CULTURE, FEMALE LABOUR FORCE PARTICIPATION, AND SELECTIVE MIGRATION

New insights from an interdisciplinary meta-analysis

EVA MARKOWSKY*

This version: December 2024

ABSTRACT

I meta-analyse 160 estimates of the relationship between female migrants' labour force participation and ancestry culture, reconciling studies from economics and sociology. The overall association between these two variables across all studies is very close to zero. However, this null result is driven by selective migration based on labour market orientation. Exploiting heterogeneity in the composition of countries of ancestry across studies, I show that the association between individual labour force participation and ancestry culture is smaller in studies that include many female migrants from countries with low gender equality.

* Universität Hamburg, eva.markowsky@uni-hamburg.de
The research was supported by the German Research Foundation. Data was self-collected and is available for replication efforts from the author.

As part of the "New Culture Economics" (Guiso *et al.*, 2006; Gershman, 2017), a growing number of researchers attempt to identify the association between cultural norms and female labour force participation by examining how female immigrants' behaviour is related to gender norms in their countries of origin.[1] Keeping the host country context fixed, this approach promises to overcome the endogeneity concerns that arise in simple country comparisons of culture and women's economic outcomes due to the inter-relatedness of cultural norms, the economic environment, and other institutions, both formal and informal.

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). However, identifying the targeted relationship in immigrant samples poses some serious empirical challenges, most notably possible biases from unobserved individual heterogeneity and selective migration. These challenges, together with substantial variance in research designs between existing studies, lead to the question of how to synthesise this body of research as a whole. The literature seems to agree that culture matters, but we do not know how much it matters, whether its influence differs across contexts, or if it is biased by empirical obstacles. In this paper, I conduct a systematic quantitative review of this body of literature in order to answer these questions. Specifically, I collect a data set of empirical estimates of the association between female migrants' labour force participation and their ancestry culture and apply meta-analytic tools to systematise and summarise the literature, culminating in an analysis of three potential sources of bias: publication selection, unobserved heterogeneity (or omitted variable bias), and selective migration.

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 individual labour market success is regressed on origin or ethnicity dummies to determine the influence of membership in a particular ethnic group (e.g., Kelley & McAllister, 1984; Reimers, 1985).[2] The major innovation of the studies following van Tubergen

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

et al. (2004) and Fernández (2007) is to include continuous measures of aggregate norms or behaviour in the country of ancestry that are directly related to female labour force participation and thus allow quantification of the relationship with cultural origin.

Fernández & Fogli (2009) relate second-generation immigrant women's hours worked and numbers of children 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. They find strong positive correlations between female immigrants' behaviours in the US and the related norms in their countries of ancestry. 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 receiving countries on immigrants' economic integration. Among other findings, they confirm a strong influence of country-of-origin female labour force participation rates on women's labour supply. The authors generally find support for destination, origin, and "community effects" on immigrants' labour market integration, the latter being the statistical effect of membership in a particular origin group within a specific destination. Numerous other sociology scholars 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, both research strands seem somewhat unconnected at the moment. 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 in a combined meta-analysis.

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" (Hamermesh, 2007), i.e., replications of an original finding using different data, sample selection criteria, or study designs. Beyond integrating evidence from the two disciplines in a meta-summary, meta-regression analysis allows me to investigate the influence of different features of the research design on studies' results and to examine the three above-mentioned sources of bias:

First, I test for publication bias in this body of literature. It may be the case that journal editors and researchers select papers for publication based on expected results (file drawer problem) or that authors chose specifications by statistical significance (p-hacking). As regards the literature reviewed here, the existence of working papers that find a *negative* correlation between immigrants' behaviour and characteristics of their home country but have remained unpublished so far (e.g., Köbrich León, 2013), might hint at a publication bias also in this scientific debate. Additionally, as Stanley & Doucouliagos (2012) point out, publication selection might arise as an "unintended consequence of good intentions or sound scientific practices" (p. 52), which is why meta-analysts routinely check for its presence in their data. Meta-analytical tools provide means to test whether published results are selective and correct for this bias in subsequent analyses.

Second, migrants of common ancestry share other characteristics besides cultural values. Fernández (2011) mentions the unobserved quality of human capital as a likely candidate; another example is economic development in the country of origin, which is also strongly related to gender norms (Falk & Hermle, 2018). The culture measure could thus pick up these related and, at worst, unobserved variables rather than cultural norms about appropriate roles for women in society. In meta-regressions, I can provide some insight into this issue by comparing the results of studies with different sets of control variables.

Third, like most research on immigrant behaviour, the literature 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 behaviour might not match dominant norms in their countries of origin, potentially biasing correlations between these two variables. 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 cultural influences. 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 than those from countries with high participation. 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.* (2020a) provide first empirical evidence for this notion, while Beblo *et al.* (2020b) conceptualise the mechanism theoretically. In the meta-setting, cultural selection of immigrants implies that studies including more respondents from low participation countries would find smaller cultural effects, on average. The composition of countries of ancestry is mostly data-driven and thus rarely addressed explicitly in the literature. However, there is considerable variance in the sets of ancestry countries included in the data sets, even between studies that look at immigrants in the same host country. These differences create an intriguing source of variation, lending itself to a meta-analytic investigation that allows systematic comparison of estimations utilising data with differential distributions of immigrants across ancestry countries.

To my knowledge, about 30 primary studies from economics and sociology 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 **??** 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. We see substantial variance in essential dimensions of the research design, allowing identification of the influence of study characteristics on the results. At the same time, the existing studies also share enough similarities to draw meaningful comparisons. It is thus possible to determine the influence of single varying characteristics (e.g., the host country context), *holding other dimensions constant*.

CONSTRUCTING THE DATA SET

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, 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 the title and abstract.

This search yielded a list of 52 studies that appeared to conduct analyses of cultural influences 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 use 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 **??**, I extract a total of 160 estimations of the relationship of interest 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 origin 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 relationship with culture 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 effect of interest 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.

Effect Sizes

The 160 estimations I draw from the primary studies all conduct regression analyses 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.[3] To make individual effect sizes comparable across different regression techniques (e.g., OLS vs. logit) and across differential methods of reporting results (e.g., coefficients vs. marginal effects), I standardize them by

3 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 software (Sribney & Wiggins, 2021).

calculating the *partial correlation coefficient* r, as defined in Equation 1, following Stanley & Doucouliagos (2012).

$$r = \frac{t}{\sqrt{t^2 + df}} \tag{1}$$

Here, t is the t- or z-statistic of each coefficient or marginal effect, while df represents the degrees of freedom of this statistic, approximated by the number of observations in the primary estimations.[4]

