Matthias Collischon*

A Research Note on the pitfalls of sample selection and what (not) to control for in gender wage gap estimations

Abstract: Estimating discriminatory gaps is one of the core topics of Sociology. A large number of studies use observational data to estimate these gaps, typically by controlling for potentially confounding factors and interpreting the coefficient of the group variable (e.g. *female*) as the degree of discrimination. However, estimations in this manner require choices both with regards to sample selection as well as the selection of variables to control for the stages where researchers assume discrimination to occur. Against the backdrop of the literature on gender wage gaps, this research note discusses these choices, their impact on the interpretation of the results and provides recommendations for applied researchers on how to deal with these choices in practice.

Keywords: Gender wage gaps, sample selection, empirical specification, discrimination, guide

Eine Forschungsnotiz zu den Implikationen der Sampleauswahl und für was man (nicht) in Schätzungen von Geschlechterlohnlücken kontrollieren sollte

Zusammenfassung: Die Schätzung von diskriminierenden Lohnlücken zwischen unterschiedlichen Gruppen ist ein wichtiger Bestandteil der soziologischen Forschung. Viele Studien nutzen zu diesem Zweck Beobachtungsdaten. Typischerweise wird hierbei auf potenzielle Störvariablen kontrolliert und der geschätzte Koeffizient der Gruppenvariable (z.B. *Geschlecht*) als Ausmaß der Diskriminierung interpretiert. Allerdings erfordern Schätzungen dieser Art Entscheidungen in Bezug auf die Auswahl des Analysesamples und der Kontrollvariablen, die Implikationen dafür haben, an welcher Stufe die Forschenden Diskriminierung vermuten. Vor dem Hintergrund der Literatur zu Geschlechterlohnlücken diskutiere ich in diesem Beitrag, wie sich diese Entscheidungen auf die Interpretation der Ergebnisse auswirken und leite Empfehlungen für angewandte Forschung ab, wie man in der Praxis mit diesen Entscheidungen umgehen sollte.

Schlagwörter: Geschlechterlohnlücken, sample selection, empirische Spezifikation, Diskriminierung, Guide

* Matthias Collischon, Institute for Employment Research (IAB), Regensburger Strasse 104, D-90478 Nuremberg, Germany, matthias.collischon2@iab.de.

 DOI: 10.5771/0038-6073-2026-1-83

1. Introduction

The investigation of gender inequalities in the labor market, especially the gender wage gap, has been a topic at the heart of sociological research for decades. A large part of the literature is interested in estimating adjusted gender wage gaps that try to estimate the degree of gender discrimination (e.g. through taste-based or statistical discrimination, Guryan and Charles 2013 or other forms of discrimination such as implicit biases, Bertrand and Duflo 2017) in pay by comparing average wages of men and women and accounting for potentially confounding factors by controlling for them in a regression analysis using observational data. For example, typically, studies (e.g. Oaxaca 1973) control for an enhanced Mincer earnings function that accounts for education, experience, and experience squared to control for differences in human capital. Analyses of this kind typically use survey or administrative data and base their conclusions on a sample of working individuals. The gender wage gap estimated this way, net of human capital and other factors, is then often interpreted as a measure of gender discrimination.

But this way to extract discriminatory gender gaps is not always as straightforward as one might think. It requires choices made by researchers on the stage at which they presume discrimination to occur, or deciding at which stage of discrimination in the labor market they are interested in. For example, should one account for occupations in this estimation, or do we underestimate the true extent of discrimination if women are also discriminated against at the access to occupations? How should we restrict the sample? Should we include all employees or only use full-time employees? And how about selection into the labor market? We observe no wages for non-employed individuals, but working itself could be an outcome of discrimination as well. There is no one-size-fits-all solution that can tell us how to specify estimating discriminatory gaps, which, consequently, leads to a debate among researchers what to control for and what samples to use.

This research note aims at providing a brief overview on the choices researchers need to make when estimating discriminatory gaps at the stages of sample selection and model specification. While this article focuses on gender gaps, the arguments are easily applicable to all analyses investigating discriminatory group differences, such as when comparing migrants with natives. I provide quick and stylized overviews on decisions that arise during research and what implications these decisions have for where researchers assume discrimination to occur. Lastly, I provide some recommendations for how to handle these decisions in research papers.

