women they encounter will likely not work. Therefore, a woman who observes another woman not working early in the century gets very little new (unexpected) information. Observing working women is informative but it becomes commonplace only later in the century when most of the learning has been completed. 3) Women observe the participation decisions of women from the previous cohort. Those women were less informed and less likely to work. 4) The “noise” in women’s participation decisions is large. Women don’t know others’ ability, don’t know whether the mother was nurtured, and don’t know how uncertain they were. Through all this noise, the belief about the mean is a weak signal. 5) The beliefs of others in your region are highly correlated with your own beliefs because people in the same region see common signals. A correlated signal contains less information than an independent one.

To quantify these claims, we simulate an economy that is an approximation to the economy where women learn from the decisions of other women. To keep the linear Bayesian updating rules, we consider an economy where women observe additional normally distributed signals whose signal-to-noise ratio is the same as the information embedded in the participation decisions they observe. This is an upper bound on how much additional information comes from others’ decisions because normally distributed signals contain more information than any other kind of signal with the same signal-to-noise ratio (Cover and Thomas 1991).

To estimate the signal-to-noise ratio of women’s employment decisions, run a regression of participation on beliefs. Since the informativeness of women’s labor decisions changes over time, there should be a separate regression run for each decade. Compute the  $R^2$ . The signal-to-noise ratio, the ratio of the explained sum of squares (signal) to unexplained (noise), is  $R^2/(1 - R^2)$ . To construct a signal with the same amount of noise, first compute the cross-sectional variance of women’s beliefs. This is the total sum of squares. Multiply this variance by  $1 - R^2$  to get the unexplained sum of squares. Create an  $m \times 1$  vector of i.i.d. normal random variables with mean zero and variance  $(1 - R^2)var(\mu_t)$ , where  $m$  is the number of women in the economy. Add this noise shock to the vector of women’s beliefs. Each woman in generation  $t + 1$  sees a subset of the signals about generation  $t$  beliefs, where the subset is the signals with indices  $j \in \mathbb{J}_i$ .

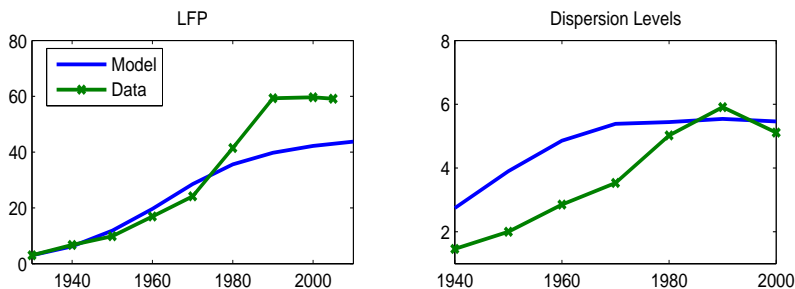


Figure 8: LABOR FORCE PARTICIPATION WHEN WOMEN OBSERVE PARTICIPATION DECISIONS OF OTHERS.

The time-series of labor force participation that results from simulating this model, with the same calibrated parameters as in table 1 of the main text, appears in figure 8. This approach generates a labor force participation rate that is only a couple of percentage points higher at the end of the sample. Thus, learning from other women’s participation choices does speed up the increase in labor force participation by speeding up learning, but its effect is small.

## G Changing Model Timing: 25 years from birth until motherhood

The model is designed to explain the labor force participation decisions of women with children under 5 years of age. The majority of these women in the census data are between the ages of 25 and 35, with an average age of 32. This 10-year interval is part of the reason why we look at 10-year periods. Whether women return to the labor force afterwards or not is not something our theory has anything to say about, nor is it relevant for the participation rates of our subgroup. What our timing assumptions miss is that it takes about 25 years between when a girl is born and when she makes her decision about maternal employment. Therefore, the decisions of mothers determine the information that others observe 25 years, not 10 years later.

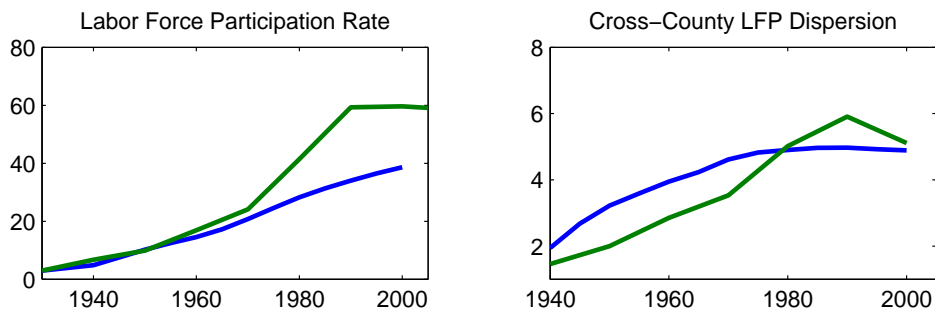


Figure 9: RESULTS WITH TWENTY-FIVE YEARS UNTIL MOTHERHOOD.