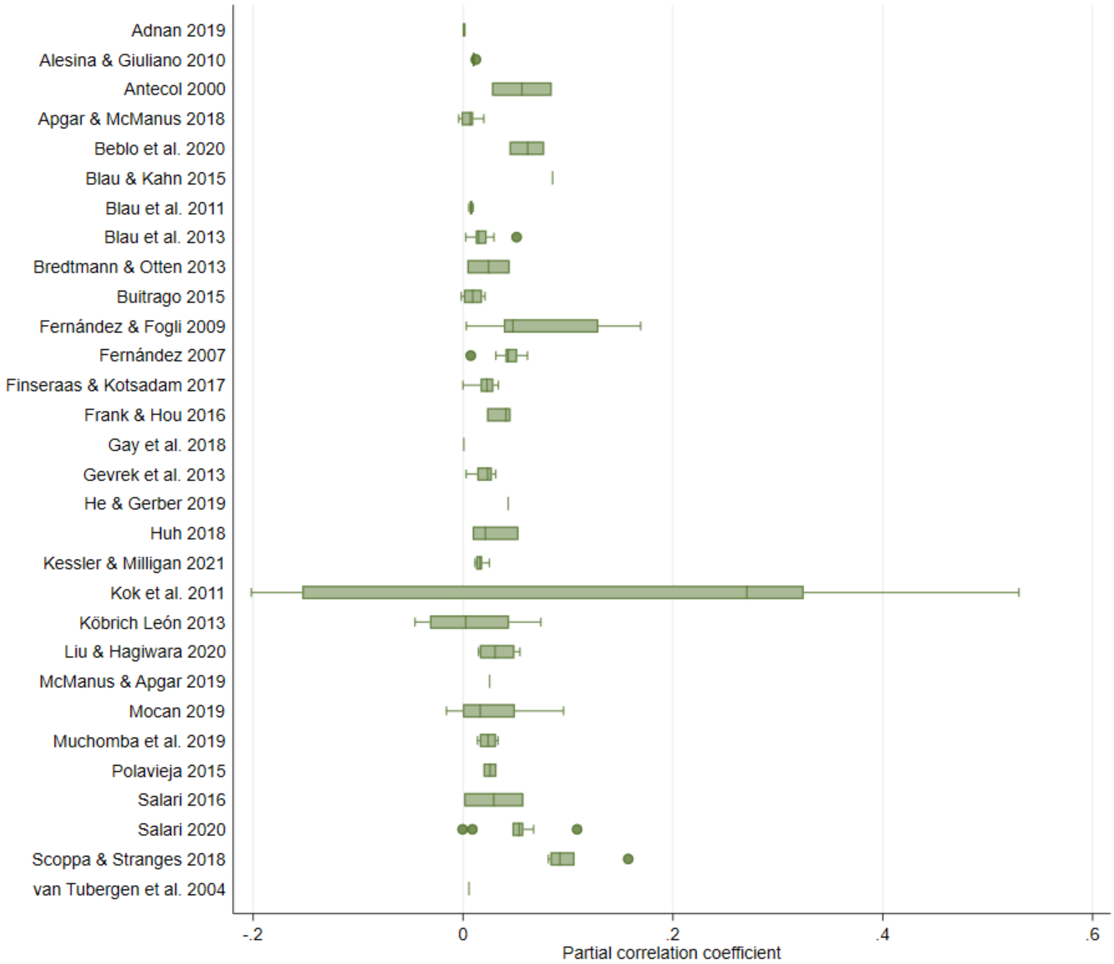
The partial correlation coefficient has the advantage of allowing comparisons across estimations from different models. It measures the direction and strength of two variables' association, holding other influences constant. Thus, the resulting value can be interpreted as a ceteris paribus correlation 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, or 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 heterogeneity in results *within* a given study, not 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 **??**, their study relates female first and second-generation immigrants' labour force participation in the Netherlands to the gender gaps in labour force

4 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.

Figure 1: Boxplot of partial correlation coefficients by study



participation in eight countries of origin.[5] 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 relies on 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 in the paper, 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, this study's resulting standardised correlation coefficients present enormous outliers in both directions. The sizeable 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 of immigrants of the first generation. From the information presented in the paper, I cannot infer the reason for these extreme results.[6]

Independent of these outlier values, Figure 1 illustrates that the estimated partial correlations vary not only between but also within studies.

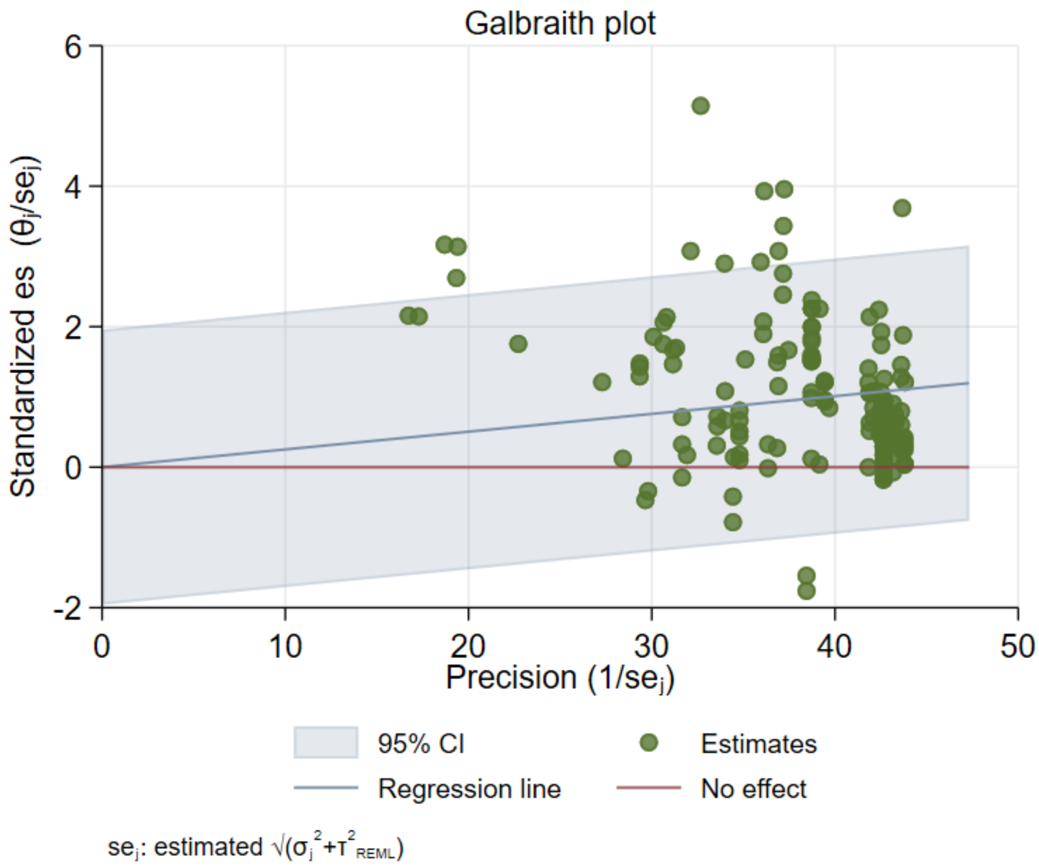
As a first approximation of the full extent of within- and between-study heterogeneity in results, Figure 2 shows a Galbraith plot of effect size estimates, standardised to z-scores, against their precision. The estimates by Kok *et al.* (2011) are excluded to ensure readability. Higher z-scores (y-axis) indicate a stronger, positive correlation between ancestry culture and individual labour supply. 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 Chapter , I will test whether these moderators can explain the variation in results.

