Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR WIRTSCHAFTS- UND
SOZIALWISSENSCHAFTEN

Culture, Female Labour Force Participation, and Selective Migration: New Meta-Analytic Evidence

Eva Markowsky

WiSo-HH Working Paper Series
Working Paper No. 65
January 2022

WiSo-HH Working Paper Series
Working Paper No. 65
January 2022

Culture, Female Labour Force Participation, and Selective Migration: New Meta-Analytic Evidence

Eva Markowsky, University of Hamburg

ISSN 2196-8128

Font used: „TheSans UHH“ / LucasFonts

Die Working Paper Series bieten Forscherinnen und Forschern, die an Projekten in Federführung oder mit der Beteiligung der Fakultät für Wirtschafts- und Sozialwissenschaften der Universität Hamburg tätig sind, die Möglichkeit zur digitalen Publikation ihrer Forschungsergebnisse. Die Reihe erscheint in unregelmäßiger Reihenfolge.

Jede Nummer erscheint in digitaler Version unter
https://www.wiso.uni-hamburg.de/de/forschung/working-paper-series/

Kontakt:
WiSo-Forschungslabor
Von-Melle-Park 5
20146 Hamburg
E-Mail: experiments@wiso.uni-hamburg.de
Web: http://www.wiso.uni-hamburg.de/forschung/forschungslabor/home/

Culture, Female Labour Force Participation, and Selective Migration: New Meta-Analytic Evidence

Eva Markowsky

January 2022

Abstract

I analyse about 160 estimates of the relationship between female migrants' labour force participation and ancestry culture, integrating studies from economics and sociology that were previously unconnected. The literature exhibits large heterogeneity in results that is only partly explained by differences in data, sample selection, and methods. Part of the excess heterogeneity seems to be driven by selective reporting with preference given to studies that find positive and statistically significant correlations between ancestry culture and female labour force participation. Differential composition of countries of ancestry proves another important source of heterogeneity. Estimations drawn from data with higher average ancestry-gender equality find lager culture effects, implying that the behaviour of women from low-gender equality countries might not be representative for their country of origing culture. I discuss how cultural selection of immigration could explain this relationship and why it can bias measured culture effects. The analysis provides valuable insights for future applications of the socio-epidemiological approach.

1 Introduction

In many countries around the globe, women still lag behind men in terms of labour market participation. While some countries have experienced tremendous progress towards gender equality in the labour market and now exhibit almost no gender gaps in participation rates, others still show large gender imbalances and little to no change over the last two decades (World Economic Forum (WEF), 2021). One proposed reason for extensive and persistent between-country gaps are cultural differences regarding the gendered division between the public and private sphere.[1] However, simple comparisons of behaviour across countries would not allow the identification of this influence due to the endogenous relationship between culture and other informal as well as formal institutions. As part of the "New Cultural Economics" (Guiso *et al.*, 2006; Gershman, 2017) literature, a growing number of researchers attempt to solve this issue by examining the degree to which norms and values in the countries of origin shape immigrants' behaviour, thus keeping the host country environment fixed. During the past years, this literature has experienced tremendous growth, following the seminal work by Raquel Fernández and Alessandra Fogli in economics (Fernández, 2007; Fernández & Fogli, 2009) and by Frank van Tubergen and co-authors in the field of sociology (van Tubergen *et al.*, 2004; van Tubergen, 2006). Differences in data, sample, and methods lead to heterogeneous estimates of the effect of culture. In this paper, I collect a data set of empirical estimates of the association between female migrants' labour force participation and their ancestry culture and meta-analyse those data to systemise and summarise this extensive body of literature.

Fernández & Fogli (2009) relate second-generation immigrant women's labour force participation and fertility in the United States to aggregate female labour force participation and fertility rates as well as attitudes on gender equality in their countries of ancestry and term this method the "epidemiological approach".[2] They find strong positive correlations between immigrant women's behaviours in the US and aggregate behaviour and norms in the countries of ancestry. After trying to rule out other potential mechanisms, particularly unobserved differences in quality of human capital between different origin groups, they interpret their findings as evidence of the influence of culture. In the following years, numerous studies applied their approach to different data (e.g., Blau *et al.*, 2011), different behavioural outcomes (e.g., Nollenberger *et al.*, 2016) and to immigrant populations in different host countries (e.g., Stichnoth & Yeter, 2016). At the same time, very similar research was conducted in the field of sociology, starting with the influential paper by van Tubergen *et al.* (2004). They combine individual-level labour market data from 18 "western" countries to compare the influence of continuous characteristics of sending and

[1] Culture, in this context, is rather broadly defined as beliefs, preferences, and norms that are differentially distributed across countries (e.g., Polavieja, 2015, p. 170)

[2] Because it studies the geographical distribution of values, just as medical epidemiology studies the distribution of disease.

receiving countries on immigrants' economic integration. They, too, find, among others, a strong influence of country-of-origin female labour force participation rates on immigrant women's labour supply. The authors generally find support for destination, origin, and "community effects" on immigrants' labour market integration, the latter being the effect of membership in a particular origin group within a specific destination. Numerous other sociology researchers followed in applying similar methods to different data sets and various research questions (e.g., Dinesen, 2013; Apgar & McManus, 2018; Hajdu & Hajdu, 2016).

Despite the apparent similarities between the economics and sociology papers in methods and data used, at the moment, both research strands seem somewhat unconnected. Given the high policy relevance and the inherently interdisciplinary nature of the topic, I strive to consolidate these two branches of the literature in this paper. To maximise comparability, I focus on studies of female labour force participation.

Table 1: Studies of origin effects on female migrants' labour force participation

	Published	Host Country	Immigrant Generation	Outcome	Culture Measure	No. of Ancestries
Adnan (2019)	IZA Discussion Paper	USA	1.5th + 2nd	participation, hours	FLFP	124
Alesina & Giuliano (2010)	J. Econ. Growth	USA	2nd	participation	Attitudes towards family	80
Antecol (2000)	Labour Econ.	USA	1st, 2nd	participation	FLFP	70
Apgar & McManus (2018)	Soc. Forces	USA	2nd	participation	female to male LFP ratio	131
Beblo et al. (2020b)	Econ. Bull.	27 European Countries	1.5th + 2nd	hours	FLFP	93
Blau & Kahn (2015)	J. Hum. Cap.	USA	1st	hours	female to male LFP ratio	27
Blau et al. (2011)	Rev. Econ. Stat.	USA	1st	hours	female to male LFP ratio	-
Blau et al. (2013)	J. Popul. Econ.	USA	2nd	hours	FLFP, avg hours	-
Bredtmann & Otten (2013)	mimeo Working Paper	24 European Countries	1st, 2nd	hours	female to male LFP ratio	58
Buitrago (2015)	Dissertation	USA	2nd	weeks, hours	FLFP	7
Fernández & Fogli (2009)	AEJ: Macro	USA	2nd	hours	FLFP	25
Fernández (2007)	JEEA	USA	2nd	participation, hours	FLFP, attitudes towards female employment	25
Finseraas & Kotsadam (2017)	Eur. Sociol. Rev.	Norway	2nd	employment	FLFP	49
Frank & Hou (2016)	Work, Empl. & Soc.	Canada	1st	participation	FLFP, attitudes towards gender roles	-
Gay et al. (2018)	Rev. Econ. Househ.	USA	1st	participation	language, FLFP	
Eylem Gevrek et al. (2013)	Int. Migr.	Canada	2nd	participation, hours	female to male LFP ratio	18
He & Gerber (2020)	Int. Migr. Rev.	USA	1st	employment	female to male LFP ratio	-
Huh (2018)	J. Popul. Res.	USA	1st	participation	Gender equality	43
Kessler & Milligan (2021)	Oxf. Econ. Pap.	Canda	2nd	hours	FLFP	25
Köbrich León (2013)	SOEP Papers	Germany	1st, 2nd	employment, hours	FLFP, gender attitudes	20
Kok et al. (2011)	IZA Discussion Paper	Netherlands	1st, 2nd	participation	gender gap in LFP	8
Liu & Hagiwara (2020)	RIETI Discussion Paper	Japan	1st	participation	FLFP, gender attitudes	10
McManus & Apgar (2019)	Demography	USA	2nd	hours	FLFP	124
Mocan (2019)	J. Comp. Econ.	25 European Countries	2nd	participation, hours	FLFP, work attitudes	46
Muchomba et al. (2020)	Fem. Econ.	USA	1st	hours	female to male LFP ratio	-
Polavieja (2015)	Am. Sociol. Rev.	23 European Countries	1st + 1.5th + 2nd	participation	Traditionalism	23
Salari (2016)	Dissertation	USA	2nd	participation	FLFP	23
Salari (2020)	Rev. Econ. Househ.	USA	2nd	participation, hours	FLFP	22
Scoppa & Stranges (2019)	Labour	Italy	1st	participation	FLFP	-
van Tubergen et al. (2004)	Am. Sociol. Rev.	18 countries	1st	participation	FLFP	-

To my knowledge, there are about 30 primary studies from economics and sociology that correlate immigrant women's labour force participation with characteristics of their countries of ancestry. Together, through multiple specifications in most papers, these studies offer over 160 estimations of the relationship between individual labour supply and country-of-ancestry characteristics. Table 1 lists essential features of these studies: The host country in which immigrants are observed, the immigrant generation that is studied, how labour force participation in the host country and country-of-origin culture are measured, and the number of ancestry countries that are represented among the respondents in each study. This heterogeneity introduces ambiguity in the interpretation of central findings. For example, Eylem Gevrek *et al.* (2013) find a much smaller relationship between second-generation immigrants' labour market participation in Canada and aggregate behaviour in their ancestry countries compared to the findings by Fernández & Fogli (2009). Across the whole literature, estimated correlations for the relationship between origin culture and destination female labour force participation vary between statistically significantly positive, statistically insignificant, almost precisely zero, and statistically significant negative (see Section 4). This variance in results might be due to actual differences in the underlying relationship, e.g., migrants in the US could be more attached to their ancestry culture than migrants in other host countries. However, it is also conceivable that the differences in results are at least partly driven by heterogeneity in other study characteristics, as illustrated in Table 1.

To disentangle the influence of study differences in data, sample, and methods on estimated culture effects, I conduct a meta-analysis of economic and sociological studies of the relationship between female migrants' labour force participation and cultural values in their ancestry countries. Meta-analysis provides tools to accumulate scientific research and to integrate and summarise primary studies in a systematic, reproducible fashion (Borenstein *et al.*, 2009). By comparing results on quantitative metrics, meta-analysis can complement narrative reviews. It helps make sense of large and intricate bodies of literature with many "scientific replications", i.e., replications of an original finding using different data, sample selection criteria, or study designs (Hamermesh, 2007). Beyond integrating evidence from the two disciplines in a meta-summary, meta-regression analysis allows me to utilise heterogeneity in the literature to determine the influence of study characteristics on results. As Table 1 shows, while differing in essential dimensions, the included studies also share enough similarities to draw meaningful comparisons. It is thus possible to determine the influence of single varying characteristics (e.g. the choice of culture measure), *holding other dimensions constant*.

Additionally, the meta-perspective provides novel insights regarding a central empirical challenge that cannot be resolved conclusively in individual studies: Like most research on immigrant behaviour, the socio-epidemiological approach likely suffers from issues of immigrant selection. Migrants are not a representative sample of the country of origin's population, but are selected in terms of skills, education, or other characteristics (Borjas, 1987; Chiswick, 1999; Docquier *et al.*, 2020). Therefore, their preferences might not match dominant norms in their countries of origin, potentially biasing correlations between these two variables. As Fernández & Fogli (2009) argue, systematic differences between emigrants and stayers would imply a *downward* bias on the correlation of interest. Since the main finding of the literature is a positive association between migrants' behaviours and their ancestry culture, most authors do not regard selection as a severe threat to their conclusions. Consequently, while many admit to a potential selection bias, researchers at most seek to mitigate these concerns by including controls at the individual or the country-of-origin level without conducting formal analyses or empirical tests of the issue.

However, a systematic underestimation of the effect of culture on migrant behaviour can have severe consequences for policy conclusions. The bias is even more grave in this context if the selection is based on cultural norms, i.e. the explanatory variable in models of culture effects. For example, Blau & Kahn (2015) suggest that women from countries with low female labour force participation rates might be more strongly selected in terms of their labour market orientation compared to women from countries with high participation. They argue that women from countries with low female representation in the labour market could actively select out of such gender-unequal environments in search of societal conditions that more closely fit their preferences. Such immigrant selection based on labour market orientation would imply a weaker or even negative relationship between behaviour in the destination country and ancestry norms among female migrants from countries with low female labour force participation compared to women from countries with high levels of female employment. Beblo *et al.* (2020b) conceptualise this relationship theoretically, while Beblo *et al.* (2020a) provide first empirical evidence for this notion.

In the meta-setting, studies that include more respondents from low participation countries would find smaller culture effects, on average. The composition of countries of ancestry is mostly data-driven and thus rarely addressed explicitly in the literature, other than some authors specifically excluding respondents from certain countries of origin (e.g., Fernández & Fogli 2009 exclude women from post-communist countries due to "profound transformations in the economies, institutions, and cultures" (p. 154) during the time frame of their analysis). These differences create an intriguing source of variation, lending itself to a meta-analytic investigation that allows systematic comparison of numerous estimations utilising data with differential distributions of immigrants across ancestry countries. This analysis additionally contributes new evidence to the scarce literature on the cultural selection of immigrants.