2. How gender wage gaps are typically estimated

As described previously, a large amount of the evidence for gender wage gaps comes from observational studies that try to elicit the extent of discrimination with regression analysis, by controlling for all confounding factors, typically in the following form:

$$\ln(w_i) = \beta_0 + \beta_1 female_i + \gamma X_i' + \epsilon_i$$

Where the natural logarithm of wages is the dependent variable, X is a set of confounding variables and $female$ is an indicator variable for the respondents' gender. β_0 is the constant of the estimation; β_1 then provides the gender gap net of confounding factors – the adjusted gender gap – and is sometimes interpreted as discrimination. ϵ_i is an idiosyncratic error term. This style of regression analysis is closely related to methods like Kitagawa-Oaxaca-Blinder (Kitagawa 1955; Oaxaca 1973; Blinder 1973) style decomposition analyses that estimate separate regressions for men and women and calculate unexplained gender gaps based on these regressions, or the pooled decomposition proposed by Fortin (2008) which uses the *female*-coefficient from a pooled regression net of all control variables as a measure for the adjusted wage gap.

In intuitive terms, what this type of regression does is to provide us with estimations of the gender wage gap that does not arise due to the observable characteristics captured by X. For example, if X contains a measure of education, the regression accounts for systematic differences in educational levels between men and women. Consequently, the estimate of β_1 provides us with an estimate of the gender wage gap holding education constant, i.e. for men and women with the same level of education. In the same way, other characteristics that differ systematically between men and women can be controlled for as well. Accounting for relevant characteristics, the remaining gap is then often used as a measure of wage discrimination of women.

While this way to estimate gender discrimination seems straightforward and transparent, several choices arise that can have important implications for the interpretation of the results. This paper especially focuses on the choice of the analytical sample, as well as the question what characteristics to control for in the next sections.

Additionally, what I want to state clearly is that discrimination is a theoretical concept that can hardly be measured directly using observational data. Even if we adjust for observable characteristics, the remaining adjusted gender gap, in statistical terms, is a descriptive measure of gender differences in wages that are not explained by the characteristics we account for. It is not necessarily discrimination. Based on theoretical considerations and assumptions (that hardly can be tested empirically), we can interpret this gap as discrimination, but, as I will show, we should be cautious in doing so. In the following sections, I assume that researchers are primarily interested in estimating gender discrimination and discuss the choices that need to be made when conducting empirical estimations against this background.

3. What choices matter?

3.1 Sample selection

The first decision researchers have to make when investigating discrimination is which sample to use. When investigating gender wage gaps, both adjusted and unadjusted, we require information on wages. However, wages are only observable for individuals who are employed, and this can pose a problem in assessing the scope of discrimination. For example, data from Eurostat for 2019 show that the raw gender wage gap in Europe varies drastically: in Estonia, it amounts to 22%, whereas it is only 5% in Italy (Eurostat 2023b) does this mean that Italy is far more gender-egalitarian and less discriminatory than Estonia? Not necessarily, it could also depend on differences in the labor force participation rate: In Estonia, 77% of women aged 20 to 64 in 2019 were employed, compared to only 53% in Italy (Eurostat 2023a). Thus, a smaller gender wage gap does not necessarily mean that there is less discrimination – it could for example just imply that the barriers to entry or preferences for work among women differ drastically between countries. Thus, the smaller wage gap in Italy could simply reflect that the women who work in Italy are a much more selective group compared to Estonia. Furthermore, selection patterns of women into the labor market differ over time (Blau and Kahn 2017), thus complicating comparisons also within countries over time.

Selection processes like these are a problem in empirical research. We also cannot simply solve this problem by setting the wages of non-employed women to "0", as the earnings of these women would not be zero if they worked. Some studies deal with this selection problem by accounting for sample selection by applying a Heckman-style (Heckman 1979) correction procedure, which can use exogenous instruments to correct for selection into the labor market in an instrumental variable-style. However, this method is oftentimes not feasible as we lack exogenous instruments for labor force participation. For example, Christofides, Polycarpou, and Vrachimis (2013) use family circumstances and non-labor income as instruments. However, it is easy to come up with explanations how these instruments could also be biased or affect wages directly, which would render them invalid instruments. The authors also show uncorrected results, which is good practice when using these corrections.

