Extended Abstract: Assessing Factors that Influence Women's Participation in the Invention Ecosystem

Michelle Saksena[1], Nicholas Rada[2], Katherine Black[3] and Lisa Cook[4]

Abstract:

To explore factors that may be contributing to the underrepresentation of women in patenting, we adopt a model that explores: (i) the relationship between local economic and inventive environments and increasing women inventor participation, and (ii) how higher education influences a county's probability of hosting its first woman inventor. To this end, we combine patent grant and inventor gender and location information from PatentsView (1990-2019) with U.S. Census and Bureau of Economic Analysis county-level higher education and economic information. Our findings indicate that a county's per-capita income and labor force size had small but positive effects on increasing the number of women inventors. The evidence favors an environment of other inventors as being relatively more influential to expanding the number of women inventors. Counties with higher patenting activity in chemistry technologies had the highest and most consistent impact on increasing women inventor counts. We further find that team size has a non-linear relationship with women inventorship. While average team size is 2.7, larger teams (4.475) have a higher propensity to have women inventors, but then decreases as team size increases thereafter. The effect of team size also varies regionally. For example, larger inventor teams in the USPTO Silicon Valley region had the highest likelihood of adding women inventors relative to the other USPTO regions. [5] Interestingly, although the size of all male inventor teams was weakly complementary to women inventor counts in all areas of the country, its effect was largest on the East Coast.

[1] Office of the Chief Economist- US Patent and Trademark Office (Michelle.saksena@uspto.gov)
[2] Office of the Chief Economist- US Patent and Trademark Office (Nicholas.rada@uspto.gov)
[3] ADDX, Office of the Chief Economist- US Patent and Trademark Office
[4] Department of Economics- Michigan State University (lisacook@msu.edu)
[5] https://www.uspto.gov/about-us/uspto-office-locations

Assessing Factors that Influence Women's Participation in the Invention Ecosystem

Motivation

There is increasing recognition among U.S. public and private sectors that a lack of diversity in intellectual property (IP) fields hinders technological progress (Bell et al. 2019). In response, various initiatives to diversify representation in IP, especially among women, minorities, and veterans, have moved to the forefront of policy discussions in recent years. [6]

Women's underrepresentation in the IP ecosystem continues to be a persistent problem as women make up about 13% of US inventors as of 2019 (Toole et al. 2020). Bunker-Whittington and Smith-Doerr (2008) investigate lower patenting rates of female life scientists (which includes chemistry, the most prolific patenting field for women) and find that, conditional on education and career history, women are less likely to patent than men. In addition, Delgado, Mariani and Murray (2019) test the hypothesis that women are more geographically constrained than men, finding that women source their knowledge for patenting more locally than men.

While it is established that there is persistent, secular underrepresentation of women in all of U.S. IP, Delgado, Mariani and Murray (2019) elude to significant geographical heterogeneity of women's participation throughout the U.S. Related studies on geographical inequality have largely focused on regional disparities as it pertains to a narrow definition of human capital accumulation, such as access to education (Logan et al. 2012) and wages and income (for example, Nunn et al. 2018). Our analysis indicates regional differences in women inventor counts as our measure for human capital. For example, figure 1a and b show three-year average counts of unique women inventors by county between 1990-1992 and 2017-2019. In 1990-1992, the most concentrated areas of women inventors were in the Silicon Valley and East Coast regions. By 2017-2019, Silicon Valley and East Coast regions maintained their dominance in women's participation, but there is expansion in the middle of the country, particularly around large cities such as Denver, Minneapolis-St. Paul and Houston, TX. We propose to extend the analysis of geographical human capital inequality measured via women's patenting rates, disaggregating based on USPTO regional areas.

[6] Despite notable growth in the last 40 years, women remain underrepresented at all points in the IP lifecycle (SUCCESS Act 2018), from applying for and being granted a patent, to staying active by inventing again, (USPTO 2019 & USPTO 2020) to seeking commercialization for their patents (Shaw and Hess 2018). This is shown in several USPTO studies, which calculate measures of women's involvement in patenting and track progress. Although the Women Inventor Rate (WIR) continues to increase, it is still only 12.84% of the total patent inventor population for 2019. This is significantly lower than other benchmarks of women's education and employment as scientists and engineers (National Science Board 2020).