Before investigating this issue in detail, I conduct a comprehensive meta-summary to quantitatively synthesise economic and sociological research on female migrants' labour force participation in relation to their ancestry culture. To ensure the validity of the summary results, I also check for bias resulting from selective reporting. This first part of the analysis provides detailed answers to the question *What is the overall effect of origin culture on female immigrants' labour force participation?*. In the next step, I answer the question *What is the influence of study characteristics on results?* by regressing individual results on observable study features across the whole sample as well as within more homogeneous subsamples of primary estimates. Finally, I test whether differential sets of countries of origin influence the results. This last part addresses the vital question *Could selective migration bias estimates of culture effects?*

The meta-summary confirms a robust positive association between female migrants' labour force participation and ancestry culture. However, the actual relationship might be weaker than a simple aggregation of primary estimates would suggest since I find evidence of a positive bias from selective reporting of results. Heterogeneity analysis reveals the influence of research design features on results and indicates differential culture effects: Higher generation immigrants are less influenced by their ancestry culture than migrants of the first generation. Cultural inputs, such as average attitudes in the country of origin, matter less for immigrants' choices than the aggregate behaviour of their non-migrant peers in the ancestry country. Similarly, contemporary behaviour in the country of origin has a greater influence than culture measured in the past, e.g., at the time of migration. Finally, controlling for potential confounders, like partner characteristics, area of residence, and quality of human capital, is consequential since omitting these variables leads to biases in the estimates of the direct influence of culture on labour force participation. Merging information on country-of-ancestry gender equality to my meta-dataset, I find evidence for the influence of the country-of-ancestry composition regarding gender equality which could indicate selective migration on cultural values. Estimations with data containing more origin countries with relatively low gender equality find weaker correlations between destination labour force participation and ancestry culture than data containing fewer countries of ancestry with poor gender equality.

The rest of the paper is organised as follows: In Section 2, I describe methodological considerations when examining cultural effects on immigrant behaviour and their implications for interpreting central findings of the literature. Section 3 describes the construction and properties of the meta data set. Results of the meta-summary and meta-regression analyses are reported in Sections 4 and 5. Finally, in Section 6 I summarise the state of research on the cultural selection of immigrants and investigate whether selective migration biases findings on culture effects. The paper closes with a discussion and concluding remarks.

2 Measuring Culture

Attempts to measure the influence of ancestry culture on women's labour market outcomes go back to the 1980s. Early studies in economics, as well as sociology, mainly apply standard regression approaches, where the outcome of interest is regressed on, among others, origin or ethnicity dummies to determine the influence of membership in a particular ethnic group (e.g., Kelley & McAllister, 1984; Reimers, 1985).[3]

The major innovation of the studies following van Tubergen *et al.* (2004) and Fernández (2007) is to include continuous measures of aggregate norms or behaviour in the country of origin that are directly related to female labour force participation, resulting in estimations of the following form:

$$Y_{is} = \alpha_0 + \alpha_1 Culture_s + \boldsymbol{\gamma} \boldsymbol{Z} + u_i \tag{1}$$

where Y is the outcome variable of interest of individual i with ancestry s, $\boldsymbol{Z}$ is a vector of control variables with corresponding coefficients-vector $\boldsymbol{\gamma}$, and $Culture$ is the measure of ancestry culture, observed at the country-of-origin level. The relationship of interest, the effect of ancestry culture, is captured by α_1.

Besides labour force participation and fertility this methodology has been applied to a large variety of outcomes, including saving rates (e.g., Fuchs-Schündeln *et al.*, 2017), political participation (e.g., Alesina & Giuliano, 2011), health behaviour (e.g., Ljunge, 2016), and trust (e.g., Moschion & Tabasso, 2014). However, in this paper, I will focus on studies of female labour force participation to ensure comparability of results.

As Fernández (2011) points out, in order to plausibly argue that α_1 in Equation 1 captures the effect of culture, one has to address at least four empirical challenges. (i) The appropriate choice of culture measure, (ii) other sources of heterogeneity in female immigrants' labour force participation that might be orthogonal or endogenous to their ancestry culture, (iii) potentially omitted variables that vary systematically across countries of origin, and (iv) the possibility of a bias stemming from selective migration. I will address how the literature deals with the first three points in more detail below, while the issue of selective migration is analysed in depth in Section 6.

Even among empirical papers focusing on female labour force participation, there is considerable heterogeneity in the choice of identifying variable to capture the cultural

[3] Similar approaches are still applied regularly in contemporary studies as well, for example, Khoudja & Platt (2018); Kislev (2017).

norms, values, beliefs, or customs related to women's participation in the workforce, as demonstrated in Table 1. Among these choices, Apgar & McManus (2018) point out an interesting differentiation: While many authors rely on *behavioural* measures as proxies for cultural norms regarding women's involvement in the work sphere, such as absolute or relative female labour force participation (e.g., Blau *et al.*, 2013; Finseraas & Kotsadam, 2017), others are rather interested in the *inputs* to these behaviours, like attitudes towards gender equality (e.g. Fernández, 2007; Mocan, 2019), or social institutions (Apgar & McManus, 2018). Recognising that participation rates might capture other non-cultural variables, like economic growth, some authors use the ratio of female to male labour force participation (e.g., Blau *et al.*, 2011). Similarly, because cultural values about the appropriate role of women in home versus market production might vary *within* countries of origin, other authors use age-specific participation rates (e.g., Bredtmann & Otten, 2013). Polavieja (2015) develops an even more detailed approach to match the relevant values as closely as possible. He uses measured attitudes of non-migrants from the same countries of ancestry that resemble the immigrants in his data on a large set of observable characteristics.

Of course, migrants differ in dimensions other than their ancestry that likely influence their labour market involvement. However, Fernández (2011) points out that many of the individual characteristics that are routinely employed as control variables in other labour market research might themselves be influenced by ancestry culture and could thus be endogenous to the correlation of interest. For example, the immigrant women's educational attainment, their region of residence, marital status, and partners' characteristics could very well be influenced by culture. Even though including these and similar variables as covariates in the estimation probably takes away some of the culture effect, Fernández (2011) still argues for their inclusion to tease out the net culture effect on labour force participation. Tthe primary studies deal with this decision in heterogeneous ways: Some control only for education, some include large sets of individual covariates.

The last point concerns another source of potential omitted variable bias, stemming from other characteristics besides cultural values shared by individuals of common ancestry. Fernández (2011) mentions the unobserved quality of human capital as a likely candidate. Most of the literature deals with this in two ways: Controlling for parents' education or controlling for internationally standardised indicators of quality of human capital at the country-of-origin level. Finseraas & Kotsadam (2017) employ a more rigid method to deal with this issue by estimating ancestry-culture effects for female-male sibling pairs in detailed Norwegian data while including sibling fixed effects. While this certainly rules out unobserved confounding factors such as quality of parents' human capital or systematic differences in work ethic, it is also very data demanding and not feasible in many settings. Another important country-of-ancestry characteristic in this context is economic development since it is strongly related to gender norms at the country level (Falk & Hermle, 2018). The culture measure could thus pick up development rather than cultural norms about appropriate roles for women in society, leading some authors to additionally control for measures such as GDP in the country of origin.

These methodological considerations are crucial to the interpretation of the literature's central results. Nevertheless, to date, there is no systematic review that discusses how the various attempts to deal with these issues influence individual conclusions. In the following, I investigate how these and other methodological decisions by researchers attempting to identify the effect of cultural norms on female labour market behaviour influence their results.

3 The Data Set

3.1 Data Collection

I aim to collect all published studies (in peer-reviewed journals or the grey literature) from economics and sociology that regress female migrants' labour force participation on a measure of culture in the country of origin, as illustrated in Equation 1, even if this relationship is not the main focus of that research paper. To identify all relevant studies, I searched for the term "(immigrant* OR migrant*) AND (origin OR ancestry OR source) AND countr* AND (cultur* OR integration)" in the most important economic and social sciences literature databases: the American Economic Association's *EconLit*, *Research Papers in Economics (RePEc/IDEAS)* as well as the *International Bibliography of the Social Sciences (IBSS)* and *Web of Science*. Additionally, I conducted "snowball" searches starting with the seminal papers by Fernández & Fogli (2009) and van Tubergen *et al.* (2004) and checked all publications that cite these papers. To cover current and unpublished papers, I further searched the programs of international economics and sociology conferences for contributions that seemed to fit the above criteria judging from title and abstract.

This search yielded a list of 52 studies that appeared to conduct analyses of cultural effects on female labour force participation, judging from the titles and abstracts. After closer screening, 22 of these had to be excluded because they focus on internal migration instead of international migrants (N = 2), because they do not utilise a quantitative measure of culture but instead include dummies for individual source countries (N = 2) or broad origin regions (N = 3), because they focus on male labour force participation (N = 2), because they do not report the coefficient of interest but, e.g., only interaction terms (N = 4), or because they turned out to not apply the targeted regression approach at all but conduct, for example, cross-country analyses (N = 8).

From the remaining papers, listed in Table 1, I extract a total of 160 estimations of culture effects using a standardised coding scheme. Most papers estimate more than one specification, e.g., by analysing different samples or testing various culture measures. For example, from the seminal paper by Fernández & Fogli (2009) I code a total of 15 separate effect sizes from specifications with different outcome variables, using different data sets, and including differential sets of controls on the individual and the source country level.

For each study, I code all reported specifications except those containing interaction terms with the coefficient of interest. In these cases, the estimated culture effect for the affected group would have to be calculated as the linear combination of two coefficients (that from the culture measure and that from the interaction term). When those are not reported in the paper, it is impossible for the meta-researcher to obtain the associated standard errors without access to the original data. For example, Fernández & Fogli (2009, p. 171) include a table with specifications where they interact the measure of ancestry culture *LFP in 1950* with a dummy (*same*) indicating whether the woman's husband has the same ethnic ancestry as she does. For those women who share their ancestry with their partner, the culture effect would be estimated as the sum of the coefficients of *LFP 1950* and *same* × *LFP 1950*, which is not reported in the paper and hence cannot be coded for the meta-analysis, because I have no way to retrieve the standard errors for the sum of the two coefficients without access to the original data.

3.2 Effect Sizes

The 160 estimations I draw from the primary studies all conduct regression analyses similar to the one outlined in Equation 1, where the labour force participation of individual female migrants is regressed on, among others, a measure of culture in their countries of ancestry.

These conditional correlations, the effect sizes the meta-analysis is based on, are captured by the coefficients of the culture measure in the primary regressions - analogous to coefficient α_1 in Equation 1.[4] To make individual effect sizes comparable across different regression techniques (e.g., OLS vs logit) and across differential methods of reporting results (e.g., coeffcients vs marginal effects), I standardise them by calculating the *partial correlation coefficient* r, as defined in Equation 2, following Stanley & Doucouliagos

[4]Five effect sizes are reported as odds ratios (OR_α) in the primary studies. I transform them into logistic regression coefficients by taking the natural logarithm $\alpha = ln(OR_\alpha)$ and calculate the coefficients' standard errors as $se(\alpha) = \frac{se(OR_\alpha)}{exp(\alpha)}$, assuming that the odds ratios' standard errors are calculated using the delta rule, as is standard practice in most statistical softwares (Sribney & Wiggins, 2021).