5 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.

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

Before that, I integrate all primary estimates in a meta-summary to obtain an overall effect size.

Figure 2: Galbraith plot



INTEGRATING RESULTS

The starting point for calculating the meta-analytic overall association between female labour force participation and ancestry culture in my data is obtaining a weighted average of all 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:

$$ES_{ij} = ES_0 + \varepsilon_{ij}, \tag{2}$$

where ES_i is the ith effect size reported in study j - the conditional correlation between female migrants' labour force participation and characteristics of their countries of ancestry, standardised to partial correlation coefficients in this case. ES_0 is the "true" effect size, modelled to be common to all observations in the metadata, estimated with sampling error ε_{ij}. Effectively, the common-effects model postulates all of the collected estimates to be drawn from the same population with a common mean. It is assumed that $\varepsilon_{ij} \sim N(0, \sigma_i^2)$.

In contrast, random-effects meta-analysis presumes estimates to be drawn from several distinct populations and allows individual estimates to vary randomly around ES_0:

$$ES_{ij} = ES_0 + \theta_{ij} + \varepsilon_{ij} \quad \text{with} \quad \theta_{ij} \sim N(0, \tau^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0, \sigma_i^2), \tag{3}$$

where the estimate-specific "true" effect size consists of two components: ES_0 and the "random effect" θ_{ij}. τ^2 is a measure of between-estimate heterogeneity, beyond the variance from sampling, independent of both ES_0 and ε_{ij}, that is to be estimated.

In both cases, ES_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 $1/SE_i^2$, whereas the random-effects model adds an estimate of between-study heterogeneity, $\hat{\tau}^2$, to the variance, resulting in weights of $1/(SE_i^2 + \hat{\tau}^2)$ (Stanley & Doucouliagos, 2012, p. 46). Stanley & Doucouliagos (2015) propose a third, alternative method, "unrestricted WLS".[7] Here, the weights are given by $1/\phi SE_i^2$, where a multiplicative constant ϕ is added to the variance of the common-effects estimator. The authors demonstrate that the unrestricted WLS method produces estimates of the overall effect size that are identical to that of the common-effects method but more appropriate (i.e., wider) confidence intervals when there is evidence of heterogeneity. Furthermore, they show that unrestricted WLS estimates reach more reliable results than random-effects models in the presence of publication bias (Stanley & Doucouliagos, 2015, 2017).

We already know that there is substantial heterogeneity in the data set from Figure 2, and we will see in the next section that there is also evidence for selective reporting. I, therefore, continue

7 While both common- and random-effects meta-analysis also employ some version of weighted least squares (WLS) estimations, they constrain the WLS's (common) variance term, σ^2, to be one. Their proposed method does not impose this restriction and is, therefore, "unrestricted" (Stanley & Doucouliagos, 2017, p. 23).

with the unrestricted WLS model. Due to the hierarchical structure of the data set, resulting in statistical dependency between effect sizes from the same study, I estimate cluster-robust variances (Hedges *et al.*, 2010).

The resulting average effect size, ES_0, is calculated at 0.021 with a 95%-confidence interval of [0.005 ; 0.037]. Excluding the effect size estimates by Kok *et al.* (2011) leads to a considerably smaller overall effect size ES_0 = 0.015 (95%CI [0.006;0.024]). For comparison, a common-effects model estimates the confidence interval for the overall effect size in the latter sample to be [0.014; 0.016]. Compared to other meta-analytical findings in economics, this mean effect size has to be considered rather small (Doucouliagos, 2011).[8]

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 firmly rejected with a Q-statistic of 2690.49 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 the meta-summary stems from potential publication bias. 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 subsequent chapters.

REPORTING BIAS

It is possible that, despite my best research effort, the metadata 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

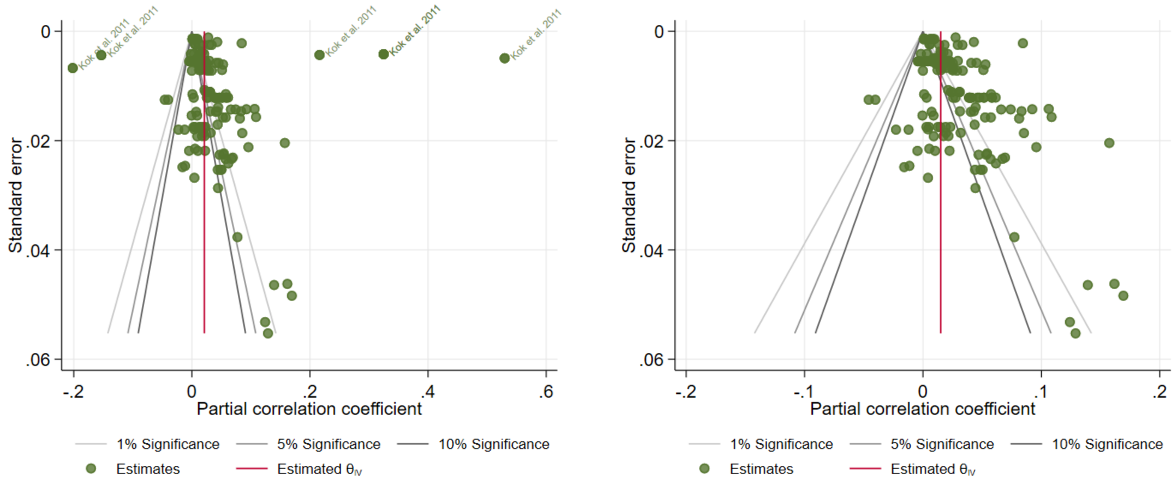
8 Based on a review of 41 meta-analyses in economics that report more than 22,000 partial correlations, Doucouliagos (2011) concludes that partial correlations smaller than 0.07 constitute small effect sizes in economics, even if the correlation is statistically significant.

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 are more likely to find their way into my data set 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 creates 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 whether a selective publication process distorts the emerging overall association between culture and female labour force participation.

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. The red vertical line represents the overall effect size, as estimated in the previous section. 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" from the sample 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 toward 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 in the bottom-left and bottom-middle parts of the graph. The grey lines represent standard significance levels. For example, the area between the two darkest lines holds all estimates that are statistically insignificant at the 10%-level. It seems that insignificant estimates of the

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.



(a) All studies

(b) Excluding Kok et al. 2011

Figure 3: Funnel graphs of effect sizes

association between culture and behaviour are underrepresented in the metadata, especially among the data points with low precision. This asymmetry could hint toward 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 Egger *et al.* (1997) and Stanley & Doucouliagos (2012):

$$ES_{ij} = \beta_0 + \beta_1 SE_{ij} + \varepsilon_{ij}, \tag{4}$$

Here, ES_{ij} is again the ith effect size 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 coefficient β_1 of the standard error variable is supposed to capture the degree of selective reporting bias. Estimated at 1.521 with s.e. = 0.438, 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 confirms the conclusion from the visual inspection of the funnel plot that there is 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). The model estimates it to be 0.010 with s.e. = 0.002. This estimate suggests that the bias from selective reporting or publication of results accounts for almost half of the "naive" overall effect size, as calculated in Chapter .

