IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14778

Gender Differences in Economics PhD Field Specializations with Correlated Choices

Eva Sierminska
Ronald L. Oaxaca

OCTOBER 2021

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14778

Gender Differences in Economics PhD Field Specializations with Correlated Choices

Eva Sierminska
LISER, DIW Berlin and IZA

Ronald L. Oaxaca
University of Arizona and IZA

OCTOBER 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Gender Differences in Economics PhD Field Specializations with Correlated Choices*

We model the process of field specialization choice among beginning economists within a multivariate logit framework that accommodates single and dual primary field specializations and incorporates correlations among field specialization choices. Conditioning on personal, economic, and institutional variables reveals that women graduate students are less likely to specialize in Labor/Health, Macro/Finance, Industrial Organization, Public Economics, and Development/Growth/International and are more likely to specialize in Agricultural/Resource/Environmental Economics. Field-specific gender faculty ratios and expected relative salaries as well as economics department rankings are significant factors for gender doctoral specialization dissimilarity. Preferences and characteristics contribute about equally to field specialization dissimilarity.

JEL Classification: J01, J16, J31

Keywords: gender, economics, specialization, salaries

Corresponding author:
Eva Sierminska
LISER
11, Porte des Sciences
L-4364 Esch-sur-Alzette
Luxembourg
E-mail: eva.sierminska@liser.lu

* We thank Karla Cardova, Shuya He, Swapnil K. Motghare and Hilal Akinci for excellent research assistance and the Sloan Foundation for their feedback and financial support. We also gratefully acknowledge the very helplful comments of Christina Gathmann, Dominique Meurs, Juan Pantano, Todd Sorensen, and the participants of the 2021 AEA meetings and 2020 EEA conference.

1 Introduction

The distribution of women across academic fields remains uneven. In the economics profession this is uniquely acute as there are disproportionately fewer women in the field and the share of women has been stagnant over time in the U.S. (Lundberg & Stearns (2019)) and in European academia (Auriol et al. (2019)). Our examination of the latest data from the American Economic Association on PhD dissertations indicates that over 40% of women who completed their dissertation in the past decade declared to be in Labor, Health or Microeconomics. Doctoral dissertation fields for the remaining women were distributed over 17 JEL fields. Among males, the top 3 fields accounting for 40% of specializations were Microeconomics, Macro/Monetary Economics, and Labor. Apart from significant gender differences in some of the other fields, Health and Macro/Monetary stand out because each is in the top 40% for either males or females, but not for both.[1]

The focus of our paper is the process that underlies the choice of field specialization among students in PhD economics programs. These early career choices have implications for gender differences in subsequent career growth and development (Fortin et al. (2021)). Our contribution is to understand what is driving gender segregation in fields of economic specialization by examining their doctoral field specialization choices. We investigate whether gender differences in fields are driven by salaries and academic employment prospects, or rather by non-economic factors. In this sense, we extend the literature on decision-making at the graduate school level, which is relatively underdeveloped given the limited data (Altonji et al. (2016)).

This research is intended to shed light on the larger issue of gender segregation in the labor market, especially as it applies more broadly to the underrepresentation of women in the STEM (Science, Technology, Engineering and Mathematics) and other highly qualified fields. Economics in itself is a very diverse field with fields of specialization varying along a spectrum from abstract theory to applied empirical economics at the borderline of other fields (e.g. health).

Primary fields are understood here to represent the most aggregated classifications for analytic purposes. We adopt an analytic framework that accommodates both single primary field and multi (primary) field specializations. The model provides for inclusion of measures of anticipated salary and academic employment by field. We rely on non-salary data obtained from Academic Analytics and EconLit as well as individual faculty salary data for public colleges and universities available from official state websites.

Our model shows statistically significant gender differences in field choices even after con-

[1]See Table A4 in the Appendix Tables.

ditioning on personal, economic, and institutional variables. Women graduate students are less likely to specialize in Labor/Health, Macro/Finance, IO, Public Economics, or Development/Growth/International and more likely to specialize in Agricultural/Environmental Economics. Gender disparities in field specializations are far more pronounced when considering both single primary and dual primary field specializations than when considering only graduate students who specialized in a single primary field in economics.

The rest of the paper is organized as follows: Section 2 discusses the literature related to research on the lack of diversity and field specialization in economics and also discusses studies that examine college major choices, Section 3 describes the data, Section 4 develops the conceptual framework, Section 5 presents the empirical results, Section 6 is a discussion section, and Section 7 presents the summary and conclusions.

2 Related Literature

Field specialization among economists is not a topic that has received a lot of attention in the literature. There is a paucity of data needed to better understand the process of field specialization and it implications for the representation of women in economics. Instead, most of the current literature has been concerned with the performance of women in academia (Ginther & Kahn (2004)) and their under-representation in economics in both the United States (Lundberg & Stearns (2019)) and Europe (Auriol et al. (2019)). This literature explores the academic experiences of women with respect to research, promotion and mentoring.[2]

Field choice, though, has been examined, but mostly at the undergraduate level with some exceptions for graduate school choice (Altonji et al. (2016)). A few of the studies mentioned, support our approach of incorporating subjective forecasts of future earnings and employment probabilities in the model (e.g. Arcidiacono et al. (2012) and Wiswall & Zafar (2015.)). In addition, Zafar (2013) explores the sources of gender gaps in the choice of college major and finds that differences in tastes and preferences constitute most of the gap. Yet, non-pecuniary determinants explain more than half of the choice for females, which is about a 25 percentage point (ppt) greater contribution than for men. However, our findings for graduate fields of specialization in economics do not exhibit such large differences among women and men.

The literature that has taken field specialization in economics into consideration finds there is a lack of diversity within the field, which has further consequences for career progression and for policy advice among other things. Fortin et al. (2021) find that female

[2]For a comprehensive and informative edited volume overviewing the literature see Lundberg (2020).

under-representation in assistant professor positions outside of the top 50 institutions can be fully explained by gender differences in field specialization. Bayer & Rouse (2016) find female and male AEA members to have different opinions on key economic issues. When comparing gender disparities in field specializations among those with single field specializations, Sierminska & Oaxaca (2021) find greater disparity among academic economists based on publications than among graduate students based on doctoral dissertations. Consequently, gender field specialization disparity among academic economists is higher in older cohorts. Chari & Goldsmith-Pinkham (2017) examine the variation in conference attendance across different NBER Summer programs and find a lower representation of women in fields categorized as Finance and Macroeconomics, and higher representation in a field characterized as Microeconomics. A similar concentration is found by Beneito et al. (2018) when they analyze the attendance of AEA meetings from 2010 to 2016 by examining JEL codes. We also find this to be the case among PhD students. Beneito et al. (2018) find that gender differences in subfields are established in the early stages of economic studies (undergraduate level), where women outperform men in microeconomics related courses and men outperform women in macroeconomic-related courses. The most common argument for the underrepresentation of women in some sub-fields has been path dependence or the lack of role models in the field.

Both Dynan & Rouse (1997) and Hale & Regev (2014) point to the presence of role models as a contributing factor for females that are considering pursuing studies in economics. Carrell et al. (2010) find that in a setting where students are randomly assigned to professors, female students perform significantly better in intro math and science courses taught by women and are more likely to pursue majors in the STEM fields (economics is not considered here).

Hale & Regev (2014) find path dependence to be an important mechanism that drives the persistence of low shares of women in academic institutions. By disentangling employer gender bias from causal effects of gender composition of faculty, they find the effect of the female faculty share to be higher when the share of female faculty is low. The effect of female faculty share declines as the share increases suggesting the elimination of gender biases when female faculty is larger. Thus, fields with more women present will experience a higher increase in women over time. Dolado et al. (2012), test the implications of theories that predict path dependence, such as gender differences in preferences for research topics, gender differences due to external factors (men avoiding female fields or women shying away from female fields), and gender attitudes in highly competitive environments such as academic research in top departments. They find that the probability that a woman works in a given field is positively related to the share of women already working in that field (path dependence) and suggest that this could be related to women not wanting to work in maledominated fields rather than vice versa. The effect is found to be weaker for younger female researchers.

3 Data

The data we use in this paper come from several sources: Academic Analytics, EconLit, and official websites of states in the U.S., serve a different purpose and are merged at different stages of the research process, in order to construct variables necessary for estimation.

Graduate students The data on graduate students come from EconLit for the years 2009 to 2018. Every year the December issue of the *Journal of Economic Literature* publishes a list of economics and economics-related doctoral degrees conferred by U.S. and Canadian universities during the previous academic year along with the JEL codes that identify the field(s) of specialization for the doctoral dissertation.[3] The graduate students' fields of specialization are therefore identified directly by the degree granting institution. Gender information is obtained from Gender-API.com for about 90% of the sample. The rest of the sample is coded manually from internet searchers, photos, etc.

Academic Analytics and EconLit The largest of the datasets comes from Academic Analytics (AA). AA is a data gathering entity that has information on over 270,000 faculty members at more than 385 subscribing U.S. PhD granting universities (out of approximately 431). The AA data are structured by departments and disciplines and include information on 1) the publication of scholarly work, 2) citations to published journal articles, 3) research funding by federal agencies, 4) honorific awards received by faculty members, 5) year of graduation and PhD granting institution, and 6) gender.[4]

The American Economic Association EconLit database also includes information on all published articles in economics along with the JEL codes. These are merged with information on academics and their publication records from AA in order to identify the main field of research specialization for each academic economist.

JEL codes JEL codes consist of one letter and usually one or two digits.[5] In our paper we focus on the letter code. There can be as many as 5 to 7 JEL codes listed on a dissertation entry (and on publications). In our research design we select the most prevalent letter to identify the field for each dissertation (the mode). Our analysis has indicated that in almost

[3]The list of the 20 JEL codes can be found at https://www.aeaweb.org/jel/guide/jel.php
[4]AA assigns gender based on an in-house algorithm created for this purpose.
[5]For example, G53 refers to Financial Literacy, a sub-field of Finance (G) and Household Finance (G5).

30 percent of cases more than one mode exists. At some point in the process, the true field of specialization is thus blurred. Consequently, we exclude doctoral dissertations that report three or more main fields (981 dissertations – about 9.5% of the original sample) and retain 1420 cases (15% of the remaining sample) that report 2 main fields (two modes).

Single/Dual field sample Our analytic sample is comprised of individuals who specialize either in one or in two primary doctoral fields. Thus, the decision setting is one in which an economics graduate student chooses one or two specializations from a menu of primary doctoral fields. Our initial sample has 9350 observations. Merging with AA leaves us 95% of the initial sample – 8853 observations with 15% of the sample reporting two fields.

Salaries For purposes of predicting field specialization among graduate students, salary information is collected from public websites of states. Many states have policies that make faculty salaries publicly available. In addition to salary records, the data typically provide the home department of the faculty member, a secondary department if applicable, the individual's position, and whether they are employed full-time. The salary information is merged with the Academic Analytics data to obtain information on the years lapsed since receipt of the PhD, the PhD granting institution, field specialization, and gender. For those individuals for whom gender information is not available, we use Gender-API.com or code them manually based on internet searches.

Rankings The rankings of Economics departments and universities are taken from the US News and World Report rankings in 2018 and are classified as described in Section 4.3.1.

4 Conceptual Framework

Field specialization among PhD economists over their working lives is an ever evolving process. The more disaggregated the identification of field specialization, the more prone field specialization is to change. From initial conditions one can imagine a career trajectory that is partly planned and partly a response to exogenous events.

What one might choose to regard as initial conditions is somewhat subjective. We define initial conditions to correspond to field specialization in economics in graduate school at the doctoral level. Clearly, there are potentially very important antecedents which culminate in the graduate school experience, e.g. family background, K-12 experiences, undergraduate education, etc. However, the data requirements for these antecedents are daunting and in some cases are addressed by other scholars.[6] Our manageable strategy is to analyze graduate school determinants of one's field of doctoral specialization.

Typically, U.S. doctoral programs in economics entail five or six years for completion and arguably for most individuals the doctoral field decision is effectively made in their third or fourth year of graduate study. Thus, a decision made in year t regarding the doctoral field of specialization is likely concluded in year $t+2$ or $t+3$ with receipt of a PhD.

We employ parametric models of doctoral field specialization choices to rationalize observed gender differences in distributions across fields of specialization at the initial career stage. Field specialization choice can be framed in terms of the most commonly accepted economic framework for choice, i.e. utility maximization. For discrete choice, the most widely accepted framework is that of random utility. As shown below, the random utility approach is only a starting point to a more comprehensive estimation strategy that is not explicitly dependent on any particular utility function.

Bounded rationality and behavioral economics can certainly cast doubt on the universal assumption of fully informed and maximizing rational actors. Our approach does not necessarily assume full information for a rational agent. What it does do is model the revealed determinants of the choices observed. Implicit in this are a variety of economic, social, and informational constraints faced by the individual. The modeling incorporates attributes of field specialization choices as well as attributes of the individuals making choices over field specialization.

Our approach groups determinants into three categories: personal characteristics, economic factors, and graduate school environmental factors. Personal characteristics include variables such as gender, year of PhD, and indicators for multiple field specializations. Economic factors include measures of field specific expected relative (to all other fields) starting salary and field specific expected probability of academic employment in an American PhD granting institution. These economic factors are choice specific. Since expected relative starting salary may be a bit myopic, we also include expected relative salary at 7 years beyond the completion of one's doctorate. Graduate school environmental factors include quality/rankings of the department and university, and gender composition of the faculty by field.

Given that the choice of economic field specialization will be considered at the initial stage of graduate school, we will be able to abstract more from subsequent life events that influence one's career trajectory with respect to academic versus nonacademic careers. Nevertheless,

[6] For example, Avilova & Goldin (2018) address the issue of choosing economics as a major in undergraduate studies. Altonji et al. (2016) presents a dynamic model of field specialization with sample selection and summarizes studies based on survey data for a few individually selected private colleges and universities.

we include a measure of the prospects for immediate postgraduate academic employment by field of specialization.[7]

The multivariate binomial field choice environment is one in which one can chose either single or multiple primary doctoral field specializations. This is an example of what more generally can be termed a "multi-response" model. We consider only cases in which individuals choose no more than two fields of specialization. The dual field specializations account for about 60% of those with more than one field of specialization.

4.1 Multivariate Binary Field Choice

Suppose there are J primary fields. A straightforward random utility rationalization for a multi-field model is one in which the individual faces J independent binary but not mutually exclusive decisions. For the ith individual, S_{ij} is an indicator denoting selection of the jth primary field, $j = 1, ..., J$. Consider the additive random utility model for each field given by:

$$U_{ij} = Y_{ij}\theta_j + Z_i\gamma_j + \varepsilon_{ij}, i = 0, ...N, j = 1, ..., J,$$

where U_{ij} is the net utility from choosing j, Y_{ij} is a vector of variables reflecting the attributes associated with the ith individual's choice of the jth field, Z_i is a vector of individual (case) specific attributes including a constant term, θ_j and γ_j are parameter vectors, and ε_{ij} is a logistic distributed error term.[8] It follows that

$$\begin{aligned} Prob(S_{ij} = 1) &= Prob(U_{ij} > 0), j = 1, ..., J. \\ &= Prob(Y_{ij}\theta_j + Z_i\gamma_j + \varepsilon_{ij}) > 0 \\ &= Prob(\varepsilon_{ij} > -(Y_{ij}\theta_j + Z_i\gamma_j)). \end{aligned}$$

The model yields J independent binomial logit equations in which individuals can exhibit multiple doctoral field specialization with the probability of each field specializations given by

$$Prob(S_{ij} = 1) = \frac{exp(Y_{ij}\theta_j + Z_i\gamma_j)}{1 + exp(Y_{ij}\theta_j + Z_i\gamma_j)}, j = 1, ..., J.$$

[7] Academic employment is still the largest source of employment at the early stage of an economist's career and overall, though we acknowledge that nonacademic employment is becoming relatively more important. Stock & Siegfried (2014) estimate that in 2011 the proportion of economics PhDs employed by academic institutions overall was 56.3%.

[8] The parameters of the utility functions are identified up to the usual scale normalization (Choe et al. (2020)).

When individuals can choose either a single or two primary doctoral fields, the maximum number of possible dual field specialization pairs is given by

$$\binom{J}{2} = \frac{J!}{(J-2)!\,2!} = \frac{(J)(J-1)}{2}.$$

After adding the J single (primary) field specializations, there are

$\dfrac{(J)(J-1)}{2} + J = \dfrac{(J)(J+1)}{2}$ possible field choice outcomes/specializations.

