DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 53, ARTICLE 29, PAGES 937–968
PUBLISHED 19 NOVEMBER 2025
https://www.demographic-research.org/Volumes/Vol53/29
DOI: 10.4054/DemRes.2025.53.29

Replication

Reassessing the U-shaped relationship between gender equality and fertility: A replication and extension of Kolk's (2019) study using comprehensive gender equality measures

Haohao Lei

Contents

1	Introduction	938
2	Background	938
2.1	Theoretical framework	938
2.2	Empirical evidence	940
3	Data	942
3.1	Fertility data	942
3.2	Introducing the Historical Gender Equality Index	942
4	Results	946
4.1	Replication with HGEI	946
4.2	Adjusting period TFR’s tempo distortion	948
5	Conclusion and discussion	951
6	Acknowledgements	954
	References	955
	Appendix	958

Demographic Research: Volume 53, Article 29
Replication

Reassessing the U-shaped relationship between gender equality and fertility: A replication and extension of Kolk's (2019) study using comprehensive gender equality measures

Haohao Lei[1]

Abstract

BACKGROUND

While theories predict that fertility initially declines and subsequently rebounds with rising gender equality, previous empirical studies have produced mixed results, partly due to differences in gender equality measurement and methodological approaches.

OBJECTIVE

This study replicates Kolk's (2019) research, which found weak evidence for a U-shaped relationship between public-sphere gender equality and fertility. The study re-evaluates these findings using data covering the period 1950–2003 and incorporating a broader, multidimensional measure of gender equality, the Historical Gender Equality Index.

RESULTS

The analysis identifies a conditional U-shaped relationship between gender equality and fertility, meaning that the expected fertility rebound at high gender equality emerges only when controlling for long-term fertility-decline trends and fertility tempo distortions.

CONCLUSIONS

The replication provides evidence that the previously identified weak support for a U-shaped relationship may stem from the choice of gender equality indicators. By employing a more comprehensive measure and controlling for period and country fixed effects, this analysis reveals a conditional U-shaped relationship. However, this modest fertility rebound at high gender equality levels is insufficient to reverse the broader, long-term decline in fertility observed across advanced societies.

CONTRIBUTION

This study reveals that the weak support for the U-shaped relationship identified in previous research stems from the selection of gender equality measures. By adopting a more appropriate measure of gender equality, this analysis aligns empirical evidence with

[1] Department of Sociology, University of Oxford; Nuffield College, University of Oxford; Leverhulme Centre for Demographic Science, University of Oxford; Oxford, UK; Max Planck Institute for Demographic Research, Rostock, Germany. Email: haohao.lei@nuffield.ox.ac.uk.

contemporary theories. Furthermore, the differing results between previous studies and this analysis emphasise the critical role of gender equality in shaping fertility trends in both the private and public spheres.

1. Introduction

Demographers have long debated whether there is a U-shaped relationship between country-level gender equality and fertility (Arpino, Esping-Andersen, and Pessin 2015; Kolk 2019; McDonald 2013). A U-shaped pattern here means that an increase in gender equality initially leads to a decrease in the total fertility rate (TFR), and then, once it reaches a certain point, further increases in gender equality lead to an increase in TFR at the country level. The recent empirical study by Kolk (2019), which focuses on this topic, suggests that there is only weak support for a U-shaped pattern between societal gender equality and the TFR when comparing societies over time. This study replicates and extends Kolk's (2019) analysis using data covering the period 1950–2003 and employing a more comprehensive gender equality indicator that captures multiple domains beyond female political empowerment. Specifically, the weak support for a U-shaped relationship between gender equality and fertility in Kolk's (2019) study is likely due to the selection of gender-equality measures. My findings demonstrate a conditional U-shaped relationship between gender equality and fertility when employing a more comprehensive measure of gender equality and accounting for secular fertility decline. However, this rebound is insufficient to offset the overall decline in fertility: even at peak equality, the overall TFR does not exhibit a reversal as relevant theories predict.