(2012).

$$r = \frac{t}{\sqrt{t^2 + df}} \tag{2}$$

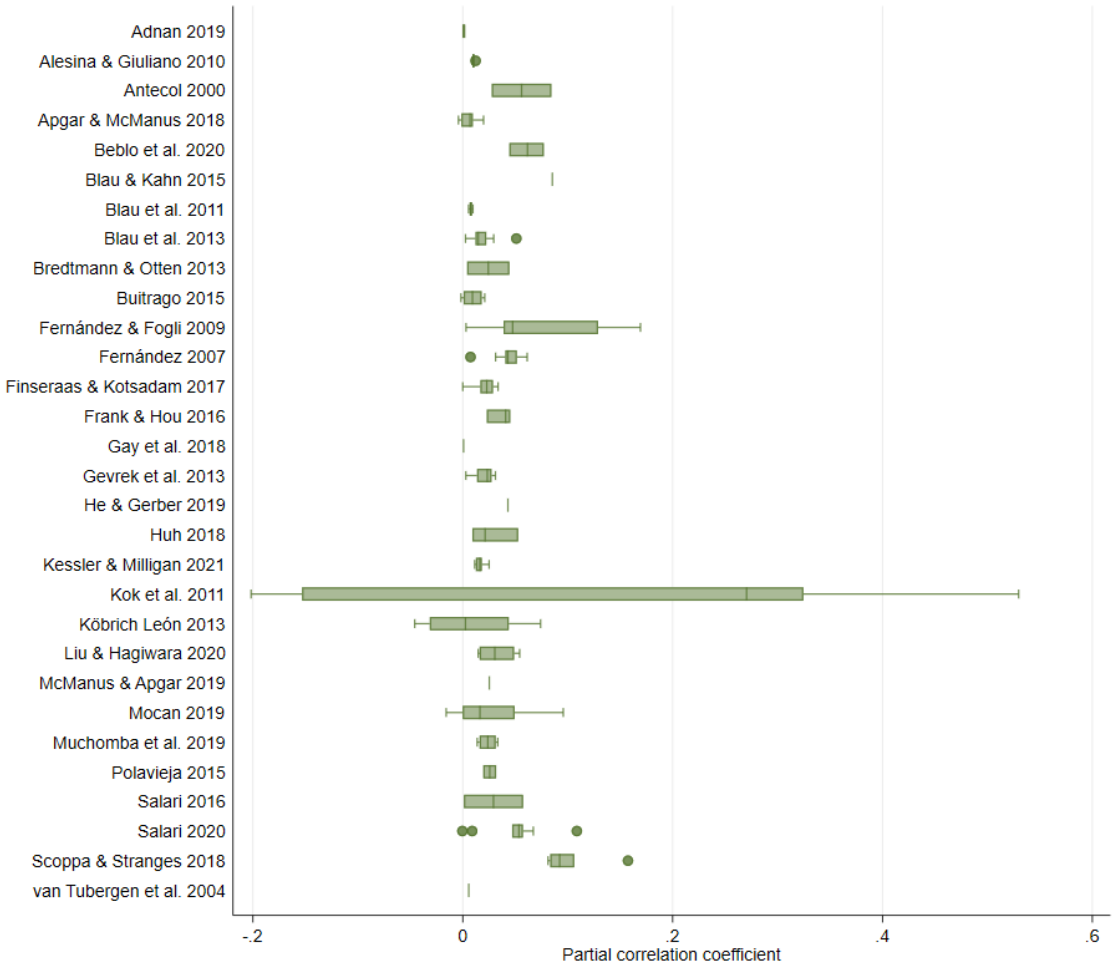
Here, t is the t- or z-statistic of each coefficient or marginal effect α_1, while df represents the degrees of freedom of this statistic, approximated by the number of observations in the primary estimations.[5] The partial correlation coefficient comes with the advantage of allowing comparisons across estimations from different models. It measures the direction and strength of two variables' association, holding other influences constant. Thus, the resulting value can be interpreted as a ceteris paribus correlation between culture and female labour force participation in the present context. The drawback of this unitless effect size measure is that it does not allow interpretation of the economic significance of the estimated effect size (Stanley & Doucouliagos, 2012, p. 25). Where the a priori expected correlation between the utilised culture measure and female labour force participation is negative (e.g., the population share of members of conservative religions, average agreement to the statement "Being a housewife is just as fulfilling as working for pay"), I multiply the partial correlation coefficient with -1 to harmonise the direction of effect sizes across studies. Positive values indicate that ancestry from a culture supportive of working women (e.g., a culture with high female labour force participation rates) is positively related to individual labour supply.

Figure 1 shows the distribution of the partial correlation coefficients by study in a box plot diagram. Note that the plotted variance relates to the different estimates *within* a given study and not to the precision of the individual estimates. Most standardised correlations between culture and female labour force participation lie between -0.05 and 0.2. The study by Kok *et al.* (2011) presents a very obvious outlier that reports estimates with both considerably lower and higher values. As indicated in Table 1, their study relates female first and second-generation immigrants' labour force participation in the Netherlands to the gender gaps in labour force participation in their eight countries of origin.[6] I extract five different estimates of effect size from that study: Separate estimations for immigrants of the first and second generation, as well as one that pools across generations, one specification that utilises an alternative measure of culture (the ratio of female to male participation) and, finally, a specification that additionally accounts for cohort-trends in native women's participation rates as a measure of "host country culture". All five estimates' standard errors are rounded to 0.00, i.e., they are among the most precisely estimated data points in the meta-data set, which in itself does not seem implausible since the authors draw from an immense data source with more than 50,000 observations. However, as we see in Figure 1, the resulting standardised correlation coefficients from this study present enormous effect sizes in both directions. The sizeable

[5] When t- or z-values are not reported, I calculate them as the ratio of coefficient or marginal effect and the corresponding standard error. In the few cases where standard errors are rounded to zero, I replace them with 0.004 to calculate the statistics.

[6] Per the above-described rule, I multiply correlation coefficients of the gender gap variable with -1, because larger gender gaps in labour force participation indicate *lower* female participation.

Figure 1: Boxplot of partial correlation coefficients by study



negative correlation of -0.2 stems from the specification using only second-generation immigrants, while the positive outlier (r = 0.53) relates to the estimation with immigrants of the first generation. From the information presented in the paper, I cannot infer the reason for these extreme results.[7] Independent of these outlier values, Figure 1 illustrates that the estimated culture effects vary not only between but also within studies.

As a first approximation of the full extent of within- and between-study heterogeneity in results, Figure 2 shows a Galbraith plot of effect size estimates, standardised to z-scores, against their precision. The estimates by Kok *et al.* (2011) are excluded to ensure readability. Higher z-scores (y-axis) indicate a stronger, positive correlation between ancestry culture and individual labour supply, and data points to the right have larger statistical power than those in the left-hand part of the x-axis. The blue regression line's slope indicates the overall effect size, to be discussed in the next section. The wide dispersion of effect size estimates to both sides of the line and beyond the 95%-confidence interval shaded in light blue suggest substantial heterogeneity in the meta-data. To explore this heterogeneity, I extract all relevant dimensions of within and between-study differences, such as data and sample characteristics or properties of the estimated models. In Section 5, I will test whether these moderators can explain the variation in results. Before that I integrate all primary estimates in a meta-summary to obtain an overall effect size.

4 Meta Summary: Integrating individual results

In this section, I carry out a series of standard meta-analytic techniques to answer the question *What is the overall effect of origin culture on female immigrants' labour force participation?* I first calculate the average effect size. Second, I test for the possibility that the summary results are biased by selective reporting of results.

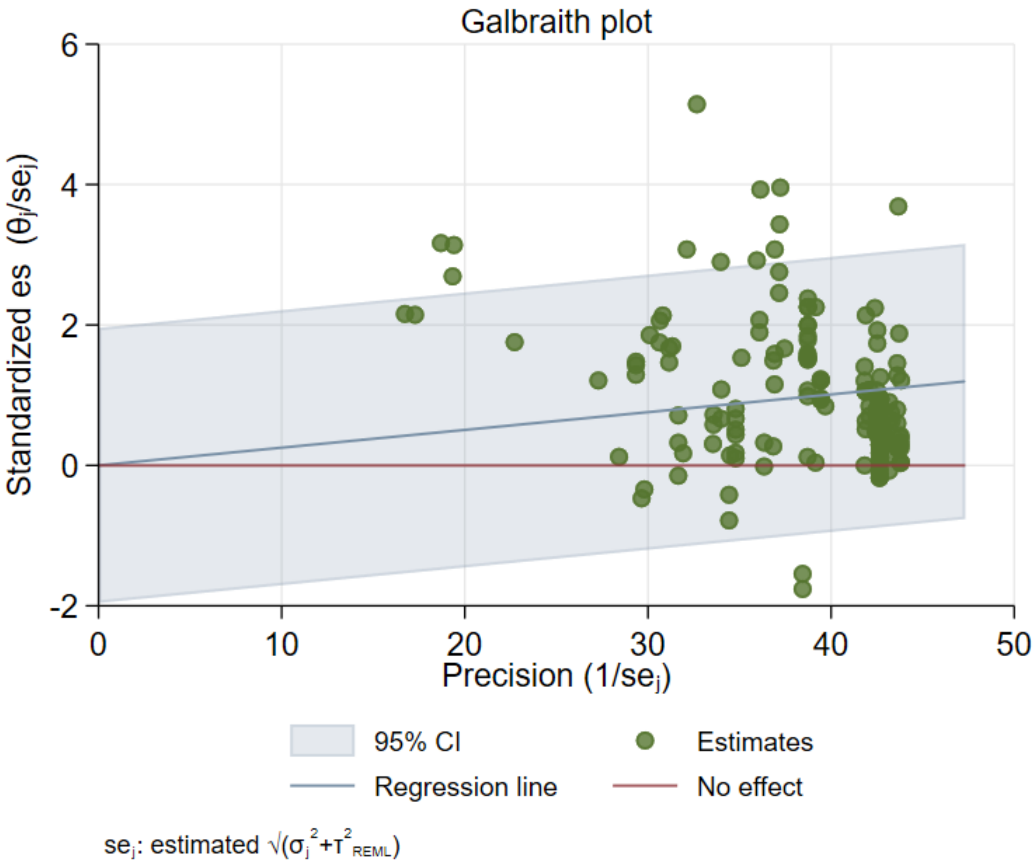
4.1 Overall Culture Effect

I start by calculating the meta-analytic overall culture effect in my data as the weighted average of primary partial correlation coefficients, where coefficients with higher precision are given larger weights. The specific weighting procedure depends on the chosen meta-analytic framework for integrating primary results. The simplest case is the common-effects model:

$$CE_i = CE_0 + \varepsilon_i, \tag{3}$$

[7] I contacted the authors to obtain the unrounded standard errors and to confirm the reported effect sizes, but they do not have access to the underlying data anymore and could not provide any additional information.

Figure 2: Galbraith plot



where CE_i ist the ith estimate of the culture effect - the conditional correlation between female migrants' labour force participation and characteristics of their countries of ancestry, standardized to partial correlation coefficients. CE_0 is the "true" culture effect that is assumed to be common to all observations in the meta data, which are estimated with sampling error ε_i. Effectively, the common-effects model assumes all of the collected estimates to be drawn from the same population with a common mean. It is assumed that $\varepsilon_i \sim N(0, \sigma_i^2)$.

In contrast, random-effects meta-analysis presumes estimates to be drawn from several distinct populations and allows individual estimates to vary randomly around CE_0:

$$CE_i = CE_0 + \theta_i + \varepsilon_i \quad \text{with} \quad \theta_i \sim N(0, \tau^2) \quad \text{and} \quad \varepsilon_i \sim N(0, \sigma_i^2), \tag{4}$$

where the estimate-specific "true" effect consists of two components: CE_0 and the sampling error θ_i. τ^2 is a measure of between-estimate heterogeneity, beyond the variance from sampling, that is to be estimated.

In both cases, CE_0 can be obtained by calculating a weighted average of the primary estimates, where estimates with higher precision are given larger weights. In the common-effects case, the weights are given by the inverse variance $1/SE_i$, whereas the random-effects model uses $1/(SE_i^2 + \hat{\tau}^2)$, where $\hat{\tau}^2$ is the estimate of τ^2 (Stanley & Doucouliagos, 2012, p. 46). Due to the vast differences in samples, culture measures, and other study design features, I cannot reasonably assume that all estimates in my metadata stem from the same population and share a common underlying effect. Random-effects meta-analysis thus seems the appropriate choice (see Borenstein *et al.*, 2009). I use the estimation method following DerSimonian & Laird (1986), but the results are robust to using other random-effects methods.[8]

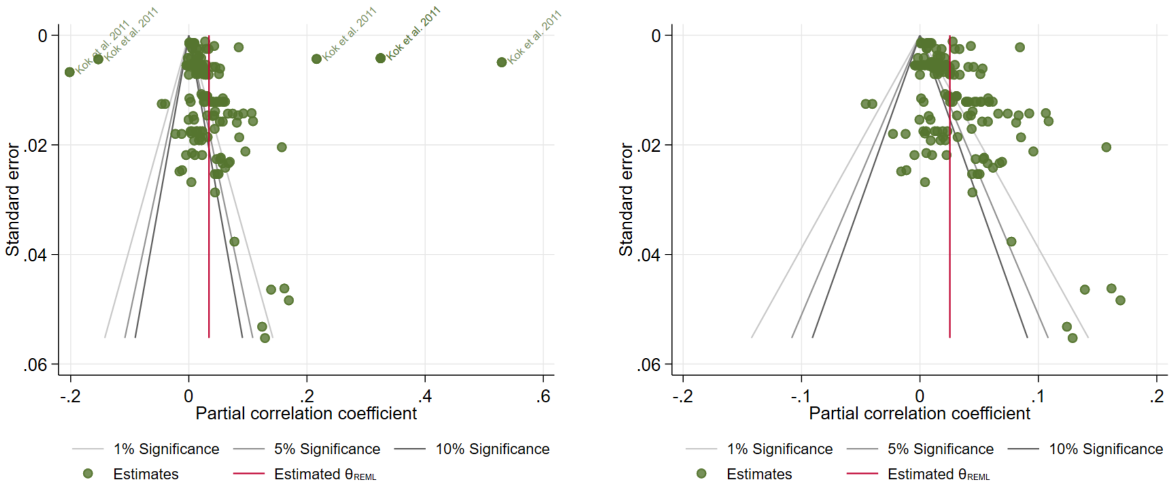
The resulting average effect size, CE_0 is calculated at 0.033 with 95%-confidence interval [0.024 ; 0.041]. Excluding the effect size estimates by Kok *et al.* (2011) leads to a somewhat smaller overall culture effect $CE_0 = 0.024$ (95%CI [0.021;0.027]). Thus, the overall association between ancestry culture and female migrants' labour market participation is small but positive and statistically significantly different from zero. However, this result is somewhat tentative since the homogeneity test is strongly rejected with a Q-statistic of 26834.72 and a corresponding p-value of < 0.0001. This implies considerable heterogeneity in results beyond what would be expected from sampling error alone. Another source of uncertainty in this result stems from the crucial assumption in random-effects meta-analysis that the observed studies report a *random* sample from the complete population of effect sizes. The estimated overall effect size will be biased if the studies in the meta-data set are selected, for example, because large and positive associations between culture and female labour force participation are more likely to be reported since they confirm the general expectation regarding the influence of culture. Therefore, I will test the data set for reporting bias before exploring the sources of between-study heterogeneity in detail in the next chapter.