A limitation of the independent binomial logit framework is that it assumes away correlated field specialization choices. A more general framework is a multivariate logit model for correlated binary choices. A relatively attractive implementation of a multivariate logit choice model with correlated choices is described in Bel et al. (2018) and Russell & Petersen (2000). Here, the multivariate logit model is not derived from a random utility model *per se* but rather from an assumed structure for conditional probabilities that have the same form as those from the independent binomial logit models rationalized by random utility but which allow for correlated choices:

$$Prob[S_{ij} = 1|I_{ij}] = \frac{exp(I_{ij})}{1 + exp(I_{ij})}, \quad j = 1, ...J$$

where $I_{ij} = Y_{ij}\theta_j + Z_i\gamma_j + \sum_{l \neq j} \psi_{jl} s_{il}$, s_{ij} is the realization of the random variable S_{ij}, and $\psi_{jl} = \psi_{lj}$ is the association parameter reflecting the underlying correlation between the choices of j and l.

If we let $S_i = (S_{i1}, ..., S_{iJ})$ denote the set of random primary field specializations and $s_i = (s_{i1}, ..., s_{iJ})$ denote the realizations of the primary field specializations, it follows from Bel et. al (2018) that the joint distribution of S_i implied by the assumed conditional probabilities is given by

$$Prob(S_i = s_i|I_{i1}, ..., I_{iJ}) = \frac{exp\left[\sum_{j=1}^{J} s_{ij}(Y_{ij}\theta_j + Z_i\gamma_j) + \sum_{l>j} s_{ij}s_{il}\psi_{jl}\right]}{\sum_{m_i \in M} exp(\mu_{m_i})},$$

where M represents the choice outcome space. The outcome space consists of potentially $(J)(J+1)/2$ field choice outcome combinations. Full Information Maximum Likelihood (FIML) estimation of the above multivariate logit model becomes unwieldly as J increases.

We adopt the particularly tractable estimation strategy developed in Bel et. al (2018) termed the Composite Conditional Likelihood (CCL) estimation method. The complexity of the estimation task is significantly reduced by maximum likelihood estimation of the conditional probabilities. The CCL log likelihood function is given by

$$lnL(\theta, \gamma, \psi; s) = \sum_{i=1}^{N} \sum_{j=1}^{J} \left\{ s_{ij} ln \left[\frac{exp(I_{ij})}{1 + exp(I_{ij})} \right] + (1 - s_{ij}) ln \left[\frac{1}{1 + exp(I_{ij})} \right] \right\}.$$

While not fully efficient, CCL yields consistent estimators. Monte Carlo experiments by Bel. et al (2018) show that the CCL finite sample bias is very close to that of FIML.[9,10]

4.1.1 Gender Differences in Specialization Preferences

It is assumed that group preferences toward fields of doctoral specialization are manifested by the parameters of the multivariate logit model with correlated binary choices. This of course does not imply that identical parameters for women and men would yield identical group field specialization outcomes. Gender differences in field specialization determinants could yield gender disparities in field specialization even if there were no group differences in the parameters. In section 5 we report decomposition exercises that shed light on the relative roles of gender differences in preferences and characteristics in generating gender differences in field specializations.

4.2 Field (Alternative) Specific Variables

The field/choice specific variables in Y_{ij} reflect the attributes of the choice and correspond to the outcomes for each individual under each field specialization. These are described in more detail below.

4.2.1 Relative salaries

It is natural to inquire to what extent expected relative salaries by field influence economics graduate students in their choice of field of doctoral specialization and whether there are gender differences in relative salary effects. The choice-specific expected salary variable in the vector Y_{ij} is not an expectation in the mathematical sense but is rather an anticipated or

[9] An alternative approach is the Generalized Extreme Value (GEV) described in Bresnahan et al. (1997). This approach is a generalization of a multinomial logit model with $(J)(J+1)/2$ possible field choice outcomes/specializations. This is not as attractive in the present case as we have J = 10 primary fields yielding 55 possible field choice outcomes.

[10] If the ψ_{jl} association parameters are jointly restricted to equal 0, the model collapses to the independent binary logits model.

perceived salary relative to all other fields. It is constructed from salary regressions estimated separately by field and gender for the samples collected from state websites:

$$ln\left(W_{ijkt}\right) = X_{ijkt}\beta_{jk} + u_{ijkt}, i = 1, ..., N_k, j = 1, ..., J, k = m, f, t = 1, , , .T_i,$$

where W_{ijkt} is academic salary, X_{ijk} is the observation vector for salary determinants, β_{jk} is the parameter vector, and u_{ijkt} is a disturbance term. The variables in X include number of years post-PhD, years post-PhD squared, indicators for rankings of the departments and universities associated with the institutions at which individuals earned their doctorates and also where they are employed, whether or not the individual has more than 1 field of specialization, and the year in which the salary is observed.

Public college and university faculty salary data reported on official state websites vary considerably in terms of the types of salaries being reported and in terms of disclosing what is being reported. Given the ambiguity that can be present regarding the type of salary being reported, it is not possible to convert the salary information to a uniform basis. Consequently, we add a set of indicators for base salary, net salary, annual salary, gross salary, and for the absence of information on which type of salary is being reported.

Although there are repeated observations for some individuals, we do not use panel data methods. Many salary determinants are largely time invariant for individuals. We do, however, cluster standard errors at the department level. Moreover, our intention is not to attempt the specification and estimation of the true data generating process underlying academic salary determination. Rather, we seek to approximate the information from which a beginning economist choosing their fields of doctoral specialization might intuitively form a forecast or prediction of field specific relative academic salaries in a situation in which the individual cannot foresee every contingency associated with a future starting job. For the attributes of the future types of academic institutions at which one could be employed or the salary type that would apply, we assign the field specific sample mean values when using the estimated field specific salary equations to forecast one's anticipated salary in academia.

Accordingly, the expected starting salary of an individual of gender $k = m, f$ in field specialization j is obtained as an out-of-sample forecast:

$$\widehat{lnW}_{ijk}^{(1)} = X_{ijk}^{(1)}\hat{\beta}_{jk}^{(1)},$$

where '(1)' indicates that number of years post-PhD is set equal to 1. Choice-specific expected starting relative salaries are constructed as follows:

$$V_{ijk}^{(1)} = \frac{\widetilde{W}_{ijk}^{(1)}}{\sum\limits_{l \neq j} \widetilde{W}_{ilk}^{(1)}/J},$$

where $\widetilde{W}_{ijk}^{(1)} = \exp\left(\widehat{lnW}_{ijk}^{(1)} + 0.5\hat{\sigma}_{u_{jk}}^2)\right)$, and $\hat{\sigma}_{u_{jk}}^2$ is the estimated variance for the log normal salary equation disturbances. Similarly, the choice-specific expected relative salaries at 7 years post-PhD (more or less for tenured positions) are constructed as

$$V_{ijk}^{(7)} = \frac{\widetilde{W}_{ijk}^{(7)}}{\sum\limits_{l \neq j} \widetilde{W}_{ilk}^{(7)}/J}.$$

Given these specifications, we are able to examine the separate influences on doctoral field specialization of immediate and longer term field specific relative salaries.

4.2.2 Probability of academic employment

For doctoral field choices the prospects for academic employment could be as important or even more important than relative salaries. We employ a logit framework to capture the subjective probability that an individual choosing primary doctoral field j will attain academic employment in a PhD granting institution.

Let $E_{ijk} = 1(i \in G_{dik} \cap A_{eik})$ where $k = m, f$, G_{di} and A_{ei} denote the individual i reference graduate doctoral dissertation institution and employment at a PhD granting university in our sample, respectively. If an individual is found in both samples, they are identified as an individual with a doctorate from institution G_{di} who is employed in an American PhD granting institution (A_{ei}). Operationally, academic employment is assigned if individual i who received their PhD in economics in year t_i (at G_{di}) appears in the PhD granting institution sample (A_{ei}) in either year $t_i + 1$ or $t_i + 2$.

The logit model is estimated separately by fields for males and females using academic economist samples ($\widetilde{N}_k$):

$$\begin{aligned} P_{ijk}^{emp} &= prob(E_{ijk} = 1 \mid H_{ijk}) \\ &= \Lambda(H_{ijk}\phi_{jk}), i = 1, ..., \widetilde{N}_k, j = 1, ..., J, k = m, f \end{aligned}$$

where $\Lambda(\cdot)$ is the logistic CDF, ϕ is a parameter vector, and H is a vector of determinants that includes indicators for the ranking of one's PhD institution and economics department (or not being ranked), and year of PhD. The graduate student (out-of-sample) predicted academic employment probabilities are calculated as

$$\widehat{P}_{ijk}^{emp} = \Lambda \left(H_{ijk}\hat{\phi}_{jk} \right), i = 1, ..., N_k, j = 1, ..., J, k = m, f.$$

4.2.3 Gender Faculty Shares by Field

We use the ratios of the number of female faculty members and male faculty members in each field relative to the number of total faculty in the department from which the individual earned their doctorate to capture the effects of the gender mix of faculty on doctoral field specialization choices.

4.3 Individual (Case) Specific Variables

4.3.1 Economics Department and University Rankings

Indicator variables are constructed to represent the economics ranking of the department from which the graduate students earned their PhDs. These indicators correspond to Top 20, Top 20 to Top 40, below Top 40 and non-ranked departments.[11] Indicator variables are also, constructed to represent the national ranking of the university from which the graduate students earned their PhDs. The university rankings correspond to Top 20, Top 20 to Top 40, below Top 40, and non-ranked. The omitted category is below Top 40 for both groups of variables.

4.3.2 Year of PhD

In the interest of parsimony, we control for the year in which one received their PhD as a continuous variable. This construction allows us to interpret the field specific coefficients as indicative of doctoral field specialization trends. An alternative would be to use year indicator variables and frame the year effects as cohort effects.

4.4 Field Specialization Disparity Measures

While every one of the J primary fields is selected by some subset of the sample, not all of the $(J)(J-1)/2$ possible doctoral dual field specializations are necessarily selected. The estimated probability that individual i of gender k would specialize in any given primary

[11] The absence of an economics department ranking includes departments such as Finance or Agriculture & Resource Economics.

doctoral field j is given by

$$\widehat{P}_{kij} = \frac{exp(\widehat{I}_{ij})}{1 + exp(\widehat{I}_{ij})}, \quad k = m, f, \quad j = 1, ...J,$$

where $\widehat{I}_{ij} = Y_{ij}\hat{\theta}_j + Z_i\hat{\gamma}_j + \sum_{l \neq j} \hat{\psi}_{jl} s_{il}$. A useful property of logit models with a constant term is that the means of the predicted probabilities match the sample proportions:

$$\overline{P}_{kj} = \frac{\sum_{i=1}^{N_k} \widehat{P}_{kij}}{N_k} \quad k = m, f, \quad j = 1, ...J.$$

where N_k is the total number of individuals in the sample.

The presence of dual field specializations implies that the sample primary field proportions sum to more than 1. To accomodate the total number of possible field specializations, including the single primary fields, let $S_{kij} \cdot S_{kil}$ denote a field specialization indicator for dual specialization in primary fields j and l, where S_{kij} and S_{kil} are indicators for specialization in fields j, l, and $k = m, f$. Note that for specializations in only a single primary field, $j = l$ and $S_{kij} \cdot S_{kij} = S_{kij}$. We introduce a new set of indicators to account for both single and dual field specializations denoted by 's':

$$\begin{aligned} \pi_{kis} &= S_{kij} \quad \text{for single field specialization} \quad j = 1, ..., J \\ &= S_{kij} \cdot S_{kil} \quad \text{for dual field specialization} \quad j \neq l. \end{aligned}$$

It follows that the predicted sample proportions for every observed field specialization s (single or dual field specializations) are given by

$$\overline{\pi}_{ks} = \frac{\sum_{i=1}^{N_k} \widehat{\pi}_{kis}}{N_k} \quad k = m, f, \quad s = 1, ...J^*,$$

where $\widehat{\pi}_{kis} = 1$ or 0 is the predicted field specialization for s, and J^* denotes the total number of actual observed field specializations, i.e. $J \leq J^* \leq (J)(J+1)/2$. For each individual, consider the set of predicted primary field specialization probabilities such that $\widehat{P}_{kij} > 0.5 \; \forall j$. If no predicted primary field specialization probability satisfies this condition, then $\widehat{\pi}_{kis} = 0$ for $s = 1, ..., J^*$. Otherwise, $\widehat{\pi}_{kis} = 1$ for some single or dual field specialization s. If only one primary field specialization satisfies the condition that its highest predicted probability exceeds 0.5, the individual is predicted to specialize in only one primary field. If exactly two predicted primary field specialization probabilities exceed 0.5, then the individual is predicted to specialize in these two fields.

When more than two predicted primary field specialization probabilities for an individual exceed 0.5, we impose a dual field weighting scheme. First, we determine all possible pairings of the total number of predicted primary field specializations. Let $K_i > 1$ represent the number of predicted dual fields for individual i. Next, we assign a probability of $1/K_i$ to each of the possible specialization pairs. If the individual actually selected one of the possible dual field specializations that we predicted, the specialization is assigned a weight of $1/K_i$, otherwise we assign a weight of 0. Consequently, the field specialization proportions are constrained to sum to 1:

$$\sum_{s=1}^{J^*} \widehat{\pi}_{kis} = 1, \quad k = m, f.$$

4.4.1 Duncan Dissimilarity Index

A straightforward measure of disparity between the distributions of some outcome for two demographic groups is given by the Duncan Dissimilarity Index (DDI):

$$D_{mf} = \frac{1}{2} \sum_{s=1}^{J^*} \left| \overline{\pi}_{ms} - \overline{\pi}_{fs} \right|,$$

If the field distributions were identical for women and men, the value of the index would be 0. At the other extreme if there were no gender overlap (complete segregation) in fields of specialization, the value of the index would be 1 (or 100 if scaled up to percentages). The value of the index represents the proportion of either gender who would have to change doctoral field specializations in order for there to be complete parity.

We can readily conduct counterfactual simulations to determine how the value of the Duncan index would change if women and men had the same values of the variables determining field specialization or faced the same (estimated) multivariate binary logit parameters when they were choosing their doctoral field of specialization,

If we adopt the estimated field specialization multivariate logit model for men as the baseline, the counterfactual Duncan Dissimilarity index value arising from gender differences in the values of the specialization determinants (characteristics) would be computed as

$$D_{mf}^{mc} = \frac{1}{2} \sum_{s=1}^{J^*} \left| \overline{\pi}_{ms} - \overline{\pi}_{fs}^{m} \right|,$$

where $\overline{\pi}^m_{fs} = \frac{\sum_{s=1}^{J^*} \widehat{\pi}^m_{fis}}{N_f}$, and $\widehat{\pi}^m_{fis} = 1$ or 0 is the counterfactual prediction that the ith female would select doctoral specialization 's' if she faced the estimated male field specialization logit model. We can also simulate the counterfactual value of the Duncan index arising from gender differences in the logit choice parameters:

$$D^{mp}_{mf} = \frac{1}{2} \sum_{s=1}^{J^*} \left| \overline{\pi}_{fs} - \overline{\pi}^m_{fs} \right| .$$

As an alternative, we can simulate the counterfactual effects of gender differences in characteristics and parameters using the estimated field specialization multivariate logit model for women as the baseline.

4.4.2 Quadratic Disparity Measure

One shortcoming of the Duncan index is that summation over the absolute value functions $|\overline{\pi}_{ms} - \overline{\pi}_{fs}|$ does not admit the construction of separable characteristics and parameter counterfactual decomposition components that sum to the sample value of the index. Consequently, we introduce the decomposable quadratic (field) disparity measure (QDM) constructed from the sum of squared gender differences in the field specialization proportions:

$$Q_{mf} = \frac{1}{2} \sum_{s=1}^{J^*} \left(\overline{\pi}_{ms} - \overline{\pi}_{fs} \right)^2 .$$

The higher the value of Q_{mf}, the greater the degree of gender field specialization disparity. It can be shown that Q_{mf} is bounded between 0 and 1, which corresponds to full parity and perfect segregaton, respectively.[12]

If we adopt the estimated multivariate binomial logit model for men as the baseline, the QDM can be decomposed as follows:

$$\begin{aligned} Q_{mf} &= \frac{1}{2} \sum_{s=1}^{J^*} \left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs} + \overline{\pi}^m_{fs} - \overline{\pi}_{fs} \right)^2 \\ &= \underbrace{Q^{mc}_{mf}}_{characteristics} + \underbrace{Q^{mp}_{mf}}_{parameters} + \underbrace{Q^{mcp}_{mf}}_{cross\,products} , \end{aligned}$$

[12]Details on the derivation of the boundaries can be found in the Technical Appendix.

where $Q_{mf}^{mc} = \dfrac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{ms} - \overline{\pi}_{fs}^{m}\right)^2$ is the estimated amount of the gender field disparity arising from gender differences in characteristics, $Q_{mf}^{mp} = \dfrac{1}{2}\sum_{s=1}^{J^*}(\overline{\pi}_{fs}^{m} - \overline{\pi}_{fs})^2$ is the estimated amount of the gender field disparity arising from gender differences in parameters, and $Q_{mf}^{mcp} = -\sum_{s=1}^{J^*}\left(\overline{\pi}_{fs}^{m} - \overline{\pi}_{ms}\right)\left(\overline{\pi}_{fs}^{m} - \overline{\pi}_{fs}\right)$ is the estimated amount of the gender field disparity arising from the cross products (interactions) of characteristics and parameters. An alternative counterfactual is obtained by adopting the estimated multivariate binomial logit model for women as the baseline.