2. Background

2.1 Theoretical framework

Gender equity or equality is one of the key determinants of fertility because gender influences reproductive behaviours through mechanisms such as parental investment (Trivers 2017) and autonomy in the household (Folbre 1983). Gender equality measures, such as increasing the percentage of educated females in the labour force, have long been believed to negatively correlate with fertility at the country level (United Nations 1995). However, recent theoretical developments in demography are making demographers rethink the linkage between gender equality and fertility: they suggest that there is a U-shaped pattern between gender equality and fertility at the country level (Anderson and

Kohler 2015; Esping-Andersen and Billari 2015; Goldscheider, Bernhardt, and Lappegård 2015; McDonald 2000a, 2000b).

McDonald (2000a, 2000b) offers a theory rooted in individual-level decision-making. McDonald (2013) indicates that the main driving force behind the initial fertility decline is the gap between the rapid increase in gender equity in individual-oriented institutions, such as the education system and labour markets, and the stagnation of gender equity in family-oriented institutions. McDonald (2013) then predicts that a country's fertility rate will bounce back to the replacement level once gender equity in family-oriented institutions catches up with that in individual-oriented institutions. If the gender equity gap between individual- and family-oriented institutions remains significant, the country will continue experiencing an extremely low fertility rate, mainly because increasing gender equity in individual-oriented institutions increases women's opportunity costs of raising a child, which leads to lower fertility intentions. Once gender equity in family-oriented institutions increases, the cost of raising a child for women will be lower because men will share part of the cost. Therefore, the country-level fertility rate may revert to the replacement level. McDonald (2013) further highlights that his theory can only be tested using aggregated data across countries and should not be used to explain individual-level fertility variation within a country.

Goldscheider, Bernhardt, and Lappegård (2015) hypothesise a theoretical framework very similar to McDonald's (2000a, 2000b). Their theory is based on a twofold gender revolution framework. According to this framework, the first part of the gender revolution changes women's gender roles and allows them to enter the public sphere by, for example, accessing the labour market and education. This access leads to a decline in fertility rates because of the increasing opportunity cost of childbearing. Then, in the second part of the gender revolution, Goldscheider, Bernhardt, and Lappegård (2015) predict there will be changes in men's gender roles, accompanied by improved gender equity in the private sphere. According to their prediction, men will share more responsibility in the private sphere by, for example, spending more time on unpaid domestic work. These changes in men's gender roles will reduce the cost of raising a child for women and increase their fertility intentions. Compared with McDonald's (2000a, 2000b) theory, the theory proposed by Goldscheider, Bernhardt, and Lappegård (2015) focuses more on the importance of men's roles; however, the fundamental logic of the theories is very similar.

Besides Goldscheider, Bernhardt, and Lappegård (2015) and McDonald (2000a, 2000b), other researchers (Anderson and Kohler 2015; Esping-Andersen and Billari 2015) have proposed similar gender-fertility reversal theories, which also emphasise mechanisms based on gender equity or equality gaps both inside and outside the household, as well as the opportunity costs of raising children. However, a notable limitation of these frameworks, excluding McDonald's (2000a, 2000b), is their conceptual ambiguity regarding the appropriate level of analysis and the definition of gender equity or equality, which hinders the consistency of empirical tests of the hypothesised U-shaped relationship.

Before discussing the specific empirical evidence regarding gender and fertility, it is necessary to distinguish between the concepts of equality and equity (Esping-Andersen and Billari 2015; Fraser 1994; McDonald 2000a, 2000b, 2013). According to McDonald's (2013) classification, gender equality can be straightforwardly measured by the outcome gap between men and women, such as the gender wage gap and the gender education participation gap. However, gender equity cannot be straightforwardly measured based on the outcome gap, as it pertains to perceptions of equal opportunity. Therefore, even with a gender gap in certain outcomes in society, a high level of gender equity can still be present as long as both genders perceive the outcome gap to be fair in terms of equal opportunities (Esping-Andersen and Billari 2015; McDonald 2013). Although McDonald's (2000a, 2000b) theory emphasises that gender equity, rather than equality, determines fertility, other theories, such as that proposed by Goldscheider, Bernhardt, and Lappegård (2015), focus more on equality.