Stanley & Doucouliagos (2012) note that this regression approach tends to underestimate the underlying relationship if a non-zero effect truly exists. They show that replacing the standard error with its squared term produces more reliable results. Doing so leads to a somewhat larger estimate of the bias-corrected overall partial correlation of 0.014 with s.e. = 0.001.

In summary, there is evidence of asymmetry in the meta-data set, biasing the average effect size away from zero. In particular, results from less precise estimations seem to be reported more often 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 the data and their influences on the overall effect size.

MEASUREMENT BIAS: EXPLORING HETEROGENEITY

In the following, I use meta-regression analyses to investigate the influence of heterogeneity in data characteristics and model specification on individual results and to explore the possibility of a bias from omitted variables on individual results. To this aim, I introduce moderators, i.e., dimensions of heterogeneity between studies and specifications, into the regression framework in Equation 4:

$$ES_{ij} = \beta_0 + \beta_1 SE_{ij} + \gamma \boldsymbol{X_{ij}} + \varepsilon_{ij}, \tag{5}$$

where $\boldsymbol{X_{ij}}$ is a vector of observable study and specification characteristics described below, and γ represents the associated vector of coefficients. I continue to estimate the model using "unrestricted" WLS (Stanley & Doucouliagos, 2017).[10]

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 2 lists the collected study and specification attributes. Panel A lists **study characteristics**, as already illustrated in Table **??**. 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 studies are published in economics outlets; seven appeared in sociology or demography journals.

Panel B provides descriptive information about the **specification characteristics**, i.e., the mod-

10 As a robustness check, Table A.2 in the Appendix reports the results when estimating with random effects, instead. We see that this alternative specification corroborates the general conclusions. Where the random effects model finds significant results and the main results do not, I argue that the WLS results are more reliable since they account for the hierarchical structure of the data.

Table 2: Descriptive statistics on study and specification moderators

Panel A: Study characteristics		Mean	Min	Max	N
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 B1: Data moderators		Mean	min	max	N
Mean year of data collection		1999	1970	2013	160
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
	Restricted: married	0.41	0	1	160
Culture measured...	as input	0.25	0	1	160
	as output	0.75	0	1	160
	lagged	0.41	0	1	160

Panel B2: Model moderators		Mean	min	max	N
Individual controls include...	education	0.79	0	1	160
	partner characteristics	0.34	0	1	160
	children	0.35	0	1	160
	area of residence	0.50	0	1	160
Origin 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.

erators 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.

The first line reports the mean year of data collection, varying between 1970 (Fernández (2007); Fernández & Fogli (2009) and Salari (2016) all work with U.S. Census data from 1970) and 2013 (He & Gerber (2020) use the American Community Survey from the years 2011-2015). Another potentially significant source of heterogeneity in the data comes from differential approaches to measuring women's 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".

Concerning the sample, I identify three crucial moderators. First, cultural influences are estimated in different immigrant generations, with most specifications focused on the second and higher generations. About 20 per cent use first-generation immigrants, and only a tiny share of specifications (4%) analyses a pooled sample of first and second-generation immigrants that I pool together with the specifications using higher-generation immigrants in the meta-regression. The second sample moderator is the average sample age. It varies between 30 and 47, with a mean of about 38. Finally, about 41 per cent of specifications restrict the sample to married women.

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. As Table 2 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 cause heterogeneity in results is the period 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 precisely, 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 heterogeneity within samples of immigrants and with the possibility of omitted variable bias.

To this aim, I code whether specifications control for the respondents' education, area of residence, the presence of children, 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. If this reasoning is correct, we should see systematically different estimates in specifications that include these controls. While most specifications (almost 80%) control for education, only a minority includes covariates of presence or number of children and partner's education or income. Half of the specifications control the women's state, region, or otherwise defined residence area.

On the country-of-ancestry level, I code a moderator for whether specifications include a proxy of the unobserved quality of human capital. If there is a positive correlation between this variable and female migrants' labour force participation as well as a positive relation between the 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 correlation with culture. In this case, we would expect studies that include the covariate to estimate systematically smaller effect sizes because their estimates are corrected for this positive omitted variable bias. This would constitute a strong case for always including information on the country-of-ancestry quality of human capital in future studies.

Finally, I document whether specifications control 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. Due to the small number of clusters (at most 29 studies), which can impose a downward bias on standard errors when using robust variance estimation, I also implement cluster wild bootstrapping as the recommended alternative for handling dependence between effect sizes (Joshi *et al.*, 2022; Roodman *et al.*, 2019).[11]

Table 3 shows the results of estimating Equation 5 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 second or higher generations. For each moderator, the table reports the estimated coefficient, the p-value associated with the t-test in cluster-robust variance estimation (in round brackets), where standard errors are clustered at the study level, and the p-value from a Wald-test and confidence intervals (in square brackets) resulting from cluster wild bootstrapping with 999 replications.

In the first line, 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 is indeed an issue in this field of research. Reassuringly, this issue seems more mitigated in the sub-sample of studies that underwent peer review, as the cor-

11 Cluster wild bootstrapping (CWB) is an established alternative to cluster-robust variance estimation to avoid inflated Type I - error rates for small numbers of clusters (Cameron *et al.*, 2008; Cameron & Miller, 2015). Recently, it has also been applied in meta-analytic settings with hierarchical data structures (e.g., Ola & Menapace, 2020; McEwan, 2015).