The decompositions allow for simulations of how any policy changes that impact the effects of gender differences in characteristics or in the values of the specialization determinants might alter field specialization dissimilarity.

5 Empirical Findings

All of our empirical findings discussed below pertain to our doctoral sample. We start with descriptive statistics on primary field specialization comparisons for women and men followed by a list of summary statistics of our main covariates, after which we provide the estimation results from our field specialization model. Next, we report counterfactual simulations of our field segregation indexes as well as decomposition exercises with the Quadratic Dissimilarity Measure. Finally, we provide a predictive performance evaluation of our model.

5.1 Field specialization according to JEL classification

In order to conduct our empirical analysis in a manageable fashion, we aggregate the 20 JEL field classifications into 10 primary fields. Our primary field groupings are guided by a desire to focus on JEL categories that reflect related areas of economic inquiry.[13] There are a total of 8,853 individuals who specialized in either a single primary field or in two primary fields: 2,585 (29.1%) are women and 6,268 (70.8%) are men.

Table 1 shows that among graduate students who received their PhD's between 2009 and 2018, men have a statistically significantly higher presence in Econometrics, Micro, and Macro/Finance (3 to 6 percentage points), while women have a significantly higher presence in Labor/Health and Dev/Growth/Int (12 and 2 percentage points, respectively). Because of dual field specializations, the percent distributions add to more than 100. Dual field specializations accounted for 15.4% of the doctoral dissertations. The dual field specialization

[13] These are Econometrics, Microeconomics (Micro), Labor/Health, Macro/Finance, Industrial Organization (IO), Environmental & Agricultural Economics (Ag/Environmental), Public Economics; Development/Growth/International (Dev/Growth/Int); History, and Other.

rate was marginally lower for women (14.7%) then that of men (15.6%) but the difference is not statistically significant. Dual field specializations vary considerably with respect to how frequent each primary field is paired with other primary fields.[14]

Table 2 reports gender differences in dual field specializations by primary field. Among women the primary fields of Labor/Health and Development/Growth/International were paired most frequently with other primary fields. Whereas for men, Micro and Macro/Finance were paired most frequently with other primary fields. Accounting for 32.3% of dual field specializations among women, Labor/Health was the top primary field choice for dual field specializations. The Labor/Health percentage for women exceed that of the men by a statistically significant 9.58 percentage points. Dev/Growth/Int was the second most frequently selected field with 32.0% of women selecting it versus 25.7% of men. Micro was a top primary field choice for dual field specializations for men and it was significantly more prevalent among men at 41.7% vs 31.8% for women.

Although Macro/Finance was the second most frequently paired field for men (33.0%) but not for women (29.4%), the gender difference in percentages was not statistically significant. Econometrics was paired with other fields less frequently for women (by 6 percentage points), while Ag /Environmental was paired more frequently for women (also by 6 percentage points). Gender differences in the remaining fields of IO, Public Economics, History, and Other were not statistically significant.

5.2 Descriptive statistics

Descriptive statistics for the case specific and choice specific covariates appearing in the field specialization model are reported in Tables 3 and 4, respectively. The case specific covariates are comprised of the economics departments (Econ) rankings, the university (Uni) rankings, and the year in which the individual earned their doctorate. For departments and universities, rankings below the top 40 are treated as the reference category. Women are less likely to receive their PhDs from a top 20 economics department or from a top 20 university. The average year in which our doctoral sample received their doctorates was 2013. On average the women graduate students received their doctorates about 1.6 months earlier than the men.

[14]In terms of the original 20 category JEL classification in the doctoral sample, women are more prevalent in Labor (J) and Health (I) by about 5 to 6 percentage points and in Economic Development (O) by 2 percentage points. Men, on the other hand, have a statistically significantly higher presence in Quantitative Methods (C), Microeconomics (D) and Macro (E) by about 3 percentage points in the first two fields and 4 in the third field. Men also have a small but significant advantage in Financial Economics (G) and IO (L). The Duncan Dissimilarity Index (DDI) for 20 groups stands at about 13.2% (see Appendix Table A4 for details).

The choice specific covariates are comprised of the faculty share of women and men in the graduate student's department in a given field ('F share field' and 'M share field', respectively), the predicted probability of academic employment by field (P(empl)), and the field specific relative salary one year post PhD (Rel sal field yr1) and seven years post PhD (Rel sal field yr7). The average faculty field shares are small and statistically the same for women and men with two exceptions: the share of female faculty in field 10 ('Other') was significantly higher for female graduate students (0.73% vs. 0.58%) and the share of male faculty in field 8 ('Dev/Growth/Int') was also significantly higher for female graduate students (0.87% vs. 0.75%). The predicted probabilities of academic employment for men are statistically significantly higher in every field except History. For women, by far the highest predicted academic employment probability is in History (0.40). For men the highest predicted employment probability is in Econometrics (0.15) followed closely by eight other fields (0.14 - 0.11). Field specific relative (to all other fields) salaries one year out of graduate school are also statistically significantly higher among men for most fields. The exceptions are Labor/Health, Ag/Environmental, Dev/Growth/Int, and Other. These fields are relatively more attractive salary-wise for women. The highest relative salaries for women are in Ag/Environmental, Labor/Health, and Dev/Growth/Int, whereas for men the highest relative salaries are in IO, Ag/Environmental, and Micro. The relative salaries 7 years out of graduate school remain th highest for men in Econometrics, Micro, and Public Economics and for women in Labor/Health, Ag/Environmental, Dev/Growth/Int.

5.3 Multi-field model

A ten-primary field aggregation implies that there are a maximum of 55 field specialization outcomes: 10 single field specializations and 45 dual field specializations. For men and women combined, 44 of the 45 possible dual field specializations are observed. The only dual field specialization not observed is Econometrics and History. When disaggregating by gender, 43 of the possible 45 dual field specializations were realized among males and 40 among females. For women the number of observations in each realized dual field combination ranged from 1 to 36, whereas for men the number of observations ranged from 1 to 98.

Estimation of the multivariate logit CCL model entails deciding which dual field combination association parameters (ψ_{jl}) should be estimated to take proper account of correlated choices (otherwise the model will not converge). These decisions are based on the number of observations for each observed field pairing. We seek a lower bound on the number of observations that a field pairing must have in order to be included in the estimated model. In order to have some minimally clear separation in the cutoff number of observations that applies equally to men and women, we require that the lower bound satisfies two criteria: (1) the lower bound must be such that it includes a minimum of two more observations than the next lowest number of observations for a field pairing; and (2) the lower bound must be the same for men and women.

Adoption of the above decision rule leads to estimation of the association parameters for all dual field pairings with 12 or more observations. For women this lower bound comprises 16 (40%) of the 40 realized dual field pairings and 294 (78%) of the 379 dual field observations. The average and median number of observations for the included field pairings are 18.4 and 15.5, respectively, and for the excluded field pairings are 3.5 and 5.0, respectively. In the case of men the lower bound of at least 12 observations comprises 26 (60%) of the 43 realized dual field pairings and 909 (93%) of the 980 dual field observations. For men the average and median number of observations for the included field pairings are 35.0 and 55.0, respectively, and for the excluded field pairings are 4.4 and 5.5, respectively.

Although not reported, all of the estimated association parameters were negative and statistically significant. The negative values for $\hat{\psi}_{jl}$ make sense because for any given binary outcome field choice, conditioning on any alternative field choice means that the odds are low that the outcome field choice will be selected. There are 9 alternative fields with which to pair plus the choice of specializing in only a primary field.

Pooled regression While the multivariate CCL logit model is estimated separately for males and females, a useful overview of gender differences in field specialization is obtained by pooling the male and female doctoral thesis samples and adding an indicator variable for females (after conditioning on covariates and taking account of correlated choices). These gender effects are reported as marginal effects in Table 5. The marginal effects for the gender indicator variable are calculated for discrete changes from 0 to 1, with all other variables set equal to their sample mean values. Gender is statistically significant for six primary fields. Other things equal, women are more likely to select Ag/Environmental as a doctoral field of specialization and less likely to select Labor/Health, Macro/Finance, IO, Public Economics, or Dev/Growth/Int. The major difference from the unconditional raw data results occurs with respect to the Labor/Health field which was selected by 33% of the women versus 21% of the men. Had other things been equal for men and women, women would have been 34% less likely to select Labor/Health as a doctoral field than men.[15] Similarly, 16% of women

[15]The largest contributing factors to the observed Labor/Health specialization for women as opposed to the negative marginal female gender effect are the gender differences in relative salaries. At the sample means, the predicted probability of women specializing in Labor/Health is increased by 92.6 percentage points over what it would be if women faced the male average 7yr relative salaries and is reduced by 6.5 percentage points over what it would be if women faced the male average 1yr relative salaries.

selected Dev/Growth/Int (a 2 ppt lead over men) in the raw results, while the estimation results show that women would be 14% less likely to select it as a doctoral field compared to men.

Separate regressions With respect to the separately estimated models for women and men, we report the signs of the significant marginal effects in Table 6, while the actual marginal effects and their associated standard errors by field specialization choice are in the Appendix.[16] For continuous covariates, these marginal effects are calculated for 1 standard deviation changes in the variables.

Individual (case) specific covariates First thing to notice in Table 6 is that economics department rankings matter less for women's field choice than for men's. Among men, we observe no significance only for one field (Ag/Environmental), while for women we observe a complete lack of significance in 5 fields (Econometrics, Micro, Public Economics, History and Other). Receiving one's PhD from a top 20 (or top 40) economics department increased the probability that a woman would specialize in Ag/Environmental (or Dev/Growth/Int); however, the probabilities that a woman would specialize in Labor/Health or IO (or Macro for top 40) decrease. For men, the top 20 or top 40 economics department effect is associated with increased probabilities of specializing in IO and Dev/Growth/Int and diminished probabilities of specializing in Micro, Labor/Health. For Other, it is negative for top 20 department and positive for department ranking between 20 and 40. This also confirms that women are statistically significantly less likely to be found in top 20 departments and if they have received their doctorates from the top 20, it is less likely that they specialized in the fields where they are over represented. Among women who received their doctorates from unranked departments, the probabilities for specializing in Macro/Finance, IO, or Dev/Growth/Int increase and decrease for specializing in Labor/Health. In the case of men, the probabilities for specializing in IO, Public Economics, Dev/Growth/Int, or History increase, while the probabilities decrease for specializing in Labor Health, Macro/Finance, or Other.

Next, we examine the effect of university rankings on doctoral field choice. These results are in relation to those from institutions ranked below the top 40. University rankings are more likely to have a significant effect on field choice compared to economics department rankings for both women and men. They still have no effect for women when choosing Econometrics or Public Economics. For women receiving their PhDs from top 20 universi-

[16] Table A5 reports the marginal effects for the individual/case specific variables, and Table A6 reports the marginal effects for the field/alternative specific variables.

ties, the probabilities of specializing in Micro, Labor/Health, or Other are higher while the probabilities of specializing in Dev/Growth/Int are lower. If the university ranked in the top 40 but less than the top 20, women had higher probabilities of specializing in Micro, Macro/Finance, or IO, but lower in Ag/Environmental or Dev/Growth/Int. For men who receive their PhD's from top 20 or mid-ranked universities, the field specialization probabilities are increased for six fields (Micro, IO, Public Economics), and reduced for 3 fields (Econometrics, Macro/Finance, or Dev/Growth/Int). The probability for field choice in Labor/Health, History and Other increases only in top universities. The only statistically significant effect for women receiving their PhD's from an unranked university was a positive effect on the probability of specializing in Labor/Health and a negative effect for Ag/Environmental or History. The probability effects from earnings one's PhD from an unranked university were far more extensive for men. Field specialization probabilities are increased for Labor/Health, Macro/Finance, IO, or Other and reduced for Micro, Ag/Environmental, Public Economics, Dev/Growth/Int, or History.

The year in which one received their doctorate reflects trends in field specialization. The estimated marginal effects for the trend variable are calculated on the natural basis of a unit change. There was a positive trend in the probability of specializing in Labor/Health for women, and negative probability trends for specializing in Micro, Macro/Finance, IO, or Ag/Environmental. Field specialization probability trends were positive for men with respect to Econometrics, Macro/Finance, IO, or Dev/Growth/Int. The only negative field specialization probability trend for men is that of History.

Field-Specific (choice) covariates Table 6 indicates that the number of women faculty in each primary field as a share of the total number of department faculty had statistically significant effects on field specialization probabilities for women in half of the fields: Econometrics, Ag/Environmental, Public Economics, Dev/Growth/Int, and Other. However, the effects are positive for only Dev/Growth/Int and therefore the larger the share of the faculty who were women in Dev/Growth/Int, the more likely it was that female graduate students would specialize in this field. The female field faculty share variable also had only statistically significant negative effects on field specialization probabilities for men in three fields: Econometrics, Microeconomics, and Ag/Environmental. Thus, we find no path dependence in the field most over-represented by women (e.g. Labor/Health), but we do in Dev/Growth/Int. We also find the share of women faculty to be a discouraging factor in fields underrepresented by women. This refers to Econometrics both by men and women and in Micro by men.

The number of men faculty in each given field as a share of the total number of department faculty had statistically significant effects in three fields for both women and men. For women, the effect was negative in all three fields: IO, Public Economics, and History. For men, a higher share of male faculty in the selected field as a share of total faculty in the department had a positive effect on selecting Macro/Finance and negative effects for selecting IO or History. Thus, we find no negative path dependence for women in the fields where they are underrepresented, but we do find positive path dependence for men in one of those fields (Macro).

The field specific expected probability of employment in a PhD granting institution has statistically significant effects for women in three cases. The higher the expected probability of female academic employment in Labor/Health or IO, the more likely a female graduate student would specialize in these fields. On the other hand, the higher the expected probability of female academic employment in Ag/Environmental, the lower the probability that a female would specialize in this field. By contrast for men, the academic field employment probability effects are statistically significant in seven field specialization probabilities. Increases in the male academic employment probabilities increased the probabilities of male graduate students specializing in Micro, Labor/Health, Macro/Finance, IO, Dev/Growth/Int, or Other, and reduced the probability of males specializing in History.

For female graduate students, the expected 1st yr and 7yrs post PhD field specific relative salary effects are statistically significant for the same six fields: Micro, Labor/Health, Macro/Finance, IO, Dev/Growth/Int, and Other. Expected field specific relative salary 1st year post PhD reduced the probability of selecting the field while the 7 years post PhD expected relative salary increased the probability of selecting the field. In the case of male graduate students, the expected field specific relative salaries exhibited a somewhat different pattern. The expected 1st year post PhD field specific relative salary increased the probabilities of specializing in Econometrics or IO while decreasing the probabilities of specializing in Labor/Health, History, or Other. Expected 7yrs post PhD field specific relative salary increased the probabilities of male graduate students specializing in Labor/Health, History, or Other while decreasing the probabilities of specializing in Econometrics or Macro/Finance. Common to all graduate students is that the expected 1st and 7yrs post PhD fields specific relative salaries exhibit opposite effects on the probabilities of field specialization. It could be that positive 1st year relative salary effects coupled with negative 7th year relative salary effects is indicative of relatively higher discount rates (less patience) whereas the opposite sign pattern is indicative of lower discount rates (more patience). This latter pattern of possibly lower relative discount rates held for both women and men in Labor/Health and in Other and it held twice as often for women than for men. Furthermore, the only suggestion of higher discount rates occurred for men in Econometrics and in IO.

5.4 Counterfactuals and Field Specialization Decompositions

Table 7 reports gender differences in the primary doctoral fields of specialization for observed and counterfactually-predicted field specialization percentages. Using the estimated multivariate (CCL) logit binomial parameters for males (females), we construct the counterfactual distributions by predicting the female (male) primary field distribution. By the property of logit models with constant terms, the predicted primary field distributions for each gender group match the sample proportions. Because of the presence of dual primary field specializations, the distributions sum to more than 100%. The counterfactual distributions also sum to more than 100% because of multi-field specializations. However, there are no restrictions in place that preclude the counterfactual predictions from implying multi-field specializations in excess of two primary fields. As a result, the predicted counterfactuals imply significant increases in multi-field specializations.