Most empirical studies concentrate on identifying the relationship between gender equality and fertility, mainly because the concept of gender equity is difficult to measure, especially at the country level, as it is closely related to individual perceptions. Furthermore, in cross-cultural comparative studies, capturing the concept of equity in various cultural contexts using survey questions is challenging due to language differences and the complex notions of gender equity (McDonald 2013). For these practical reasons, in this study I will follow Kolk's (2019) research question and focus on gender equality instead of equity.

2.2 Empirical evidence

Empirical evidence that focuses on the country-level relationship between gender equality and fertility is limited. Arpino, Esping-Andersen, and Pessin (2015) examine data from 27 countries at three points and find support for a U-shaped relationship between the TFR and country-level gender-egalitarian views. In addition, a smaller gender difference in gender-egalitarian views is associated with a sharper reversal in fertility rates. Arpino, Esping-Andersen, and Pessin (2015) introduce an innovative approach; however, their measure relies on the proportion of respondents holding gender-egalitarian views rather than employing a standardised gender equality index, limiting comparability and precision. Additionally, since Arpino, Esping-Andersen, and Pessin (2015) rely on data from only three distinct time points rather than continuous longitudinal data, it remains unclear whether their identified U-shaped pattern represents a stable long-term trend or temporary fluctuations. Therefore, before Kolk (2019) published his study, there was a significant empirical gap in understanding how country-level gender equality is associated with TFR longitudinally.

Kolk (2019) addressed this gap by examining both cross-sectional and longitudinal evidence. His study analyses data from 35 countries and regions, encompassing a total of 1,993 cumulative years of data. Kolk (2019) specifically focuses on gender equality rather than equity and uses the Women's Political Empowerment Index (WPEI) as an indicator of gender equality. The V-Dem Institute develops this indicator by capturing annual information about women's civil liberties, civil society participation, and political participation in 170 countries. As a longitudinal indicator, the WPEI includes annual gender equality data from 1900 to 2012 and is one of the most comprehensive longitudinal indicators of women's empowerment (Coppedge et al. 2024). In a robustness check, Kolk (2019) suggests that the WPEI is strongly correlated with widely used cross-sectional gender equality indicators, such as the Gender Inequality Index, Gender Equality Index, Global Gender Gap Index, and the Index of Conditions for Work and Family Reconciliation, for 2010.

After examining the cross-sectional association between fertility and the WPEI, Kolk (2019) identifies an overall negative association but highlights the curvilinear relationship between fertility and the WPEI after 2000. Therefore, Kolk (2019) concludes that the post-2000 cross-sectional pattern for fertility and gender equality is consistent with the predictions of fertility-equality reversal theories. Kolk (2019) then provides a descriptive visualisation of the longitudinal association between fertility and the WPEI within countries, which indicates an opposite result. The longitudinal association is negative in most of the 35 countries, with no reversal in TFR, with the exception of Belgium, Denmark, France, and the Netherlands. This within-country pattern suggests that the post-2000 fertility–equality reversal pattern may be purely driven by between-country variation. In short, there is no evidence to suggest that, within the same society, increases in gender equality will lead to TFR increases.

According to Kolk (2019), the OLS regression analysis of the WPEI on the TFR also supports this conclusion. The model based on the complete sample suggests a negative cross-sectional relationship between fertility and the WPEI. After controlling for period trends, the between-country comparison model indicates a U-shaped pattern between the TFR and the WPEI. When focusing only on samples after 1990 and controlling for period trends, the U-shaped pattern between the TFR and the WPEI becomes particularly significant. However, when the model controls for time-invariant country characteristics the association between the TFR and the WPEI becomes negative, and no U-shaped relationship occurs. Only in the model where fertility's period trend is controlled does there seem to be a slightly higher TFR at the highest WPEI level. However, this pattern does not occur after 1990. Based on this evidence, Kolk (2019) concludes that there is only weak support for a U-shaped pattern between societal gender equality and fertility within a society over time.