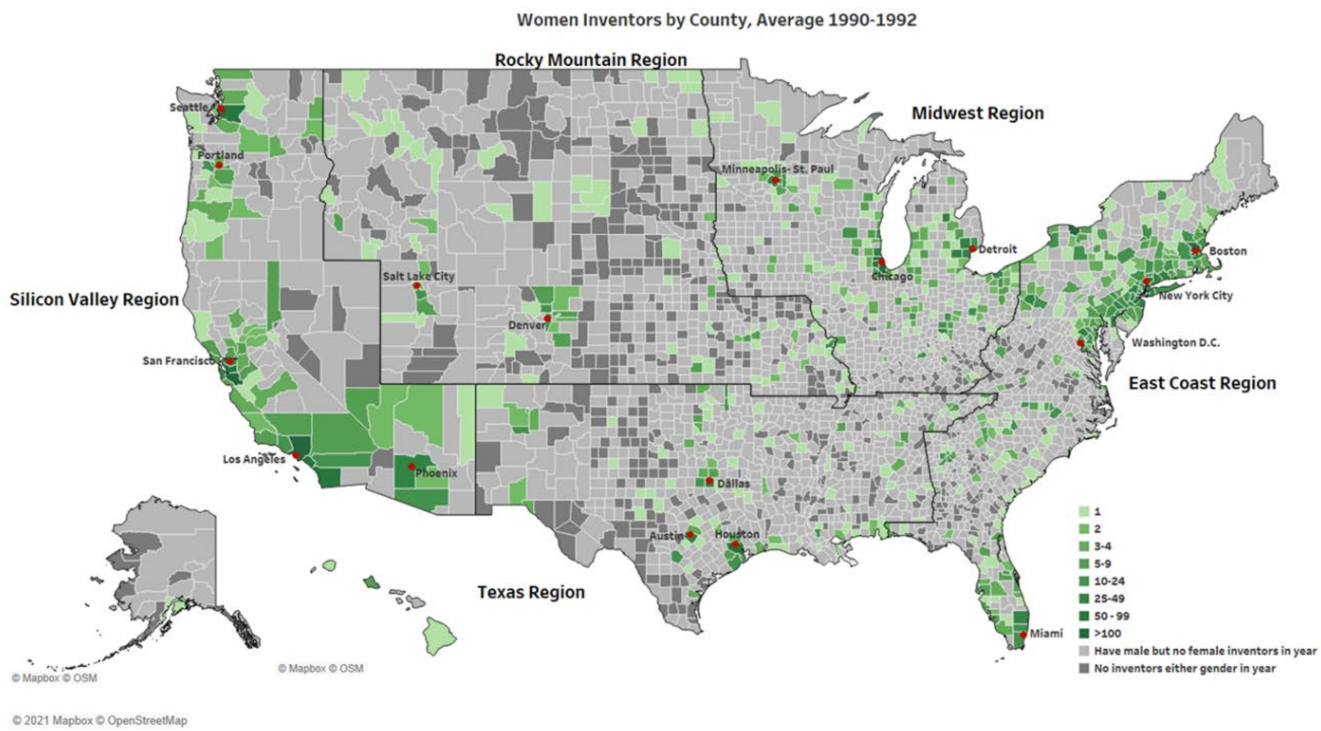


Figure 1a. Unique Women Inventors by USPTO Region, (1990-1992)

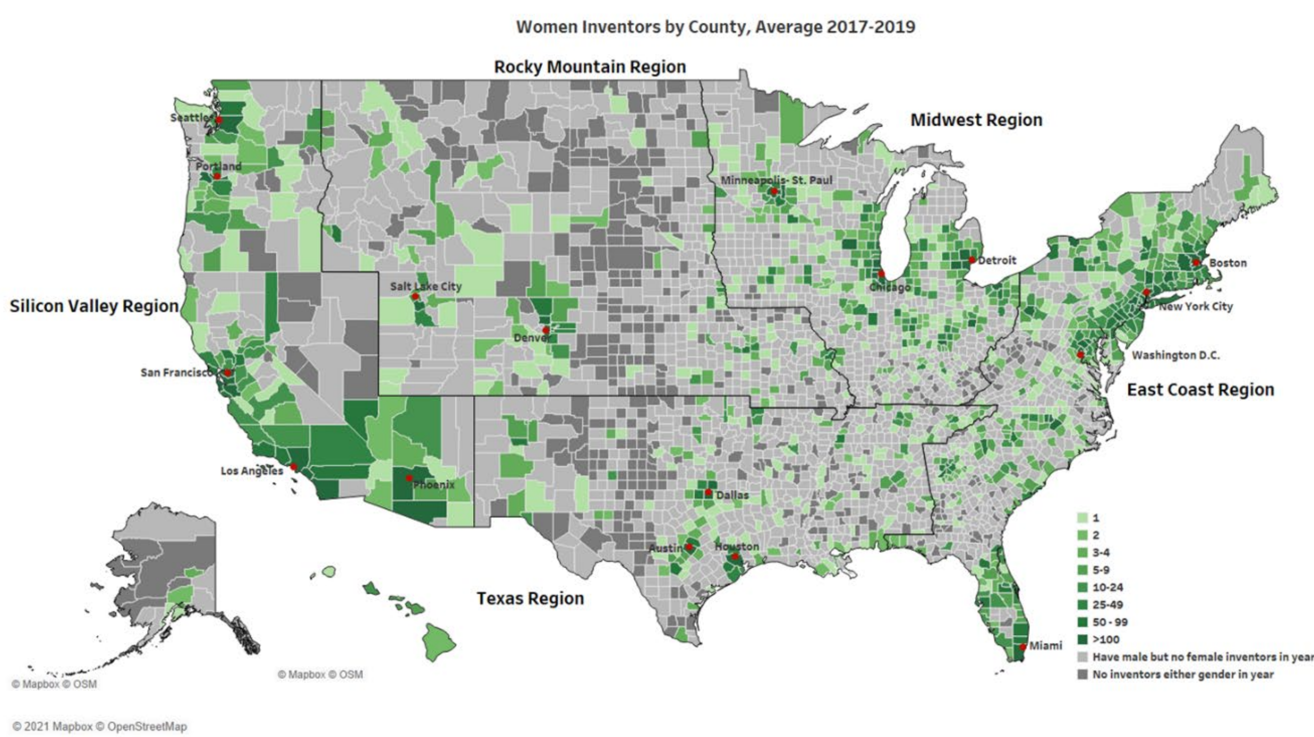


Figure 1b. Unique Women Inventors by USPTO Region, (2017-2019)