In addition to significant changes in the predicted number of primary fields that would be combined into multi-field specializations, the counterfactuals reveal significant changes to the composition of the primary fields. These changes partly reflect the counterfactual predictions of multi-field specializations in excess of two primary fields. First consider the counterfactual in which women adopt the same primary field preferences as the men but retain their original characteristics profiles. There would be no women with specializations in Micro, Labor/Health, Dev/Growth/Int, or History. At the other extreme, all women would include IO among their doctoral field specializations. With the exception of Macro/Finance, the counterfactual proportions of women in the remaining fields would be considerably larger than those of the men and the observed proportions for women. Another way of looking at this counterfactual is to consider what would be the result if men retained their primary field preferences but their characteristics were those of women.

Next, consider the counterfactual in which men adopt the primary field preferences of women but retain their original characteristics profiles. There would be no men specializing in Micro, IO, Dev/Growth/Int, History, or Other. At the other extreme, all male graduate students would specialize in Econometrics, Labor/Health, and Ag/Environmental. In addition to a 3-primary field specialization, a large percentage of men would also specialize in Macro/Finance. In this case, we can also interpret this counterfactual as if women had male characteristics. They would all specialize in Econometrics, Labor, and almost all in Ag/Environmental.

Table 8 reports our findings for the dissimilarity measures. Based on the sample proportions for the 54 field specialization combinations actually observed, the Duncan Dissimilarity Index (DDI) and the scaled (by 100) Quadratic Dissimilarity Measure (QDM) are 15.27 and 0.82, respectively. Beyond the observed sample proportions, we also obtain model predictions for the DDI and the QDM over the maximum possible 55 mutually exclusive field specialization combinations (10 single primary and 45 dual primary field specializations).

As would be expected on the basis of the counterfactuals applied to the 10 primary fields reported in Table 7, the counterfactuals arising from the assumption of either gender group as the baseline dramatically increase the field specialization dissimilarity measures. The predicted value of the DDI indicates that 27% of either the women or the men would have to change field specializations in order for there to be complete parity in the doctoral field distributions. The counterfactuals imply nearly complete gender field segregation. These effects are about the same for counterfactual characteristics and parameters. So for example, if females had the same estimated choice parameters as males, 92.9% of either gender would have had to switch doctoral fields in order for there to be equal field distributions. This is then associated with gender differences in characteristics as the same parameter estimates are being used for both genders. On the other hand if we assign the male coefficients to the characteristics of females, we estimate the effects of gender differences in parameters to be 99.7%. In effect, we are holding characteristics constant by counterfactually assigning males the characteristics of females and measuring the effect of gender differences in parameters/preferences. The results from using the female baseline are very similar.

Although the DDI does not lend itself to decomposition analysis, we are able to decompose the predicted QDM value of 2.64. With the male baseline, characteristics account for 18.98 and parameter differences account for 25.64. The cross-product term accounts for -41.93 of the QDM. This negative value arises because the counterfactual field proportion minus the actual gender field proportion are of the same sign for every field. In the decomposition, these positive cross-products are summed and multiplied by minus 1. If we net out the interacation effect, characteristics and parameters account for 42.5% and 57.5% of net QDM gap, respectively. When using the female baseline for the counterfactuals, these magnitudes are reversed with characteristics having a larger effect (56.3%) and parameters having a smaller effect (43.7%).

5.5 Model evaluation

We evaluate the validity and usefulness of our model based on a simple measure that assesses model performance in predicting doctoral field specialization choices.[17] Table 9 reports the predictive accuracy rates corresponding to primary fields chosen as either a single or a dual field specialization and all field specializations (both single primary field specializations and dual field specializations). The overall averaged primary field predictive accuracy rates are

[17]The derivations of the predictive accuracy measures are presented in the Technical Appendix.

very close for women and men at 90 - 92%. The individual primary field rates are close for women and men with the exceptions that the model predicts the Micro specialization choice much better for men. Micro and History exhibit the highest accuracy rates for men at 99%. For women the highest predictive accuracy rate is that of History at 99%. The overall averaged accuracy rates for the 55 theoretically possible field specializations are also very close for women and men at 88 - 89 %.

While predictive accuracy is bounded between 0 and 100% and higher values indicate that the estimated model is more accurate, there is no clear accuracy rate threshold by which to judge the usefulness of the model. One standard is to ask how the calculated accuracy rates compare with those based on a naïve model. For the primary field predictions, a naïve model would hold that the binary choice of each primary field specialization is decided by the flip of a fair coin. Thus, the probability that an individual would choose any given field j is 0.5. This means that the expected accuracy rate for any given primary field and all primary fields taken together is 50%. Clearly, the model's predicted primary field accuracy rates for men and women far exceed this threshold of purely random primary field specialization choices. For the 55 mutually exclusive field specializations, a naïve model would hold that each outcome is equally likely with probabilities of 1/55. Thus, the overall averaged predictive accuracy rate would be $100 \times 1/55 = 1.8\%$. Again, the model predictive accuracy rate far exceed the naive threshold.

6 Discussion

While the multivariate logit model of binomial choices among a set of primary fields takes account of correlated choices, the estimated model does not have the property that the averaged mutually exclusive $(J)(J+1)/2$ predicted field specializations match the observed sample proportions. This is only true for predictions of the J primary fields. Consequently, the restriction that all predicted possible field specialization proportions add to 1 must be imposed post estimation. An alternative would be estimation of a multinomial/conditional logit model defined over $(J)(J+1)/2$ mutually exclusive outcomes. Because of the potential dimensionality issue with J primary fields, this is not a practical alternative. The problem is compounded by allowing for more than two-field specializations.

Inherently, identification of causation is a major challenge when attempting to explain outcomes in a process embedded in a dynamic sequence of educational and career decisions. While the choice of doctoral field specializations is arguably the outset of a professional economist's career, there are many antecedents to this early career outcome. The decisions and outcomes from one's undergraduate course of study clearly have some bearing on one's graduate school experience and choices. In turn, family background and K-12 experiences are antecedents to one's undergraduate experience. One can easily make the case for a plethora of selection effects along the way. There is likely no completely effective way to manage these issues when modeling the choice of doctoral field specializations. We have been able to at least condition on a variety of circumstances that seem reasonable to consider as sources that are systematically associated with early career specialization in economics and how these differ between men and women.

7 Summary and Conclusions

The central objective of this paper is to examine the process of doctoral field specialization among beginning economists and its implications for gender differences in field specialization. To this end, we model field specialization choices using a multivariate logit model for correlated binomial choices. On the empirical side, we merge several datasets including dissertation data, faculty list data, as well as academic salaries. Our model strongly dominates naïve models in terms of predictive accuracy. For primary field predictions, the model accuracy rate is 90% to 92% versus an expected rate of 50% for the naïve model. In the case of predicting the 55 primary plus dual field specializations, the model accuracy rate is 88% to 89% versus an expected rate of 1.8% for the naïve model.

After conditioning on personal, graduate school environmental, and economic factors, we find statistically significant gender effects in doctoral field choices among economics graduate students. Women are more likely to choose Ag/Environmental fields and less likely to choose Labor/Health, Macro/Finance, IO, Public Economics, or Dev/Growth/Int. The result for Labor/Health is particularly noteworthy as descriptive statistics show women dominating the Labor/Health field. This result points to the importance of controlling for gender differences in the determinants of field specialization choices as these can override gender differences in preferences.

The field specialization choice model estimated separately for women and men includes both individual specific and choice (field) specific covariates. The individual specific covariates comprise department and university rankings as well as the year in which one received their doctoral degree. Women who received their PhD from a top 20 or a top 20-40 economics department were less likely to have a specialization in Labor/Health or IO. Men who received their PhD from a top 20 or top 40 economics department were also less likely to have a specialization in Labor/Health; however, they were more likely to specialize in IO or Dev/Growth/Int. In turn, women who received their PhD's from a top 20 or top 40 university were less likely to specialize in Dev/Growth/Int but more likely to specialize in

Micro. Men who received their PhD's from a top 20 or top 40 university were less likely to specialize in Econometrics, Macro/Finance, or Dev/Growth/Int but more likely to specialize in Micro, IO, or Public Economics.

The growth trends were positive for the fields in which men and women are over represented – Labor/Health for women and Econometrics and Macro/Finance for men. The year in which one received their PhD has an effect on the field of specialization possibly related to funding trends or the popularity of other factors. As specified in our model, the continuous year variable is interpreted as picking up trends. The inclusion of year indicators would perhaps better reflect cohort effects – a topic to be further investigated.

Overall, it seems that women are statistically significantly less likely to be found in the top 20 departments. If they do graduate from these, they are less likely to be in the fields in which they are overrepresented. If women are graduating in fields in which they are underrepresented it is mostly from unranked departments. It would seem that if there were more women graduate students in top 20 or top 20-40 economics departments compared to those departments with a ranking below the top 40, selecting Labor/Health as a field would be less frequent and we would observe a more equal distribution of women across fields. Similarly for men, since Labor/Health is one of the most popular fields for both. Possibly, Labor/Health's popularity is due to it being perceived as a more forgiving field in terms of job prospects when it comes to school rankings and there are more opportunities to specialize in the field outside the top schools. The trend is to continue to see more women in Labor/Health and less in Econometrics and Macro/Finance – not diminishing gender field specialization dissimilarity. If current trends persist, additional gender gaps in field specialization may develop in IO and Development/Growth/Int.

The field specific explanatory variables comprise female and male faculty in each field as shares of the total faculty, the probabilities of academic employment and expected relative salaries 1-year and 7-years post PhD. The only field for which the female faculty share had a positive effect for women specializing in that field was Dev/Growth/Int. There were no fields for which the male faculty share had a positive effect for women. On the other hand, there were several fields for which the female or male faculty share had negative effects for either males or females. The only field in which the male faculty share had a positive effect on the probability that a male would select that field was Macro/Finance.

In other words, we find no path dependence (measured by the share of women in the field in the department) in the field mostly over-represented by women (e.g. Labor/Health), but we did in Dev/Growth/Int. We did find the share of women to be a discouraging factor in fields underrepresented by women. This refers to Econometrics both for men and women and in Micro for men. We find no negative path dependence (measured by the share of men in the fields in the department) for women in the fields where they are underrepresented, but we do find positive path dependence for men in one of those fields (Macro).

We find that increased probabilities of academic employment in Labor/Health and IO increase the probability of specializations in these fields for both women and men, though the effect is stronger for men (and also significant several other fields). The probabilities of academic employment themselves are also higher for men than for women. Thus, this is not one of the factors that fundamentally affects field specialization for women in our model.

The predicted field 1yr post-PhD relative salaries for women often have negative effects on the probability of selecting a given field, while the 7yr post-PhD relative salaries exhibit positive effects on field specializations. This may suggest a longer run and less myopic view for women. Generally, field specific relative salaries seem to matter slightly more often for women than for men.

The Duncan Dissimilarity index calculated for the sample proportions of the 10 single and 44 dual field specializations attains a value of 15.3%, and the predicted value when restricting field specializations to not exceed 2 is 27.1%. The counterfactual Duncan Dissimilarity measures indicate that field specialization segregation would be virtually complete if women faced male parameters and vice versa. This suggests substantial gender differences in characteristics and parameters of the model. Gender heterogeneity in parameters and characteristics is also implied by the quadratic dissimilarity measure decompositions.

If we consider the primary fields in which women are underrepresented (Econometrics, Micro, and Macro/Finance), we find that growth trends reinforce this underrepresentation. In some cases, department and university rankings have positive effects on women specializing in Micro or Macro/Finance. Nevertheless, there are not very many factors that encourage women to select fields where they are underrepresented. Overall it seems that if women and men had more similar characteristics they would be less likely to select fields of specialization where they dominate (e.g. more in top schools, more equal salaries).

The next step for this research would be to examine gender differences in the career paths of economists. How do gender differences in field specialization evolve over time? What are the implications of this evolution for career advancement, promotion, and compensation?

References

Altonji, J., Arcidiacono, P. & Maurel, A. (2016), The analysis of field choice in college and graduate school, *in* 'Handbook of the Economics of Education', Elsevier, pp. 305–396.

Arcidiacono, P., Hotz, V. J. & Kang, S. (2012), 'Modeling college major choices using elicited measures of expectations and counterfactuals', *Journal of Econometrics* **166**(1), 3–16.

Auriol, E., Friebel, G. & Wilhelm, S. (2019), Women in European Economics.

Avilova, T. & Goldin, C. (2018), 'What can uwe do for economics?', *AEA Papers and Proceedings* **108**, 186–90.
URL: *http://www.aeaweb.org/articles?id=10.1257/pandp.20181103*

Bayer, A. & Rouse, C. E. (2016), 'Diversity in the economics profession: A new attack on an old problem', *Journal of Economic Perspectives* **30**(4), 221–42.
URL: *http://www.aeaweb.org/articles?id=10.1257/jep.30.4.221*

Bel, K., Fok, D. & Paap, R. (2018), 'Parameter estimation in multivariate logit models with many binary choices', *Econometric Reviews* **37**(5), 534–550.

Beneito, P., Boscá, J., Ferri, J. & García, M. (2018), Women across Subfields in Economics: Relative Performance and Beliefs, Technical report.
URL: *http://documentos.fedea.net/pubs/dt/2018/dt2018-06.pdf*

Bresnahan, T. F., Stern, S. & Trajtenberg, M. (1997), 'Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s', *The RAND Journal of Economics* **28**, S17.

Carrell, S. E., Page, M. E. & West, J. E. (2010), 'Sex and science: How professor gender perpetuates the gender gap', *The Quarterly Journal of Economics* **125**(3), 1101–1144.
URL: *http://dx.doi.org/10.1162/qjec.2010.125.3.1101*

Chari, A. & Goldsmith-Pinkham, P. (2017), Gender representation in economics across topics and time: Evidence from the nber summer institute, Working Paper 23953, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w23953*

Choe, C., Jung, S. & Oaxaca, R. (2020), 'Identification and decompositions in probit and logit models', *Empirical Economics* .

Dolado, J., Felgueroso, F. & Almunia, M. (2012), 'Are men and women-economists evenly distributed across research fields? Some new empirical evidence.', *SERIEs* **3**(3), 367–393.
URL: *http://dx.doi.org/10.1007/s13209-011-0065-4*

Dynan, K. & Rouse, C. E. (1997), 'The underrepresentation of women in economics: a study of undergraduate economics students.', *Journal of Economic Education* **28(4)**, 350–368.

Fortin, N., Lemieux, T. & Rehavi, M. (2021), 'Gender differences in fields of specialization and placement outcomes among phd in economics', *AEA Papers & Proceedings* **111**.

Ginther, D. K. & Kahn, S. (2004), 'Women in Economics: Moving Up or Falling Off the Academic Career Ladder?', *Journal of Economic Perspectives* **18**(3), 193–214.
URL: *http://www.aeaweb.org/articles.php?doi=10.1257/0895330042162386*

Hale, G. & Regev, T. (2014), 'Gender ratios at top PhD programs in economics', *Economics of Education Review* **41**, 55–70.
URL: *http://ideas.repec.org/a/eee/ecoedu/v41y2014icp55-70.html*

Lundberg, S., ed. (2020), *Women in Economics*, CEPR Press.

Lundberg, S. & Stearns, J. (2019), 'Women in economics: Stalled progress', *Journal of Economic Perspectives* **33**(1), 3–22.
URL: *http://www.aeaweb.org/articles?id=10.1257/jep.33.1.3*

Russell, G. J. & Petersen, A. (2000), 'Analysis of cross category dependence in market basket selection', *Journal of Retailing* **76**(3), 367–392.

Sierminska, E. & Oaxaca, R. (2021), 'Field specialization among begining economists: Are there gender differences?', *AEA Papers & Proceedings* **111**, 1–7.

Stock, W. & Siegfried, J. (2014), *Career Patterns of Economics Ph.D.s: A Decade of Outcomes for the Class of 1997*, Cambridge University Press.

Wiswall, M. & Zafar, B. (2015.), 'Determinants of college major choice: Identification using an information experiment', *The Review of Economic Studies* **82**(2), 791–824.

Zafar, B. (2013), 'College major choice and the gender gap', *Journal of Human Resources* **48**(3), 545–595.