3. Data

3.1 Fertility data

For the replication, I accessed the same fertility data source as Kolk (2019) from the Human Fertility Database. For countries with vague definitions or overseas territories, I followed Kolk's classification methods: the United Kingdom refers to England and Wales, France refers to mainland France excluding overseas territories, and Germany refers to Western Germany until reunification (1990) and then all of Germany. Finally, in my replication I removed observations with missing TFR. Most of the data excluded due to missing values are from before 1960. The countries I included in my main analysis are: Australia, Austria, Belgium, Bulgaria, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Lithuania, the Netherlands, New Zealand, Norway, Poland, Portugal, Russia, Slovakia, Slovenia, South Korea, Spain, Sweden, Switzerland, Ukraine, the United Kingdom, and the United States.

3.2 Introducing the Historical Gender Equality Index

Most theories predicting a U-shaped relationship between gender equality and fertility emphasise the gender equality gap both inside and outside the household. McDonald (2013) uses the terms 'family-oriented institutions' and 'individual-oriented institutions', whereas Goldscheider, Bernhardt, and Lappegård (2015) use the terms 'private sphere' and 'public sphere'. However, the gender equality indicator used in Kolk's (2019) study, the WPEI, focuses only on gender equality outside the household, or more specifically, female political empowerment. This indicator does not capture any information regarding gender equality inside the household. Kolk (2019, p.35) suggests that female political empowerment may encourage institutions that support greater private-sphere gender equality. However, this assumption is not explicitly supported by existing theory or empirical evidence. More importantly, gender equality is multidimensional (UNDP 2010), and only focusing on improvements in female political empowerment may not capture all information about changes in gender equality. Therefore, the absence of a U-shaped relationship between the WPEI and TFR cannot disprove the gender–fertility reversal theories.

To extend Kolk's (2019) original paper and address some of the limitations associated with the WPEI, I propose introducing a new longitudinal gender indicator into the analysis: the Historical Gender Equality Index (HGEI), developed by Dilli, Carmichael, and Rijpma (2019). As a longitudinal indicator, the HGEI covers 129 countries from 1950 to 2003. It addresses key limitations of the WPEI by providing multidimensional data that better aligns with theoretical conceptions of gender equality.

The HGEI captures four key dimensions of gender equality: health (life expectancy and child sex ratios), household autonomy (age gap at first marriage), political power (women's parliamentary representation), and socioeconomic resources (average years of schooling and labour force participation). Formally, the HGEI is constructed using the following equation:

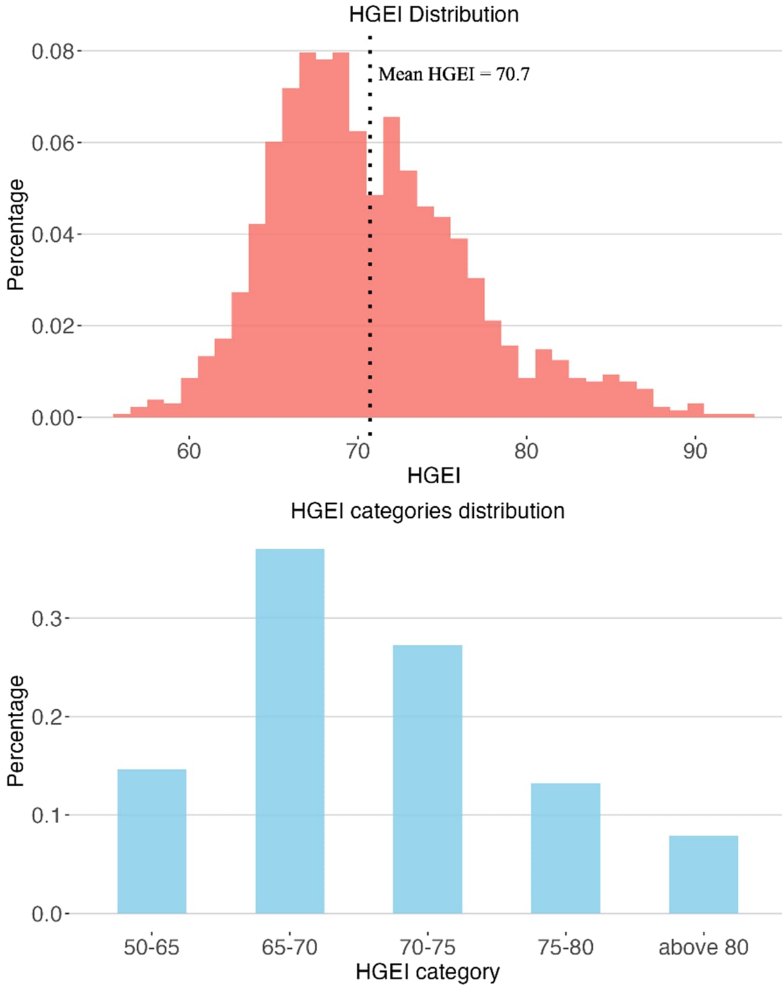
$$\begin{aligned} \mathrm{HGEI}_{c,t} = 100 \Bigg[& 0.09 \min\left(\frac{\mathrm{LifeExp_F}}{\mathrm{LifeExp_M}}, 1\right) + 0.16 \min\left(\frac{\mathrm{SexRatio_F^{0\text{-}5}}}{\mathrm{SexRatio_M^{0\text{-}5}}}, 0.944\right) \\ & + 0.25 \min\left(\frac{\mathrm{AgeFirstMarriage_F}}{\mathrm{AgeFirstMarriage_M}}, 1\right) \\ & + 0.25 \min\left(\frac{\mathrm{ParliamentSeats_F}}{\mathrm{ParliamentSeats_M}}, 1\right) \\ & + 0.12 \min\left(\frac{\mathrm{YearsSchooling_F}}{\mathrm{YearsSchooling_M}}, 1\right) \\ & + 0.13 \min\left(\frac{\mathrm{LFParticipation_F}}{\mathrm{LFParticipation_M}}, 1\right) \Bigg] \end{aligned}$$

The HGEI measures women's relative position compared to men's by calculating female-to-male ratios for each component. Health indicators, such as sex ratio at birth and life expectancy, are adjusted for biological differences between the sexes. The interpretation of the index is straightforward: an HGEI value below 1 indicates inequality disadvantaging women, a value of 1 reflects perfect gender equality, and a value above 1 suggests inequality disadvantaging men. Its multidimensional structure allows researchers to capture gender inequality comprehensively, and its longitudinal design enables analysis across a sufficiently long time span.

The HGEI has some limitations. First, using the gender gap in marriage age as a proxy for household autonomy is not ideal because the postponement of marriage itself is related to fertility rates in ways not directly tied to gender equality. Nonetheless, the relative gender gap in marriage age is independent of the absolute level of postponement of marriage. Second, the HGEI does not include information regarding a country's gender inequality levels after 2003. As Kolk's (2019) original study suggests, much of the reversal in fertility in terms of gender equality occurred in the post-2000 period. The lack of data from the post-2000 years may underestimate the strength of the potential U-shaped relationship between gender equality and fertility. Nevertheless, it remains relevant to examine whether fertility reversals began before 2000 alongside improvements in private-sphere gender equality.

I accessed the HGEI data from the Clio Infra project website. Due to the fact that the HGEI only covers countries' gender equality from 1950 to 2003, compared to Kolk's (2019) original research based on the WPEI the total cumulative years of data decrease from 2,054 to 1,280. In addition, three countries and region in Kolk's (2019) original study – Iceland, Belarus, and Taiwan – are not included in this extension due to a lack of HGEI data; therefore, the total number of countries included in this analysis decreases from 35 to 32. Figure 1 shows the distribution of the HGEI data. The distribution of the HGEI is slightly right-skewed, with a mean value of 70.70. Meanwhile, most of the samples included in the analysis have an HGEI category between 65 and 70. A relatively small percentage of the samples belong to the highest gender-equality level, with an HGEI above 80.