Table 3: Meta-regression results

	(1) Baseline	(2) US	(3) Journal	(4) Econ	(5) Gen ≥ 2
SE of effect size	1.827	3.421	1.311	2.048	1.784
	(0.00)	(0.00)	(0.10)	(0.00)	(0.01)
	[0.02]	[0.08]	[0.37]	[0.05]	[0.00]
	[0.502, 2.910]	[-1.563, 5.498]	[-1.237, 3.188]	[0.0389, 3.783]	[0.288, 2.885]
Study moderators					
Working Paper	-0.029	0.016		0.003	-0.007
	(0.01)	(0.07)		(0.77)	(0.29)
	[0.12]	[0.48]		[0.84]	[0.34]
	[-0.0626, 0.0129]	[-0.0617, 0.0530]		[-0.0564, 0.0296]	[-0.0352, 0.00907]
Publication year	-0.000	-0.000	0.000	0.000	-0.001
	(0.82)	(0.27)	(0.65)	(0.42)	(0.19)
	[0.49]	[0.39]	[0.58]	[0.55]	[0.41]
	[-0.000588, 0.00230]	[-0.000954, 0.000857]	[-0.000385, 0.00198]	[-0.000672, 0.00253]	[-0.00140, 0.000695]
Sociology	-0.015	0.029	-0.017		0.014
	(0.24)	(0.00)	(0.19)		(0.21)
	[0.54]	[0.03]	[0.48]		[0.42]
	[-0.0602, 0.0264]	[0.00316, 0.0490]	[-0.0576, 0.0309]		[-0.0247, 0.0349]
Europe (Ref: US)	-0.018		0.002	-0.011	-0.019
	(0.16)		(0.90)	(0.64)	(0.06)
	[0.37]		[0.91]	[0.73]	[0.20]
	[-0.0612, 0.0331]		[-0.0374, 0.0345]	[-0.0578, 0.0688]	[-0.0407, 0.0131]
Other	-0.028		-0.030	-0.015	-0.004
	(0.01)		(0.00)	(0.02)	(0.45)
	[0.08]		[0.16]	[0.23]	[0.55]
	[-0.0808, 0.00631]		[-0.0750, 0.0234]	[-0.0378, 0.0138]	[-0.0207, 0.0119]
Data moderators					
Participation (Ref: Working hours)	-0.004	-0.002	-0.022	-0.003	-0.002
	(0.31)	(0.43)	(0.11)	(0.35)	(0.57)
	[0.19]	[0.25]	[0.09]	[0.12]	[0.54]
	[-0.0265, 0.0200]	[-0.0338, 0.0780]	[-0.0811, 0.00229]	[-0.0217, 0.0358]	[-0.0128, 0.0104]
Employment	-0.011	-0.014	-0.022	-0.018	0.005
	(0.28)	(0.02)	(0.14)	(0.15)	(0.34)
	[0.66]	[0.06]	[0.36]	[0.43]	[0.48]
	[-0.0367, 0.0240]	[-0.0322, 0.000682]	[-0.0652, 0.0238]	[-0.0988, -0.0583]	[-0.0165, 0.0266]
Other outcome	-0.000	0.002	-0.022	0.004	0.001
	(0.96)	(0.39)	(0.16)	(0.45)	(0.76)
	[0.83]	[0.19]	[0.45]	[0.38]	[0.88]
	[-0.0485, 0.0894]	[-0.176, 0.136]	[-0.101, 0.0756]	[-0.196, 0.123]	[-0.0228, 0.0254]
1st generation	0.030	0.045	0.031	0.043	
	(0.03)	(0.00)	(0.06)	(0.00)	
	[0.11]	[0.01]	[0.54]	[0.06]	
	[-0.00298, 0.122]	[0.00834, 0.129]	[-0.0147, 0.0557]	[-0.000442, 0.173]	
Mean age	-0.004	-0.003	-0.003	-0.005	0.000
	(0.08)	(0.04)	(0.20)	(0.02)	(0.63)
	[0.25]	[0.05]	[0.59]	[0.16]	[0.59]
	[-0.0103, 0.00181]	[-0.00796, 0.00000599]	[-0.0121, 0.00319]	[-0.0107, 0.000909]	[-0.00240, 0.00192]
Restricted: married	-0.016	-0.009	-0.025	-0.014	0.003
	(0.13)	(0.14)	(0.09)	(0.15)	(0.30)
	[0.29]	[0.11]	[0.13]	[0.33]	[0.32]
	[-0.0635, 0.0142]	[-0.0358, 0.0470]	[-0.0822, 0.00283]	[-0.0437, 0.0632]	[-0.0177, 0.0230]
Input measure	-0.019	-0.006	-0.020	-0.013	-0.015
	(0.01)	(0.41)	(0.04)	(0.25)	(0.00)
	[0.03]	[0.52]	[0.09]	[0.38]	[0.02]
	[-0.0371, -0.00178]	[-0.0353, 0.0218]	[-0.0536, 0.0123]	[-0.0782, 0.120]	[-0.0200, -0.00617]
Lagged culture	-0.028	-0.021	-0.031	-0.030	-0.015
	(0.01)	(0.00)	(0.06)	(0.01)	(0.00)
	[0.04]	[0.04]	[0.05]	[0.03]	[0.09]
	[-0.0653, -0.000748]	[-0.0488, -0.00277]	[-0.110, 0.000149]	[-0.0725, -0.0132]	[-0.0533, 0.00260]
Mother's ancestry					0.010
					(0.07)
					[0.20]
					[-0.0112, 0.0214]
Father's ancestry					0.003
					(0.47)
					[0.66]
					[-0.00774, 0.0131]
Model moderators					
Education	0.004	-0.010	0.002	-0.001	-0.013
	(0.99)	(0.03)	(0.68)	(0.84)	(0.04)
	[0.99]	[0.06]	[0.71]	[0.87]	[0.05]
	[-0.0133, 0.00924]	[-0.0245, 0.000392]	[-0.0128, 0.0102]	[-0.0188, 0.0103]	[-0.0280, 0.00180]
Children	-0.008	-0.031	0.004	-0.040	-0.006
	(0.38)	(0.00)	(0.71)	(0.00)	(0.36)
	[0.53]	[0.01]	[0.77]	[0.01]	[0.54]
	[-0.0393, 0.0164]	[-0.0550, -0.0107]	[-0.0304, 0.0393]	[-0.0615, -0.00801]	[-0.0213, 0.0618]
Partner characteristics	0.032	0.010	0.032	0.005	0.009
	(0.00)	(0.33)	(0.04)	(0.44)	(0.08)
	[0.02]	[0.45]	[0.02]	[0.52]	[0.05]
	[0.00165, 0.0589]	[-0.0327, 0.0816]	[0.00210, 0.0901]	[-0.0159, 0.0330]	[0.0000189, 0.0274]
Area of residence	-0.012	-0.016	-0.010	-0.004	0.007
	(0.14)	(0.03)	(0.55)	(0.56)	(0.08)
	[0.38]	[0.05]	[0.67]	[0.66]	[0.25]
	[-0.0342, 0.0211]	[-0.0457, -0.0000459]	[-0.0786, 0.0297]	[-0.0285, 0.0173]	[-0.0114, 0.0135]
Quality of human capital	-0.009	-0.012	-0.012	-0.007	-0.003
	(0.02)	(0.00)	(0.00)	(0.04)	(0.00)
	[0.11]	[0.01]	[0.01]	[0.25]	[0.01]
	[-0.0207, 0.00450]	[-0.0191, -0.00218]	[-0.0196, -0.00354]	[-0.0190, 0.00815]	[-0.00782, -0.00192]
GDP	0.003	-0.009	-0.002	-0.010	-0.009
	(0.66)	(0.00)	(0.67)	(0.02)	(0.01)
	[0.72]	[0.02]	[0.72]	[0.24]	[0.03]
	[-0.0147, 0.0203]	[-0.0169, -0.00148]	[-0.0132, 0.0170]	[-0.0261, 0.0107]	[-0.0149, -0.00110]
Observations	160	93	138	132	116
No. of studies	29	17	23	22	16
adj R^2	0.600	0.862	0.637	0.744	0.598

Notes: Dependent variable: Effect size (r). t-test - p-values in round brackets from cluster-robust variance estimation. Wald test - p-values and confidence intervals in square brackets from cluster wild bootstrapping with 999 replications. 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.

relation between effect size and standard error is not statistically significant here. However, the fact that this association is also visible, albeit only significant at the 10%-level, when restricting to studies on immigrants in the United States (Column 2) contradicts the idea that there is a genuine relationship between sample sizes and effect sizes that could be driven by differential integration mechanisms in countries with differing sizes of immigrant populations.