Table 1: Gender Differences in Doctoral Field Specializations (%)

	Women	Men	Diff.	Std. Error	P-value
Econometrics (C)	4.45	7.56	-3.11***	0.582	0.000
Micro (D)	13.81	18.24	-4.43***	0.876	0.000
Labor/Health (I,J)	33.38	21.12	12.26***	1.000	0.000
Macro/Finance (E,G)	16.60	22.64	-6.04***	0.948	0.000
IO (L)	6.81	7.50	-0.69	0.608	0.257
Environ & Agric (Q)	10.21	9.67	0.54	0.696	0.434
Public (H)	3.52	4.04	-0.52	0.452	0.253
Dev/Growth/Int (O,F)	16.83	14.89	1.94**	0.845	0.021
Econ History (B,N)	1.01	1.34	-0.33	0.259	0.197
Other (P,A,K,M,R,Y,Z)	8.12	8.62	-0.49	0.651	0.450
Total	114.74	115.60			
Observations	2585	6268			

Note: *** statistically significant at the 1% level; ** statistically significant at the 5% level

Table 2: Gender Differences in Doctoral Dual Field Specialization by Primary Field (%)

	Women	Men	Diff.	Std. Error	P-value
Econometrics (C)	10.50	16.87	-6.37***	2.161	0.003
Micro (D)	31.76	41.72	-9.96***	2.945	0.001
Labor/Health (I,J)	32.28	22.70	9.58***	2.628	0.000
Macro/Finance (E,G)	29.40	33.03	-3.63	2.817	0.198
IO (L)	15.49	16.46	-0.98	2.224	0.661
Environ & Agric (Q)	16.80	10.43	6.37***	1.978	0.001
Public (H)	9.71	11.25	-1.54	1.876	0.413
Dev/Growth/Int (O,F)	32.02	25.66	6.36**	2.695	0.018
Econ History (B,N)	1.31	2.45	-1.14	0.873	0.191
Other (P,A,K,M,R,Y,Z)	20.73	19.43	1.31	2.406	0.589
Observations	381	978			

Note: *** statistically significant at the 1% level; ** statistically significant at the 5% level

Table 3: Summary statistics for women and men (case specific variables

	Women	Men	Diff.	Std. Error	P-value
Econ Top 20	0.37	0.42	-0.05***	0.011	0.000
Econ 20-40	0.20	0.21	-0.01	0.009	0.477
Econ NR	0.10	0.09	0.01	0.007	0.453
Uni Top 20	0.23	0.28	-0.05***	0.010	0.000
Uni 20-40	0.16	0.16	0.00	0.009	0.557
Uni NR	0.02	0.02	0.00	0.003	0.116
Year	2013.29	2013.42	-0.14**	0.065	0.034
Observations	2585	6268			

Note: Econ -refers to rankings of Economics departments; Uni - refers to ranking of universities; NR - not ranked
Year - Year of graduation

Table 4: Summary statistics for women and men (choice specific variables)

	Women	Men	Diff.	Std. Error	P-value
F share field 1	0.47x10-4	0.26x10-4	0.21x10-4	0.003	0.197
F share field 2	0.00	0.09x10-4	-0.09x10-4	0.001	0.364
F share field 3	0.24	0.24	0.00	0.029	0.882
F share field 4	0.60	0.66	-0.06	0.061	0.342
F share field 5	0.29	0.29	0.00	0.039	0.997
F share field 6	0.21	0.23	-0.02	0.025	0.442
F share field 7	0.46	0.50	-0.04	0.054	0.422
F share field 8	0.17	0.17	0.00	0.036	0.929
F share field 9	0.32	0.36	-0.04	0.070	0.611
F share field 10	0.73	0.58	0.15**	0.060	0.014
M share field 1	0.08	0.05	0.03	0.019	0.156
M share field 2	0.06	0.07	-0.01	0.025	0.637
M share field 3	1.34	1.38	-0.04	0.092	0.717
M share field 4	4.58	4.45	0.13	0.247	0.601
M share field 5	1.55	1.74	-0.19	0.123	0.125
M share field 6	0.75	0.74	0.01	0.086	0.840
M share field 7	1.95	1.98	-0.03	0.169	0.854
M share field 8	0.87	0.75	0.12*	0.061	0.053
M share field 9	0.47	0.53	-0.06	0.051	0.241
M share field 10	2.06	1.94	0.12	0.136	0.345
P(empl) field 1	0.10	0.15	-0.06***	0.002	0.000
P(empl) field 2	0.11	0.14	-0.04***	0.002	0.000
P(empl) field 3	0.09	0.12	-0.03***	0.002	0.000
P(empl) field 4	0.10	0.11	-0.01***	0.002	0.000
P(empl) field 5	0.09	0.11	-0.02***	0.002	0.000
P(empl) field 6	0.13	0.14	-0.01***	0.001	0.000
P(empl) field 7	0.05	0.13	-0.08***	0.002	0.000
P(empl) field 8	0.10	0.11	-0.01***	0.002	0.000
P(empl) field 9	0.40	0.13	0.27***	0.005	0.000
P(empl) field 10	0.08	0.09	-0.01***	0.001	0.000
Rel sal field 1 yr1	0.68	0.71	-0.03***	0.004	0.000
Rel sal field 2 yr1	0.53	1.28	-0.76***	0.003	0.000
Rel sal field 3 yr1	1.04	0.95	0.09***	0.002	0.000
Rel sal field 4 yr1	0.60	0.61	-0.01***	0.002	0.002
Rel sal field 5 yr1	0.62	2.74	-2.12***	0.008	0.000
Rel sal field 6 yr1	4.49	2.13	2.36***	0.013	0.000
Rel sal field 7 yr1	0.46	0.70	-0.24***	0.005	0.000
Rel sal field 8 yr1	1.09	0.37	0.72***	0.003	0.000
Rel sal field 9 yr1	0.10	0.22	-0.12***	0.001	0.000
Rel sal field 10 yr1	1.62	0.84	0.79***	0.012	0.000
Rel sal field 1 yr7	0.70	0.86	-0.16***	0.004	0.000
Rel sal field 2 yr7	0.68	1.26	-0.58***	0.003	0.000
Rel sal field 3 yr7	1.53	1.03	0.49***	0.003	0.000
Rel sal field 4 yr7	0.66	0.66	-0.00	0.002	0.395
Rel sal field 5 yr7	0.39	2.43	-2.04***	0.007	0.000
Rel sal field 6 yr7	4.03	2.41	1.62***	0.011	0.000
Rel sal field 7 yr7	0.51	0.68	-0.17***	0.005	0.000
Rel sal field 8 yr7	1.08	0.36	0.71***	0.003	0.000
Rel sal field 9 yr7	0.43	0.26	0.17***	0.002	0.000
Rel sal field 10 yr7	0.94	0.58	0.36***	0.007	0.000
Observations	2585	6268			

Note: Field (JEL) 1-Econometrics; 2-Micro; 3-Labor/Health; 4-Macro/Finance; 5-IO; 6-Environmental & Agriculture; 7-Public Economics; 8-Development/Growth/International; 9-Economic History; 10-Other.
F/M share field X - refers to share of female/male faculty in field X in the department.
P(empl) field X - probability of academic employment in field X (as listed above)
Rel sal field X yr 1/ Rel sal field X yr 7 - relative salary in field X (as above) 1 year and 7 years after graduation, respectively

Table 5: Marginal effects - multifield (doctoral sample)

	(1) Econ	(2) Micro	(3) Labor	(4) Macro	(5) IO	(6) Ag	(7) Public	(8) Dev	(9) Hist	(10) Oth
Female (d)	-0.042	-0.072	-0.341***	-0.042***	-0.459***	0.302***	-0.060**	-0.144***	0.304	0.051
	(0.074)	(0.101)	(0.004)	(0.009)	(0.147)	(0.041)	(0.029)	(0.031)	(0.442)	(0.036)
N	8,853	8,853	8,853	8,853	8,853	8,853	8,853	8,853	8,853	8,853

Note: Labor: Labor/Health; Macro: Macro/Finance; Ag: Agriculture & Environmental Economics;
Dev: Development/Growth/International
Marginal effects; Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
(d) for discrete change of dummy variable from 0 to 1
The association parameters for the following fields are included (column labels are used e.g.(1st field, 2nd field)): 12,13,14; 23,24,25,26,27,28,210; 34,35,36,37,38,310; 45,46,47,48,410; 56,58,510; 68,610; 78, 710; 810.

Table 6: Statistically significant marginal effects for women and men (doctoral sample)

		Individual (Case) Specific							Choice Specific				
		Econ Top 20	Econ Top 20-40	Econ Not Ranked	Univ Top 20	Univ Top 20-40	Univ Not Ranked	PhD Year	Faculty Share (W)	Faculty Share (M)	Prob Acad Empl	Rel Sal Year 1	Rel Sal Year 7
Econometrics	Women								-				
	Men		-		-	-		+	-			+	-
Micro	Women				+	+		-				-	+
	Men	-	-		+	+	-		-		+		
Labor/Health	Women	-	-	-	+		+	+			+	-	+
	Men	-	-	-	+		+				+	-	+
Macro	Women		-	+		+		-				-	+
	Men		-	-	-	-	+	+		+	+		-
IO	Women	-	-	+		+		-		-	+	-	+
	Men	+	+	+	+	+	+	+		-	+	+	
Ag/Environmental	Women	+				-	-	-	-		-		
	Men						-		-				
Public Economics	Women								-	-			
	Men		+	+	+	+	-						
Dev/Growth/Int	Women		+	+	-	-			+			-	+
	Men	+	+	+	-	-	-	+			+		
History	Women						-			-			
	Men		-	+	+		-	-		-	-	-	+
Other	Women				+				-			-	+
	Men	-	+	-	+		+				+	-	+

Note: Omitted categories include: Economics departments ranked below top 40; Universities ranked below top 40.
For continuous variables: marginal effects of a 1 standard deviation change are calculated; Variables are standardized to mean =0 and sd=1 except for the year variable. Full results in Table A5 and Table A6.

Table 7: Actual and counterfactual predicted probabilities for primary fields of specialization.

	Male	Female (male coeff)	Male (female coeff)	Female
Econometrics (C)	7.56	46.46	100.00	4.45
Micro (D)	18.24	0.00	0.00	13.81
Labor/ Health (I,J)	21.12	0.00	100.00	33.38
Macro/Finance (E, G)	22.64	13.04	41.45	16.60
IO (L)	7.50	100.00	0.00	6.81
Environ & Agric (Q)	9.67	38.22	99.82	10.21
Public (H)	4.04	10.10	7.77	3.52
Dev/Growth/Int (O, F)	14.89	0.00	0.00	16.83
Econ History (B,N)	1.34	0.00	0.00	1.01
Other (P,A,K,M,R,Y,Z)	8.62	19.57	0.00	8.12
	115.60	227.39	349.04	114.74

Table 8: Counterfactual dissimilarity measures: Duncan Dissimilarity Index (DDI) and Quadratic Field Disparity Measure (QDM).

		characteristics	parameters		
DDI= 27.09	DDI_mf_c	92.92	99.65	DDI_mf_p	Male baseline
	DDI_fm_c	98.72	95.42	DDI_fm_p	Female baseline
		characteristics	parameters	interaction	
QDM = 2.64		18.98	25.64	-41.93	Male baseline
		42.54	*57.46*		*%*
		23.62	18.31	-39.28	Female baseline
		56.33	*43.67*		*%*

Note: Raw DDI = 15.27; Raw QDM = 0.82; DDI_mf_c: if women faced the same parameters as men DDI_mf_p: if men possessed the same characteristics as women.

DDI_fm_c: if men faced the same parameters as women DDI_fm_p: if women possessed the same characteristics as men.

Table 9: Predictive accuracy rates (%).

	Women	Men
Econometrics	95.55	90.86
Micro	89.86	98.55
Labor/Health	79.92	86.73
Macro/Finance	83.40	88.40
IO	93.19	92.20
Ag/Environmental	89.75	90.33
Public Economics	96.48	95.69
Dev/Growth/Int	85.03	91.53
History	98.99	98.66
Other	91.33	91.18
Overall for the Primary Fields	90.35	92.41
Overall for Multifield (55) specializations	88.52	87.52
N	2585	6268

Table A1: Fields defined by JEL codes (10 fields)

Fields	JEL codes	Detailed JEL codes
Econometrics	C	C. Mathematical and Quantitative Methods
Micro	D	D. Microeconomics
Labor	I, J	I. Health, Education and Welfare, J. Labor and Demographic Economics
Macro/Finance	E, G	E. Macroeconomics and Monetary Economics; G.Financial Economics;
IO	L	L. Industrial Organization
Environmental & Agricultural	Q	Q. Agricultural & Natural Resource Economics, Environmental Econ
Public	H	H. Public Economics
Development/Growth/International	F, O	O. Economic Development, Innovation, Technological Change, and Growth; F. International Economics
Economic History	B, N	B. History of Economic Thought, Methodology, and Heterodox Approaches; N. Economic History
Other	P,A,K,M,R,Y,Z	P. Economic Systems; A. General Econ and Teaching; Z. Other Topics; ; K. Law and Economics; R. Urban, Reg, Real Estate & Transportation Economics

Table A2: Distribution of multifields across the doctoral dissertation sample (2009-2018)

Males	Econ	Micro	Labor	Macro	IO	Env	Public	Dev	Hist	Other
Econ	309									
Micro	91	826								
Labor	11	39	1152							
Macro	38	98	32	1264						
IO	3	52	19	14	397					
Env	3	24	4	9	28	572				
Public	2	24	26	22	2	1	220			
Dev/Gr/Int	10	29	48	78	21	26	15	909		
Hist	0	5	3	6	0	1	1	3	79	
Other	7	46	40	26	22	6	17	21	5	540
Column Sum	474	1143	1324	1419	470	606	253	933	84	540
Total Obs		6268								
Total Fields Reported		7246								

Females	Econ	Micro	Labor	Macro	IO	Env	Public	Dev	Hist	Other
Econ	75									
Micro	14	250								
Labor	4	20	764							
Macro	16	30	12	375						
IO	2	12	1	6	138					
Env	1	7	8	9	14	239				
Public	1	4	20	3	2	2	86			
Dev/Gr/Int	1	18	36	26	7	16	5	422		
Hist	0	1	1	1	1	0	0	1	26	
Other	1	15	21	9	14	7	0	12	0	210
Column Sum	115	357	863	429	176	264	91	435	26	210
Total Obs		2585								
Total Fields Reported		2966								

Note: 1359 out of 8853 doctoral dissertations in our sample report 2 fields (15%). For now we have excluded those reporting three or more (981 - 9.5% of original sample)

Table A3: Distribution of primary and multi - fields across the doctoral dissertation sample (shares out of total fields reported, 2009-2018)

Males											
	Econ	Micro	Labor	Macro	IO	Env	Public	Dev	Hist	Other	
Econ	4.26										
Micro	1.26	11.40									
Labor	0.15	0.54	15.90								
Macro	0.52	1.35	0.44	17.44							
IO	0.04	0.72	0.26	0.19	5.48						
Env	0.04	0.33	0.06	0.12	0.39	7.89					
Public	0.03	0.33	0.36	0.30	0.03	0.01	3.04				
Dev/Gr/Int	0.14	0.40	0.66	1.08	0.29	0.36	0.21	12.54			
Hist	0.00	0.07	0.04	0.08	0.00	0.01	0.01	0.04	1.09		
Other	0.10	0.63	0.55	0.36	0.30	0.08	0.23	0.29	0.07	7.45	Total
Column sum (%)	6.54	15.77	18.27	19.58	6.49	8.36	3.49	12.88	1.16	7.45	100
Total Fields Reported		7246									

Females											
	Econ	Micro	Labor	Macro	IO	Env	Public	Dev	Hist	Other	
Econ	2.53										
Micro	0.47	8.43									
Labor	0.13	0.67	25.76								
Macro	0.54	1.01	0.40	12.64							
IO	0.07	0.40	0.03	0.20	4.65						
Env	0.03	0.24	0.27	0.30	0.47	8.06					
Public	0.03	0.13	0.67	0.10	0.07	0.07	2.90				
Dev/Gr/Int	0.03	0.61	1.21	0.88	0.24	0.54	0.17	14.23			
Hist	0.00	0.03	0.03	0.03	0.03	0.00	0.00	0.03	0.88		
Other	0.03	0.51	0.71	0.30	0.47	0.24	0.00	0.40	0.00	7.08	Total
Column sum (%)	3.88	12.04	29.10	14.46	5.93	8.90	3.07	14.67	0.88	7.08	100
Total Fields Reported		2966									

Note: 1359 out of 8853 doctoral dissertations in our sample report 2 fields (15%). For now we have excluded those reporting three or more (981 - 9.5% of original sample)

Table A4: Difference in specialization among women and men according to 20 JEL categories (doctoral sample, 2009-2018).

