

Pamela Campa, SITE  
Michał Myck, CenEA  
Monika Oczkowska, CenEA  
Maria Perrotta Berlin, SITE  
Jesper Roine, SITE  
March 2024

## Closing the Gender Data Gap

High-quality data plays a crucial role in enhancing our comprehension of evolving social phenomena, and in designing effective policies to address existing and future challenges. This particularly applies to the gender dimension of data, given the profound impact of the pervasive so-called “gender data gap”. In recent decades, data recovered from archives, high quality surveys, and census and administrative data, combined with innovative approaches to data analysis and identification, has become pivotal for the progress of documenting structural gender differences. Nonetheless, before we can close the gender gaps on the labour market, within households, in politics, academia and other areas, researchers and policy-makers must first ensure a closure of the gender data gap.



## Introduction

Any progress in our understanding of social phenomena hinges on the availability of data, and there is no doubt that recent advances in economics and other social sciences would not have been possible without countless high quality data sources. As we argue in this policy brief, this applies also, and perhaps particularly, to the documentation of different dimensions of gender inequalities and the analysis to identify their causes. Over the last few decades innovative ways to document historical developments, combined with improvements in the access to existing data, as well as new approaches to data collection, have become cornerstones in the progress made in our understanding of the various expressions of gender inequality. In the economic sphere this has covered themes such as labor market status, earning and income levels, wealth accumulation over the life course, education investments, pensions, as well as consumption patterns and time allocation – in particular caregiving and household work. Researchers have also been able to empirically study gender inequalities in politics, culture, crime, the justice system and in academia itself.

Groundbreaking studies in gender economics, including those by Claudia Goldin, the recent Nobel Prize laureate, would not have been possible without high quality data and innovative ways aimed at closing the “gender data gap”, a term coined by Caroline Criado Perez, in her bestseller “Invisible women” (Criado Perez, 2020). In the introduction to the book she notes that “(...) the chronicles of the past have left little space for women’s role in the evolution of humanity, whether cultural or biological. Instead, the lives of men have been taken to represent those of humans overall.” (p. XI). The gender data gap is the result of deficits of informative data sources on women, which has been augmented by frequent lack of differentiation of information by sex/gender in available sources. Closing the gender gaps along the dimensions already identified in existing

studies will require a continuous monitoring of evidence, thus closing the gender data gap in the first place. New studies focused on greater equality and on the effectiveness of various implemented policies will continue to rely on good data. Thankfully, few new datasets currently ignore the gender of the respondents. However as our understanding of the biological and cultural aspects of sex and gender grows, the way data is collected will need to be modified.

As we prepare for the new challenges ahead of those designing data collection efforts and examining the data, we believe it is important to give credit to the authors of some of the groundbreaking studies that paved the way to the current pool of evidence on gender inequality. Around the time of the International Women’s Day, we recall several empirical studies in gender economics that, in our opinion, merit special attention due to either their innovative approaches to data collection, their unique access to original data sources, or their methodological novelty. These studies bring valuable insights into specific dimensions of gender inequality. This short list is naturally a subjective choice, but we believe that all of these studies deserve credit not only among researchers within gender economics, but also among those more broadly interested in the recent progress in the understanding of different aspects of gender inequality.

## From Data to Policy Recommendations

Over the last few decades substantial efforts have been made to provide empirical evidence concerning **historical trends** in inequalities between men and women on the labor market. Seminal work in this field was conducted by Claudia Goldin in the 1970s and 80s, culminating in the publication of the path-breaking book *Understanding the Gender Gap: An Economic History of American Women* (Goldin, 1990). The book fundamentally changed the view of women’s role in the labor market. Empirically Goldin shows that female labor force participation has been significantly higher in historical times than