In this analysis, we investigate factors that capsulize conditions within the female inventor's environment. Specifically, we examine the effects of team size, the number of all male inventor teams and local presence and concentration of R&D by technology fields on women's representation in patenting. Following past literature, we assume that team size is a proxy for capital investment (Breitzman and Thomas, 2015) where small team sizes approximate lower capital investment. There is evidence suggesting that gender inclusiveness is particularly useful with development of the most novel technologies (Díaz-García, González-Moreno and Saez-Martinez, 2013), where perhaps smaller teams are more prevalent and where capital investment is lower because of perceived high investment risk. Following this literature, we hypothesize that women inventor counts have a non-linear relationship with team size. Considering the regional clustering of technology fields (Kerr and Robert-Nicoud, 2020), we suspect that team size will differentially effect women's representation by region and by technology field. In addition, we believe that higher educational attainment will have a positive effect on accommodating women inventors in a county.

Data

For patent, inventor, and inventor teams, we draw from the PatentsView (PV) database that contains information on the gender and location of inventor-patentees.[7] Economic variables specified at the county level are sourced from the Census, Bureau of Labor Statistics (BLS) and the Bureau of Economic Analysis (BEA).[8]

Table 1 provides summary statistics for our variables of concern. Our data span through years 1990 to 2019. Our dependent variable is county-level women inventor counts. We control for average county per capita income and labor force numbers and are interested in the effects of team size, number of all male inventor teams, and technology field on the propensity to generate additional women inventors at the county-level. We are also interested in understanding the effect of educational attainment on generating a county's first woman inventor.

Data Construction

Using PV data we construct our dependent variable, *number of women inventors*, by county and year. PV provides gender attributed, location specific patent grant data for all patents from 1976-2019. We summed the number of unique female inventors for every county and year combination to generate *number of women inventors*. Similarly, PV lists all inventor names for every patent. We identified team size of every patent and averaged this to get the average *team size* for every county-year combination. Because PV provides gender attributed data, we were also able to identify all male inventor teams. Because the location of inventors within a team can vary, we counted the team in every inventor location reported in each patent.

PV also provides data for the Cooperative Patent Classification (CPC) for every patent, we used the eight CPC sections to categorize every patent in our dataset to a technology field[9]. For the indicator variables, a county-year observation would be assigned a technology field if at least one patent in a county-year was

---

[7] See www.PatentsView.org.

[8] See Table 1 for variables, data coverage and source.

[9] The eight CPC sections are: *Human necessities, Performing operations; transporting, Chemistry; metallurgy, Textiles; paper, Fixed constructions, Mechanical engineering; lighting; heating; weapons; blasting engines or pumps, and Physics, Electricity IPC. For more information see:* https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/classification/cpc.html.

designated a particular CPC section. The *%CPC(1-8)* variables were created as the percentage of patents within each county-year under one of the nine CPC sections.

Labor force are county-year level data which gives the number of people who are working or actively looking for work. The data come from the BLS Local Area Unemployment Statistics (LAUS) program.

Per capita income data is sourced from the Bureau of Economic Analysis. County-year level averages of per capita income are used in our analysis.

Using IPUMS NHGIS data, we combine Census decennial data and 5-Year American Community Survey (ACS) estimates to construct our education variables. For the intervening years for which we do not have education estimates (1991-1999), we used a linear interpolation to infer educational attainment values. Starting in 2005, Census changed their data collection process to the American Community Survey (ACS), which collects data in all counties every 5 years. Although a new 5-year ACS is published annually, due to guidance from Census concerning overlapping 5 year estimates, we use a stepwise construction for the years 2005-2019[10]. Specifically, the 2005 5-year ACS education estimates for years 2005-2009, the 2010 5-year ACS is used from years 2010-2014, and the 2015 estimate for 2015-2019. For the years in between the decennial and implementation of the 5 year ACS (2000-2004), we linearly interpolated from the 2000 decennial Census to the 2005 5 year ACS estimates. This gives the number of women with a bachelors, masters and PhDs for every county and between 1990-2019.

Summary Statistics

Table 1 indicates that the average inventor team size is 2.7, while there are 72 all male inventor teams in a typical county. In terms of technology concentrations, patents in *Human necessities*, *Performing operations; transporting* had the highest average county-wide technology concentrations at 23%. *Textiles; paper was the* lowest while *Chemistry* and *Mechanical* patents were on average 10% and 11% of technology concentration respectively for a typical county. In terms of the percentage of counties with at least one patent in any of the eight technology fields, we see that nearly 67% of year-county observations had patents in *performing operations; transporting* technology fields, while only 15% of counties have patents in *textiles; paper*. Patents in *Chemistry; metallurgy* which is known to have the highest rates of women inventors were present in 43% of counties, while *Mechanical engineering; lighting; heating; weapons; blasting engines or pumps,* known to have one of the lowest levels of female representation, is present in 50% of year-county observations (Toole et al., 2019). Finally, as expected the average number of Bachelors graduates was highest at almost 7,500 per county with a precipitous drop in the number of PhD holders at almost 350 graduates for an average county.