	Women	Men	Diff.	Std. Error	P-value
General Economics and Teaching (A)	0.12	0.18	-0.06	0.093	0.522
History of Ec Thought, Method. & Heter. App. (B)	0.04	0.18	-0.14	0.086	0.112
Mathematical and Quant Methods (C)	4.68	7.56	-2.88***	0.585	0.000
Microeconomics (D)	13.27	16.19	-2.92***	0.842	0.001
Macro and Monetary Econ (E)	6.23	10.02	-3.79***	0.665	0.000
International Economics (F)	5.22	5.47	-0.25	0.528	0.636
Financial Economics (G)	7.47	8.76	-1.29**	0.648	0.046
Public Economics (H)	3.48	3.32	0.16	0.422	0.699
Health, Education and Welfare (I)	11.76	7.32	4.44***	0.654	0.000
Labor and Demog Economics (J)	15.63	10.08	5.55***	0.749	0.000
Law and Economics (K)	1.32	1.56	-0.25	0.283	0.381
Industrial Organization (L)	5.53	6.48	-0.95*	0.564	0.094
Business Adm, Marketing, Personnel Econ (M)	0.97	1.07	-0.10	0.237	0.668
Econ History (N)	0.74	0.54	0.19	0.180	0.286
Ec Develop, Innovation, Tech. Change & Growth (O)	8.01	6.13	1.88***	0.583	0.001
Economic Systems (P)	1.59	1.39	0.20	0.279	0.478
Agric & Nat Resource Economics, Envir. Ec. (Q)	8.24	7.59	0.65	0.626	0.303
Urban, Reg, Real Estate & Transp Econ (R)	4.26	4.12	0.14	0.467	0.766
Misc. Categories (Y)	0.66	0.83	-0.17	0.206	0.403
Other: Cultural, Sociol & Anthrop Econ. (Z)	0.81	1.21	-0.40	0.243	0.100
Observations	2585	6268		DDI=13.2%	

Note: Total sample: n = 8853; DDI – Duncan Dissimilarity Index

Table A5: Individual (case) specific marginal effects - multifield -women and men Part 1 (doctoral sample)

	Econometrics			Microeconomics			Labor/Health			Macro/Finance			IO		
Women															
	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Econ Top 20	0.016	0.049		-0.070	0.151		-0.339	0.082	***	-0.089	0.064		-1.860	0.572	***
Econ 20-40	0.008	0.015		-0.029	0.044		-0.098	0.032	***	-0.042	0.025	*	-0.131	0.041	***
Econ NR	-0.023	0.050		-0.027	0.085		-0.104	0.031	***	0.073	0.026	***	0.067	0.027	**
Uni Top 20	0.014	0.027		0.089	0.013	***	0.106	0.015	***	0.052	0.050		0.016	0.018	
Uni 20-40	-0.004	0.014		0.069	0.010	***	-0.007	0.011		0.039	0.014	***	0.041	0.015	***
Uni NR	-0.002	0.005		-0.002	0.010		0.030	0.010	***	0.011	0.010		0.003	0.007	
Year	-0.004	0.009		-0.036	0.006	***	0.019	0.007	***	-0.022	0.009	***	-0.022	0.007	***
Men	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Econ Top 20	0.041	0.044		-0.310	0.013	***	-0.105	0.043	**	-0.044	0.046		0.216	0.051	***
Econ 20-40	-0.055	0.026	**	-0.071	0.012	***	-0.438	0.088	***	-0.272	0.036	***	1.056	0.113	***
Econ NR	-0.005	0.017		-0.003	0.004		-0.432	0.050	***	-0.321	0.017	***	0.064	0.007	***
Uni Top 20	-0.061	0.013	***	0.216	0.006	***	0.104	0.014	***	-0.894	0.039	***	0.194	0.016	***
Uni 20-40	-0.250	0.025	***	0.011	0.007	*	-0.020	0.028		-0.243	0.019	***	0.336	0.045	***
Uni NR	0.005	0.008		-0.076	0.004	***	0.301	0.038	***	0.100	0.011	***	0.095	0.018	***
Year	0.038	0.008	***	-0.002	0.004		-0.009	0.020		0.082	0.012	***	0.066	0.012	***

	Agric			Public			Development			History			Other		
Women															
	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Econ Top 20	0.130	0.046	***	0.014	0.049		0.051	0.047		0.002	0.008		0.031	0.072	
Econ 20-40	0.013	0.015		0.001	0.015		0.099	0.024	***	-0.006	0.006		0.019	0.025	
Econ NR	-0.005	0.014		-0.005	0.012		0.096	0.016	***	0.001	0.003		-0.001	0.006	
Uni Top 20	-0.040	0.027		-0.012	0.025		-0.053	0.027	**	-0.006	0.008		0.071	0.013	***
Uni 20-40	-0.055	0.034	*	0.001	0.004		-0.066	0.016	***	0.005	0.010		0.001	0.010	
Uni NR	-0.186	0.020	***	0.000	0.004		0.002	0.004		-0.023	0.007	***	-0.001	0.005	
Year	-0.027	0.007	***	-0.009	0.008		-0.003	0.005		-0.002	0.003		0.003	0.004	
Men	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Econ Top 20	0.054	0.079		0.003	0.033		0.049	0.023	**	0.006	0.006		-0.186	0.036	***
Econ 20-40	0.027	0.027		0.142	0.036	***	0.042	0.024	*	-0.023	0.010	**	0.172	0.053	***
Econ NR	0.026	0.026		0.030	0.008	***	0.199	0.012	***	0.006	0.003	*	-0.090	0.018	***
Uni Top 20	-0.163	0.168		0.103	0.016	***	-0.385	0.026	***	0.007	0.002	***	0.116	0.017	***
Uni 20-40	-0.107	0.102		0.140	0.024	***	-0.506	0.024	***	-0.003	0.003		-0.004	0.040	
Uni NR	-0.032	0.012	***	-0.009	0.004	***	-0.016	0.006	**	-0.029	0.007	***	0.069	0.028	**
Year	0.032	0.022		-0.010	0.007		0.082	0.009	***	-0.010	0.004	***	0.019	0.031	

Note: Marginal effects of a 1 standard deviation change; Variables are standardized to mean =0 and sd=1 except for the year variable.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Sample size: women=2585; Men= 6268

Table A6: Field (alternative) specific marginal effects - multifield -women and men Part 2 (doctoral sample)

	Econometrics			Microeconomics			Labor/Health			Macro/Finance			IO		
Women	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Share of W	-0.018	0.003	***				0.005	0.009		0.006	0.007		0.004	0.006	
Share of M	0.000	0.002		-0.001	0.004		-0.011	0.014		0.006	0.013		-0.007	0.004	*
Prob(empl)	-0.044	0.046		0.093	0.140		0.157	0.082	**	0.038	0.033		1.731	0.548	***
Rel Sal Yr 1	-0.321	0.289		-0.478	0.044	***	-0.390	0.061	***	-1.522	0.204	***	-0.689	0.086	***
Rel Sal Yr 7	0.298	0.326		0.580	0.029	***	0.568	0.049	***	1.534	0.235	***	0.825	0.113	***
Men	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Share of W	-0.015	0.002	***	-0.010	0.001	***	0.004	0.004		0.008	0.009		-0.005	0.007	
Share of M	0.000	0.002		-0.007	0.005		-0.007	0.009		0.012	0.007	*	-0.005	0.003	**
Prob(empl)	-0.016	0.037		0.031	0.008	***	0.099	0.033	***	0.090	0.046	**	0.066	0.033	**
Rel Sal Yr 1	2.345	1.223	*	-0.031	0.320		-10.818	1.376	***	1.336	1.160		2.401	1.355	*
Rel Sal Yr 7	-2.632	1.222	**	0.284	0.321		11.393	1.360	***	-2.243	1.179	**	-1.297	1.285	

	Agric			Public			Development			History			Other		
Women	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Share of W	-0.032	0.016	**	-0.265	0.037	***	0.010	0.002	***	0.001	0.001		-0.005	0.003	*
Share of M	0.001	0.007		-0.053	0.025	**	0.002	0.004		-0.087	0.028	***	-0.001	0.005	
Prob(empl)	-0.115	0.029	***	-0.010	0.018		0.045	0.034		-0.002	0.007		-0.086	0.064	
Rel Sal Yr 1	0.102	0.262		-0.455	0.320		-0.802	0.056	***	0.021	0.062		-0.607	0.358	*
Rel Sal Yr 7	-0.171	0.274		0.451	0.354		0.922	0.054	***	-0.032	0.074		0.682	0.358	*
Men	dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se		dy/dx	se	
Share of W	-0.029	0.012	***	0.001	0.006		0.000	0.002		0.001	0.001		-0.004	0.005	
Share of M	0.000	0.008		-0.006	0.007		0.006	0.004		-0.004	0.002	***	0.001	0.003	
Prob(empl)	-0.007	0.030		-0.011	0.028		0.033	0.017	**	-0.010	0.005	**	0.071	0.022	***
Rel Sal Yr 1	-0.227	1.207		2.505	2.222		0.112	0.533		-0.856	0.324	***	-11.207	1.490	***
Rel Sal Yr 7	0.374	1.091		-2.729	2.221		-0.758	0.534		0.887	0.324	***	11.185	1.534	***

Note: Marginal effects of a 1 standard deviation change; Variables are standardized to mean =0 and sd=1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Sample size: women=2585; Men= 6268

Table A7: Difference in specialization among women and men (salary sample, multifield)

	Women	Men	Diff.	Std. Error	P-value
Econometrics (C)	5.95	8.15	-2.20***	0.784	0.005
Micro (D)	11.69	17.74	-6.06***	1.090	0.000
Labor/Health (I,J)	30.50	16.96	13.54***	1.159	0.000
Macro/Finance (E,G)	11.76	16.93	-5.17***	1.074	0.000
IO (L)	4.98	8.46	-3.48***	0.786	0.000
Environ & Agric (Q)	20.06	17.43	2.62**	1.129	0.020
Public (H)	3.32	4.34	-1.03*	0.586	0.080
Dev/Growth/Int (O,F)	11.96	9.70	2.27**	0.888	0.011
Econ History (B,N)	2.56	3.44	-0.88*	0.523	0.091
Other (P,A,K,M,R,Y,Z)	7.26	8.74	-1.48*	0.818	0.070
Observations	7200		DDI=17.4%		

Note: DDI – Duncan Dissimilarity Index

Table A8: Difference in specialization among women and men (salary sample, primary field)

	Women	Men	Diff.	Std. Error	P-value
Econometrics (C)	4.77	6.41	-1.64**	0.703	0.020
Micro (D)	10.93	16.61	-5.69***	1.062	0.000
Labor/Health (I,J)	30.50	16.96	13.54***	1.159	0.000
Macro/Finance (E,G)	11.20	15.99	-4.79***	1.050	0.000
IO (L)	3.87	6.45	-2.57***	0.694	0.000
Environ & Agric (Q)	19.71	17.31	2.40**	1.125	0.033
Public (H)	1.45	3.32	-1.87***	0.497	0.000
Dev/Growth/Int (O,F)	9.54	8.36	1.18	0.825	0.151
Econ History (B,N)	2.28	2.03	0.25	0.420	0.554
Other (P,A,K,M,R,Y,Z)	5.74	6.55	-0.81	0.719	0.259
Observations	7200		DDI=17.4%		

Note: DDI – Duncan Dissimilarity Index

Table A9: Salary regression - multifield-women and men Part 1 (salary sample)

	(1) Econ-M	(2) Econ-F	(3) Micro-M	(4) Micro-F	(5) Lbr-M	(6) Lbr-F	(7) Macro-M	(8) Macro-F	(9) IO-M	(10) IO-F
ysphd	0.0440***	0.0840*	0.0187***	0.0560***	0.0161**	0.0621**	0.0159***	0.00816	0.00967	0.0162
	(4.92)	(2.25)	(4.22)	(4.65)	(2.68)	(2.73)	(3.93)	(1.62)	(0.97)	(0.71)
ysphd_sqr	-0.000561**	-0.00197	-0.000179**	-0.00138**	-0.000178	-0.00158*	-0.000144***	-0.0000817*	-0.000102	-0.00100
	(-3.00)	(-1.66)	(-3.21)	(-3.44)	(-1.61)	(-2.20)	(-3.36)	(-2.15)	(-0.97)	(-1.28)
PhD Ec T 20	0.00211	-0.367	0.182	-0.302*	-0.00392	-0.0378	-0.0163	-0.563	-0.0799	0.0120
	(0.01)	(-0.96)	(1.56)	(-2.53)	(-0.05)	(-0.23)	(-0.13)	(-1.52)	(-0.45)	(0.15)
PhD Ec 20-40	-0.161	-0.236*	-0.0376	-0.128	0.0727	-0.101	-0.169	-0.0355	-0.401*	-0.0466
	(-1.38)	(-2.12)	(-0.33)	(-1.62)	(0.87)	(-0.49)	(-1.48)	(-0.25)	(-2.30)	(-0.20)
PhD Ec NR	-0.101	-1.152***	-0.0941	-0.221	0.0448	0.0586	-0.231	-0.480*	-0.117	-0.940***
	(-0.72)	(-4.85)	(-0.47)	(-1.86)	(0.44)	(0.29)	(-1.39)	(-2.29)	(-0.40)	(-8.72)
PhD Uni T 20	-0.0670	-0.0727	-0.156	0.133	0.0105	0.0736	-0.245	0.547	-0.0375	0.262**
	(-0.78)	(-0.62)	(-1.52)	(1.28)	(0.12)	(0.64)	(-1.80)	(1.49)	(-0.20)	(2.94)
PhD U 20-40	-0.392*	0.0131	0.0662	0.0631	0.0384	0.172	-0.00880	0.150	-0.0157	-0.0883
	(-2.16)	(0.08)	(0.61)	(0.29)	(0.42)	(1.61)	(-0.09)	(1.03)	(-0.08)	(-1.26)
PhD Uni NR	0.0187	0	0.250	0.0815	-0.304	0.0305	0.0883	0.295	-0.0149	0
	(0.12)	(.)	(1.56)	(0.69)	(-1.40)	(0.19)	(0.47)	(1.71)	(-0.04)	(.)
Econ Top 20	0.500**	0.154	0.371*	0.339***	0.513***	0.375	0.426***	0.697***	-0.156	0.472***
	(2.96)	(0.34)	(2.42)	(3.71)	(4.33)	(1.80)	(3.42)	(4.29)	(-0.27)	(11.93)
Ec 20-40	0.153	0.0388	0.0657	0.188**	0.283***	0.222	0.0833	0.203	-0.310	1.878***
	(1.28)	(0.19)	(0.63)	(2.81)	(3.58)	(1.80)	(0.75)	(1.54)	(-0.91)	(8.49)
Econ NR	-0.137	-1.232	-0.116	-0.191*	-0.233*	-0.191	-0.372**	0.217	-0.192	0.345
	(-1.04)	(-2.09)	(-0.82)	(-2.10)	(-2.51)	(-1.65)	(-2.63)	(1.33)	(-0.97)	(1.33)
Uni T 20	0.00707	-0.0334	0.196	0	0.268*	0.356	0.289*	0	0.262	0
	(0.08)	(-0.08)	(1.55)	(.)	(2.29)	(1.35)	(2.02)	(.)	(0.73)	(.)
Uni 20-40	0.0520	-0.928***	0.175	-0.179*	-0.0241	0.131	0.139	0.126	0.366	-1.492***
	(0.69)	(-3.96)	(1.94)	(-2.31)	(-0.26)	(0.75)	(1.03)	(0.85)	(1.52)	(-7.63)
Uni NR	-0.490*	0.164	-0.434	-0.00158	-0.129	-0.138	0.140	0.0263	-0.418*	-0.544***
	(-2.15)	(0.82)	(-1.81)	(-0.02)	(-1.04)	(-0.76)	(0.84)	(0.11)	(-2.02)	(-5.68)
Year	0.0205	0.0427	-0.00624	0.0281	0.00559	0.00877	0.00206	0.0306	-0.0173	0.0565**
	(0.80)	(1.62)	(-0.36)	(1.85)	(0.36)	(0.59)	(0.14)	(1.97)	(-0.82)	(3.27)
Dual fields	-0.225**	-0.0936	-0.0503	0.199*	-0.143	0.260*	-0.205*	0.0707	-0.138	0.137*
	(-3.00)	(-0.45)	(-0.50)	(2.23)	(-1.50)	(2.27)	(-2.16)	(0.44)	(-0.80)	(2.25)
Base sal	-0.0511	-0.0557	-0.171	0.0445	-0.154*	-0.361*	-0.122	0.271	0.120	-0.380*
	(-0.26)	(-0.15)	(-1.25)	(0.52)	(-2.00)	(-2.35)	(-1.02)	(1.28)	(0.35)	(-2.09)
Net sal	-0.196	-1.157	-0.415	-0.0353	-0.348	-0.461	-0.0359	-0.0198	-0.532	0.141
	(-0.61)	(-1.28)	(-1.85)	(-0.28)	(-1.96)	(-1.71)	(-0.23)	(-0.13)	(-1.80)	(0.90)
Annual sal	-0.302	-0.694	-0.282	-0.204*	-0.287	-0.431	-0.0674	0.0915	-0.786	-0.0199
	(-1.74)	(-0.91)	(-1.58)	(-2.21)	(-1.93)	(-1.67)	(-0.43)	(1.05)	(-1.57)	(-0.12)
Gross sal	-0.164	-0.163	-0.159	-0.238*	-0.0187	-0.134	-0.00708	-0.0969	-0.628	0.0230
	(-1.83)	(-0.56)	(-1.75)	(-2.28)	(-0.19)	(-1.02)	(-0.06)	(-0.65)	(-1.23)	(0.38)
Sal Not Id	0.154	-1.064	-0.233	-0.166	-0.124	-0.243	-0.128	-0.129	-0.317	-2.094***
	(1.25)	(-1.34)	(-1.57)	(-1.09)	(-0.76)	(-1.06)	(-0.69)	(-0.72)	(-1.00)	(-6.17)
Constant	-29.59	-73.30	24.35	-44.98	0.547	-6.054	7.736	-50.29	47.45	-102.4**
	(-0.57)	(-1.39)	(0.70)	(-1.47)	(0.02)	(-0.20)	(0.26)	(-1.61)	(1.12)	(-2.93)
N	469	86	1,021	169	976	441	974	170	481	72
Pseudo R-sq										