Besides the classical Heckman-approach, other possibilities to tackle selection exist. Blundell et al. (2007) use a bounding approach and estimate wage distribution for the UK with different assumptions on selection into employment. Mulligan and Rubinstein (2008) use the Heckman approach with number of young children as an instrument, but additionally use an identification at infinity-approach to account for selection into employment. This approach corrects for selection patterns into employment assuming that selection works the same way for different demographic groups. Olivetti and Petrongolo (2008) estimate median wage gaps for various countries, correcting for selection into employment. As the study estimates median gaps, their imputation method only relies on predicting whether wages of non-employed individuals would be below or above the median of observed wages.

Overall, while there might not be one one-size-fits-all solution to selection into employment, researchers should be aware of this issue. One should at least acknowledge it, as only investigating gender wage gaps among employed individuals ignores these potential selection processes that affect men and women differently.

Next, researchers also make choices about their sample among employed individuals. The most common question here is whether to investigate gender wage gaps among all employees, or whether to exclude specific groups, such as marginally employed individuals (e.g. German *Minijobbers*) or part-time employees. This topic applies to estimating adjusted and unadjusted gaps. Again, no one-size-fits all solution exists here, but it is important to be mindful and transparent about the choices we make as researchers as they also implicitly affect where we assume that discrimination occurs. For example, studies with German administrative data often only contain full-time employees as the German data lack information on hours worked which make adequate comparisons among part-time employees with large heterogeneities in working time basically impossible. However, in Germany in 2021, 49% of women work part-time, but only 12% of men (BMFSFJ 2023). Thus, analyses only for full-time employees miss selection procedures – either due to discrimination or due to individual preferences – into full-time employment. It is understandable that the data do not allow for more all-encompassing analysis, but researchers should acknowledge these issues. For example, Huffman, King, and Reichelt (2017) analyze the effect of organizational policies on gender gaps with German administrative data using only full-time employees in the main analysis, but mention that the results hold when including part-time work as well. Arulampalam et al. (2007) investigate gender gaps using quantile regression methods without correction for selection into the labor market, but show results for the public and private sector separately and make their sample selection criteria transparent.

Table 1: Sample restrictions and implicit assumptions on the stages of discrimination

Sample restriction	(implicit) Assumption on stage of discrimination	Example Study
None	Discrimination can happen at the entry to the labor market and in the labor market	Christofides, Polycarpou, and Vrachimis (2013)
Employees only	Discrimination occurs among employees, not at the extensive margin	Most studies, e.g. Combet and Oesch (2019)
Full-time employees only	Discrimination occurs among employees, access to full-time employment is equal for men and women	Huffman, King, and Reichelt (2017)

Furthermore, there are more detailed choices that researchers make about e.g. the inclusion of self-employment, age restrictions or restrictions for specific years or regions when constructing the sample for the analysis. These are highly-dependent on the specific circumstances and there is no right or wrong for these restrictions, but, again, researchers should make these restrictions transparent and discuss their implications for the interpretation and context of the gender wage gap they estimate. Table 1 shows a non-exhaustive overview on specific sample restrictions, their implications for the assumed stage of discrimination and an example study with a specific sample restriction.

Overall, what should researchers do regarding sample selection? If possible, the decisions should be based on theoretical reasoning. In some cases, for example if we are interested in promotions and the theoretical considerations only apply to employees, this provides an argument for only focusing on employees in the analysis. However, in many cases, it is not clear at which stage e.g. taste based discrimination (Becker 1957) arises. Thus, it is important to make the choices and limitations of the study transparent and make clear for what population the results are valid.

3.2 Specification

Besides the question of which sample to use for estimations, the empirical specification is also strongly tied to where researchers assume discrimination to occur. The most basic version for estimating gender wage gap is simply comparing average wages for men and women, as for example done in the unadjusted gender wage gap estimates for each country released by Eurostat (2023b). These estimates close no channel through which discrimination can occur, but they also do not account for structural differences in e.g. education or labor market experience, that can also simply be shaped by choices and preferences and are not necessarily the outcome of discrimination.