Figure 1: Descriptive statistics for HGEI

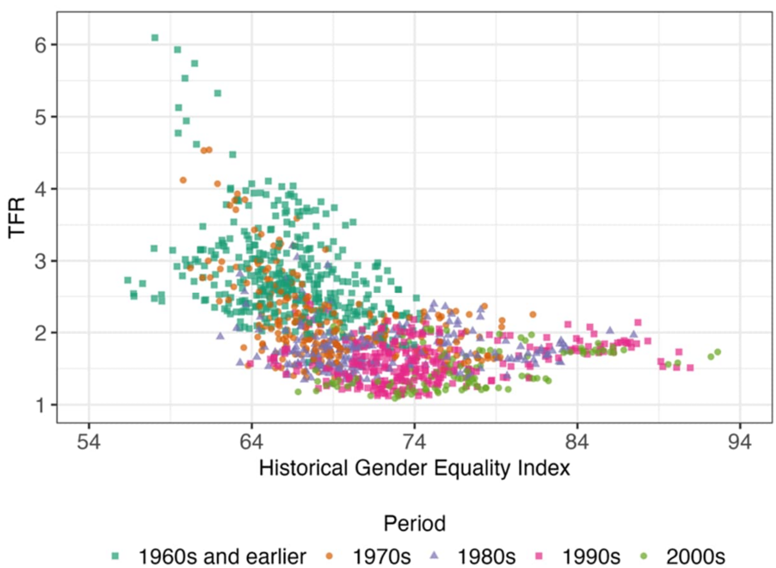


4. Results

4.1 Replication with HGEI

In Figure 2, I follow Kolk's (2019) study and visualise the relationship between the HGEI and TFR. The overall relationship is strongly negative at low-to-medium HGEI levels and then flattens out at high HGEI, and there is a slight attenuation of the negative slope among the highest-equality observations. This curvilinear, slightly U-shaped pattern appears more pronounced here than in Kolk's original analysis. I then used an OLS regression analysis to test whether this curvilinear relationship between the HGEI and TFR is purely driven by between-country effects.

Figure 2: Association between HGEI and TFR in 32 countries for all periods



To test the U-shaped relationship between the HGEI and country-level period fertility, I estimated a sequence of ordinary least squares (OLS) models that progressively include year and country fixed effects. The basic specification is:

$$TFR_{c,t} = \alpha + \beta_1 D_{c,t}^{50-65} + \beta_2 D_{c,t}^{65-70} + \beta_3 D_{c,t}^{75-80} + \beta_4 D_{c,t}^{\geq 80} + \varepsilon_{c,t}$$

where $TFR_{c,t}$ denotes the total fertility rate in country *c* and year *t*, $D^{g}_{c,t}$ are dummy variables indicating the Historical Gender-Equality Index (HGEI) interval g (the 70–75 band serves as the reference group), and $\varepsilon_{c,t}$ is an idiosyncratic error term.

In Model 1.2, I add year fixed effects:

$$TFR_{c,t} = \alpha + \sum_{g} \beta_g \, D^{g}_{c,t} + \delta_t + \varepsilon_{c,t}$$

Here, the term $\sum_g \beta_g D^{g}_{c,t}$ represents the set of estimated coefficients for each HGEI interval g, with the 70–75 category omitted as the reference group. This allows the model to compare the fertility level in each HGEI interval relative to the 70–75 band. Each coefficient β_g captures the average difference in TFR between the respective HGEI category and the reference group, holding other factors constant.

Next, I control for all time-invariant country heterogeneity by including country fixed effects μ_c in model 1.3:

$$TFR_{c,t} = \alpha + \sum_{g} \beta_g \, D^{g}_{c,t} + \mu_c + \varepsilon_{c,t}$$