Methods

The present analysis assesses how economic, technology, and team factors influence women inventor-patentees, as well as differences in how various levels of education influence the probability that a U.S. county has a woman inventor-patentee. The analysis uses the zero-inflated negative binomial (*zinb*) model. A strength of the *zinb* model is that it accommodates data with excessive zeros (Lambert, 1992; Lord et al., 2004; Raihan et al., 2019). This is of particular applicability to our analysis, as the majority of U.S. counties in our sample do not have a woman inventor in 1990-2019. Alternatively, some counties (ir)regularly hosted

[10] See https://www.census.gov/programs-surveys/acs/guidance/estimates.html.

women inventors. Thus, there are two different processes in our data – one in which counties may or may not be attempting (unsuccessfully) to create an environment conducive to women inventors, and one in which the innovation ecosystem is present and attracts, produces, and supports women inventors.

The *zinb* model is uniquely suited to account for these different processes. It is a mixed model, combining a logistic function to estimate how education influences the probability that a county will have its first woman inventor, and a negative binomial function to estimate how economic factors influence the number of women inventors. To generate additional insight, we further specify a model that allows for comparison of the economic and educational factors across USPTO regions[11].

As noted by table 1, the dependent variable, women inventor counts, is overdispersed. Overdispersion occurs when the data's distributional variance exceeds its mean. In the event that the data is not overdispersed, a zero-inflated Poisson (*zip*) model is most efficient (Cameron and Trivedi, 2005). We test the women inventor data for overdispersion in three ways. As previously mentioned, we examine whether the mean-to-variance ratio > 1 among the number of women inventor counts. Second, our model estimates an alpha (α) parameter that tests for overdispersion after including explanatory variables in our *zinb* model of women inventors. Third, we test whether the zero-inflated poisson (*zip*) model is more appropriate for these count data. All three tests confirm our application of the *zinb*. Essentially, the *zinb* inverts to a *zip* model when the mean equals the variance.

Empirical Model

We allow counts of women inventors (*y*) in county *i* to be distributed,

$$y_i \sim Poisson(\mu_i), \tag{1}$$

where $\mu_i = \exp(\mathbf{X}_i\mathbf{B} + v_i)$ is the mean woman inventor frequency, and $e^{v_i} \sim Gamma\left(\frac{1}{\alpha}, \alpha\right)$ (Stata, 2019, p.1635). Equation (1) states that counts of women inventors may be explained by a vector of independent variables(**X**), estimable parameters $(\boldsymbol{\beta})$, an unobserved parameter v_i, and the overdispersion parameter α. When $\alpha = 1$, *y* is Poisson distributed and the *zip* model is most efficient. When $\alpha > 1$, *y* is distributed by a negative binomial process and the *zinb* model is more appropriate than the *zip*.

[11] The USPTO Regions are the Following: East Coast, Midwest, South/Texas, Rocky Mountain (Mountain) and Silicon Valley/West Coast.

Table1. Pooled summary statistics

	μ	σ	Min	Max
Dependent Variable (PatentsView)				
# of women inventors	7.1	46.5	0.0	2,956.0
Economic variables				
Labor force (BLS)	64,863.2	178,897.4	261.2	5,121,584.0
Per capita income (BEA)	30,437.8	12,390.3	7,096.0	230,141.0
Inventor team variables (PatentsView)				
Team size	2.7	1.3	1.0	23.0
# of all male teams	72	401	0	21,415
% of counties with Cooperative Patent Classification (CPC) concentrations (PatentsView)				
Human necessities	23%	28%	0%	100%
Performing operations; transporting	23%	28%	0%	100%
Chemistry; metallurgy	10%	18%	0%	100%
Textiles; paper	2%	8%	0%	100%
Fixed constructions	7%	18%	0%	100%
Mechanical engineering; lighting; heating; weapons; blasting engines or pumps	11%	21%	0%	100%
Physics	14%	21%	0%	100%
Electricity	11%	19%	0%	100%
Technology indicators by CPC				
Human necessities	64%	48%	0%	100%
Performing operations; transporting	67%	47%	0%	100%
Chemistry; metallurgy	43%	49%	0%	100%
Textiles; paper	15%	36%	0%	100%
Fixed constructions	38%	49%	0%	100%
Mechanical engineering; lighting; heating; weapons; blasting engines or pumps	50%	50%	0%	100%
Physics	51%	50%	0%	100%
Electricity	45%	50%	0%	100%
Women's education (Census)				
Bachelors	7,483.1	23,522.4	17.0	761,572.0
Masters	3,192.2	10,135.7	0.0	287,419.0
PhDs	348.5	1,308.4	0.0	40,577.0
N= 63,946				

There are three elements to the *zinb* model. The first is a probability density function (PDF) of observing county *i* with zero women inventors (*y*):

$$\Pr(y_i = 0) = F_i + (1 - F_i)f(y_i = 0) \tag{2}$$

(Stata, 2019, pp. 1635, 2859; Raihan et al., 2019). Note, there are two terms in (2), indicating that there are two possible explanations for observing a county with zero women inventors. The first term, F_i, is the probability of observing a *structural-zero* county and is determined by a logistic distribution function where $F_i = \frac{\lambda_i}{1+\lambda_i}$ and $\lambda_i = \exp(\boldsymbol{Z_i}\Delta)$. *Structural-zero* counties are those that have not produced a women inventor. The second term in (2), $(1 - F_i)f(y_i = 0)$ follows a negative binomial distribution of women inventors. These counties host women inventors. Some of these counties, however, do not consistently host women inventors each year. Hence, they are termed *observational zero* counties.