4.2 Reporting Bias

It is possible that, despite my best research effort, the meta data set does not contain every existing study of the relationship of interest. My meta-analytical conclusions will be biased if the likelihood of retrieving a given study is systematically related to its results. Research has shown that in economics, as well as in other sciences, more extensive, more statistically significant results are more likely to be published in peer-reviewed journals (e.g., Card & Krueger, 1995; Brodeur *et al.*, 2016). Such a mechanism creates problems when published studies have a higher likelihood of finding their way into my data set

[8] Veroniki *et al.* (2016) recommend using the restricted maximum likelihood (REML) estimator for continuous data but also caution that it might be biased in small samples. Since my meta-data set is relatively small (30 studies, 160 effect size estimates), I use the most common DerSimonian and Laird method. Both methods result in identical estimates of the overall effect sizes up to the third decimal and very similar confidence intervals, while the REML-method estimates slightly larger between-study heterogeneity τ^2 and thus a somewhat larger (0.003) confidence interval. The analysis is implemented using the "meta summarize" command in Stata 17.

Figure 3: Funnel graphs of culture effects on female labour force participation



(a) All studies

(b) Excluding Kok et al. 2011

or when the "missing" results are not published as part of the grey literature either but instead remain in the proverbial file drawers. A related source of selection can occur when a theory or existing evidence create a strong expectation of the sign or magnitude of the researched relationship, deterring reporting or publishing of results that are "unexpected" by that logic (Imai *et al.*, 2021). I, therefore, test for the presence of reporting bias before attempting to draw definite conclusions about the overall culture effect.

A typical test is the visual assessment of a scatter plot of effect sizes against their standard errors, as shown in Figure 3, where the scale of the vertical axis is reversed such that more precise estimates lie towards the top of the graph. Since smaller studies need larger effect sizes to obtain statistically significant results, reported effect sizes often differ with sample size, resulting in asymmetry in the scatter plot. In the absence of selective reporting or publication of results, i.e., when estimates are "missing" at random, the less precise estimates (with larger standard errors) are expected to be relatively widely dispersed at the bottom of the plot. In contrast, the more precisely estimated effect sizes should cluster around the "true" value at the top of the graph, resulting in the characteristic funnel shape.

Figure 3a shows that the plot does not resemble the expected shape a lot due to the highly precise estimates by Kok *et al.* (2011) at the top of the graph dispersing far wider than the ones with larger standard errors towards the bottom. To increase readability, I, therefore, plot the same relationship in Figure 3b when excluding this study.[9] The scatter now roughly resembles the expected funnel shape, but there is a noticeable lack of small estimates of low precision. The grey lines represent standard significance levels.

[9]Since the estimates by Kok *et al.* (2011) are not only outliers but also leverage points with high precision and thus large weights in meta-analytical settings, I exclude them from the following analyses.

For example, the area between the two darkest lines holds all statistically insignificant estimates at the 10%-level. It seems that insignificant estimates of the culture effect are underrepresented in the metadata, especially among the data points with low precision. This asymmetry could hint towards a publication bias in the literature.

As a more formal test of funnel plot asymmetry, I regress the individual effect sizes on their standard errors, following Stanley & Doucouliagos (2012):

$$CE_{ij} = \beta_0 + \beta_1 SE_{ij} + \varepsilon_{ij}, \tag{5}$$

Here, CE_{ij} is again the ith measured culture effect taken from the jth study while SE_{ij} represents the associated standard error as a measure of that effect size's precision. Since standard errors in such a meta regression cannot reasonably be expected to be independently and identically distributed, I follow the recommendations by Stanley & Doucouliagos (2012, 2014) and estimate weighted least squares (WLS), weighting by each effect size's inverse variance.

The results are reported in the first column of Table 2. The coefficient β_1 of the standard error-variable is supposed to capture the degree of selective reporting bias. It is positive and statistically significantly different from zero, suggesting that less precise estimates (i.e., those with larger standard errors) tend to report larger effect sizes. This indicates selective reporting of results, with preference given to estimates supporting a more considerable positive correlation between female migrants' labour force participation and ancestry culture.

The regression approach to inspecting funnel asymmetry has the additional advantage that the constant β_0 from this model delivers an estimate of the "true" effect size - corrected for selective reporting. Statistically, it represents an extrapolated effect size measured with the highest possible precision and thus zero standard errors (Imai *et al.*, 2021). Therefore, the constant in Column 1 suggests that the selection-corrected conditional correlation between ancestry culture and female labour force participation lies considerably lower than the one obtained by the random-effects meta-summary at about 0.01.

To further inspect the robustness of the regression test results, I conduct additional analyses, as suggested by Stanley & Doucouliagos (2012): In the remaining columns of Table 2, I account for potential dependence of effect sizes that stem from the same study by clustering standard errors at the study level (Column 2) or by including study fixed effects (Column 3). Including fixed effects leads to an even higher estimate of the degree of positive selective reporting and an even lower estimate of the bias-corrected overall effect size. The $H0$ of the funnel-asymmetry test that there is no relationship between effect size magnitude and precision is thus rejected quite clearly, confirming the conclusions from the visual funnel plot inspection.

Table 2: Regression test of selective reporting

	(1)	(2)	(3)
SE of effect size	1.521***	1.521**	3.063***
	(0.438)	(0.695)	(1.024)
Constant	0.010***	0.010*	0.006*
	(0.002)	(0.006)	(0.003)
Observations	160	160	160
Study FE	No	No	Yes
Clustered s.e.	No	Yes	No

Note: Dependent variable: Effect size (r). *** p<0.01, ** p<0.05, * p<0.10

In summary, there is evidence of selective reporting, biasing the average culture effect away from zero. In particular, results from less precise estimations seem to be reported more likely when they show a positive, relatively large association between ancestry country characteristics and migrant women's behaviour. This finding is corroborated by regression-based tests of the relationship between effect size magnitude and precision.

However, the detected asymmetry is no final indication of selective reporting since it could also reflect a genuine relationship between effect sizes and precision. For instance, it is possible that studies with smaller samples indeed come to systematically different conclusions than those with large samples, e.g., because countries with less immigration (i.e. small samples of immigrants in population surveys) reach higher levels of immigrant integration or because the studies with small samples share some other characteristic that goes along with differences in results. Therefore, in the next section, I will investigate possible sources of heterogeneity in my data and their influences on the overall effect size. To ensure my meta-regression results are not biased by selective reporting, I will account for effect size precision in these analyses.

5 Explaining Heterogeneity

In the following, I utilise meta-regression analyses to investigate the influence of heterogeneity in data characteristics and model specification on individual results while still controlling for precision. This section provides answers to the question *What is the influence of study characteristics on results?* To this aim, I introduce moderators, i.e., dimensions of heterogeneity between studies and specifications, into the regression framework in Equation 5:

$$CE_{ij} = \beta_0 + \beta_1 SE_{ij} + \boldsymbol{\beta X_{ij}} + \varepsilon_{ij}, \tag{6}$$

where $\boldsymbol{X_{ij}}$ is a vector of observable study and specification characteristics described below and $\boldsymbol{\beta}$ represents the associated vector of coefficients. I continue to estimate the model using "unrestricted" WLS (Stanley & Doucouliagos, 2017), i.e., I use the inverse variance for weighting individual effect sizes (as in the previous section), with cluster-robust standard errors.[10] Study-fixed effects cannot be included in this model since several moderators only vary at the study level.

5.1 Sources of Heterogeneity

When collecting observable dimensions of heterogeneity in this literature, I differentiate between the characteristics of studies and specifications. Each study reports at least one, but in most cases several, distinct specifications, e.g., from estimating the relationship in different data or testing various outcome variables. Table 3 lists the collected study and specification attributes. Panel A lists **study characteristics**, as already illustrated in Table 1 in Section 3.1. Since I exclude the outliers and leverage points reported by Kok *et al.* (2011), the data set contains 29 studies. The most obvious difference is the host country that is considered. Seventeen of the studies in my final sample analyse the behaviour of migrants in the United States. Three focus on Canada, one each on Germany, Italy, and Norway. Four papers utilise a pooled dataset of several European countries, and one study, van Tubergen *et al.* (2004), is based on a pooled sample of 17 North American and European countries and Australia. Six studies are not (yet) published in peer-reviewed journals at the time of data collection but are only available as working papers or as part of a dissertation. Most of the studies are published in economics outlets; only seven appeared in sociology or demography journals.

Panel B provides descriptive information about the **specification characteristics**, i.e., the moderators that vary between specifications and therefore potentially within studies. I differentiate between variables that describe differences in the *data* and those that concern differences in the estimated *models*, starting with the data moderators.

A potentially meaningful source of heterogeneity in the data comes from differential ap-

[10] While common-effects meta analysis assumes σ^2 to be equal to one, WLS, while producing identical estimators, does not impose this restriction and is, therefore, "unresticted" (Stanley & Doucouliagos, 2015, p. 2117). Stanley & Doucouliagos (2017, 2015) show that unrestricted WLS delivers less biased estimators compared to common- as well as random-effects models in the presence of selective reporting and when there is heterogeneity. In Appendix-Table A.2, I demonstrate that my main results are robust to estimating DerSimonian and Laird-random-effects models instead. In the instances where the latter produces significant effects while my main analysis does not, I tend to trust the unrestricted WLS more, since this method allows accounting for dependent effect sizes from the same study by estimating cluster-robust variances (see also Imai *et al.*, 2021).

Table 3: Descriptive statistics on study and specification moderators

		∅	min	max	N
Panel A: Study characteristics					
Host country:	USA	0.59	0	1	29
	European	0.24	0	1	29
	Other	0.17	0	1	29
Reviewed publication		0.8	0	1	29
Publication year		2015	2000	2021	29
Field:	Economic	0.76	0	1	29
	Sociology, Demography	0.24	0	1	29
Panel B: Specification characteristics					
B1: Data moderators		∅	min	max	N
Dependent variable:	Working hours	0.49	0	1	160
	Participation	0.35	0	1	160
	Employment	0.09	0	1	160
	Fulltime employment	0.04	0	1	160
	Other	0.03	0	1	160
Sample:	1st generation	0.22	0	1	160
	1st & 2nd generation	0.04	0	1	160
	2nd & higher generation	0.74	0	1	160
	Mean age	38.27	30	47	160
Culture measured...	as input	0.25	0	1	160
	as output	0.75	0	1	160
	lagged	0.41	0	1	160
B2: Model moderators		∅	min	max	N
Individual controls include...	education	0.79	0	1	160
	partner characteristics	0.34	0	1	160
	area of residence	0.50	0	1	160
Source country controls include...	avg. quality of human capital	0.17	0	1	160
	GDP	0.26	0	1	160

Notes: The table lists characteristics (mean, min, max, and the number of non-missing observations) of all studies and specifications included in the meta-regression analyses. Panel A starts with general information on the studies; Panel B1 continues with data moderators; Panel B2 describes model moderators.

proaches to measuring labour force participation. While most specifications use working hours, others focus on binary labour market outcomes: Participation for the largest part, defined as being employed or actively searching for work. Only a few specifications are focused on employment or fulltime-employment. Here, the "other" category is made up of two specifications each, that measure "weeks working per year" and "number of days employed in previous year". In the regression analyses, I aggregate the different measures into an indicator that equals one when one of the binary measures is used as opposed to the continuous ones.

Concerning the sample, I identify two crucial moderators. First, cultural effects are estimated in different immigrant generations, with most specifications focused on the second generation. About 20 per cent use first-generation immigrants, and only a tiny share of specifications (4%) analyses a pooled sample of the first and second-generation immigrants that I pool together with the specifications on the first generation in the meta-regression. The second sample moderator is the mean sample age. It varies between 30 and 47 with a mean of about 38.

As already discussed in Section 2, specifications further differ in the measure of culture they employ. Table A.1 in the appendix lists all source country characteristics that are used for that purpose in the primary studies. While the list is far too long to compare between specifications with every possible measure, I follow Apgar & McManus (2018) and differentiate between "input" and "output" measures of ancestry culture, according to the categorisation in Table A.1. I count attitudes, institutions, and religion as input and all measures relating to aggregate behaviour in the country of ancestry as output. The United Nation's Gender Empowerment Measure (GEM) is quite clearly an output measure since it combines the behavioural components *proportion of women's seats in parliament*, *share of women in positions of economic decision making*, and *women's share of income earned* (Bose, 2015). The sub-indices of the Social Institutions and Gender Index (SIGI) by Branisa *et al.* (2009, 2013) are explicitly aimed towards comparing gendered institutions (Bose, 2015), and I, therefore, treat them as inputs. As Table 3 shows, most specifications use cultural output measures; Only about a quarter of all effect sizes stem from estimations using inputs. The estimated coefficients of this moderator will inform us about whether cultural inputs or behavioural outputs have a more decisive influence on migrants' decision making.