previously believed. Before Goldin, researchers mainly studied twentieth century data. Based on this it looked as if women's participation in the labour market is positively correlated with economic growth. Goldin's work showed instead that women were more likely to participate in the labour force prior to industrialization, and that early expansion of factories made it more difficult to combine work and family. Seen over the full 200 year period, from before industrialization to today, the pattern of women's labour market participation is in fact U-shaped, pointing to the importance of various societal changes that alter incentives and possibilities for women's work. Goldin's contribution is however not just about getting the empirical picture right. At least equally important is the recognition of women as individual economic agents, who make forward looking decisions under various institutional constraints and limitations related to social norms about identity and family, as well as education opportunities and labor market options. While some decision can be modeled as taken by "the economic *man*", others by *households*, it may seem surprising that studying *women's* decisions was for so long neglected.

Institutional, cultural and economic factors behind historical trends have become the focus of much of the literature trying to identify the forces driving gender disparities. Some of the most original work considers the role that "chance" plays in determining individual decisions related to gender – how having a first-born son (e.g. Dahl and Moretti, 2008) or having twins (Angrist and Evans, 1998), both of which can be considered random, – affect choices related to partnership, future fertility and the labor market. Others examine the influence of gender imbalances caused by major historical events. Brainerd (2017) investigates the consequences of extremely unbalanced sex ratios in cohorts particularly affected by the massive loss of lives during World War II in the Soviet Union. By exploiting a unique historical data source derived from the first postwar census, combined with statistics registry

records from archives, Brainerd provides evidence that the war-induced scarcity of men profoundly affected women's outcomes on the marriage market. Women were more likely to never get married, give birth out of wedlock and get divorced. On top of that, unbalanced sex ratios affected married women's intrahousehold bargaining power and resulted in lower fertility rates and a higher rate of marriages with a large age gap between spouses. The post-war institutional setup increased the cost of divorce and withdrew legal obligations to support children fathered out of wedlock, which exacerbated the consequences from the shortage of men by further reducing the rates of registered marriages and increasing marital instability.

The examples above highlight how conditions beyond individuals' control can contribute to social gender imbalances, or shed light on existing gender biases. How these 'exogenous' circumstances translate into economic inequalities and what additional factors drive disparities has been the focus of much academic work on gender inequalities. One of the most challenging questions has been that of demonstrating that **discrimination** of women, rather than women's characteristics or choices, are behind the growing body of evidence on economic gender inequality. In this respect Black and Strahan (2001) provide important convincing conclusions by using significant changes in the level of regulation in the US banking sector. Increasing competition between banks lowered banks' profits, and led to a reduced ability of managers to 'divide the spoils', and thus to discriminate between different types of employees. The authors used information on wages within specific industries (including banking) from one of the oldest ongoing surveys in the world – the US Current Population Survey (CPS). By exploiting detailed individual data covering a period of several decades the authors show that higher levels of banking sector regulations (prior to deregulation) facilitated greater premia paid out to male compared to female employees. Thus, increased competition in



the banking sector brought favorable changes to women's pay conditions as well as their position in banks' management.

While long running surveys such as the CPS continue to serve as invaluable sources of information on the relative conditions of men and women, the growing availability of administrative data has opened new opportunities for documentation of inequalities and identification of the reasons behind these. For instance, the ability to track individuals throughout their work history before and after the arrival of their first child has allowed researchers to compare the trajectories of women's and men's earnings, wages and working hours. This comparison has revealed the existence of the so-called "child penalty", with women experiencing a drop in their labor market position relative to their male partners after the birth of their first child, and with the gap persisting for many years. Strikingly, this penalty has been estimated in some of the most gender-equal countries in the world, such as Sweden (Angelov et al., 2016) and Denmark (Kleven et al., 2019), two countries which have spearheaded collecting and making rich administrative data available to researchers.