In this context, I want to at least point to a broader debate on what to control for when estimating regressions using observational data in general that has emerged recently. One important tool in this debate are directed acyclical graphs (DAGs, see Huntington-Klein 2025, chapters 6 to 8, and Rohrer 2018) to differentiate, among others, between spurious correlations (the classical omitted variable bias) and mediation (a variable on the path from treatment to outcome). In the case of gender, as at least biological gender is determined before labor market outcomes, most potential control variables are mediators, and, in the following, I will discuss them as such and not as omitted variables.

The most common specification for empirical estimations that aim at estimating adjusted wage gaps condition on human capital variables, such as labor market experience and education. The implicit assumption here is that these variables are more likely the outcome of individual choices and preferences than discrimination, but one should also keep in mind that measures like labor market experience may also, to some degree, be the outcome of discrimination if women have a harder time finding employment compared to men.

The most debated points are whether to adjust for characteristics of specific jobs, such as industry, occupations and job level. The problem with these variables is that they clearly correlate with wages, but can be outcomes of discrimination themselves. For example, managers gain high wages, but access to management positions for women may be limited by discrimination. Goldin (2014) argues that men discriminate women in the access to typical "male" occupations in order to guard their occupational prestige. This argument would speak against controlling for occupation in empirical studies, as occupations themselves could be the outcome of discrimination. In this line of thinking, these variables would be *bad controls*.

However, these variables are likely also the outcome of individual preferences and choices, not just discrimination. In this case, one would want to condition on these variables. In most cases, the choice which job one performs is likely, in the aggregate, a mixture of both: discrimination and individual preferences. This is why especially whether or not adjusting for the aforementioned characteristics requires a transparent explanation by the researcher why or why not she controls for these characteristics and what this means for the channel through which discrimination can occur. For example, Cha and Weeden (2014) investigate the effect of overtime work on the gender wage gap and do not condition on occupation (although they estimate subsample analysis by occupation) as overtime work may be tightly linked to specific occupations that also have barriers to access for women.

Nisic (2017) investigates regional differences in the gender wage gap and controls for occupation and professional position. She argues that "because the core hypothesis states that regional differences in the wage gap will arise holding everything else constant, a large set of control variables was included". On the one hand, controlling for occupation and professional position in this case also closes differences in the access for these positions as a potential channel for discrimination that could differ between urban and rural regions. On the other hand, if there are differences in preferences between men and women between urban and rural areas, not controlling for these variables would overestimate the magnitude of discrimination. Thus, the decision whether or not to control for occupations is not always straightforward.

One middle-of-the-road solution could be to condition on broad measures of occupation, but not on e.g. job level when we assume that the general occupational choice is more the outcome of preferences than discrimination, but job level within occupations is restricted for women.

Linked Employer-Employee data can also be used to adjust for establishment fixed effects in wage gap estimations. Using these fixed effects, one rules out the access to specific firms as a channel of discrimination and estimates intra-establishment gender wage gaps. This should only be done when one particularly interested in intra-firm gaps. For example, Zimmermann and Collischon (2023) study the impact of female-friendly organizational policies on gender wage gaps. In the baseline estimation, they condition on firm fixed effects, time-varying firm characteristics, human capital variables and occupations (as broad codes). That means that the effect of organizational policies on gender wage gaps in this case assumes that they affect the wages of currently employed women within the firm, but do not work through the access to firms or through other variables.

Table 2: Estimation specification and implicit assumptions on the stages of discrimination

Control Variables	(implicit) Assumption on stage of discrimination	Example Study
None	Discrimination can occur through human capital attainment, the entry to jobs, at the entry to firms/industries and within jobs	Gender wage gaps e.g. by Eurostat
Human capital	Discrimination can occur through the entry to jobs, at the entry to firms/industries, and within jobs	Cha and Weeden (2014)
Human Capital + Occupations	Discrimination can occur at the entry to firms/industries and within jobs	Nisic (2017)
Human Capital + Occupation + Industry/Firm-FE	Discrimination can occur within jobs	Zimmermann and Collischon (2023)

Table 2 shows a non-exhaustive list of control variables and which respective channels of discrimination they close. Again, I also provide example studies corresponding to the choice of control variables. As displayed in the discussions in this Section, the choice of control variables is highly dependent on the context and to assumptions made by the researcher.