Another characteristic of the culture measure that could potentially influence results and cause heterogeneity is the time span in which the culture proxy is measured. Fernández & Fogli (2009) argue for using past values of the variable of interest since these more accurately describe the cultural environment in the country of ancestry at the time of emigration. About 41 per cent of all specifications follow this reasoning and use lagged culture variables, measured at the time of migration for first-generation migrants, around the year of birth for the second generation, or, less precise due to data restrictions, one to two decades lagged.

The last set of moderators concerns systematic differences in the estimated *models*. The applied estimation technique (e.g., OLS vs probit) is almost perfectly correlated with the chosen dependent variable (continuous hours vs binary outcomes) and thus not coded as a separate moderator. However, I investigate the influence of included control variables on the individual and the country-of-origin level to capture how individual specifications deal with the empirical challenges mentioned by Fernández (2011), as discussed in Section 2.

To this aim, I code whether specifications control for the respondents' education, area of residence, or their partner's characteristics. These are the potential confounders mentioned by Fernández (2011) as being related to female migrants' labour force participation but also likely in themselves influenced by ancestry culture. We should see systematically different estimates of culture effects in specifications that include these controls if this reasoning is correct. While most specifications (almost 80%) control for education, only a minority includes covariates of partner's education or income, most likely due to missing information in many data sources. Half of the specifications control for the women's state, region, or otherwise defined area of residence.

On the country-of-ancestry level, I code a moderator for whether specifications include a proxy of the unobserved quality of human capital as a variable that is often mentioned as another suspect for omitted variable bias in the literature (Fernández, 2011). If there is a positive correlation between this variable and female migrants' labour force participation as well as a positive relation between quality of human capital and the culture measure (e.g., aggregate female labour force participation in the country of origin), then omitting the human capital variable leads to an overestimation of the culture effect. In this case, we would expect studies that include the covariate to estimate systematically smaller culture effects because their estimates are corrected for this positive omitted variable bias. This would constitute a strong case for always including such a variable in future studies.

Finally, I document whether specifications control for the economic development of countries of origin due to the strong positive relationship between economic growth and gender equality (Falk & Hermle, 2018). I find that 17 per cent of specifications explicitly deal with unobserved human capital differences of the respondents in their data sets and 26 per cent control for GDP.

The fact that moderators can vary both within and between studies results in a hierarchical data structure, which I address by continuing to estimate with standard errors clustered at the study level.

5.2 Meta-Regression

Table 4 shows the results of estimating Equation 6 including the moderators described above. To investigate the robustness of the influences of single moderators, I estimate the model first in the complete set of specifications and then in a series of more homogeneous sub-samples of specifications: Only those which analyse immigrants in the United States, only those that underwent a peer-review process in scientific journals, only specifications reported in economics studies, and finally only those that study immigrants of the 2nd or higher generations.

First, we see that even when accounting for every observable dimension of heterogeneity, there is a positive association between effect sizes' magnitudes and their standard errors, indicating that selective reporting might indeed be an issue in this field of research. Further, the fact that this association is also visible when restricting to studies on immigrants in the United States (Column 2) contradicts the idea that there is a genuine relationship between sample size and estimate of the culture effect that could be driven by differential integration mechanisms in countries with differing sizes of immigrant populations. Additionally, on average, specifications reported in unpublished working papers appear to report lower culture effects than those published in peer-reviewed journal articles. This can be interpreted as further proof of journals', reviewers', or editors' unwillingness to publish unexpected results, but it could also mean that the studies finding lower culture effects are of lower quality that does not survive the rigorous review process.

The other study attributes, publication year, field, and host country, do not exhibit a clear, systematic relationship with the effect sizes' magnitudes, except that the five studies (29 specifications) conducted outside the US and Europe tend to find more negligible culture effects.

Using a binary dependent variable instead of continuous hours seems to lead to smaller estimates, but this relationship is not statistically significant.

Expectedly, migrants of the first generation are more strongly oriented towards ancestry norms than the second and higher generations. There seems to be a negative relationship between mean sample age and effect sizes, suggesting that younger women might be more influenced by ancestry culture. However, the effect is small and not robust across

Table 4: Meta-regression results

	(1) Baseline	(2) US	(3) Journal	(4) Econ	(5) Gen ≥ 2
SE of effect size	1.860***	3.532***	1.596**	2.158**	1.463**
	(0.589)	(0.805)	(0.621)	(0.911)	(0.548)
Ref: Journal					
Working paper	-0.031***	0.007		-0.019*	-0.008
	(0.009)	(0.010)		(0.010)	(0.005)
Publication year	0.000	-0.001**	0.001	-0.001	-0.000
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Ref: Economics					
Sociology	-0.007	0.025***	-0.010		0.003
	(0.010)	(0.006)	(0.012)		(0.005)
Ref: US					
Europe	-0.013		-0.010	-0.005	-0.011**
	(0.012)		(0.012)	(0.025)	(0.005)
Other	-0.027**		-0.029**	-0.008	-0.009**
	(0.010)		(0.010)	(0.010)	(0.004)
Binary dependent variable	-0.004	-0.007	-0.010	-0.010	-0.005
	(0.005)	(0.006)	(0.011)	(0.008)	(0.004)
1st generation	0.021*	0.039***	0.026	0.024*	
	(0.012)	(0.010)	(0.016)	(0.013)	
Mean age	-0.003	-0.002*	-0.004	-0.003	-0.001**
	(0.002)	(0.001)	(0.003)	(0.002)	(0.000)
Ref: Output measure					
Input measure	-0.020***	-0.000	-0.021**	-0.017**	-0.012***
	(0.006)	(0.008)	(0.008)	(0.006)	(0.003)
Lagged culture	-0.032**	-0.020***	-0.046**	-0.030**	-0.014***
	(0.013)	(0.005)	(0.021)	(0.012)	(0.004)
Education	-0.003	-0.023***	-0.004	-0.007	-0.007**
	(0.005)	(0.008)	(0.007)	(0.005)	(0.003)
Partner characteristics	0.030***	0.031***	0.033**	0.015	0.008**
	(0.009)	(0.006)	(0.014)	(0.009)	(0.003)
Area of residence	-0.019***	-0.034***	-0.023***	-0.013	0.000
	(0.007)	(0.006)	(0.008)	(0.008)	(0.002)
Quality of human capital	-0.010**	-0.015***	-0.007*	-0.013**	-0.004***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.001)
GDP	0.003	0.003	0.000	0.003	-0.010***
	(0.005)	(0.005)	(0.007)	(0.006)	(0.002)
Observations	160	93	138	132	116
No. of studies	29	17	23	22	16
adj R^2	0.574	0.830	0.587	0.633	0.594

Notes: Dependent variable: Effect size (r). Standard errors are clustered at the study level. Columns 1 - 5 report results for all primary estimations, for those conducted with immigrants to the United States, those published in peer-reviewed journals, those conducted by economists, and those analysing immigrants of the 2nd and higher generation. *** p<0.01, ** p<0.05, * p<0.10

all subsamples.

Measuring country-of-origin culture with input instead of output measures and using lagged culture proxies show a persistent negative correlation with effect size magnitude. Taken together, these two coefficients suggest that female migrants are rather influenced by the *contemporary behaviour* of their peers in the country of origin than past behaviours, stated values, religion, or gendered institutions.

Turning to the influence of model characteristics, we see significant negative effects of controlling for education and area of residence, indicating that omitting these variables could lead to a positive bias in the estimated culture effect. However, the respective coefficients are not statistically significant across all subsamples. Meanwhile, the effect of controlling for partner characteristics goes in the opposite direction: Specifications that include this covariate find *larger* effect sizes, on average, suggesting underestimated culture effects when omitting it. For a negative omitted variable bias to occur, there must be a negative correlation either between the partner's education or income and the culture measure or between partner characteristics and the outcome variable. The latter would be consistent with the intra-household division of labour: The individual migrant woman works less the higher the education or income of her partner. Whatever the reasons for the negative bias, the finding underlines the significance of accounting for partner characteristics when estimating cultural effects on women's labour force participation.

Finally, controlling for the quality of human capital does robustly lead to moderately smaller estimates of culture effects, underlining the importance of this potentially omitted variable. As Fernández (2011) points out, migrants from one country of origin share more than a common ancestry culture. From these results, it seems that the quality of human capital is one such shared factor and omitting it from the estimation leads to overestimated culture effects. Economic development in the country of origin, on the contrary, does not seem to play an essential role as a control variable.

According to the adjusted R^2 values at the bottom of the table, between 57 and 83 per cent of the variation in primary estimates are explained by included moderators, with the highest explanatory power obtained in the sub-sample of specifications that look at immigrants in the United States. In the next section, I use my unique meta-data to explore an additional source of heterogeneity that cannot be tested in primary analyses: the influence of differential sets of ancestry countries included in the analyses.

6 The Role of Country-of-Origin Composition

Another important potential source of variation in estimated culture effects previously mostly neglected is the set of countries of origin included in the analysis. Depending on the studied host country, utilised data, and time frame of the analysis, there are considerable differences in how many and which countries of ancestry are included. Even within the papers focused on immigrants in the United States, the number of countries of origin ranges between seven (from Buitrago 2015 who focuses on immigrants from Latin American countries) and 131 (Apgar & McManus 2018 obtain a large sample by pooling across 20 years of data from the Current Population Survey (CPS) and impose no restrictions on countries of origin). As I established in Section 1, differential sets of origin countries that are considered can imply different selection patterns for immigration. Before investigating this relationship in detail in the meta-data, I provide a brief review of the related literature on cultural selection of immigrants and discuss how this phenomenon relates to estimating culture effects in immigrant samples.

6.1 Cultural Selection of Immigrants

The fact that migrants are rarely a random sample from their country of origin is commonly acknowledged in the migration literature. Many studies show that migrants are selected in terms of education and skills (e.g., Docquier *et al.*, 2007; Grogger & Hanson, 2011; Belot & Hatton, 2012).

Selection based on cultural values has been less present in the public and academic discussions, despite mounting empirical evidence that migrants also differ from stayers in their preferences, norms, attitudes, and beliefs. For example, van Dalen *et al.* (2005) demonstrate that aspiring migrants from Ghana, Morocco, Egypt, and Senegal hold values that are more "in keeping with the Western world" (p. 774) than their compatriots who report no intention to move abroad. Several other studies document migrant selectivity on diverse cultural dimensions, such as risk aversion (Jaeger *et al.*, 2010), moral values (Casari *et al.*, 2018; Turati, 2021), political attitudes (Berlinschi & Fidrmuc, 2018), individualism (Knudsen, 2019), and religiosity (Docquier *et al.*, 2020).[11]

Most relevant for female migrants' labour force participation studies is research on the relationship between gender equality and migrant selectivity. There are two aspects to this relationship, namely, (i) the influence of gender equality at the country level on emigration rates of differentially skilled women and men and (ii) migrants' selection in their attitudes towards gender equality. On the macro level, the former relationship is

[11] As a consequence, migration has important implications for cross-country cultural change, as Rapoport *et al.* (2020) demonstrate both theoretically and empirically.

hard to pinpoint because gender inequality can act as a push factor, incentivising women to leave the country and, at the same time, restricting their freedom of movement and, therefore, their migration decision. Macro-level studies thus provide mixed evidence on the link between gender equality and female migration, where the result seems to depend on the specific dimension of (in)equality (e.g., labour market outcomes vs formal and informal institutions) that is considered (Bang & Mitra, 2011; Naghsh Nejad & Young, 2014; Baudassé & Bazillier, 2014; Ferrant & Tuccio, 2015). On the micro-level, Ruyssen & Salomone (2018) use data from the Gallup World Polls (GWP) to examine the relationship between *perceived* discrimination and migration *intentions* of women in 148 countries. They find a strong effect of perceived gender discrimination on women's stated intention to leave the country.

To my knowledge, the study by Docquier *et al.* (2020) constitutes the only specific test of migrant selection based on individual gender *attitudes*. The authors also use GWP data to compare aspiring emigrants to those who prefer to stay in their country of birth by their levels of religiosity and attitudes towards gender equality. The analysis is restricted to the native working-age population in Middle Eastern and North African (MENA) countries because citizens of the MENA region hold less gender-egalitarian attitudes on average and are more religious than common regions of destination for migrants from these countries. Concerning gender norms, the authors find more gender-egalitarian attitudes among aspiring migrants compared to non-migrants for the young (between 15 and 30 years of age), for single women, for people living in rural areas, and in countries where the Shia branch of Islam dominates the Sunnis.[12]

Judging from the limited literature on this issue, it seems plausible that female migrants from countries with high levels of gender discrimination, especially the highly skilled ones, are positively selected on attitudes favouring gender equality and are therefore not representative of the related cultural norms in their country of ancestry. Accordingly, analyses of cultural effects on female labour market integration might reach different conclusions, depending on the distribution of societal gender norms in the sets of origin countries that are included in their investigations. If this were true, then some part of the variance in results among these studies might not stem from actual differences in the underlying relationship but from differential sets of origin countries that are included in the studies' samples - a decision that is data-driven in most cases and implicitly treated as random

[12] A different but related approach is taken by Fuchs *et al.* (2021) who compare attitudes towards gender equality (and other social values) between natives and refugees from seven different countries in Germany. Controlling for individual characteristics, they find that migrants from Afghanistan, Eritrea, Syria, Iran, and Iraq show *stronger* support for economic gender equality than native Germans. Against the background of well-established cross-country differences in these attitudes between Germany and the analysed countries of ancestry of the opposite direction, the authors argue that their findings might be interpreted as evidence of positive selection on gender norms. However, they cannot rule out social desirability or cultural (over-)assimilation as drivers of their results.

in the literature.