**Model** This model is one where a child grows up for 25 years and realizes her potential wage at 25. At the same time, the woman marries and starts having children. She is a married woman with a child under 5 years of age until age 35, when she drops out of our sample. We stagger families so that every year an equal number of children are born. The parameters are all equal to our benchmark parameters. Signals are drawn from wage and maternal employment decisions of women from the current and last 10 cohorts. The labor force participation rates in 9 include only the cohorts that are between 25-35 years old. This model has three features that help to slow the increase in participation. One feature is a longer childhood. Information generated from a woman participating today will not be revealed for 25 years. A second feature is that participation rates include not only the current cohort, but also 10 years of older cohorts who made their participation decisions with less information and are therefore less likely to participate. A third feature is that signals are drawn from both current and 10 years of past cohorts. The potential wage and maternal employment decisions of an older woman are less likely to be informative. What we learn from this is the more realistic modeling of the timing of childbirth and introducing overlapping cohorts helps to add more persistence to the learning model.

## H Occupation Choice Appendix

Allowing women the option to participate in a time-intensive, high-wage career, to have a normal career, or to nurture children, results in more women choosing the high-wage career over time. As the composition of career choices changes, wages rise and the labor supply elasticity falls. This appendix details the solution and calibration of this model extension.



**Equilibrium** Substituting (11) and (12) into expected utility produces the following optimization problem. Choose  $n_{it}, h_{it} \in \{0, 1\} : n_{it} + h_{it} \leq 1$  to maximize:

$$\frac{((n_{it} + h_{it}\tilde{w}_{it})w_{it} + \omega_{it})^{1-\gamma}}{1-\gamma} + \frac{\beta}{1-\gamma} E_{it} \left[ e^{(a_{i,t+1} - n_{it}\theta - h_{it}\tilde{\theta})(1-\gamma)} \right] + (1 - n_{it} - h_{it})L. \quad (13)$$

Beliefs  $\theta \sim N(\mu_{i,t}, \sigma_{i,t}^2)$  and  $\tilde{\theta} \sim N(\tilde{\mu}_{i,t}, \tilde{\sigma}_{i,t}^2)$  are formed according to the rules in (4) and (5). Because the unknown components of  $\theta$  and  $\tilde{\theta}$  are independent, updating occurs separately for high-intensity and low-intensity careers. Distributions of observed wage outcomes indexed by  $\mathbb{J}_{i,t}$  are consistent with distribution of optimal labor choices  $n_{i,(t-1)}$  and  $h_{i,(t-1)}$ .

**Solving the model** Bayesian updating with  $J$  signals is equivalent to running the following regression of children's potential wages on mothers' labor choices  $W_{it} - \mu_a = N_{it}\theta + H_{it}\tilde{\theta} + \varepsilon_{it}$ , where  $W_{it}$ ,  $N_{it}$  and  $H_{it}$  are  $J \times 1$  vectors  $\{\log w_{j,t}\}_{j \in \mathbb{J}_i}$ ,  $\{n_{i,t-1}\}_{j \in \mathbb{J}_i}$  and  $\{h_{i,t-1}\}_{j \in \mathbb{J}_i}$ . Then, agents form a linear combination of the OLS-estimated  $\hat{\theta}$  or  $\tilde{\theta}$  and the prior beliefs  $\mu_t, \tilde{\mu}_t$ . Let  $\bar{h}_{i,t}$  be the sum of the high-intensity careers chosen by the set of families that  $(i, t)$  observes:  $\bar{h}_{i,t} = \sum_{j \in \mathbb{J}_i} h_{i,t}$ . The resulting estimate of  $\tilde{\theta}$  is normally distributed with mean  $\tilde{\mu}_{i,t} = \sum_{j \in \mathbb{J}_i} (\log w_{j,t} - \mu_a) h_{j,t} / \bar{h}_{i,t}$  and variance  $\tilde{\sigma}^2 = \sigma_a^2 / \bar{h}_{i,t}$ . For each possible career choice, we compute the expectation of (13), conditional on time  $t$  information  $(\mu_{it}, \sigma_{it}, \tilde{\mu}_{it}, \tilde{\sigma}_{it})$ . The expected value of staying out of the labor force,  $EUO$  and of working in a low-intensity career  $EUW$  are given by (7) and (8). The expected utility of a high-intensity career is