In the second panel, we see that none of the study attributes exhibits a clear, systematic, and robust relationship with the effect sizes' magnitudes. Hence, whether a study is published in a peer-reviewed journal, is conducted in Europe, the United States, or other countries, is authored by economists or sociologists, or when it was published does not contribute to explaining the variation in results. This indicates that heterogeneity is caused by data and model differences rather than the more general study differences.

Moving to the data moderators, we do indeed find meaningful influences on results. Both the sample specification and the culture measure are systematically related to the size of the estimated culture effects.

Unsurprisingly, migrants of the first generation are more strongly oriented towards ancestry norms than the second and higher generations, even though the difference is not statistically significant in all sub-samples in the bootstrapping procedure. The small, negative relationship between mean sample age and effect sizes, suggesting that younger women might be more influenced by ancestry culture, is also not robust across sub-samples.

Meanwhile, measuring country-of-origin culture with input instead of output measures shows a stronger negative correlation with estimated culture effects but also not consistently at conventional levels of statistical significance. Using lagged culture proxies shows a persistent negative correlation with effect size magnitude across all sub-samples. 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.

In the sub-sample of specifications where the sample consists of second and higher-generation immigrants I add two additional moderators that describe how cultural ancestry was assigned to these native-born women. With "either" parent forming the reference category, these moderators indicate that cultural effects are somewhat more pronounced when the mother's ancestry is used to assign cultural background. However, the relationship does not hold up to standard levels of statistical significance in the bootstrap procedure. Using the father's ancestry does not seem to produce different results than using either parent's cultural background.[12]

Turning to the influence of model characteristics, we see that specifications that control for education do not find systematically different results than those that do not. Only in the studies conducted in the US and those authored by economists there is a small but statistically significant negative correlation. The same is true for controlling for the presence or number of children but with more sizeable coefficients. The negative relationship with accounting for area of residence is only significant in the US sub-sample and of similar size as the one with education.

Meanwhile, the effect of controlling for partner characteristics goes in the opposite direction: Specifications that include covariates for partner's income or education find *larger* effect sizes, on average. For a negative omitted variable bias to occur, there must be a negative correlation either between the partner's characteristics and the culture measure or between the partner's characteristics and the outcome variable, female labour force participation. The latter would be consistent with intra-household division of labour: The individual migrant woman works less the higher the education or income of her partner. This association, however, is *not* statistically significant in the US and the Economics sub-samples.

Finally, controlling for the quality of human capital seems to lead to consistently but moderately smaller estimates of the influence of culture, underlining the importance of this potentially omitted variable. As Fernández (2011) points out, migrants from the same 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 cultural influences. Economic development in the country of origin seems to play a similar role as a control variable. In both cases, the negative correlations reach statistical significance in three out of the five sub-samples.

According to the adjusted R^2 values at the bottom of the table, between 60 and 86 per cent of the variation in primary estimates are explained by the included moderators, with the highest

12 14% assign mother's ancestry, 54% use father's, and the remaining 31% use either, i.e., the parent born abroad.

explanatory power obtained in the sub-sample of specifications that look at immigrants in the United States.

Summing up, the meta-regression analyses in this section inform us about important sources of heterogeneity in the estimated effect sizes and give some insight into the role of omitted variable bias in this literature. The measured relationship between culture and female labour force participation is more substantial in migrants of the first generation, and the behaviour of these migrants' contemporary peers in the country of origin has a stronger influence than cultural input measures, such as religion, stated values, and gendered institutions, and aggregate behaviour measured in the past. This indicates that immigrant populations might "update" their cultural values.

Furthermore, the geographical distribution of immigrants within host countries, quality of human capital, and economic growth in the origin country present important co-determinants of female labour market integration. Omitting these variables leads to an overestimation of the association with culture. To a lesser extent, the same might be true for education and individual fertility, but the meta-analytic evidence is less robust in these cases. Conversely, omitting partner characteristics leads to underestimating the association of interest, which fits theories of the division of labour in the household.

In the next section, I use my unique meta-data to explore another source of heterogeneity that cannot be tested in primary analyses: the influence of differential sets of ancestry countries included in the analyses.

SELECTIVITY BIAS: COUNTRY-OF-ORIGIN COMPOSITION

The studies included in the meta-data set exhibit remarkable variance in the sets of countries of origin that are represented in the underlying data. To some degree, this is expected, as it follows from the fact that studies focus on immigrants in different host countries that are characterised by differential immigrant inflows and populations. However, even within the papers focused on 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 imposing no restrictions on countries of origin).

I argue that this heterogeneity is meaningful in the light of selective migration. 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 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).

Most relevant for female migrants' labour force participation 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 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), single women, people living in rural areas, and in countries where the Shia branch of Islam dominates the Sunnis.[13]

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 countries 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 in most of 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 labour supply model is based on an identity economics framework

13 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 a positive selection based on gender norms. However, they cannot rule out social desirability or cultural (over-)assimilation as drivers of their results.

(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. It shows that 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 by replicating the approach by Fernández & Fogli (2009) in European data and varying the set of included countries of origin. In doing so, we demonstrate that the relationship between immigrant women's decision-making and their cultural ancestry is stronger among immigrants from countries with high levels of gender equality.

Exploring Cultural Selection with Metadata

Using my meta-data set, I can exploit the variation in included countries of ancestry to test this relationship with more statistical power. To this aim, I complement the meta-data with information on gender equality in countries of ancestry.

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 primary studies. Including information in 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 dimensions of gender equality. Even though, for example, female LFP rates, which 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.[14] 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.

14 See World Economic Forum (WEF) (2021) for details on the components and the calculation.

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 some studies, information on included countries of ancestry is missing, so these studies have to be excluded from the subsequent analyses.[15] This leaves me with 19 studies reporting 122 effect sizes with complete information on countries of origin.