Again, as is the case with sample selection, theory-guided choice of specifications is generally the best way. If a theory predicts discrimination to occur only within occupations, then one should control for occupations in the estimation. However, this is oftentimes not the case. Thus, in general, I recommend showing different specifications in robustness checks to investigate how sensitive the results are to these assumptions. One possibility to achieve this easily and comprehensively are specification curves (Simonsohn, Simmons, and Nelson 2020) which can be used to show a large set of coefficient estimates from different estimations.

4. Conclusion and Recommendations for applied researchers

This paper provides a brief discussion on the implicit assumptions that researchers make when estimating gender wage gaps, both adjusted and unadjusted. Both sample selection and the selection of control variables have consequences for where discrimination is assumed to occur. For example, using a sample of full-times employees only provides wage gap estimations for these groups, but closes the channel of the access of full-time jobs as an explanation for wage inequalities. Regarding the choice of control variables, conditioning on occupations provides a gender gap estimate within occupations and (implicitly) ignores potential gender differences in the access to jobs.

For applied researchers, here are some short recommendations of how to deal with these topics in studies:

1. **If possible, make decisions based on theory.** Sometimes, theoretical considerations call for the use of specific samples (e.g. employees when investigating effects of promotions) or specifications (e.g. theories on within-job discrimination imply controlling for occupations). Researchers should motivate their decisions for the empirical analysis based on the large body of theoretical considerations that exists in the literature on gender differences. For example, when investigating taste-based discrimination, it is reasonable to control for other characteristics that are relevant for productivity, such as education or experience, while this might not be the case for other research questions.
2. **Be explicit about the sample you investigate.** A selection-correction for entry into the labor market is hardly feasible in most cases, but researchers should at least acknowledge which stages are potential blind spots for the paper, or which stages are not investigated. For gender wage gaps, this for example means discussing whether selection into the labor market is relevant. Researchers should clarify whether they expect any bias by only investigating employed individuals for example, with regards to the expected gendered selection patterns into employment.
3. **Clearly define what kind of gap you are investigating.** Researchers are oftentimes only interested in specific aspects of gender gaps, for example gender gaps within firms or within the same occupation. That is fine, but it should be clear, in the connection of theory and the empirical application, what the estimand (Lundberg, Johnson, and Stewart 2021) of the paper is.

4. **Show multiple specifications.** At least in robustness checks, I recommend showing how different specifications affect the results. For example, showing results with and without controlling for occupations gives lower and upper bound estimates of gender gaps when it is unclear to what degree occupations could be an outcome of discrimination. One possibility to show the results from multiple specifications intuitively are specification curves (Simonsohn, Simmons, and Nelson 2020).

Again, there is no one-size-fits all solution to the specific issues when investigating discriminatory gaps and there are oftentimes good reasons for specific choices. When doing research, it is important to make these choices salient and transparent and be clear about what a study is really doing.

Acknowledgements:

I thank Martin Popp, Jens Stegmaier and Florian Zimmermann for their helpful comments.

Conflict of Interest Statement

The corresponding author states that there is no conflict of interest.

References

Arulampalam, Wiji; Booth, Alison L.; Bryan, Mark L. (2007): Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. In: *ILR Review* 60 (2): 163–186. DOI: 10.1177/001979390706000201.

Becker, Gary Stanley. 1957. *The Economics of Discrimination*. Economics research studies of the Economics Research Center of the University of Chicago. Chicago, Ill.: Univ. of Chicago Press.

Bertrand, M.; Duflo, E. (2017): Field Experiments on Discrimination. In: Handbook of Field Experiments, Bd. 1: Elsevier (Handbook of Economic Field Experiments): 309–393.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations". *Journal of Economic Literature* 55 (3): 789–865. doi: 10.1257/jel.20160995.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates". *The Journal of Human Resources* 8 (4): 436-455. doi: 10.2307/144855.

Blundell, Richard; Gosling, Amanda; Ichimura, Hidehiko; Meghir, Costas (2007): Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds. In: *Econometrica* 75 (2): 323–363. DOI: 10.1111/j.1468-0262.2006.00750.x.

BMFSFJ. 2023. "Teilzeitquote von sozialversicherungspflichtig beschäftigten Frauen und Männern im Alter von 15 bis unter 65 Jahren nach Ländern". Retrieved April 11, 2023 (https://www.daten.bmfsfj.de/daten/daten/teilzeitquote-von-sozialversicherungspflichtig-beschaeftigten-frauen-und-maennern-im-alter-von-15-bis-unter-65-jahren-nach-laendern-131862).