In Beblo *et al.* (2020b), my co-authors and I present a theoretical model to illustrate the importance of country-of-origin gender equality when studying the labour force participation of immigrants. Our formal model of labour supply is based on an identity economics framework (Akerlof & Kranton, 2000) where individual utility is determined by effort in the labour market and the monetary returns to that effort as well as an identity component. The identity component is a function of the difference between individual effort and the "ideal" effort level for each gender, as determined by social norms about the appropriate roles of women and men in the labour market. The central assumption of the identity economics framework is that individuals experience a utility penalty when their behaviour deviates from the socially prescribed "ideal". When allowing people to choose whether to live in a society with conservative or progressive norms through migration, we show that women from countries with restrictive gender norms gain more from migrating when their returns to labour market effort are high. Thus, women with high labour market aptitude will want to migrate to countries with higher gender equality, while those with low aptitude prefer staying. Consequently, women migrating from countries with (relatively) gender-equal norms are equally likely to be of high or low labour market aptitude. In contrast, those from countries with low gender equality are positively selected in terms of their aptitude, and we would thus expect their behaviour to reflect ancestry norms to a lower degree.

In Beblo *et al.* (2020a), we provide first circumstantial evidence for this notion. We regress the labour force participation and fertility of second-generation immigrants in European host countries on labour force participation and fertility rates in their countries of ancestry. Economically and statistically significant effects of ancestry culture on female migrants' labour force participation and fertility are found only for women descending from countries in which gender equality is relatively high. The effect is far less pronounced or absent among women with ancestry from low gender equality countries, measured by the United Nations' Gender Empowerment Measure (GEM). These findings fit well with the notion of female migrants from low gender equality countries being positively selected in their labour market orientation and thus working *more* than their counterparts in the ancestry country.

Discussions of such systematic differences in the strength of the measured association between migrant behaviour and ancestry culture are to date mostly absent from the literature on ancestry effects and warrant further examination. This examination also provides an empirical contribution to the literature on the cultural selection of immigrants where the idea of systematic differences across countries of origin in the degree of cultural selectivity of emigrants does not enter the discussion either. Until now, most studies on cultural selection focus on one or few individual source countries and include country fixed effects in the cases with multiple origins to control for all country differences. However, at the aggregate level, these differences might be meaningful for migrant selectivity, as the exercise by Beblo *et al.* (2020a) demonstrates.

6.2 Testing Cultural Selection with Metadata

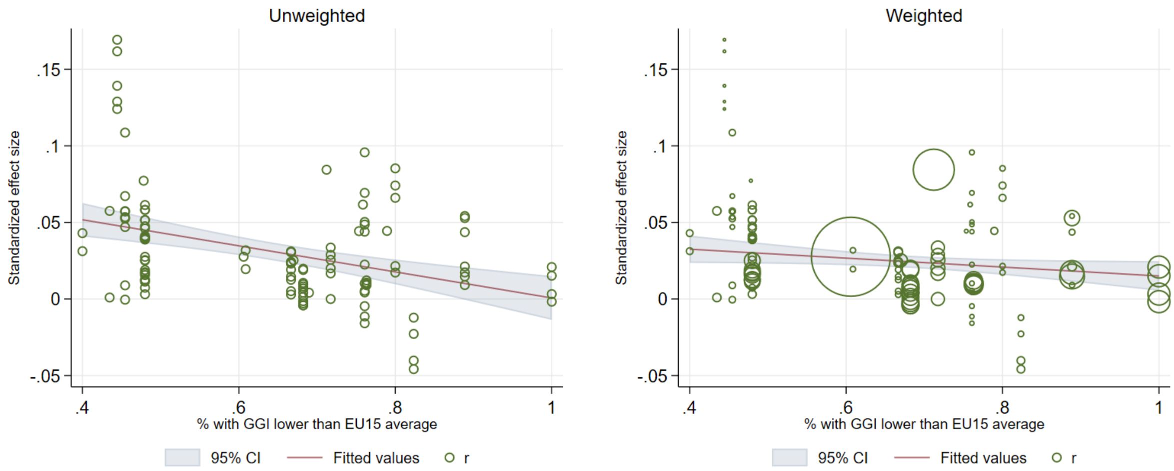
I complement the meta-data set with an indicator of gender equality for each country of ancestry included in each primary study, to test the relationship between culture effects and country-of-ancestry gender equality empirically and to answer the last research question *Could selective migration bias the average culture effect?*

I use the Gender Gap Index (GGI) provided by the World Economic Forum (WEF) (2021) as a comprehensive index of gender parity across different dimensions (political empowerment, economic participation, educational attainment, health) with high coverage. Another reason for choosing the GGI above alternative measures of gender equality is that it is not used as a culture measure in any of the primary studies. Including information into the meta-analysis that is already part of the primary regression might lead to endogeneity issues. By employing the GGI, I hope to mitigate this concern since the indicator combines several different dimensions of gender equality. Even though, for example, female LFP rates, that are regularly used as culture measures, also enter into the calculation of the GGI, this is only one of a rather long and comprehensive list of components.[13] Theoretically, the GGI ranges between zero and one, with higher scores indicating higher gender equality and one being the theoretical ideal of gender parity (World Economic Forum (WEF), 2021, p. 75). Data has been published yearly since 2006.

Where possible, I match the gender equality information to each specification's countries of origin for the year of (primary) data measurement. For studies where the primary data was collected before 2006, I assign the GGI scores from 2006 as the earliest available data point. I take the (rounded) mean year of data measurement for specifications with pooled data across multiple periods. For example, Mocan (2019) pools data from 2004 to 2013, so these countries of origin are assigned gender indicators from 2009 - rounded from the average of 2008.5. Huh (2018) works with data from 2006, so I assign each of her 43 countries of origin the gender equality info from 2006. For ten studies containing 45 specifications, information on included countries of ancestry is missing, so these studies have to be excluded from the subsequent analyses.[14]

[13] See World Economic Forum (WEF) (2021) for details on the components and the calculation.
[14] I unsuccessfully contacted the authors in an effort to obtain the missing information.

Figure 4: Scatter plots of effect sizes against shares of countries of origin with low gender equality



(a) Unweighted **(b)** Weighted by inverse variance

Ideally, I would like to calculate the share of immigrants from ancestries with low gender equality scores for each specification. Unfortunately, many papers do not report numbers of observations by countries of origin, leading to a very small number of specifications (about 70) for which I can obtain this value. Alternatively, I calculate the share of countries of origin with low gender equality that are included in each specification.

Starting from the observation that in the seminal paper by Fernández & Fogli (2009), 17 out of the 25 included countries of ancestry are European, I use the average GGI score among EU15-countries, i.e., the 15 nations that constituted the European Union prior to the accession of 2004 (OECD, 2007), as a threshold for high gender equality.[15] For each specification, I calculate the share of included countries of origin with GGI scores below the EU15 average in the year of primary data collection. This share ranges from 40 percent in two specifications reported by Fernández (2007) to 100 percent in the four specifications provided by Buitrago (2015) who focuses on second-generation immigrants with hispanic origins in the United States. The mean across all 123 specifications with non-missing values is 0.65 with a standard deviation of 0.16.

To gather a first impression of the relationship between estimates of the culture effect on female labour force participation and shares of countries of ancestry with low gender equality, Figure 4a presents a scatter plot where effect sizes are plotted against the percentage of countries of origin with low gender equality included in the specification underlying each effect size. The line shows predicted values of effect sizes and the associated 95% confidence intervals. From the scatter plot and fitted line, we see that effect sizes

[15] Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden, and the United Kingdom.

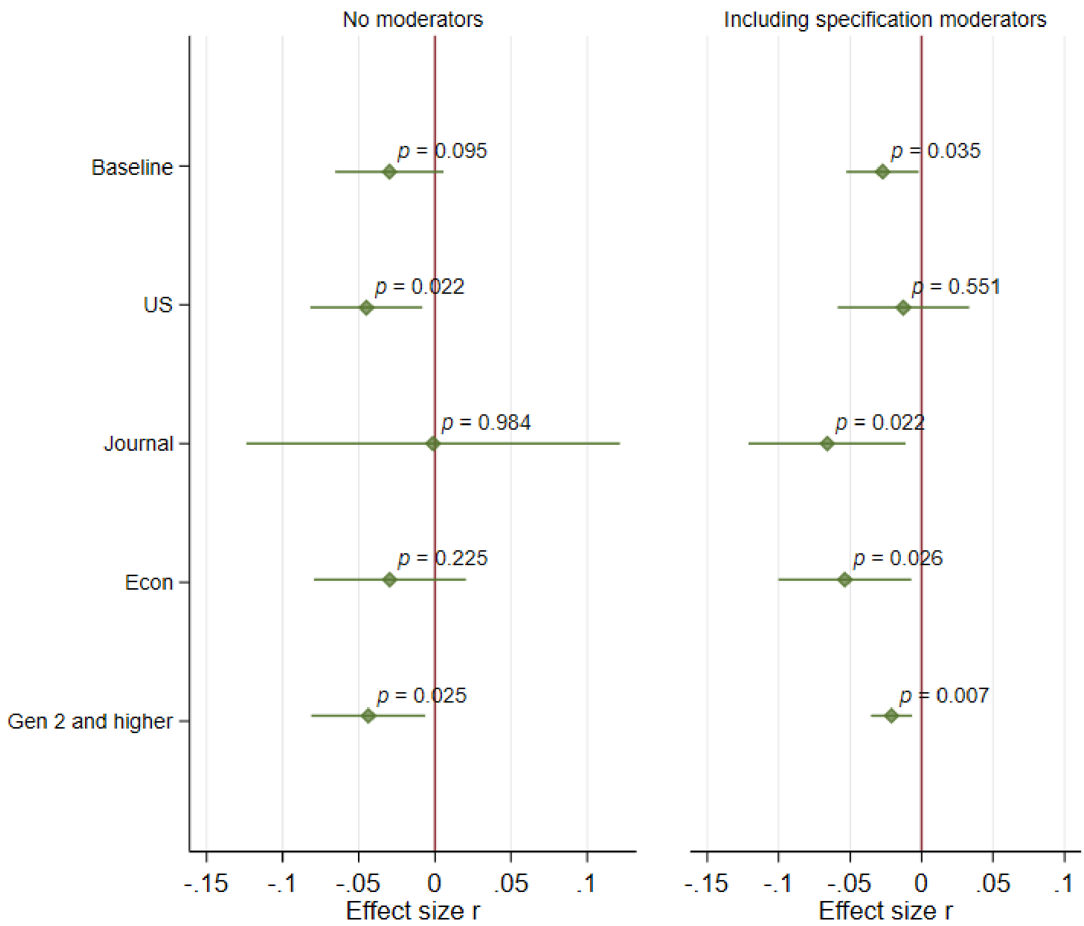
are smaller, the higher the share of countries of origin with low gender equality in the primary data set. This pattern fits the hypothesis that women from countries of origin with restrictive gender norms behave less in accordance with country-of-ancestry norms and that these behavioural differences are reflected in systematically lower estimates of culture effects. In Figure 4b, the identical relationship is plotted when weighting each effect size by its inverse variance, as in the regression analyses in Section 5. Although the fitted line slope decreases compared to the unweighted version, the relationship between shares of countries of origin with low gender equality and estimated culture effects on female labour force participation is still clearly negative.

To substantiate the analysis of this relationship, I examine it in a meta-regression framework similar to those in Section 5, and regress effect sizes (standardized to partial correlation coefficients) on the share of countries of origin with low gender equality while controlling for precision. I continue to estimate WLS models, using the effect sizes' inverse variances as weights, with cluster-robust standard errors (see Sections 4 and 5).

The left-hand panel of Figure 5 reports the coefficients of the variable capturing the share of countries of origin with low gender equality in the primary data with associated 95%-confidence intervals and p-values when performing the analysis in the same sub-samples as in the previous section. The first coefficient relates to the whole sample of specifications from studies that include information on countries of origin (N=123). We see that the meta-regression analyses corroborate the conclusion from the scatter plots: Effect sizes that stem from estimations with data that contains relatively more countries with low gender equality are systematically smaller than those estimated in data with low shares of countries of origin with restrictive gender norms. This effect is even more pronounced when restricting the sample to specifications that analyse immigrant behaviour in the United States (N=66) and only looking at specifications focusing on immigrants of the second or higher generations (N=102). When restricting the sample to specifications stemming from Economics outlets (N=101), however, the H0 that the effect is zero cannot be rejected. In the subsample of specifications that stem from studies published in peer-reviewed journals (N=103), the point estimate is almost zero, but it is estimated with high uncertainty.