Another area where individual register data has proven invaluable is in the study of the so-called "glass ceiling", i.e., the sharply increasing differences between men and women when it comes to pay as well as representation in the very top of the income distribution. In a seminal study by Albrecht et al. (2003), individual earnings for men and women were compared and differences were found to be markedly higher (with men earning much more) when comparing men in the top of the male income distribution with women in the top of the female income distribution. Also making use of Swedish registry data, Boschini et al. (2020) study a related question, namely the evolution of the share of women in the top of the income distribution. In line with other glass-ceiling results, they demonstrate that the share of women in the top is small, and that it gets smaller the higher one looks, although it has increased

over time. Decomposing incomes into labor earnings and capital income they also show that while women seem to be catching up in the labor income distribution, they clearly lag in the capital income distribution. Also, the income profile of the partners of high-income men and high-income women are strikingly different. Most high-income women have high-income partners, while the opposite is not true for high-income men.

Differences in the economic position of men and women reflected in the above examples can have their origin much before the time individuals enter the labor market. They can be driven by differences in schooling opportunities, as well as other forms of early life investments, to the extent that even much of what is perceived as choices or preferences later in life are in fact results of these subtle early life disadvantages for women. While these have largely diminished in the global North, there is a growing number of studies documenting these differences in the global South. Jayachandran and Pande (2017) examine the impact of son preference, a widespread cultural practice for example in India, on child health and development. The study leverages a simple, standardized, and broadly available indicator – the height of children – which is measured at routine health checks and included in many population surveys, such as the Demographic and Health Surveys (DHS). Additionally, their use of a natural experiment, based on the birth order of children, helps to establish a causal relationship between eldest son preference and nutritional disparities that have long-term developmental consequences among subsequent children, not only for girls but for Indian children on average. Findings like these underscore the importance of gender equality not only as a fundamental value but also as a crucial factor in promoting growth and development at the societal level.

The social costs of gender inequality have also motivated the growing research interest in gender-based violence and crime. Given the specific challenges associated with these topics – such as the clandestine and underreported nature



of these acts but also the consideration for victims' confidentiality and safety – studies in this area has required researchers to develop and apply innovative tools and data collection methods. In this framework list experiments have emerged as a methodology allowing respondents to disclose sensitive or socially undesirable attitudes indirectly, reducing the likelihood of the so-called social desirability bias in survey reporting. In a list experiment, respondents are presented with a set of statements or behaviors and asked to indicate their agreement or engagement with these. Among listed items, one is considered “sensitive” and is included only for a randomly selected subset of respondents. By comparing the average number of items agreed with by the entire sample to a control group that did not get the sensitive item, researchers can estimate the proportion of respondents who agreed with or engaged in the sensitive behavior or opinion. Kuklinski et al. (1997) is one of the pioneering contributions in this area, estimating the proportion of voters who harbored racial prejudices but who may have been unwilling to admit it in a direct survey question. List experiments have since become a widely used tool in political science and economics and have helped in the advancement of our understanding of gender-based violence (Peterman et al., 2018). Given the strong assumptions underlying the analysis the method has not become the “statistical truth serum” it was at some point considered to be. However, list experiments have broadened the analytical opportunities in an area plagued by significant informational and data challenges.

While worldwide gender gaps in economic opportunities and especially in education and health have rapidly declined (and sometimes reversed) in the last decades, larger differences remain in political empowerment (see e.g., WEF Gender Gap Report 2023). Another Nobel Prize laureate in economics, Esther Duflo, in her joint work with Raghavendra Chattopahyay (2004), have pioneered a highly prolific area of research on the impacts of women as policymakers. In their study, they leverage a unique policy experiment in

India that randomized the gender of the leader of Village Councils, and a detailed dataset based on extensive surveys administered to both Village Council leaders and villagers. The surveys allowed for estimation of the investments in different public goods in 265 Village Councils, as well as the preferences over each of these public goods among female and male villagers. Combining the randomization and this rich dataset, the authors establish that political leaders prioritize public goods that are more relevant to the needs of their own gender, suggesting that women's under-representation in politics might result in women's and men's preferences being unequally represented in policy decisions.