$$EUH_{it} = \frac{(w_{it}\tilde{w} + \omega_{it})^{1-\gamma}}{1-\gamma} + \frac{\beta}{1-\gamma} \exp \left( (\mu_a - \tilde{\mu}_{i,t})(1-\gamma) + \frac{1}{2}(\sigma_a^2 + \tilde{\sigma}_{i,t}^2)(1-\gamma)^2 \right). \quad (14)$$

The optimal career choice for woman  $i$  in generation  $t$  is: (i) if  $EUO_{it} > EUW_{it}$  and  $EUO_{it} > EUH_{it}$ , then stay home; (ii) if  $EUW_{it} > EUO_{it}$  and  $EUW_{it} > EUH_{it}$ , then work in a low-intensity career; (iii) otherwise, if  $EUH_{it} > EUO_{it}$  and  $EUH_{it} > EUW_{it}$ , then work in a high-intensity career.

**Numerical example** The census variable OCC1990 defines our high intensity occupations. This variable starts in 1950 and the earlier classification (OCC1950) is not comparable. We consider high intensity the managerial and professional specialty occupation with the exclusion, (following Goldin and Katz 2002) of non-college teachers and those in health assessment and treating occupations (nurses, dieticians, therapist, and physicians' assistants).

Duxbury and Higgins (2003) report that along many dimensions, professional careers are about twice as straining on households. The likelihood of having to do overnight job-related travel increases. 19% of non-professional and 40% of professional women report spending one night a month away from home. 30% of non-professional and 60% of professional women bring work home. Finally, non-professional women do about 11.4 hours of unpaid overtime work per month, while professional women work about 17.7 unpaid hours.

The new model introduces five additional parameters: the true cost  $\tilde{\theta}$ , the initial beliefs  $\tilde{\mu}_0$  and  $\tilde{\sigma}_0$ , the wage premium  $\tilde{w}$  and the leisure cost  $L_{hi}$  of high-intensity maternal employment. Based on these facts, we double the leisure cost  $L_{hi}$  and assuming convex nurture costs, we quadruple the true cost for children  $\tilde{\theta} = 0.16$ . As before, we calibrate to unbiased initial beliefs ( $\tilde{\mu}_0 = \tilde{\theta}$ ) and we keep initial uncertainty at the same level as before ( $\tilde{\sigma}_0 = 1.38$ ) to match the same target: the initial labor force participation rate among married women with children under 5, in all careers.

This leaves only the wage premium for high-intensity careers. According to the census, women in the occupations we categorize as high-intensity earn 30% more, on average. Therefore, we use a 30% wage premium.

**Simulation results** In figure 10, labor force participation rises more at the end. The high-intensity participation rises gradually in the model, like in the data, but overshoots at the end. Women's average

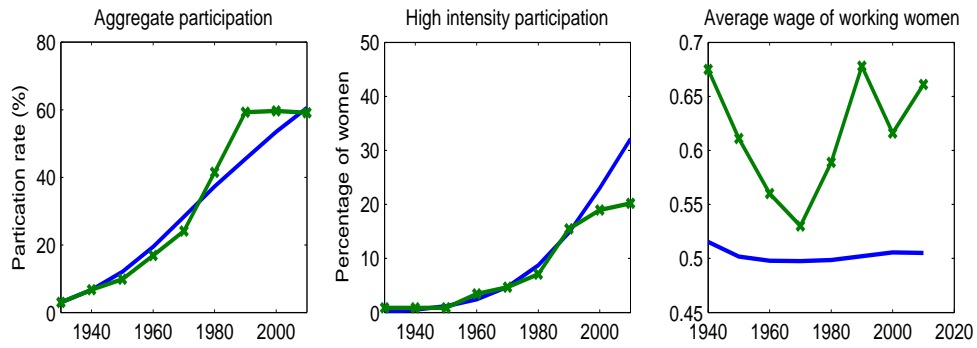


Figure 10: PARTICIPATION AND WAGES IN THE OCCUPATION CHOICE MODEL.

wage falls and then rises, although the magnitude is much less than in the data. Although a more careful calibration is in order before making any conclusions, these results suggest that the power of this mechanism to explain the rise in wages is quite modest, but its ability to explain trends in occupational sorting could be substantial.