The large standard errors could be caused by high levels of heterogeneity *within* the examined subsamples. To explore this possibility, I additionally introduce the specification moderators presented in Section 5 in the right-hand panel of Figure 5. Since the set of countries of origin, and consequently the identifying variable in this analysis, shows almost no variation at the study level, I cannot include the study moderators. We see that regressing the model with specification moderators leads to more pronounced point estimates on the variable of interest, on average. The coefficients retrieved in all five

Figure 5: Coefficients of the variable capturing the share of countries of origin with low gender equality in different meta-samples



samples are now negative, and the confidence intervals are smaller.[16] However, in the subsample of studies conducted in the United States, the coefficient of interest is substantially smaller and statistically insignificant when including moderators. Apparently, within this more homogeneous subsample, there is no significant correlation between estimated culture effects and gender equality at the country of ancestry level once other sources of heterogeneity, like the choice of the dependent variable or the included controls, are taken into account.

Summing up, I find systematically smaller estimates of cultural effects on female immigrants' labour force participation in studies that include relatively high shares of countries of ancestry with low levels of gender equality. This finding fits the conjecture that female immigrants from low-gender equality countries are selected on cultural norms, i.e., they are more likely to reject the gender norms in their countries of origin and, consequently, their behaviour does not adhere to these norms. Thus, selective migration based on cultural values can bias estimates of cultural effects in applications of the socio-epidemiological approach.

7 Conclusion

A large and growing body of empirical literature in economics and sociology is researching the influence of cultural norms on immigrants' behaviour by drawing partial correlations between migrants' outcomes in the destination and aggregate characteristics of their countries of ancestry. The present study provides an interdisciplinary quantitative review of this literature focusing on studies of female labour force participation. This comprehensive synthesis complements the existing narrative reviews and contributes to the theoretical and empirical advancement of the analysis of culture.

Integrating primary results confirms the presence of a positive and robust correlation. However, it also shows that the relationship might be weaker than previously thought since there is evidence of selective reporting: Assuming the symmetric distribution of obtained results, there is a lack of negative or small effect sizes among results that are *published* - as journal articles or as working papers - and thus included in the meta-analysis. Results from smaller studies that are estimated with relatively low precision seem more likely to be published when they obtain large and positive effect sizes, i.e., when they confirm the collective priors of the literature. Correcting this bias leads to an estimate of the overall culture effect of about one-third the size of the uncorrected estimate (partial correlation of 0.01 compared to 0.03).

[16] Since all meta-regressions in this section are conducted in relatively small samples, hypothesis tests are performed with quite low statistical power and one should interpret the coefficients and confidence intervals with the appropriate caution.

Additionally, this literature contains considerable methodological heterogeneity, calling for meta-regression analyses to test the influence of study attributes on obtained results. The results of these meta-regressions point towards some influential study features: On average, immigrant women's labour force participation is most strongly impacted by aggregate behaviour of their contemporaries in the country of origin, compared to past behaviours or aggregate attitudes and gendered institutions. Additionally, omitted variables bias poses a real threat to the estimation of unbiased culture effects and, consequently, carefully choosing relevant control variables is of great essence. The meta-regression results point out education, area of residence, partner characteristics, and quality of human capital, in particular, but further candidates for omitted variable bias are plausible that have not yet entered the primary literature and, therefore, cannot be included in the present meta-study. For example, migrants of common origin might face similar barriers or support to labour market entry, like discrimination, heritage language skills, or ethnic social networks. These mechanisms require further research, both primary and meta-analytically.

Another important finding of the present study concerns the possible influence of selective migration based on cultural values: Larger culture effects are obtained by estimations that include fewer countries with low gender equality, suggesting that the behaviour of women from low gender equality countries is less related to ancestry culture than that of women from high gender equality countries. The underlying relationship might be women with high labour market orientation actively selecting out of restrictive environments. While this negative bias does not pose a threat to the main conclusion of the literature, that *culture matters* (Fernández, 2011), recognising that it does not matter to the same degree for every immigrant seems important when drawing policy conclusions.

Innovative advancements within the socio-called epidemiological approach deal with this challenge in creative ways (e.g., Finseraas & Kotsadam, 2017) and in future research, these extensions should probably receive more attention. However, since these approaches are rather demanding of the data (e.g., panel structure, rich information on household composition), there might also be merit in finding alternative ways of adequately accounting for selection in these contexts. Additionally, the analyses presented here seem to call for more empirical research in the cultural selection of immigration where the degree of selectivity differs systematically across countries or otherwise differentiated geographic or cultural regions. So far, these differences are being controlled for in research on this matter rather than being explicitly investigated.

All in all, this paper attempts to bring together two new and vital areas of research:

Culture effects on immigrants' decision making and cultural selection of immigration. While the meta-perspective offers new and intriguing insights for both areas, it also has limitations. Due to missing information on numbers of respondents by origin in many studies, I could only calculate shares of countries of ancestry with low gender equality included in every specification, where shares of respondents from these countries would have been desirable. More comprehensive efforts to obtain missing data from authors of primary studies might provide a remedy for this shortcoming in the future. Additionally, as statistical power is relatively low in the meta-regression analyses, there might be systematic relationships where I did not find statistically significant results, especially when splitting the sample into even smaller subsamples. Null results in this study, therefore, have to be interpreted with particular caution.

Nonetheless, robust systematic relationships between study features and results could be identified even in this relatively small sample. These findings provide invaluable insights for future applications of the socio-epidemiological approach and advance the analysis of cultural effects on economic decision making. Most importantly, the present analysis points out three sources of biases in estimated culture effects that researchers have to be aware of: First, omitted variables bias can lead to over- and underestimating the influence of culture. The study also shows that this bias can be mitigated effectively by carefully choosing appropriate control variables. Second, selective migration based on cultural norms presents a downward bias on estimated cultural effects in this context. This influence is harder to appease, and more research is needed on the underlying relationship. At the very least, researchers should make an effort to check for potential selection issues in their data. Third and finally, selective reporting of results exerts a positive bias on the overall findings of the literature. This source of error can only be corrected in the publication process by collective efforts from researchers, reviewers, and editors. Awareness of the issue is probably the most fundamental prerequisite for solving it. Hopefully, this meta-analysis provides the first step towards this goal.

References

Adnan, Wifag. 2019. *Culture or Context? Revisiting the Role of Culture on Economic Outcomes.* IZA Discussion Paper 12692. IZA Institute of Labor Economics, Bonn.

Akerlof, George A, & Kranton, Rachel E. 2000. Economics and Identity. *The Quarterly Journal of Economics*, **115**(3), 715–753.

Alesina, Alberto, & Giuliano, Paola. 2010. The Power of the Family. *Journal of Economic growth*, **15**(2), 93–125.

Alesina, Alberto, & Giuliano, Paola. 2011. Family Ties and Political Participation. *Journal of the European Economic Association*, **9**(5), 817–839.

Antecol, Heather. 2000. An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates. *Labour Economics*, **7**(4), 409–426.

Apgar, Lauren, & McManus, Patricia A. 2018. Cultural Persistence and Labor Force Participation among Partnered Second-Generation Women in the United States. *Social Forces*, **98**(1), 211–244.

Bang, James T, & Mitra, Aniruddha. 2011. Gender Bias and the Female Brain Drain. *Applied Economics Letters*, **18**(9), 829–833.

Baudassé, Thierry, & Bazillier, Rémi. 2014. Gender Inequality and Emigration: Push Factor or Selection Process? *International Economics*, **139**, 19–47.

Beblo, Miriam, Görges, Luise, & Markowsky, Eva. 2020a. Gender Equality as a Confounder in the Epidemiological Approach. *Economics Bulletin*, **40**(2), 1292–1299.

Beblo, Miriam, Görges, Luise, & Markowsky, Eva. 2020b. Gender Matters in Language and Economic Behaviour – Can We Measure a Causal Cognition Effect of Speaking? *Labour Economics*, 101850.

Belot, Michèle V. K., & Hatton, Timothy J. 2012. Immigrant Selection in the OECD: Immigrant Selection in the OECD. *The Scandinavian Journal of Economics*, **114**(4), 1105–1128.

Berlinschi, Ruxanda, & Fidrmuc, Jan. 2018. *Comfort and Conformity: A Culture-based Theory of Migration.* CESifo Working Papers 7294. München.

Blau, Francine D., & Kahn, Lawrence M. 2015. Substitution between Individual and Source Country Characteristics: Social Capital, Culture, and US Labor Market Outcomes among Immigrant Women. *Journal of Human Capital*, **9**(4), 439–482.

Blau, Francine D, Kahn, Lawrence M, & Papps, Kerry L. 2011. Gender, Source Country Characteristics, and Labor Market Assimilation Among Immigrants. *The Review of Economics and Statistics*, **93**(1), 43–58.

Blau, Francine D., Kahn, Lawrence M., Liu, Albert Yung-Hsu, & Papps, Kerry L. 2013. The Transmission of Women's Fertility, Human Capital, and Work Orientation Across Immigrant Generations. *Journal of Population Economics*, **26**(2), 405–435.

Borenstein, Michael, Hedges, Larry V., Higgins, Julian PT, & Rothstein, Hannah R. 2009. *Introduction to Meta-Analysis.* 2nd ed. edn. Oxford: John Wiley & Sons.

Borjas, George J. 1987. Self-Selection and the Earnings of Immigrants. *The American Economic Review*, **77**(4), 531–553.

Bose, Christine E. 2015. Patterns of Global Gender Inequalities and Regional Gender Regimes. *Gender & Society*, **29**(6), 767–791.

Branisa, Boris, Klasen, Stephan, & Ziegler, Maria. 2009. *The Construction of the Social Institutions and Gender Index (SIGI).* IAI Discussion Papers 184. Georg-August-Universität Göttingen, Ibero-American Institute for Economic Research (IAI), Göttingen.

Branisa, Boris, Klasen, Stephan, & Ziegler, Maria. 2013. Gender Inequality in Social Institutions and Gendered Development Outcomes. *World Development*, **45**(May), 252–268.

Bredtmann, Julia, & Otten, Sebastian. 2013. *The Role of Source- and Host-Country Characteristics in Female Immigrant Labor Supply.* Munich Personal RePec Archive 68213.

Brodeur, Abel, Lé, Mathias, Sangnier, Marc, & Zylberberg, Yanos. 2016. Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics*, **8**(1), 1–32.

Buitrago, Manuel. 2015. *Culture, Employment, and Volatility: Three Essays on Hispanic Labor.* Ph.D. thesis, American University, Washington D.C.

Card, David, & Krueger, Alan B. 1995. Time-Series Minimum-Wage Studies: A Meta-Analysis. *The American Economic Review*, **85**(2), 238–243.

Casari, Marco, Ichino, Andrea, Michaeli, Moti, De Paola, Maria, Marandola, Ginevra, & Scoppa, Vincenzo. 2018. *Civicness Drain.* IZA Discussion Paper 11955. IZA Institute of Labor Economics, Bonn.

Chiswick, Barry. 1999. Are Immigrants Favorably Self-Selected? *American Economic Review*, **89**(2), 181–185.

DerSimonian, Rebecca, & Laird, Nan. 1986. Meta-Analysis in Clinical Trials. *Controlled Clinical Trials*, **7**(3), 177–188.

Dinesen, Peter Thisted. 2013. Where You Come From or Where You Live? Examining the Cultural and Institutional Explanation of Generalized Trust Using Migration as a Natural Experiment. *European Sociological Review*, **29**(1), 114–128.

Docquier, Frédéric, Lohest, Olivier, & Marfouk, Abdeslam. 2007. Brain Drain in Developing Countries. *The World Bank Economic Review*, **21**(2), 193–218.

Docquier, Frédéric, Tansel, Aysit, & Turati, Riccardo. 2020. Do Emigrants Self-Select Along Cultural Traits? Evidence from the MENA Countries. *International Migration Review*, **54**(2), 388–422.

Eylem Gevrek, Z., Gevrek, Deniz, & Gupta, Sonam. 2013. Culture, Intermarriage, and Immigrant Women's Labor Supply. *International Migration*, **51**(6), 146–167.

Falk, Armin, & Hermle, Johannes. 2018. Relationship of Gender Differences in Preferences to Economic Development and Gender Equality. *Science*, **362**(6412).

Fernández, Raquel. 2007. Women, Work, and Culture. *Journal of the European Economic Association*, **5**(2-3), 305–332.

Fernández, Raquel. 2011. Does Culture Matter? Handbook of Social Economics, vol. 1A. Amsterdam: North-Holland.

Fernández, Raquel, & Fogli, Alessandra. 2009. Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *American Economic Journal: Macroeconomics*, **1**(1), 146–177.

Ferrant, Gae Lle, & Tuccio, Michele. 2015. South–South Migration and Discrimination Against Women in Social Institutions: A Two-way Relationship. *World Development*, **72**, 240–254.

Finseraas, Henning, & Kotsadam, Andreas. 2017. Ancestry Culture and Female Employment—An Analysis Using Second-Generation Siblings. *European Sociological Review*, **33**(3), 382–392.