## Conclusions and Recommendations

The narrowing gender gap in political representation across various levels of government, the growing influence of women in other areas such as public institutions, administration etc., and the heightened awareness of the crucial role gender equality plays in socio-economic progress all bode well for improvements in access to high-quality gender-differentiated data sources. Before we can recognize and close gender gaps identified from high-quality data, the gender data gap needs to firstly be closed. Governments and public institutions should make their increasing amounts of digitized information available for research purposes. Funding should be available to collect data through surveys, and these could in turn be combined with details available in administrative sources to take advantage of the breadth of survey data and the precision of official statistics. Information needs to be collected on a frequent and regular basis to make sure that the consequences of various major developments, such as legal changes, conflicts or natural disasters, can be identified. Innovative data sources, for instance information from mobile apps or social media, can provide additional useful insights into socio-economic trends, old and new dimensions of inequalities and regular timely updates on different aspects of gender disparities.





These new data sources can become the basis for future innovative studies on gender inequalities, contributing to a better understanding of the mechanisms behind these inequalities, and providing evidence for policies and other efforts to effectively close the remaining gaps. Already now there is enough evidence to conclude that closing these gaps is not only just but that it also constitutes a fundamental basis for continued inclusive economic development.

## Post Scriptum

Contributing to the existing pool of data sources we are happy to share a regional dataset with information on gender norms and gender-based violence: the *FROGEE Survey 2021*. The data was collected using the CATI method (phone interviews) in autumn 2021 in Belarus, Georgia, Latvia, Poland, Russia, Sweden and Ukraine. In each country interviews were conducted with between 925 and 1000 adults. The survey covered areas such as: basic demographics, material conditions, labor market status, gender norms, attitudes towards harassment and violence, awareness of violence against women and awareness of legal protection for gender violence victims.

The data collection was funded by the Swedish International Development Cooperation Agency (SIDA) as part of the FREE Network's **FROGEE project**. The dataset and supporting materials are freely available for research purposes. For more information see: **FROGEE Survey on Gender Equality**.

## References

- Angrist, D. J., and Evans, N. W. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(2), 450-477.
- Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a glass ceiling in Sweden? *Journal of Labor Economics*, 21(1), 145-177.
- Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3), 545-579.
- Black, S. E., and Strahan, P. E. (2001). The division of spoils: Rent-sharing and discrimination in a regulated industry. *American Economic Review*, 91(4), 814-831.
- Boschini, A., Gunnarsson, K., and Roine, J. (2020). Women in top incomes: Evidence from Sweden 1971–2017. *Journal of Public Economics*, 181, 104-115.
- Brainerd, E. (2017). The lasting effect of sex ratio imbalance on marriage and family: Evidence from World War II in Russia. *The Review of Economics and Statistics*, 99(2), 229-242.
- Chattopadhyay, R., and Duflo, E. (2004). Women as policymakers: Evidence from a randomized policy experiment in India. *Econometrica*, 72(5), 1409-1443.
- Criado Perez, C. (2020). *Invisible women*. Vintage, London.
- Dahl, G. B., and Moretti, E. (2008). The demand for sons. *Review of Economic Studies*, 75(4), 1085-1120.
- Goldin, C. (1990). *Understanding the Gender Gap: An Economic History of American Women*. Oxford University Press.
- Kleven, H., Landais, C., and Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4), 181-209.
- Kuklinski, J. H., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. R., & Mellers, B. (1997). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science*, 402-419.
- Jayachandran, S., and Pande, R. (2017). Why are Indian children so short? The role of birth order and son preference. *American Economic Review*, 107(9), 2600-2629.
- Peterman, A., Palermo, T. M., Handa, S., Seidenfeld, D., and Zambia Child Grant Program Evaluation Team (2018). List randomization for soliciting experience of intimate partner violence: Application to the evaluation of Zambia's unconditional child grant program. *Health Economics*, 27(3), 622-628.