Note: t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A10: Salary regression - multifield-women and men Part 2 (salary sample)

	(1) Ag-M	(2) Ag-F	(3) Public-M	(4) Public-F	(5) Dev-M	(6) Dev-F	(7) Hist-M	(8) Hist-F	(9) Oth-M	(10) Oth-F
ysphd	0.0191*	-0.0229	0.0704***	0.00458	0.0430***	0.00618	0.0113	0.221**	0.0241	0.00901
	(2.33)	(-1.16)	(4.46)	(0.35)	(4.47)	(0.19)	(0.91)	(3.46)	(1.48)	(0.50)
ysphd_sqr	-0.000247	0.00114*	-0.00137***	0.000332	-0.000651**	0.0000514	-0.000219	-0.00415**	-0.000520	0.000152
	(-1.40)	(2.19)	(-3.80)	(1.13)	(-3.17)	(0.06)	(-0.85)	(-3.28)	(-1.43)	(0.36)
PhD Ec T 20	-0.104	-0.0452	-0.0215	-0.728***	-0.0146	-0.548	-0.0432	0	-0.116	0.412
	(-1.01)	(-0.19)	(-0.09)	(-24.67)	(-0.12)	(-1.12)	(-0.21)	(.)	(-0.88)	(1.61)
PhD Ec 20-40	-0.131	-0.0625	0.415*	-0.291***	-0.0915	-0.436	0.146	0	-0.185	0.00192
	(-1.21)	(-0.25)	(2.19)	(-9.67)	(-0.66)	(-0.91)	(0.68)	(.)	(-1.11)	(0.01)
PhD Ec NR	-0.151	-0.0322	0.102	0	-0.0390	-0.453	-0.240	0	-0.154	-0.0349
	(-1.19)	(-0.14)	(0.36)	(.)	(-0.47)	(-1.08)	(-1.18)	(.)	(-0.84)	(-0.22)
PhD Uni T 20	0.180	0.148	0.329	0.543***	-0.0753	0.448*	-0.0158	-0.162	0.0146	-0.743**
	(1.62)	(0.84)	(1.90)	(43.59)	(-0.61)	(2.37)	(-0.09)	(-1.21)	(0.10)	(-3.15)
PhD U 20-40	0.236	-0.146	0.493	0.192***	-0.0690	0.394	0.0475	0.544	0.171	0.379**
	(1.95)	(-0.79)	(0.89)	(7.19)	(-0.61)	(1.44)	(0.25)	(2.25)	(1.59)	(3.13)
PhD Uni NR	0.0463	0.216	-0.132	0	0.0145	0.133	0.138	0	0.209	0
	(0.27)	(0.61)	(-0.41)	(.)	(0.10)	(0.66)	(1.09)	(.)	(1.17)	(.)
Econ Top 20	0.441**	0.399	0.591***	0	0.432***	0.412	0.164	0	0.544**	0.772***
	(2.84)	(1.81)	(4.35)	(.)	(3.63)	(1.38)	(0.88)	(.)	(3.19)	(4.37)
Ec 20-40	0.318**	0.342	-0.0114	0	0.0385	0.186	0.374	0	0.107	0.246
	(3.06)	(1.38)	(-0.07)	(.)	(0.29)	(1.10)	(1.86)	(.)	(0.42)	(1.05)
Econ NR	0.0953	0.106	-0.0644	-0.306***	-0.204	-0.0847	0.260	0.384*	-0.0680	0.0966
	(0.89)	(0.93)	(-0.36)	(-20.02)	(-1.60)	(-0.59)	(1.11)	(3.00)	(-0.55)	(0.50)
Uni T 20	0.154	0	0	0	0.243	-0.179	0	0	-0.223	0
	(0.86)	(.)	(.)	(.)	(1.91)	(-0.47)	(.)	(.)	(-1.09)	(.)
Uni 20-40	-0.0253	0.0672	0.0403	0	0.162*	0.0420	0.205	0	0.379**	0.102
	(-0.22)	(0.28)	(0.30)	(.)	(2.05)	(0.23)	(0.92)	(.)	(2.74)	(0.46)
Uni NR	0.215	0.0190	0.236	0	-0.344*	-0.246	-0.669	-0.591***	-0.199	0
	(1.00)	(0.07)	(0.93)	(.)	(-2.35)	(-0.89)	(-1.95)	(-4.89)	(-1.01)	(.)
Year	-0.0276	0.00327	-0.00868	0.0190**	-0.000829	0.00221	0.0457**	-0.0340**	-0.0331	0.0271*
	(-0.93)	(0.32)	(-0.49)	(3.31)	(-0.05)	(0.08)	(3.39)	(-3.93)	(-1.80)	(2.25)
Dual fields	-0.146	-0.102	-0.280*	-0.0388	-0.182	0.255	-0.138	0	-0.169	0.807***
	(-1.88)	(-0.77)	(-2.18)	(-1.75)	(-1.93)	(1.98)	(-0.86)	(.)	(-1.73)	(4.57)
Base sal	-0.00931	0.152	-0.350**	-0.118*	-0.301*	-0.474	-0.202	0.709***	-0.277*	-0.209
	(-0.08)	(0.95)	(-2.74)	(-2.76)	(-2.14)	(-1.45)	(-1.16)	(13.08)	(-2.05)	(-1.10)
Net sal	-0.599	-0.439	-1.145***	-0.145***	-0.321	-0.467	0.0506	0	-1.017*	-2.062***
	(-1.41)	(-1.24)	(-3.54)	(-11.72)	(-1.28)	(-0.88)	(0.22)	(.)	(-2.59)	(-3.78)
Annual sal	-0.178	-0.0990	-1.192***	0.293**	-0.375	-0.486	0.100	-0.0622	-0.577	-1.212***
	(-1.12)	(-0.63)	(-3.92)	(4.32)	(-1.67)	(-1.28)	(0.65)	(-0.63)	(-1.69)	(-4.25)
Gross sal	-0.0744	0.144	-0.662*	0	-0.0503	-0.0414	-0.0156	1.030**	-0.483	-1.251*
	(-0.53)	(0.80)	(-2.28)	(.)	(-0.31)	(-0.16)	(-0.11)	(4.55)	(-1.71)	(-2.21)
Sal Not Id	-0.0301	0.321	-0.445	0	0.0399	-0.265	-0.624**	0	-0.503	-1.274*
	(-0.15)	(1.26)	(-1.32)	(.)	(0.12)	(-0.60)	(-3.07)	(.)	(-1.47)	(-2.71)
Constant	67.17	4.790	29.25	-26.40*	13.32	7.640	-80.34**	76.79**	78.90*	-42.12
	(1.12)	(0.23)	(0.83)	(-2.27)	(0.40)	(0.14)	(-2.96)	(4.22)	(2.12)	(-1.73)
N	1,003	290	250	48	557	173	198	37	503	105
Pseudo R-sq										

Note: t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A11: Salary regression - primary field-women and men Part 1 (salary sample)

	(1) Econ-M	(2) Econ-F	(3) Micro-M	(4) Micro-F	(5) Lbr-M	(6) Lbr-F	(7) Macro-M	(8) Macro-F	(9) IO-M	(10) IO-F
ysphd	0.0546***	0.0685	0.0199***	0.0587***	0.0161**	0.0621**	0.0160***	0.00728	0.00408	0.0431
	(5.07)	(0.98)	(4.54)	(4.85)	(2.68)	(2.73)	(3.78)	(1.33)	(0.34)	(0.83)
ysphd_sqr	-0.000771***	-0.00166	-0.000182**	-0.00147***	-0.000178	-0.00158*	-0.000143**	-0.0000733	-0.0000783	-0.00149
	(-3.59)	(-0.94)	(-3.32)	(-3.66)	(-1.61)	(-2.20)	(-3.23)	(-1.75)	(-0.61)	(-0.88)
PhD Ec T 20	-0.0252	0.430	0.171	-0.310*	-0.00392	-0.0378	-0.0260	-0.507	0.0174	-0.112
	(-0.15)	(0.69)	(1.40)	(-2.50)	(-0.05)	(-0.23)	(-0.21)	(-1.30)	(0.09)	(-0.98)
PhD Ec 20-40	-0.140	0.371***	-0.0585	-0.0834	0.0727	-0.101	-0.166	-0.0154	-0.313	-0.509
	(-1.19)	(5.50)	(-0.46)	(-0.97)	(0.87)	(-0.49)	(-1.40)	(-0.11)	(-1.56)	(-1.34)
PhD Ec NR	0.0485	0	-0.115	0.0776	0.0448	0.0586	-0.204	-0.536*	-0.509	0
	(0.39)	(.)	(-0.53)	(0.78)	(0.44)	(0.29)	(-1.17)	(-2.42)	(-1.91)	(.)
PhD Uni T 20	-0.0206	0.480*	-0.183	0.125	0.0105	0.0736	-0.231	0.443	0.0120	0.292
	(-0.21)	(2.66)	(-1.75)	(1.10)	(0.12)	(0.64)	(-1.68)	(1.10)	(0.06)	(1.14)
PhD U 20-40	-0.375	0.157	0.0474	0.0711	0.0384	0.172	-0.0208	0.115	0.0347	0.146
	(-1.55)	(0.80)	(0.42)	(0.31)	(0.42)	(1.61)	(-0.21)	(0.78)	(0.18)	(0.70)
PhD Uni NR	-0.106	0	0.228	-0.203*	-0.304	0.0305	0.0397	0.322	0.656	0
	(-0.57)	(.)	(1.37)	(-2.57)	(-1.40)	(0.19)	(0.19)	(1.71)	(1.66)	(.)
Econ Top 20	0.485*	0.192	0.412*	0.372***	0.513***	0.375	0.442**	0.658***	-0.488	0.380***
	(2.48)	(0.42)	(2.60)	(3.86)	(4.33)	(1.80)	(2.97)	(4.20)	(-0.66)	(5.86)
Ec 20-40	0.147	0.157	0.0793	0.196**	0.283***	0.222	0.0610	0.142	-0.273	1.972*
	(1.13)	(1.10)	(0.72)	(3.22)	(3.58)	(1.80)	(0.53)	(0.95)	(-0.78)	(2.17)
Econ NR	-0.150	0.985	-0.114	-0.201*	-0.233*	-0.191	-0.367*	0.213	0.0532	-0.0918
	(-0.82)	(1.19)	(-0.77)	(-2.23)	(-2.51)	(-1.65)	(-2.55)	(1.31)	(0.32)	(-0.14)
Uni T 20	0.0802	0.398	0.206	0	0.268*	0.356	0.259	0	0.529	0
	(1.19)	(0.75)	(1.63)	(.)	(2.29)	(1.35)	(1.53)	(.)	(1.01)	(.)
Uni 20-40	0.0567	-0.543	0.202*	-0.157	-0.0241	0.131	0.121	0.171	0.501	-1.636*
	(0.59)	(-0.81)	(2.13)	(-2.00)	(-0.26)	(0.75)	(0.75)	(1.26)	(1.72)	(-2.77)
Uni NR	-0.494	-2.280	-0.425	0.0397	-0.129	-0.138	0.128	0.0493	-0.631*	-0.307
	(-1.86)	(-1.22)	(-1.74)	(0.35)	(-1.04)	(-0.76)	(0.75)	(0.21)	(-2.56)	(-0.86)
Year	0.0203	0.0579	-0.00857	0.0295	0.00559	0.00877	0.000215	0.0344*	0.00987	0.0439
	(0.65)	(1.41)	(-0.48)	(1.84)	(0.36)	(0.59)	(0.01)	(2.02)	(0.55)	(1.68)
Dual fields	0	0	-0.0916	0.228*	-0.143	0.260*	-0.247	-0.0647	-0.129	0.00367
	(.)	(.)	(-0.80)	(2.56)	(-1.50)	(2.27)	(-1.86)	(-0.30)	(-0.57)	(0.03)
Base sal	-0.0251	1.626	-0.202	0.0234	-0.154*	-0.361*	-0.110	0.351	0.289	-0.597
	(-0.11)	(1.26)	(-1.38)	(0.28)	(-2.00)	(-2.35)	(-0.87)	(1.66)	(0.75)	(-1.71)
Net sal	-0.214	-0.0976	-0.446	-0.0981	-0.348	-0.461	-0.0577	-0.0677	-0.155	0
	(-0.59)	(-0.42)	(-1.87)	(-0.72)	(-1.96)	(-1.71)	(-0.36)	(-0.45)	(-0.43)	(.)
Annual sal	-0.301	0.727	-0.323	-0.222*	-0.287	-0.431	-0.0728	0.0324	-0.418	-0.225
	(-1.53)	(1.40)	(-1.65)	(-2.49)	(-1.93)	(-1.67)	(-0.44)	(0.28)	(-0.79)	(-0.75)
Gross sal	-0.134	-0.543	-0.186	-0.263*	-0.0187	-0.134	-0.0201	-0.181	-0.402	0.0531
	(-1.18)	(-0.79)	(-1.95)	(-2.62)	(-0.19)	(-1.02)	(-0.18)	(-1.11)	(-0.63)	(0.16)
Sal Not Id	0.138	0	-0.226	-0.250	-0.124	-0.243	-0.126	-0.158	-0.0961	-2.329*
	(0.83)	(.)	(-1.41)	(-1.57)	(-0.76)	(-1.06)	(-0.67)	(-0.88)	(-0.23)	(-2.62)
Constant	-29.33	-106.4	29.09	-47.67	0.547	-6.054	11.46	-57.84	-7.761	-76.84
	(-0.46)	(-1.28)	(0.80)	(-1.48)	(0.02)	(-0.20)	(0.37)	(-1.69)	(-0.21)	(-1.46)
N	369	69	956	158	976	441	920	162	365	56
Pseudo R-sq										

Note: t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A12: Salary regression - primary-women and men Part 2 (salary sample)