Once the model determines that a county has or currently hosts women inventors, a second PDF is used to calculate the probability of observing the number of women inventors in non-zero counties, given by,

$$\Pr(y_i > 0) = (1 - F_i)f(y_i). \tag{3}$$

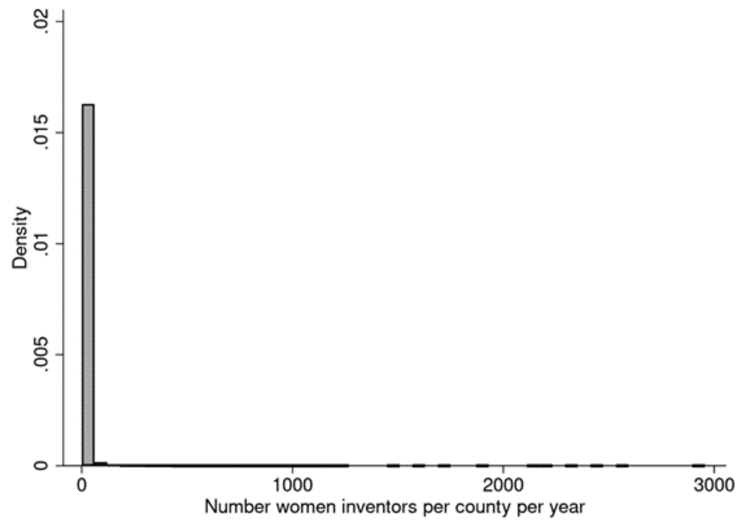


Figure 2. Distribution of women inventor counts by county and year

Notably, nested in equation (3) is the third element of the *zinb* model, link function, $f(y_i)$. The link function is used to express mean rate μ and overdispersion parameter α as a function of the regression's independent variables with a negative binomial distribution, $\boldsymbol{\Gamma}(.)$, and is given by,

$$f(y_i) = \Pr(Y = y_i | \mu_i, \alpha) = \frac{\boldsymbol{\Gamma}(m+y_i)}{\boldsymbol{\Gamma}(y_i+1)\boldsymbol{\Gamma}(m)} p_i^m (1-p_i)^{y_i} \tag{4}$$

where $m = 1/\alpha, p_i = 1/(1+\alpha\mu_i)$. Substituting (4) into (2) and (3), adding (2) and (3) together, and taking logs forms the following log-likelihood function:

$$ln\mathrm{L} = \sum_{i \epsilon S} w_i \ln\{F_i + (1-F_i)\, p_i^m\} + \sum_{i \notin S} w_i\{\ln(1-F_i) + \boldsymbol{\Gamma}(m+y_i)$$

$$-\boldsymbol{\Gamma}(y_i+1) - \boldsymbol{\Gamma}(m) + m\mathrm{ln}p_i + y_i(1-p_i)\}. \tag{5}$$

The *zinb* model estimates simultaneously the probabilities given in (2) and (3) in a single log-likelihood function, nesting the negative binomial distribution of non-zero counties, $\boldsymbol{\Gamma}(.)$, inside the logistic distribution,F_i . *S* in equation (5) is the set of counties that do not have women inventors ($y_{it} = 0$), and w_i are weights (Stata 2019, p.2859).

Empirical Application

We first ask how average education levels in a county influence the probability that a county hosts its first woman inventor. Specifically, we test three categories of female educational attainment: number of women with bachelor's degrees (*bachelors*), master's degrees (*masters*), and PhDs (*phd*). We expect that education has a positive and increasing influence on the likelihood a county hosts a woman inventor.

We specify the logistic distribution function shown in equation (2) to account for time (1990-2019), although as shown above the empirical model implemented pools the data's time series dimension:

$$\lambda_{it} = \exp(\boldsymbol{Z_{it}}\Delta) = \exp\left(z'_{itj}\delta\right) = \exp(\delta_0 + \delta_{fbachelors} fbachelors_{it} + \delta_{fmasters} fmasters_{it}$$

$$+\delta_{fphd} fphd_{it} + \delta_t YearFE_t + \delta_l StateFE_l), \tag{6}$$

where subscripts *j = bachelors, masters and phd* and *i*=1,…,51 for all U.S. states plus the District of Columbia. Equation (6) tests how education influences a county's probability of never hosting a woman inventor after accounting for time-invariant heterogeneity at the state level, as well as common shocks to all U.S. counties in a given year. In addition, we cluster the errors at the county level.