## Pamela Campa

Stockholm Institute of Transition Economics  
pamela.campa@hhs.se  
<https://www.hhs.se/sv/persons/c/campa/pamela/>

Pamela Campa is an Associate Professor of Economics at the **Stockholm Institute of Transition Economics** (SITE) at Stockholm School of Economics. Before joining SITE she was an Assistant Professor at the **University of Calgary**. Pamela holds a PhD in Economics from the **Institute for International Economic Studies (IIES)** at Stockholm University. She is a Research Affiliate at the **Center for Economic Policy Research (CEPR)**, the **Mistra Center for Sustainable Markets (MISUM)** and **Dondena Gender Initiative**.

Her research interests are in Political Economy, Environmental Economics and Gender Economics.



## Michał Myck

Centre for Economic Analysis (CenEA)  
mmyck@cenea.org.pl  
[www.cenea.org.pl](http://www.cenea.org.pl)

Michał Myck is the Director of CenEA. He previously worked at the Institute for Fiscal Studies in London and at the DIW-Berlin. He received his B.A. and M.Phil. degrees at the University of Oxford (PPE – 1997, Economics – 1999), Ph.D. degree at the University of Warsaw (2006) and completed his habilitation at the Free University Berlin (2015). His work focuses on labor and health economics and applied micro-econometrics.



## Monika Oczkowska

Centre for Economic Analysis (CenEA)  
moczowska@cenea.org.pl  
[www.cenea.org.pl](http://www.cenea.org.pl)

Monika Oczkowska works at CenEA as a Senior Research Economist since March 2012. She received a M.A. degree in Economics in 2012, and BA degrees in Economics (2010) and International Relations (2011) at the University



of Szczecin. Her research interests include the economics of ageing, well-being, gender and health.



## Maria Perrotta Berlin

Stockholm Institute of Transition Economics  
maria.perrotta@hhs.se  
<https://www.hhs.se/en/persons/p/perrottaberlin-maria/>

Maria Perrotta Berlin is Assistant Professor and Policy and Communication Manager at the Stockholm Institute of Transition Economics (SITE). She earned her Ph.D. from the Institute for International Economic Studies (IIES) at Stockholm University. Maria's main research interests are development and political economics, with a focus on gender issues and environmental policy.



## Jesper Roine

Stockholm Institute of Transition Economics  
jesper.roine@hhs.se  
<https://www.hhs.se/sv/persons/r/roine-jesper>

Jesper Roine is Professor at the Stockholm School of Economics and the Deputy Director at the Stockholm Institute of Transition Economics (SITE). Most of his research concerns income and wealth distribution and long-run development, but he has also worked on political economy and the impact of natural resources on the economy.

In addition to his research, Roine participates widely in the policy debate. He is one of the founders and regular contributor to the economics blog [ekonomistas.se](http://ekonomistas.se) (written in Swedish). Roine has written a number of reports on issues such as globalization and the income distribution, the future of the Swedish welfare state, and how the Swedish government should manage its mineral resources. He is currently a member of the government's commission on the future of work.

## [freepolicybriefs.com](http://freepolicybriefs.com)

---

The Forum for Research on Eastern Europe and Emerging Economies (FREE) is a network of academic experts on economic issues in Eastern Europe and the former Soviet Union at BEROE (Minsk), BICEPS (Riga), CEFIR (Moscow), CenEA (Szczecin), ISET-PI (Tbilisi), KSE (Kiev) and SITE (Stockholm). In 2019 the FREE Network, with support of the Swedish International Development Cooperation Agency (Sida) initiated the Forum for Research on Gender Economics (FROGEE). Publications under the FROGEE initiative contribute to the discussion on gender inequality in the region of Central and Eastern Europe. Opinions expressed in all FREE Network publications are those of the authors; they do not necessarily reflect those of the FREE Network, its research institutes or Sida.