Frank, Kristyn, & Hou, Feng. 2016. Beyond Culture: Source Country Female Labour Force Participation and the Earnings of Immigrant Women. *Work, Employment and Society*, **30**(3), 410–435.

Fuchs, Lukas M., Fan, Yu, & Scheve, Christian. 2021. Value Differences between Refugees and German Citizens: Insights from a Representative Survey. *International Migration*, **59**(5), 59–81.

Fuchs-Schündeln, Nicola, Masella, Paolo, & Paule-Paludkiewicz, Hannah. 2017. Cultural Determinants of Household Saving Behavior. *Journal of Money, Credit and Banking*, **52**(5), 1035–1070.

Gay, Victor, Hicks, Daniel L., Santacreu-Vasut, Estefania, & Shoham, Amir. 2018. Decomposing Culture: An Analysis of Gender, Language, and Labor Supply in the Household. *Review of Economics of the Household*, **16**(4), 879–909.

Gershman, Boris. 2017. Long-Run Development and the New Cultural Economics. *Pages 221–262 of:* Cervelatti, Matteo, & Sunde, Uwe (eds), *Demographic Change and Long-Run Development*. CESifo Seminar Series, vol. 221. Cambridge & London: MIT Press.

Grogger, Jeffrey, & Hanson, Gordon H. 2011. Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics*, **95**(1), 42–57.

Guiso, Luigi, Sapienza, Paola, & Zingales, Luigi. 2006. Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives*, **20**(2), 23–48.

Hajdu, Gábor, & Hajdu, Tamás. 2016. The Impact of Culture on Well-Being: Evidence from a Natural Experiment. *Journal of Happiness Studies*, **17**(3), 1089–1110.

Hamermesh, Daniel S. 2007. Viewpoint: Replication in Economics. *Canadian Journal of Economics/Revue canadienne d'économique*, **40**(3), 715–733.

He, Qian, & Gerber, Theodore P. 2020. Origin-Country Culture, Migration Sequencing, and Female Employment: Variations among Immigrant Women in the United States. *International Migration Review*, **54**(1), 233–261.

Huh, Yunsun. 2018. Family Typology and Gender Empowerment: The Labour Market Performance of Married Immigrants. *Journal of Population Research*, **35**, 237–288.

Imai, Taisuke, Rutter, Tom A, & Camerer, Colin F. 2021. Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets. *The Economic Journal*, **131**(636), 1788–1814.

Jaeger, David A, Dohmen, Thomas, Falk, Armin, Huffman, David, Sunde, Uwe, & Bonin, Holger. 2010. Direct Evidence on Risk Attitudes and Migration. *The Review of Economics and Statistics*, **92**(3), 684–689.

Kelley, Jonathan, & McAllister, Ian. 1984. Immigrants, Socio-Economic Attainment, and Politics in Australia. *The British Journal of Sociology*, **35**(3), 387–405.

Kessler, Anke S, & Milligan, Kevin. 2021. Acculturation, Education, and Gender Roles: Evidence from Canada. *Oxford Economic Papers*, **73**(2), 509–533.

Khoudja, Yassine, & Platt, Lucinda. 2018. Labour Market Entries and Exits of Women from Different Origin Countries in the Uk. *Social Science Research*, **69**, 1–18.

Kislev, Elyakim. 2017. Deciphering the 'Ethnic Penalty'of Immigrants in Western Europe: A Cross-Classified Multilevel Analysis. *Social Indicators Research*, **134**(2), 725–745.

Knudsen, Anne Sofie Beck. 2019. *Those Who Stayed: Individualism, Self-Selection and Cultural Change During the Age of Mass Migration.* Mimeo. Department of Economic History, Lund University.

Köbrich León, Anja. 2013. *Does Cultural Heritage Affect Employment Decisions–Empirical Evidence for First-and Second-Generation Immigrants in Germany.* SOEP Papers on Multidisciplinary Panel Data Research 553. German Socio-Economic Panel Study (SOEP) DIW, Berlin.

Kok, Suzanne, Bosch, Nicole, Deelen, Anja, & Euwals, Rob. 2011. *Migrant Women on the Labour Market: On the Role of Home- and Host-Country Participation.* IZA Discussion Paper 5817. IZA Institute of Labor Economics, Bonn.

Liu, Yang, & Hagiwara, Risa. 2020. *Labor Force Participation of Married Female Immigrants: Evidence from a Low Female-LFPR Host Country.* RIETI Discussion Paper Series 20-E-019. The Research Institute of Economy, Trade and Industry.

Ljunge, Martin. 2016. Migrants, Health, and Happiness: Evidence That Health Assessments Travel with Migrants and Predict Well-Being. *Economics & Human Biology*, **22**, 35–46.

McManus, Patricia A., & Apgar, Lauren. 2019. Parental Origins, Mixed Unions, and the Labor Supply of Second-Generation Women in the United States. *Demography*, **56**(1), 49–73.

Mocan, Naci. 2019. Taxes and Culture of Leisure: Impact on Labor Supply in Europe. *Journal of Comparative Economics*, **47**(3), 618–639.

Moschion, Julie, & Tabasso, Domenico. 2014. Trust of Second-Generation Immigrants: Intergenerational Transmission or Cultural Assimilation? *IZA Journal of Migration*, **3**(10), 1–30.

Muchomba, Felix M., Jiang, Nan, & Kaushal, Neeraj. 2020. Culture, Labor Supply, and Fertility Across Immigrant Generations in the United States. *Feminist Economics*, **26**(1), 154–178.

Naghsh Nejad, Maryam, & Young, Andrew T. 2014. *Female Brain Drains and Women's Rights Gaps: A Gravity Model Analysis of Bilateral Migration Flows*. IZA Discussion Paper 8067. IZA Institute of Labor Economics, Bonn.

Nollenberger, Natalia, Rodríguez-Planas, Núria, & Sevilla, Almudena. 2016. The Math Gender Gap: The Role of Culture. *American Economic Review*, **106**(5), 257–261.

OECD. 2007. *EU15 Definition*. https://stats.oecd.org/glossary/detail.asp?ID=6805 (accessed 2021-08-06).

Polavieja, Javier G. 2015. Capturing Culture: A New Method to Estimate Exogenous Cultural Effects Using Migrant Populations. *American Sociological Review*, **80**(1), 166–191.

Rapoport, Hillel, Sardoschau, Sulin, & Silve, Arthur. 2020. *Migration and Cultural Change*. CESifo Working Papers 8547. Munich Society for the Promotion of Economic Research - CESifo, München.

Reimers, Cordelia W. 1985. Cultural Differences in Labor Force Participation Among Married Women. *The American Economic Review*, **75**(2), 251–255.

Ruyssen, Ilse, & Salomone, Sara. 2018. Female Migration: A Way Out of Discrimination? *Journal of Development Economics*, **130**, 224–241.

Salari, Mahmoud. 2016. *Essays in Applied Economics and Data Analysis*. Ph.D. thesis, Texas Tech University, Lubbock.

Salari, Mahmoud. 2020. Culture and Heritage Language: A Study of Female Labor Force Participation. *Review of Economics of the Household*, **18**(2), 285–306.

Scoppa, Vincenzo, & Stranges, Manuela. 2019. Cultural Values and Decision to Work of Immigrant Women in Italy. *LABOUR*, **33**(1), 101–123.

Sribney, William, & Wiggins, Vince. 2021. *Faq: Std Errors, Confidence Intervals, & Significance Tests*. https://www.stata.com/support/faqs/statistics/delta-rule/ (accessed 2021-07-28).

Stanley, T. D., & Doucouliagos, Hristos. 2014. Meta-Regression Approximations to Reduce Publication Selection Bias. *Research Synthesis Methods*, **5**(1), 60–78.

Stanley, T. D., & Doucouliagos, Hristos. 2015. Neither Fixed nor Random: Weighted Least Squares Meta-Analysis. *Statistics in Medicine*, **34**(13), 2116–2127.

Stanley, T. D., & Doucouliagos, Hristos. 2017. Neither Fixed nor Random: Weighted Least Squares Meta-Regression. *Research Synthesis Methods*, **8**(1), 19–42.

Stanley, Tom D, & Doucouliagos, Hristos. 2012. *Meta-Regression Analysis in Economics and Business*. Routledge Adances in Research Methods, no. 5. London, New York: Routledge.

Stichnoth, Holger, & Yeter, Mustafa. 2016. Cultural Influences on the Fertility Behavior of First- and Second-Generation Immigrants. *Journal of Demographic Economics*, **82**(3), 281–314.

Turati, Riccardo. 2021. *Do You Want to Migrate to the United States? Migration Intentions and Cultural Traits in Latin America.* Working Paper. Departament d´Economia Aplicada, Universitat Autònoma de Barcelona.

van Dalen, Hendrik P., Groenewold, George, & Schoorl, Jeannette J. 2005. Out of Africa: What Drives the Pressure to Emigrate? *Journal of Population Economics*, **18**(4), 741–778.

van Tubergen, Frank. 2006. *Immigrant Integration: A Cross-National Study.* El Paso, TX: LFB Scholarly Publishing LLC.

van Tubergen, Frank, Maas, Ineke, & Flap, Henk. 2004. The Economic Incorporation of Immigrants in 18 Western Societies: Origin, Destination, and Community Effects. *American Sociological Review*, **69**(October), 704–727.

Veroniki, Areti Angeliki, Jackson, Dan, Viechtbauer, Wolfgang, Bender, Ralf, Bowden, Jack, Knapp, Guido, Kuss, Oliver, Higgins, Julian PT, Langan, Dean, & Salanti, Georgia. 2016. Methods to Estimate the Between-Study Variance and Its Uncertainty in Meta-Analysis. *Research Synthesis Methods*, **7**(1), 55–79.

World Economic Forum (WEF). 2021. *Global Gender Gap Report 2021.* Tech. rept. Geneva.

A Appendix

Table A.1: Culture measures employed by the primary studies

Input measures	**Output measures**
% Protestant rel. to catholic	FLFP
% Jewish rel. to catholic	FLFP rel. to male FLP
% Muslim rel. to catholic	Gender gap in LFP
% Orthodox rel to catholic	GEM
% Hindu rel. to catholic	Female annual / weekly working hours
% Minority religions rel. to catholic	
% Unaffiliated rel. to catholic	
% Conservative religions	
SIGI subindices	
WVS attitudes towards gender roles	
WVS attitudes towards importance of work / leisure	
WVS attitudes towards family	
ESS traditional values	

Table A.2: Robustness check: Meta-regression results from estimating random-effects models

	(1) Baseline	(2) US	(3) Journal	(4) Econ	(5) Gen ≥ 2
SE of effect size	1.445*** (0.300)	2.392*** (0.365)	1.498*** (0.329)	1.466*** (0.308)	1.596*** (0.329)
Ref: Journal					
Working paper	-0.021*** (0.005)	-0.000 (0.006)		-0.017*** (0.006)	-0.008 (0.007)
Publication year	0.000 (0.000)	-0.001*** (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Ref: Economics					
Sociology	-0.010** (0.005)	0.025*** (0.006)	-0.008 (0.005)		0.004 (0.006)
Ref: US					
Europe	-0.005 (0.005)		0.001 (0.006)	-0.005 (0.007)	-0.012** (0.006)
Other	-0.013*** (0.004)		-0.013*** (0.004)	-0.009* (0.005)	-0.009* (0.005)
Binary dependent variable	-0.003 (0.004)	-0.012** (0.005)	0.000 (0.005)	-0.004 (0.004)	-0.003 (0.004)
1st generation	0.026*** (0.004)	0.030*** (0.005)	0.022*** (0.006)	0.028*** (0.005)	
Mean age	-0.002*** (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.002** (0.001)	-0.001 (0.001)
Ref: Output measure					
Input measure	-0.015*** (0.004)	-0.006 (0.006)	-0.018*** (0.005)	-0.012** (0.005)	-0.011** (0.005)
Lagged culture	-0.025*** (0.005)	-0.016*** (0.004)	-0.016*** (0.006)	-0.027*** (0.005)	-0.011** (0.004)
Education	-0.010** (0.004)	-0.016*** (0.004)	-0.006 (0.004)	-0.014*** (0.005)	-0.010* (0.005)
Partner characteristics	0.008* (0.005)	0.018*** (0.007)	0.009* (0.005)	0.004 (0.005)	0.010* (0.005)
Area of residence	0.003 (0.004)	-0.017*** (0.005)	0.001 (0.005)	0.007 (0.005)	0.001 (0.004)
Quality of human capital	-0.001 (0.004)	-0.012*** (0.003)	-0.003 (0.005)	-0.001 (0.004)	-0.005 (0.004)
GDP	-0.003 (0.004)	-0.005 (0.003)	-0.003 (0.004)	-0.002 (0.005)	-0.010*** (0.004)
Observations	160	93	138	132	116

Notes: Dependent variable: Effect size (r). Method: DerSimonian and Laird- random effects. Columns 1 - 5 report results for all primary estimations, for those conducted with immigrants to the United States, those published in peer-reviewed journals, those conducted by economists, and those analysing immigrants of the 2nd and higher generation. *** p<0.01, ** p<0.05, * p<0.10