As demonstrated by equation (3), for non-zero counties, we characterize the environment faced by women inventors. In particular, we specify two equations to investigate different facets of how the environment affects women inventors. First, we assess the relationship between team size and women inventors, as well as the relationship between technology classes and women inventors. To that end, we specify

$$\mu_{it} = \exp(\mathbf{X_{it}B}) = \exp(x'_{itj}\beta = \beta_0 + \beta_{LF} LF_{it} + \beta_{PCI} PCI_{it} + (\beta_{TS} TS_{it} + \beta_{TS^2} TS^2_{it}) + \beta_{AM} AM_{it} + \beta_{\%cpc} \%cpc_{it} + \beta_{dcpc} dcpc_{it} + \beta_t Year_t + \beta_{state} State_l) \tag{7}$$

where subscript *i* refers to counties, *t* refers to time (1990-2019), and *j = Laborforce (LF), Per-capita Income (PCI), inventor team size (TS), all male inventor teams (AM), technology field concentration (%cpc) and presence of technology field (dcpc)*. To control for unobserved heterogeneity, we include year and i=1,…,51 state fixed effects, and we cluster errors at the county level. Equation (7) tells us the correlative relationship between a given county's economic environment and its number of women inventors, controlling for other factors. The relationships inferred in (6) and (7) reflect a national average. We therefore add a second specification, one that assess the relationship between the economic and team size variables with women inventors regionally across the U.S. In particular, we add regional slope dummies to specific variables of interest to test whether labor force, per capita income, team size and technology field and concentration influence the probability that a county produces its first woman inventor differentially by USPTO region (*R*):

$$\mu_{it} = \exp(x'_{itj}\beta) = \exp(\beta_0 + (\beta_1 LF_{it} + \beta_2 LF_{it} * R_{r-1}) + (PCI_{it}\beta_3 + \beta_4 PCI_{it} * R_{r-1})$$

$$+(\beta_5 TS_{it} + \beta_6 TS_{it} * R_{r-1}) + \sum\nolimits_{c=2}^{9} \beta_{(c+5)} \%cpc_{(c+5)it} + \sum\nolimits_{c=1}^{9} \beta_{(c+14)} dcpc_{(c+14)it}$$

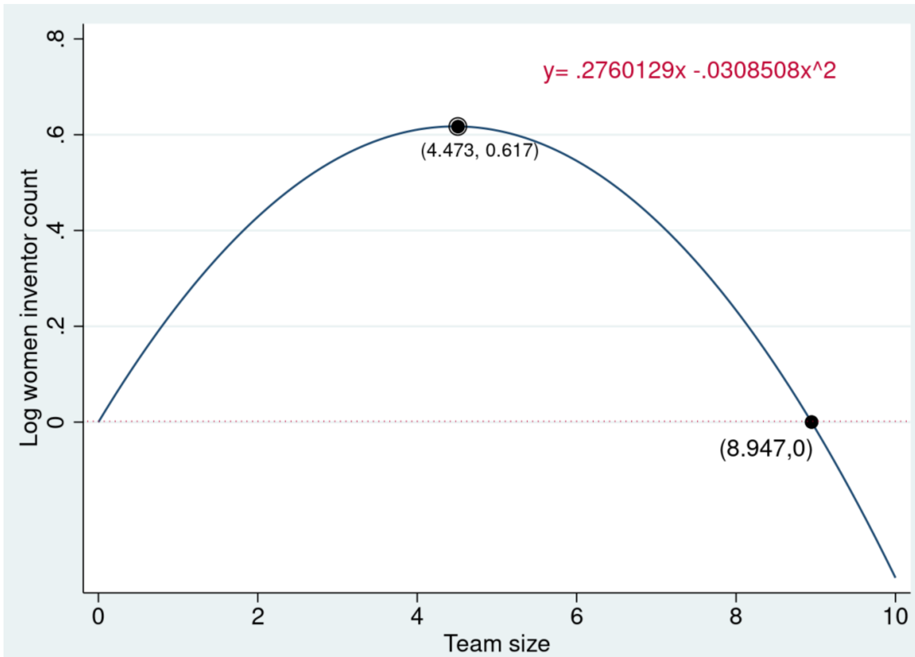
$$+\beta_t YearFE_t + \beta_l StateFE_l), \tag{8}$$

where USPTO region subscript *r = Eastern, Midwest, South, Mountain, Western.* We omit the Eastern region in regional assessments, thus all regional estimates are relative to the Eastern region. CPC technology fields are subscripted by c. We omit the *human necessities* concentration variable, %cpc1, thus all cpc concentration estimates are relative to *human necessities*, but include all cpc indicator variables (dcpc1-9) since the multiple technology fields can exist concurrently in a given county and year.

We estimate equations (6) and (7) simultaneously using the log-likelihood function specified by (5) for the national model and then again (6) and (8) for the regional model using the STATA package *zinb*.

Results

Recalling our hypothesis, our primary variables of interest are team size, number of all male inventor teams, and the technology class (CPC) indicator and concentration variables. The results from table 2 indicate that the linear and quadratic terms on team size are all highly statistically significant. The parameter coefficients for team size indicate positive and then negative values. Plotting the logged coefficients indicates a local maximum team size of approximately 4.5. Thus, these data indicate that teams up to 4.5 inventors have an increasing effect on the likelihood of generating women inventors (figure 3). We also find that the number of all male teams shows a small but complementary effect on the creation of women inventors. The exp(β) coefficient on all male teams indicates that for every 1400 all male teams there is one women inventor.



Finally, the CPC indicator variables all show a positive and statistically significant effect at least at the 10% level for all technology classes. This indicates that just the existence of one patent in any of the eight CPC fields has a strong effect on creating more women inventors. The effect was strongest within *mechanical engineering* and lowest in *physics*. Additionally, the %CPC variables indicate the marginal effect of increasing the concentration of technology development for a given CPC field relative to the *human necessities* field. *Chemistry* had the highest marginal effect relative to *human necessities* while counties with more technology development in *fixed constructions* showed the largest decreases associated with women inventors. The effects of having other inventors within the county are larger than the small but positive effects from increasing the local labor force or per-capita incomes.