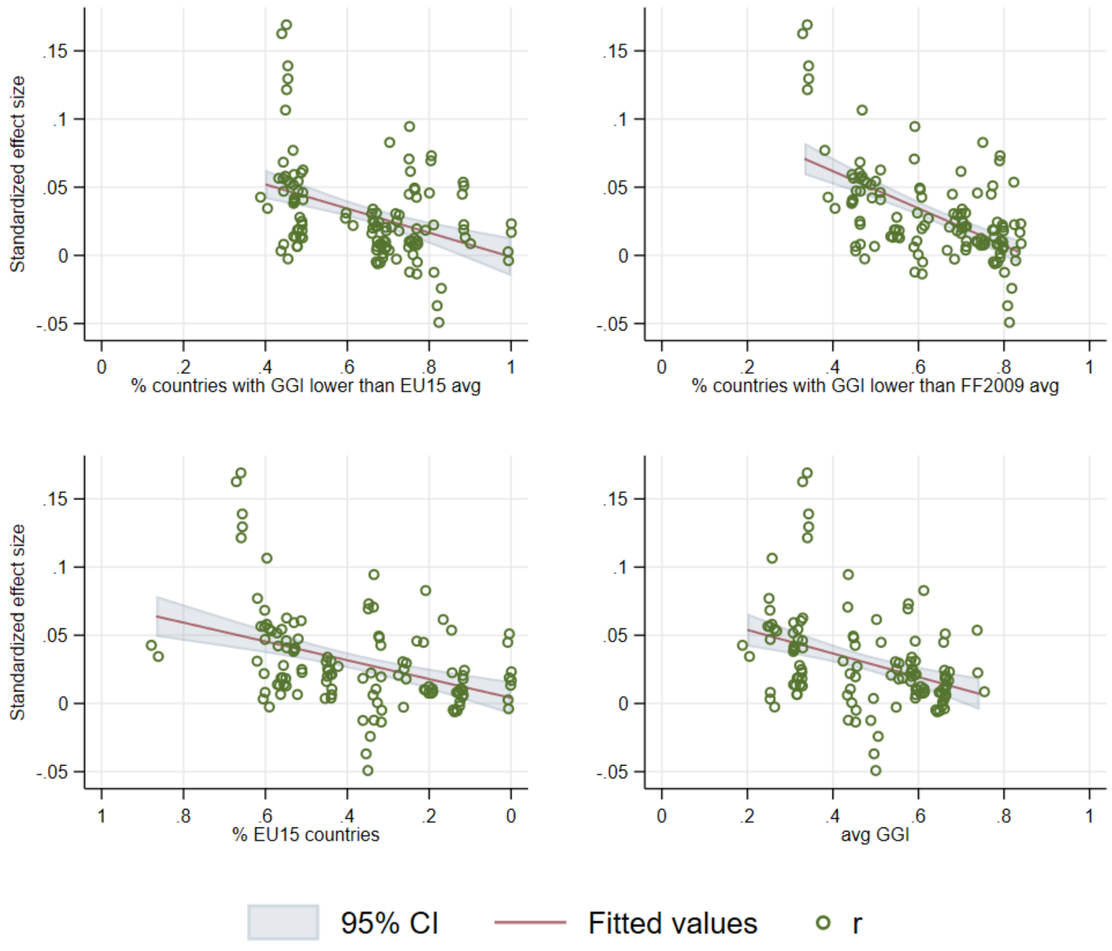
Ideally, I would like to calculate the share of immigrants from ancestries with low gender equality scores for each specification. Unfortunately, many papers do not report the numbers of observations by countries of origin, leading to a very small number of specifications (75 effect sizes within 12 studies) for which I can obtain this value. Alternatively, I calculate the share of countries of origin with low gender equality in each specification.

If selective migration based on gender norms imposes a downward bias on the measured relationship between immigrant women's labour force participation and cultural values in their countries of ancestry, we should observe smaller effect sizes among the studies that include more countries of origin of low gender equality in their analysis.

Figure 4 plots the relationship between partial correlation coefficients and shares of low-gender equality countries of origin with differing thresholds for low gender equality. 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.[16] 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 per cent in two specifications reported by Fernández

15 I unsuccessfully contacted the authors in an effort to obtain the missing information.

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



gal

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

(2007) to 100 per cent 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 122 specifications with non-missing values is 0.65, with a standard deviation of 0.16.

The result is plotted in the top-left corner of the figure. We see a clear negative association between shares of low-gender equality ancestries and effect sizes. The remaining three plots of Figure 4 show the same relationship using alternative thresholds for low vs high gender equality: the average of the countries included in the seminal study by Fernández & Fogli (2009), the share of EU15 countries, assuming higher average gender equality in Europe compared to the rest of the world[17], and the average GGI in the included countries of origin. All four panels show a clear negative association, implying a lower conditional correlation between individual behaviour and cultural values in the origin among women from countries with restrictive gender norms. As robustness checks of this finding, I perform two additional analyses: First, in Figure A.1 in the Appendix, I plot the same relationship in the smaller sample of effect sizes for which I have complete information on the numbers of observations from each included country of origin. Here, the effect sizes are plotted against the share of *respondents* from countries of origin with low levels of gender equality. The negative association is also visible in this sample, even though the slope is less steep. Second, since there is almost no variation in the set of included countries of ancestry at the study level, I calculate an average effect size per study and plot this against the share of countries with low gender equality included in this study in Figure A.2 in the Appendix. We see a strong negative association in this plot as well.

Summing up, I find systematically smaller estimates of the relationship between culture and female labour force participation in studies that include higher 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 regression approach made popular by Fernández & Fogli (2009) and van Tubergen *et al.* (2004).

17 In this case the x-axis is reversed because more European ancestries represent higher shares of countries with *high* gender equality.

I corroborate this vital finding by re-calculating the overall effect size using unrestricted WLS while controlling for the share of countries of origin with low gender equality in the underlying specification. The resulting new estimate of the overall association between ancestry culture and female labour force participation in the destination country lies between 0.037 and 0.06, depending on the applied threshold for low gender equality. This exercise confirms that a naive meta-summary of the literature underestimates the association between cultural ancestry and female labour force participation due to selective migration based on social norms with respect to gender equality. While the resulting estimate for the "corrected" overall effect size is still small, it increases the initial estimate that did not account for this particular source of selection by a factor of two to four.

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 economic analysis of culture.

Integrating primary results confirms the presence of a positive and robust correlation. However, it also shows that the correlation is relatively small compared to partial correlations in different areas of economic analyses. Additionally, 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.

Furthermore, this body of 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 the 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 correlations with culture, 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.

The most important finding of the present study concerns the possible influence of selective migration based on cultural values: Larger estimates of the relationship with culture 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 mechanism might be women with high labour market orientation actively selecting out of restrictive environments. While this negative bias does not threaten 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 so-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 away" in research on cultural selection of immigrants rather than being explicitly investigated.

These findings provide invaluable insights for future applications of the socio-epidemiological approach and advance the analysis of cultural influences on economic decision-making. Most importantly, the present analysis points out three sources of biases that researchers must 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 the estimated relationship and heavily depends on the context that is being investigated. This influence is harder to appease, and more research is needed on the underlying relationship. 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. Collective awareness of the issue is probably the most fundamental prerequisite for solving it.

REFERENCES

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

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.

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.

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.

Cameron, A. Colin, & Miller, Douglas L. 2015. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, **50**(2), 317–372.

Cameron, A. Colin, Gelbach, Jonah B., & Miller, Douglas L. 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, **90**(3), 414–427.

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.

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.

Doucouliagos, Hristos. 2011. *How Large Is Large? Preliminary and Relative Guidelines for Interpreting Partial Correlations in Economics*. School Working Paper SWP 2011/5. Deakin University Faculty of Business and Law School of Accounting, Economics and Fianance.

Egger, Matthias, Smith, George Davey, Schneider, Martin, & Minder, Christoph. 1997. Bias in Meta-Analysis Detected by a Simple, Graphical Test. *BMJ*, **315**(7109), 629–634.