	(1) Ag-M	(2) Ag-F	(3) Public-M	(4) Public-F	(5) Dev-M	(6) Dev-F	(7) Hist-M	(8) Hist-F	(9) Oth-M	(10) Oth-F
ysphd	0.0187*	-0.0223	0.105***	-0.0564	0.0435***	-0.0432	0.0419**	0.230*	0.00945	0.00948
	(2.27)	(-1.05)	(5.37)	(-1.53)	(4.88)	(-1.43)	(3.65)	(3.26)	(0.50)	(0.70)
ysphd_sqr	-0.000234	0.00113*	-0.00192***	0.00123	-0.000695***	0.00157	-0.00114***	-0.00432*	-0.000215	0.000173
	(-1.31)	(2.08)	(-4.53)	(1.17)	(-3.50)	(1.81)	(-4.75)	(-3.10)	(-0.53)	(0.54)
PhD Ec T 20	-0.104	-0.0364	-0.197	-1.001***	-0.00831	-0.687	-0.289	0	-0.119	0.549***
	(-1.01)	(-0.15)	(-0.94)	(-18.09)	(-0.08)	(-1.23)	(-1.35)	(.)	(-0.67)	(4.50)
PhD Ec 20-40	-0.130	-0.0510	0.127	-0.0781	0.0794	-0.500	0.0313	0	-0.267	-0.0148
	(-1.20)	(-0.20)	(0.44)	(-0.24)	(0.82)	(-0.85)	(0.14)	(.)	(-1.06)	(-0.16)
PhD Ec NR	-0.152	-0.0185	-0.394**	0	0.0101	-0.572	-0.204	0	-0.0319	0
	(-1.19)	(-0.08)	(-3.23)	(.)	(0.13)	(-1.08)	(-0.78)	(.)	(-0.16)	(.)
PhD Uni T 20	0.175	0.148	0.211	0.834***	-0.0825	0.519	0.0824	-0.179	0.0138	-0.882***
	(1.59)	(0.85)	(1.20)	(9.94)	(-0.73)	(1.75)	(0.30)	(-1.23)	(0.07)	(-5.20)
PhD U 20-40	0.234	-0.148	1.950*	0	-0.136	0.152	0.111	0	0.126	0.378**
	(1.95)	(-0.79)	(2.33)	(.)	(-1.65)	(0.43)	(0.41)	(.)	(0.81)	(3.27)
PhD Uni NR	0.0635	0.214	0.507	0	-0.0349	0.241	-0.132	0	0.252	0
	(0.36)	(0.60)	(1.65)	(.)	(-0.24)	(0.88)	(-0.42)	(.)	(1.19)	(.)
Econ Top 20	0.434**	0.394	0.650***	0	0.448***	0.513	0.108	0	0.537*	0.442**
	(2.81)	(1.79)	(4.30)	(.)	(3.58)	(2.01)	(0.91)	(.)	(2.14)	(3.37)
Ec 20-40	0.328**	0.353	-0.234	0	0.0509	0.126	0.403*	0	-0.0737	-0.0550
	(3.14)	(1.40)	(-1.26)	(.)	(0.37)	(0.50)	(2.41)	(.)	(-0.23)	(-0.46)
Econ NR	0.0984	0.110	-0.0617	0	-0.0744	0.118	0.00505	-0.175	-0.0970	-0.159
	(0.91)	(1.00)	(-0.31)	(.)	(-0.97)	(0.41)	(0.02)	(-1.39)	(-0.56)	(-1.10)
Uni T 20	0.160	0	0	0	0.207	-0.182	0	0	-0.393	0
	(0.89)	(.)	(.)	(.)	(1.75)	(-0.40)	(.)	(.)	(-1.62)	(.)
Uni 20-40	-0.0131	0.0722	0.176	0	0.179**	0.315	0.250	0	0.400*	-0.0825
	(-0.11)	(0.31)	(1.28)	(.)	(2.64)	(1.23)	(1.78)	(.)	(2.18)	(-0.52)
Uni NR	0.209	0.0149	-0.136	0	-0.476**	-0.495	-0.270	0	-0.283	0
	(0.96)	(0.06)	(-0.42)	(.)	(-3.27)	(-1.65)	(-1.28)	(.)	(-1.20)	(.)
Year	-0.0280	0.00182	-0.0114	0.0248	0.00371	0.0135	0.0425**	-0.0327*	-0.0176	0.0308**
	(-0.94)	(0.17)	(-0.49)	(0.82)	(0.21)	(0.54)	(3.52)	(-3.53)	(-0.91)	(3.49)
Dual fields	-0.136	-0.126	0	0	-0.140	0	0	0	-0.104	0.916***
	(-1.73)	(-0.76)	(.)	(.)	(-1.78)	(.)	(.)	(.)	(-0.49)	(17.04)
Base sal	-0.00166	0.144	-0.0537	0	-0.214	-0.678*	-0.295	0.717***	-0.220	-0.0984
	(-0.01)	(0.84)	(-0.28)	(.)	(-1.70)	(-2.11)	(-1.62)	(12.14)	(-1.16)	(-0.79)
Net sal	-0.595	-0.436	-1.777***	-0.424	-0.292	-0.670	0.0618	0	-1.094	-2.909***
	(-1.41)	(-1.23)	(-4.38)	(-2.32)	(-1.12)	(-1.46)	(0.32)	(.)	(-1.94)	(-13.68)
Annual sal	-0.178	-0.0910	-1.840***	0	-0.331	-0.364	0.124	-0.0511	-0.655	-1.675***
	(-1.12)	(-0.54)	(-5.07)	(.)	(-1.44)	(-1.31)	(1.01)	(-0.47)	(-1.44)	(-10.27)
Gross sal	-0.0727	0.142	-1.613**	0	0.0507	0.0543	0.0533	0.483***	-0.521	-2.020***
	(-0.51)	(0.79)	(-3.54)	(.)	(0.30)	(0.22)	(0.37)	(14.98)	(-1.51)	(-10.19)
Sal Not Id	-0.0298	0.316	-1.153*	0	0.0981	-0.228	-0.758**	0	-0.541	-1.963***
	(-0.15)	(1.22)	(-2.34)	(.)	(0.27)	(-0.68)	(-3.23)	(.)	(-1.19)	(-8.10)
Constant	67.89	7.705	35.19	-37.27	4.027	-14.90	-73.92**	74.61**	47.97	-48.63*
	(1.13)	(0.36)	(0.75)	(-0.61)	(0.11)	(-0.29)	(-3.05)	(3.86)	(1.22)	(-2.73)
N	996	285	191	21	480	138	117	33	377	83
Pseudo R-sq										

Note: t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Technical Appendix

TA.1 Quadratic Disparity Measure

In our multivariate binomial logit model, the quadratic (field) disparity measure (QDM) is calculated as

$$Q_{mf} = \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{ms} - \overline{\pi}_{fs}\right)^2.$$

The higher the value of Q_{mf}, the greater the degree of gender field specialization disparity. If the field combination distributions were identical for men and women, $\overline{\pi}_{ms} = \overline{\pi}_{fs}, \forall s$, the lower bound value for Q_{mf} is 0. We can establish an upper bound value for Q_{mf} by examining cases in which there is no gender overlap in fields of specialization (complete segregation). Any particular field 's' is completely segregated if the proportion for males (females) in field 's' is greater than zero and the corresponding proportion of females (males) in field 's' is zero. Complete field segregation means that every field combination is either all male or all female.

We establish the conditions that a complete segregation case must satisfy and use these results to derive the upper bound value for Q_{mf}. First, define the indicator variable $D_{ps} = 1(\overline{\pi}_{ms} > 0, \overline{\pi}_{fs} = 0)$, or equivalently, $(1 - D_{ps}) = 1(\overline{\pi}_{ms} = 0, \overline{\pi}_{fs} > 0)$. Accordingly, we can express a complete segregation case as one in which $Q_{mf} = Q^{cs}_{mf}$, where

$$\begin{aligned}
Q^{cs}_{mf} =& \frac{1}{2}\sum_{s=1}^{J^*}\left(D_{ps}\overline{\pi}_{ms} - (1 - D_{ps})\overline{\pi}_{fs}\right)^2 \\
=& \frac{1}{2}\left\{\sum_{s=1}^{J^*}(D_{ps}\overline{\pi}_{ms})^2 + \sum_{s=1}^{J^*}\left[(1 - D_{ps})\overline{\pi}_{fs}\right]^2 + 2\left[\sum_{s=1}^{J}(D_{ps})(1 - D_{ps})\overline{\pi}_{ms}\overline{\pi}_{fs}\right]^2\right\} \\
=& \frac{1}{2}\left\{\sum_{s=1}^{J^*}(D_{ps}\overline{\pi}_{ms})^2 + \sum_{s=1}^{J^*}\left[(1 - D_{ps})\overline{\pi}_{fs}\right]^2\right\} \\
=& \frac{1}{2}\left\{\sum_{s=1}^{J^*}D_{ps}(\overline{\pi}_{ms})^2 + \sum_{s=1}^{J^*}(1 - D_{ps})\left(\overline{\pi}_{fs}\right)^2\right\},
\end{aligned}$$

since $(D_{ps})(1 - D_{ps}) = 0$, $(D_{ps})^2 = D_{ps}$, and $(1 - D_{ps})^2 = 1 - D_{ps}$. Note that because there is at least one 's' such that $D_{ps} = 1$ and at least one 's' such that $1 - D_{ps} = 1$, it follows that $\sum_{s=1}^{J^*}\overline{\pi}^2_{ms} \leq 1$ and $\sum_{s=1}^{J^*}\overline{\pi}^2_{fs} \leq 1$. Therefore, $0 \leq Q^{cs}_{mf} \leq 1$.

We term the upper bound for Q_{mf} "perfect segregation" and derive this upper bound as

$$\lim_{\substack{\overline{\pi}_{mj}\to 1 \\ \overline{\pi}_{fl}\to 1}} Q_{mf} = \lim_{\substack{\overline{\pi}_{mj}\to 1 \\ \overline{\pi}_{fl}\to 1}} Q^{cs}_{mf} = 1,$$

for $j \neq l$. Thus, $0 \leq Q_{mf} \leq 1$. Accordingly, Q_{mf} can be viewed as the percentage of perfect field segregation that exists, i.e. $100 \times Q_{mf}$. As a practical matter $Q_{mf} = Q^{cs}_{mf} = 1$ if there are only two fields, one of which is all male and the other is all female.

We conduct counterfactuals to determine how the value of the QDM would change if men and women, when they were choosing their doctoral field of specialization, had the same values of the variables determining field choice or faced the same multivariate logit parameters. These counterfactual calculations are then used to decompose the sample value of Q_{mf}.

If we adopt the estimated multivariate binomial logit (with correlated choices) model for men as our baseline, the QDM can be decomposed as follows:

$$
\begin{aligned}
Q_{mf} &= \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs} + \overline{\pi}^m_{fs} - \overline{\pi}_{fs}\right)^2 \\
&= \frac{1}{2}\sum_{s=1}^{J^*}\left[\left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs}\right)^2 + (\overline{\pi}^m_{fs} - \overline{\pi}_{fs})^2 + 2\left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs}\right)\left(\overline{\pi}^m_{fs} - \overline{\pi}_{fs}\right)\right] \\
&= \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs}\right)^2 + \frac{1}{2}\sum_{s=1}^{J^*}(\overline{\pi}^m_{fs} - \overline{\pi}_{fs})^2 + \sum_{s=1}^{J^*}\left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs}\right)\left(\overline{\pi}^m_{fs} - \overline{\pi}_{fs}\right) \\
&= \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs}\right)^2 + \frac{1}{2}\sum_{s=1}^{J^*}(\overline{\pi}^m_{fs} - \overline{\pi}_{fs})^2 - \sum_{s=1}^{J^*}\left(\overline{\pi}^m_{fs} - \overline{\pi}_{ms}\right)\left(\overline{\pi}^m_{fs} - \overline{\pi}_{fs}\right) \\
&= \underbrace{Q^{mc}_{mf}}_{characteristics} + \underbrace{Q^{mp}_{mf}}_{parameters} + \underbrace{Q^{mcp}_{mf}}_{cross\, products},
\end{aligned}
$$

where $Q^{mc}_{mf} = \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{ms} - \overline{\pi}^m_{fs}\right)^2$ is the estimated amount of the gender field disparity arising from gender differences in characteristics, $Q^{mp}_{mf} = \frac{1}{2}\sum_{s=1}^{J^*}(\overline{\pi}^m_{fs} - \overline{\pi}_{fs})^2$ is the estimated amount of the gender field disparity arising from gender differences in parameters, and $Q^{mcp}_{mf} = -\sum_{s=1}^{J^*}\left(\overline{\pi}^m_{fs} - \overline{\pi}_{ms}\right)\left(\overline{\pi}^m_{fs} - \overline{\pi}_{fs}\right)$ is the estimated amount of the gender field disparity arising from the cross products of characteristics and parameters.

An alternative counterfactual is obtained by adopting the estimated multivariate binomial logit model for women as our baseline. In this case the quadratic field disparity measure can be decomposed as follows:

$$
\begin{aligned}
Q_{mf} &= \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{fj} - \overline{\pi}^f_{mj} + \overline{\pi}^f_{mj} - \overline{\pi}_{mj}\right)^2 \\
&= \frac{1}{2}\sum_{s=1}^{J^*}\left[\left(\overline{\pi}_{fj} - \overline{\pi}^f_{mj}\right)^2 + (\overline{\pi}^f_{mj} - \overline{\pi}_{mj})^2 + 2\left(\overline{\pi}_{fj} - \overline{\pi}^f_{mj}\right)(\overline{\pi}^f_{mj} - \overline{\pi}_{mj})\right] \\
&= \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{fj} - \overline{\pi}^f_{mj}\right)^2 + \frac{1}{2}\sum_{s=1}^{J^*}(\overline{\pi}^f_{mj} - \overline{\pi}_{mj})^2 + \sum_{s=1}^{J^*}\left(\overline{\pi}_{fj} - \overline{\pi}^f_{mj}\right)(\overline{\pi}^f_{mj} - \overline{\pi}_{mj}) \\
&= \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{fj} - \overline{\pi}^f_{mj}\right)^2 + \frac{1}{2}\sum_{s=1}^{J^*}(\overline{\pi}^f_{mj} - \overline{\pi}_{mj})^2 - \sum_{s=1}^{J^*}\left(\overline{\pi}^f_{mj} - \overline{\pi}_{fj}\right)(\overline{\pi}^f_{mj} - \overline{\pi}_{mj}) \\
&= \underbrace{Q^{fc}_{mf}}_{characteristics} + \underbrace{Q^{fp}_{mf}}_{parameters} + \underbrace{Q^{fcp}_{mf}}_{cross\,products},
\end{aligned}
$$

where $Q^{fc}_{mf} = \frac{1}{2}\sum_{s=1}^{J^*}\left(\overline{\pi}_{fj} - \overline{\pi}^f_{mj}\right)^2$ is the estimated amount of the gender field disparity arising from gender differences in characteristics, $Q^{fp}_{mf} = \frac{1}{2}\sum_{s=1}^{J^*}(\overline{\pi}^f_{ms} - \overline{\pi}_{ms})^2$ is the estimated amount of the gender field disparity arising from gender differences in parameters, and $Q^{fcp}_{mf} = -\sum_{s=1}^{J^*}\left(\overline{\pi}^f_{ms} - \overline{\pi}_{fs}\right)(\overline{\pi}^f_{ms} - \overline{\pi}_{ms})$ is the estimated amount of the gender field disparity arising from the cross products of characteristics and parameters.

The decompositions allow us to simulate how any policy changes that impact the effects of gender differences in characteristics or in the values of the specialization determinants might alter field specialization dissimilarity.

TA.2 Predictive Accuracy

For the purpose of evaluating the validity and usefulness of the multivariate logit model with correlated choices, we employ a simple measure of predictive accuracy to compare model accuracy in predicting each of $(J)(J+1)/2$ theoretically possible field specializations.

Model Accuracy Rate We construct indicators for each possible field specialization $s = 1, ..., (J)(J+1)/2 = 55, k = m, f$:

$$\pi_{kis} = 1(S_{kij} \cdot S_{kil}), i = 1, ..., \widehat{N}_k,^{18}$$

where $\widehat{N}_k \geq N_k$ is the number of predicted choices for group k. Let K_i equal the number of predicted dual specializations for the ith individual. The predicted choice for each field specialization is denoted by $\widehat{\pi}_{iks}$ and is determined according to

$$\begin{aligned} \widehat{\pi}_{kis} &= 1 \text{ or } 0 \quad \text{if } K_i = 1 \\ &= \frac{1}{K_i} \text{ or } 0 \quad \text{if } K_i > 1. \end{aligned}$$

For those for whom the model predicts more than one dual field specialization, we define an indicator variable for those fields in which the model predicts that an individual would specialize in more than 1 dual field and for which the individual did not actually select the given dual field:

$$d_{kis} = 1(0 < \widehat{\pi}_{kis} = \frac{1}{K_i} < 1 \quad \& \quad \pi_{kis} = 0).$$

For any given field specialization s, the total number of correct predictions is denoted by n_{ks} where

$$n_{ks} = \sum_{i=1}^{\widehat{N}_k} \left[(\pi_{kis} \cdot \widehat{\pi}_{kis}) + (1 - \pi_{kis})(1 - \widehat{\pi}_{kis}) - d_{kis}(1 - \widehat{\pi}_{kis}) \right].$$

The term $-d_{kis}(1 - \widehat{\pi}_{kis})$ ensures that there is no over-counting of the number of correct predictions and that the partial correct prediction receives a weight of $\widehat{\pi}_{kis} = \frac{1}{K_i}$.

The total number of correct predictions for gender group k is therefore

$$\begin{aligned} n_k &= \sum_{s=1}^{55} n_{ks} \\ &= \sum_{s=1}^{55} \sum_{i=1}^{\widehat{N}_k} \left[(\pi_{kis} \cdot \widehat{\pi}_{kis}) + (1 - \pi_{kis})(1 - \widehat{\pi}_{kis}) - d_{kis}(1 - \widehat{\pi}_{kis}) \right]. \end{aligned}$$

The overall econometric model accuracy rate is accordingly

$$R^{EM} = \frac{n_k}{(55)(\widehat{N}_k)}.$$

[18] Note that when $l = j$, π_{kis} refers to specialization in a primary field only.