Surprisingly, we only found that the number of women with PhDs had a statistically significant effect on the relationship with a county hosting its first woman inventor.

Table 2. National model results

National Model			
Variable	β	exp(β)	p-value
Negative Binomal			
Economic variables			
Labor force	1.64E-06	1.000002	0.00
Per capita income (USD)	0.000019	1.000019	0.00
Inventor team variables			
Team size	0.276013	1.317865	0.00
Team size squared	-0.03085	0.96962	0.00
Number of all male teams	0.000681	1.000681	0.02
% of counties with Cooperative Patent Classification (CPC) concentrations			
Performing operations; transporting	-0.9668	0.3803	0.00
Chemistry; metallurgy	0.669494	1.953248	0.00
Textiles; paper	-1.47264	0.229321	0.00
Fixed constructions	-2.06514	0.1268	0.00
Mechanical engineering; lighting; heating; weapons; blasting engines or pumps	-1.68458	0.185522	0.00
Physics	0.321866	1.3797	0.09
Electricity	-0.01247	0.987607	0.95
Technology indicators by CPC			
Human necessities	0.46729	1.595664	0.00
Performing operations; transporting	0.624052	1.866477	0.00
Chemistry; metallurgy	0.438068	1.54971	0.00
Textiles; paper	0.493947	1.638771	0.00
Fixed constructions	0.61252	1.845075	0.00
Mechanical engineering; lighting; heating; weapons; blasting engines or pumps	0.690094	1.993904	0.00
Physics	0.244283	1.276705	0.00
Electricity	0.385308	1.470067	0.00
Logit			
Bachelors	-0.00046	0.999537	0.07
Masters	-0.00077	0.999231	0.20
PhD	-0.00482	0.995192	0.04

Table 3 indicates that team size is a significant variable in all regions though there was no significant regional variation in the Midwest. Team size had the largest positive impact in the West Coast Region and the lowest in the South. The number of all male teams was also significant, but only the West Coast indicated significant regional differentiation from the East coast. While the presence of all male inventor teams was

Table 3. Regional model preliminary results

Regional model		
	exp(β)	p-value
Negative Binomial		
Labor force		
East coast	1.000002	0.00
Midwest	1.000002	0.42
South	1.000002	0.72
Mountain	1.000004	0.00
West Coast	1.000002	0.38
Per Capita Income (USD)		
East Coast	1.000015	0.00
Midwest	1.000015	0.49
South	1.000015	0.39
Mountain	1.0000023	0.00
West Coast	1.000015	0.63
Team size		
East Coast	1.212372	0.00
Midwest	1.212372	0.45
South	1.097086213	0.00
Mountain	1.134009487	0.06
West Coast	1.364214319	0.07
Number of all male teams		
East Coast	1.001505	0.00
Midwest	1.001505	0.35
South	1.001505	0.18
Mountain	1.001505	0.74
West Coast	1.000282663	0.00
% of counties with Cooperative Patent Classification (CPC) concentrations		
Performing operations; transporting	0.3763327	0.00
Chemistry; metallurgy	2.218898	0.00
Textiles; paper	0.3363812	0.00
Fixed constructions	0.1533107	0.00
Mechanical engineering; lighting; heating; weapons; blasting engines or pumps	0.1946372	0.00
Physics	1.226501	0.14
Electricity	0.9786238	0.89
Technology indicators by CPC		
Human necessities	1.610621	0.00
Performing operations; transporting	1.92963	0.00
Chemistry; metallurgy	1.533529	0.00
Textiles; paper	1.505251	0.00
Fixed constructions	1.722936	0.00
Mechanical engineering; lighting; heating; weapons; blasting engines or pumps	1.965105	0.00
Physics	1.357513	0.00
Electricity	1.482946	0.00
Logit		
Bachelors	0.999495028	0.04
Masters	0.999589084	0.43
PhD	0.993760149	0.00

weakly complementary to women inventor counts in all areas of the country, its effect was strongest in the East Coast by a factor of over 500% when compared to other regions.[12]

The CPC indicator variables are consistent with the national model and all indicate positive and significant effects on women inventor counts when a technology field is present in a county. Similar to the national model, *mechanical engineering* had the largest effect on women inventor counts while *physics* had the lowest effect. However, the %CPC variables are quite comparatively different to the results from the national model. Recall that the national model indicated that all the marginal effects from increasing concentration of any field was complementary to women inventor counts. However, in the regional model, only *chemistry* and *physics showed* positive and significant effects on women's representation, while the remainder CPC fields indicated an exclusionary effect on women inventor counts.

Like the national model, both labor force and per capita income show weakly positive effects toward women inventor counts. In the logit model, we observe that the number of women with bachelors and PhDs increases the likelihood of a county hosting its first woman inventor.