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.

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.

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.

Hedges, Larry V., Tipton, Elizabeth, & Johnson, Matthew C. 2010. Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates. *Research Synthesis Methods*, **1**(1), 39–65.

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.

Joshi, Megha, Pustejovsky, James E., & Beretvas, S. Natasha. 2022. Cluster Wild Bootstrapping to Handle Dependent Effect Sizes in Meta-Analysis with a Small Number of Studies. *Research Synthesis Methods*, **n/a**(n/a).

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

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.

McEwan, Patrick J. 2015. Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, **85**(3), 353–394.

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

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).

Ola, Oreoluwa, & Menapace, Luisa. 2020. A Meta-Analysis Understanding Smallholder Entry into High-Value Markets. *World Development*, **135**(Nov.), 105079.

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

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

Roodman, David, Nielsen, Morten Ørregaard, MacKinnon, James G., & Webb, Matthew D. 2019. Fast and Wild: Bootstrap Inference in Stata Using Boottest. *The Stata Journal*, **19**(1), 4–60.

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.

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.

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	

Notes: 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.

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	1.670***	2.675***	1.776***	1.411***	1.894***
	(0.326)	(0.386)	(0.354)	(0.324)	(0.377)
Working paper	-0.023***	0.003		-0.019***	-0.007
	(0.006)	(0.007)		(0.006)	(0.009)
Publication year	0.000	-0.001***	-0.000	0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Sociology	-0.011*	0.030***	-0.013**		0.015
	(0.006)	(0.007)	(0.006)		(0.013)
Europe	-0.008		0.002	-0.001	-0.020**
	(0.006)		(0.007)	(0.008)	(0.008)
Other	-0.011**		-0.011**	-0.010**	-0.004
	(0.004)		(0.005)	(0.004)	(0.007)
Participation	-0.005	-0.013**	-0.003	-0.004	-0.002
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
Employment	0.002	-0.010	0.006	-0.009	0.006
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
Other outcome	0.011	0.003	0.002	0.020**	0.002
	(0.007)	(0.007)	(0.010)	(0.009)	(0.006)
1st generation	0.026***	0.033***	0.023***	0.028***	
	(0.005)	(0.005)	(0.006)	(0.006)	
Mean age	-0.002**	-0.001	-0.001	-0.003***	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Only married	0.000	-0.006	-0.005	0.000	0.002
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Input measure	-0.015***	-0.012**	-0.018***	-0.012**	-0.014***
	(0.004)	(0.006)	(0.006)	(0.006)	(0.005)
Lagged culture	-0.024***	-0.017***	-0.010*	-0.029***	-0.014***
	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)
Mother's ancestry					0.010
					(0.008)
Father's ancestry					0.003
					(0.008)
Education	-0.012***	-0.013***	-0.008*	-0.011**	-0.014**
	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Children	0.004	-0.012*	0.013**	-0.005	-0.006
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)
Partner characteristics	0.007	0.011*	0.012**	0.005	0.010*
	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)
Area of residence	0.005	-0.011*	0.001	0.002	0.008
	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
Quality of human capital	-0.000	-0.012***	-0.006	-0.001	-0.003
	(0.004)	(0.003)	(0.005)	(0.004)	(0.004)
GDP	-0.003	-0.008**	-0.002	-0.006	-0.009**
	(0.004)	(0.004)	(0.004)	(0.004)	(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

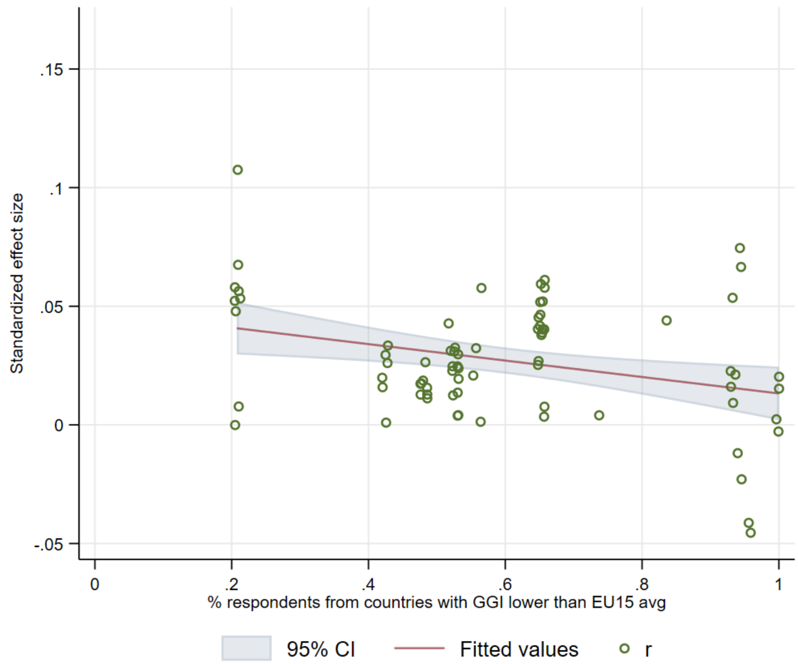


Figure A.1: Scatter plot of effect sizes against shares of respondents from countries of origin with low gender equality

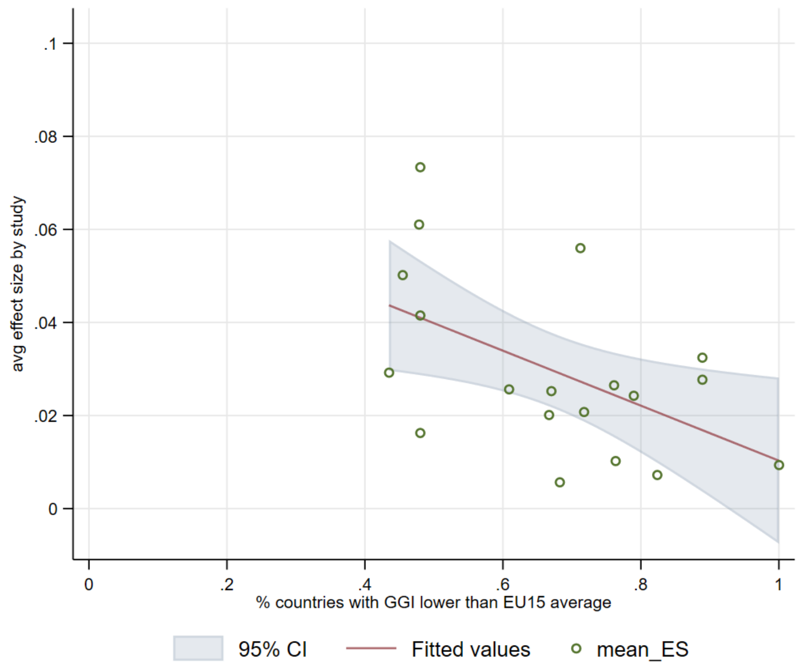


Figure A.2: Scatter plot of average effect size per study against shares of countries of origin with low gender equality