Cha, Youngjoo, and Kim A. Weeden. 2014. "Overwork and the Slow Convergence in the Gender Gap in Wages". *American Sociological Review* 79 (3): 457–484. doi: 10.1177/0003122414528936.

Christofides, Louis N., Alexandros Polycarpou, and Konstantinos Vrachimis. 2013. “Gender wage gaps, ‘sticky floors’ and ‘glass ceilings’ in Europe”. *Labour Economics* 21: 86–102. doi: 10.1016/j.labeco.2013.01.003.

Combet, Benita, and Daniel Oesch. 2019. “The Gender Wage Gap Opens Long before Motherhood. Panel Evidence on Early Careers in Switzerland”. *European Sociological Review* 35 (3): 332–345. doi: 10.1093/esr/jcz009.

Collischon, Matthias (2019) “Is There a Glass Ceiling over Germany?” *German Economic Review* 20 (4): e329-e359. doi: 10.1111/geer.12168.

Eurostat. 2023a. “Employment – annual statistics”. Retrieved March 12, 2026 (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Employment_-_annual_statistics).

Eurostat. 2023b. “Gender pay gap in unadjusted form”. Retrieved April 3, 2023 (https://ec.europa.eu/eurostat/databrowser/view/SDG_05_20/default/table?lang=en).

Fortin, N. M. (2008). “The gender wage gap among young adults in the United States: The importance of money versus people”. *Journal of Human Resources*, 43(4): 884-918.

Goldin, Claudia. 2014. “A pollution theory of discrimination: male and female differences in occupations and earnings” Pp. 313–48 in *Human capital in history: The American record*, edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo. University of Chicago Press.

Guryan, J., & Charles, K. K. (2013). “Taste-based or statistical discrimination: the economics of discrimination returns to its roots.” *The Economic Journal*, 123(572): F417-F432.

Heckman, James J. 1979. “Sample Selection Bias as a Specification Error”. *Econometrica* 47 (1): 153-161. doi: 10.2307/1912352.

Huffman, Matt L., Joe King, and Malte Reichelt. 2017. “Equality for Whom? Organizational Policies and the Gender Gap across the German Earnings Distribution”. *ILR Review* 70 (1): 16–41. doi: 10.1177/0019793916673974.

Huntington-Klein, Nick (2025): The Effect : an introduction to research design and causality / Nick Huntington-Klein. Second edition. Boca Raton, FL: CRC Press. Available online at https://theeffectbook.net/.

Kitagawa, Evelyn M. 1955. “Components of a Difference Between Two Rates”. *Journal of the American Statistical Association* 50 (272): 1168-1194. doi: 10.2307/2281213.

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. “What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory”. *American Sociological Review* 86 (3): 532–565. doi: 10.1177/00031224211004187.

Mulligan, Casey B.; Rubinstein, Yona (2008): Selection, Investment, and Women's Relative Wages Over Time *. In: *The Quarterly Journal of Economics* 123 (3): 1061–1110. DOI: 10.1162/qjec.2008.123.3.1061.

Nisic, Natascha. 2017. “Smaller Differences in Bigger Cities? Assessing the Regional Dimension of the Gender Wage Gap”. *European Sociological Review* 33 (2): 292–304. doi: 10.1093/esr/jcx037.

Oaxaca, Ronald. 1973. “Male-Female Wage Differentials in Urban Labor Markets”. *International Economic Review* 14 (3): 693. doi: 10.2307/2525981.

Olivetti, Claudia; Petrongolo, Barbara (2008): Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. In: *Journal of Labor Economics* 26 (4): 621–654. DOI: 10.1086/589458.

Rohrer, Julia M. (2018): Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. In: *Advances in Methods and Practices in Psychological Science* 1 (1): 27–42. DOI: 10.1177/2515245917745629.

Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson. 2020. "Specification curve analysis". *Nature human behaviour* 4 (11): 1208–1214. doi: 10.1038/s41562-020-0912-z.

Zimmermann, Florian, and Matthias Collischon. 2023. "Do Organizational Policies Narrow Gender Inequality? Novel Evidence from Longitudinal Employer–Employee Data". *Sociological Science* 10: 47–81. doi: 10.15195/v10.a2.