Discussion

The results from our national model indicate that there is a non-linear relationship with women inventor counts and team size. If team size is to be interpreted as a proxy for R&D capital investment as Breitzman and Thomas (2015) assert, then representation of women in IP is closely correlated to capital investment. Figure 3 indicates that team size has an increasing effect on women's inventor counts for teams as large as 4.5, but this promoting effect steadily decreases for larger teams. Our results appear to be consistent with the literature: gender diversity is prevalent within new and novel technologies (Díaz-García, González-Moreno and Saez-Martinez, 2013), where perhaps investment is smaller due to higher risk (hence, smaller team sizes). While team size was complementary to women inventor counts in all USPTO regions, the West and East coast regions indicated the largest magnitude in this effect.

Interestingly, while the West coast boasted the highest positive impact from increasing team size, it indicated the weakest complementary impact from the number of all male teams on women inventor counts. In fact, the complementary impact of this variable in the West coast was over 5 times weaker than in the East coast. This suggests that though both the East and West coast regions are established technology hubs, the East coast appears more inclusive to women inventors.

Finally, both the national and regional models indicated that the presence of any technology field benefited women inventor counts. However, the marginal effects from increasing R&D concentration by CPC field varied largely between models and was in fact exclusionary in some fields in the regional model. This supports the notion that there is substantial regional variation in the effect of the intensity of R&D by technology field. Comparing both models, we found that *chemistry* was consistently engendering of female inventors. Chemistry R&D is a well-established technology field with a history of high female inventor representation (Toole et. al, 2019).

[12] The relative complementarity of all male team inventors is equal to the following: $\frac{ratio\ of\ \#\ of\ new\ all\ male\ teams\ in\ the\ West\ coast\ to\ 1\ women\ inventor}{ratio\ of\ \#\ of\ all\ male\ teams\ in\ the\ East\ Coast\ to\ 1\ women\ inventor} = \frac{1/(\beta_{AM,West}-1)}{1/(\beta_{AM,East}-1)}\frac{1/(1.0002855-1)}{1/(1.001532-1)} = 5.37 = 537\%$. This indicates for every new female inventor in the West Coast, there are 537% more all male teams also present relative to the East Coast.

Finally, it was surprising to find that the educational attainment variables were weakly significant in both the national and regional models. It suggests that educational attainment is only a small factor in determining whether a county is determined to be a zero women inventor county. It alludes to the possibility that other factors, other than education are stronger determinants of women inventor counts. Perhaps unexplored factors like conditions in the working environment, post educational attainment, are stronger determinants in ensuring that a given county women inventors.

In summary, the results from our analysis suggest that women inventor rates are differentially effected by regional effects of technology R&D concentration. It also appears that technology maturity and thus capital investment promote women representation in IP for among small IP teams.

References

Breitzman, Anthony, and Patrick Thomas. "Inventor team size as a predictor of the future citation impact of patents." Scientometrics 103, no. 2 (2015): 631-647.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. "Who becomes an inventor in America? The importance of exposure to innovation." The Quarterly Journal of Economics 134, no. 2 (2019): 647-713.

Delgado, M., M. Mariani and F.E. Murray. "The Role of Location on the Inventor Gender Gap: Women are Geographically Constrained". Paper presented at DRUID19 Conference, Frederiksberg, Denmark. 2019

Díaz-García, Cristina, Angela González-Moreno, and Francisco Jose Saez-Martinez. "Gender diversity within R&D teams: Its impact on radicalness of innovation." Innovation 15, no. 2 (2013): 149-160.

Kerr, William R., and Frédéric Robert-Nicoud. "Tech clusters." Journal of Economic Perspectives 34, no. 3 (2020): 50-76.

Kim, Jinyoung, and Gerald Marschke. "Teams in R&D: evidence from US inventor data." Manuscript (2015).

Lambert, D., 1992. Zero-inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics* 34(1): 1-14.

Logan, John R., Elisabeta Minca, and Sinem Adar. "The geography of inequality: Why separate means unequal in American public schools." Sociology of education 85, no. 3 (2012): 287-301.

Lord, D., S. Washington, and J. Ivan, 2004. Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. Presented paper at the 83rd Annual Meeting of Transportation Research Board, April 21st. Available at: https://ceprofs.civil.tamu.edu/dlord/Papers/Lord_et_al._AA%26P_03225_March_24th.pdf.

Nunn, Ryan, Jana Parsons, and Jay Shambaugh. "The geography of prosperity." Place-Based Policies for Shared Economic Growth (2018): 11-42.

Raihan, A., P. Alluri, W. Wu, and A. Gan, 2019. Estimation of bicycle crash modification factors (CMFs) on urban facilities using zero inflated negative binomial models. *Accident Analysis and Prevention* 123: 303-313.

Stata, 2019. *Stata Base Reference Manual, Release 16*. A Stata Press Publication, College Station, Texas.

Toole, Andrew A., Amanda Myers, Charles AW Degrazia, Stefano Breschi, Edoardo Ferrucci, Francesco Lissoni, Ernest Miguelez, Valerio Sterzi and Gianluca Tarasconi. Progress and Potential: A Profile of Women Inventors on U.S. Patents. (2019) HAL.

Toole, Andrew A., Michelle J. Saksena, Charles AW Degrazia, Katherine P. Black, Francesco Lissoni, Ernest Miguelez, and Gianluca Tarasconi. Progress and Potential: 2020 update on US women inventor-patentees. (2020) HAL.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. "The increasing dominance of teams in production of knowledge." Science 316, no. 5827 (2007): 1036-1039.