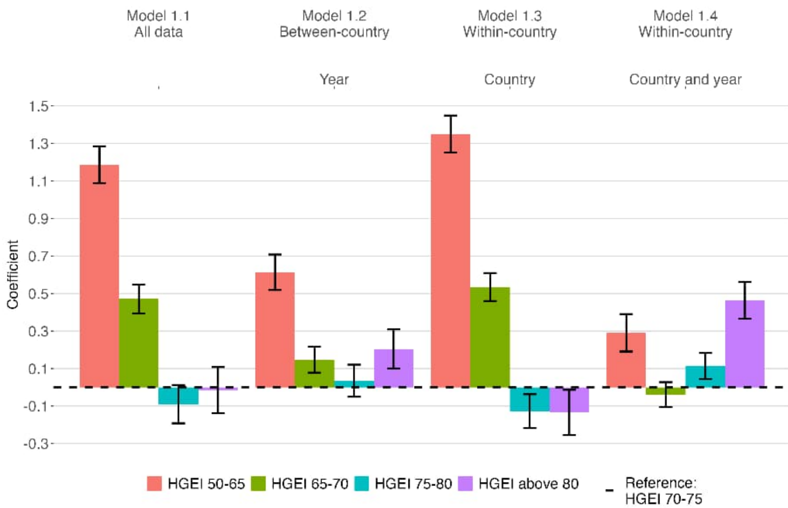
Finally, in Model 1.4, I include both year and country fixed effects:

$$TFR_{c,t} = \alpha + \sum_{g} \beta_g \, D^{g}_{c,t} + \mu_c + \delta_t + \varepsilon_{c,t}$$

The coefficients β_g in Models 1.1 to 1.4 quantify how deviations of a country's HGEI from the reference level are associated with deviations in fertility, and are the key coefficients of interest in my analysis. The models' specification is also documented in Table A-3 in the Appendix.

Figure 3 shows the coefficients β_g in Models 1.1 to 1.4. In Model 1.1, there is no U-shaped pattern between the HGEI and TFR across countries using all the samples. A slight U-shaped pattern emerges once period fertility trends are controlled for in Model 1.2, although this reversal is not statistically significant at the 5% level. Similar to Kolk's (2019) original results, Model 1.3 shows that no U-shaped pattern is initially observed when all the time-invariant country characteristics are controlled for in the model. However, after the period trends are controlled for in Model 1.4, the within-country U-shaped pattern becomes sharp and significant. The results of these models provide conditional support for the theory that there is a U-shaped relationship between gender equality and fertility once the overall declining fertility trend is adjusted for.

Figure 3: OLS regression between HGEI and TFR between and within countries



4.2 Adjusting period TFR's tempo distortion

Besides the gender equality measure, the fertility indicator is another issue in the analysis. My analysis follows Kolk's (2019) original study and uses the period TFR as the main fertility indicator, which includes a tempo distortion. Many researchers have argued that the reversal of the period TFR in the 1990s and 2000s in European countries may be the result of tempo distortion embedded in the period TFR indicator (Goldstein, Sobotka and Jasilioniene 2009; Philipov and Kohler 2001). The actual period quantum of fertility did not increase during that time. The U-shaped pattern observed during this period may be the result of tempo distortion rather than an actual reversal in the period quantum of fertility (Ryder 1956, 1959, 1964, 1983). Therefore, it is essential to test whether the conditional curvilinear association between gender equality and fertility rate identified in my replication is driven by the tempo distortion embedded in the period TFR indicator.

In this part of the robustness check I apply the tempo-adjusted TFR, developed by Bongaarts and Feeney (1998), to Kolk's (2019) original study and my extension to assess the extent to which the results remain robust after accounting for tempo distortion in the fertility rate. I accessed the tempo-adjusted TFR from the Human Fertility Database. The model specification is documented in Table A-3 in the Appendix. As shown in Figure 4 and Table A-1 in the Appendix, the tempo-adjusted TFR is not available for some of the countries included in my replication. These countries were therefore removed from the sample, decreasing the sample size from 32 countries to 24. Furthermore, since the tempo-adjusted TFR is only available for relatively recent years in most countries, the years of observation for each country also decreased significantly. The total cumulative years of data decreased from 1,280 to 649. Table A-4 in the Appendix reports the exact years for which tempo-adjusted TFR data are available for each country.

Figure 4: Number of country-years included in each analysis, by country

Figure 5 shows the distribution of the HGEI included in the analysis after the exclusion of samples. According to Figure 5, the mean HGEI is 73.30, and only a small percentage of samples have an HGEI between 50 and 65. This is because the tempo-adjusted TFR is only available for recent years, during which gender equality has remained at a high level.

Figure 5: Descriptive statistics for HGEI in the robustness check

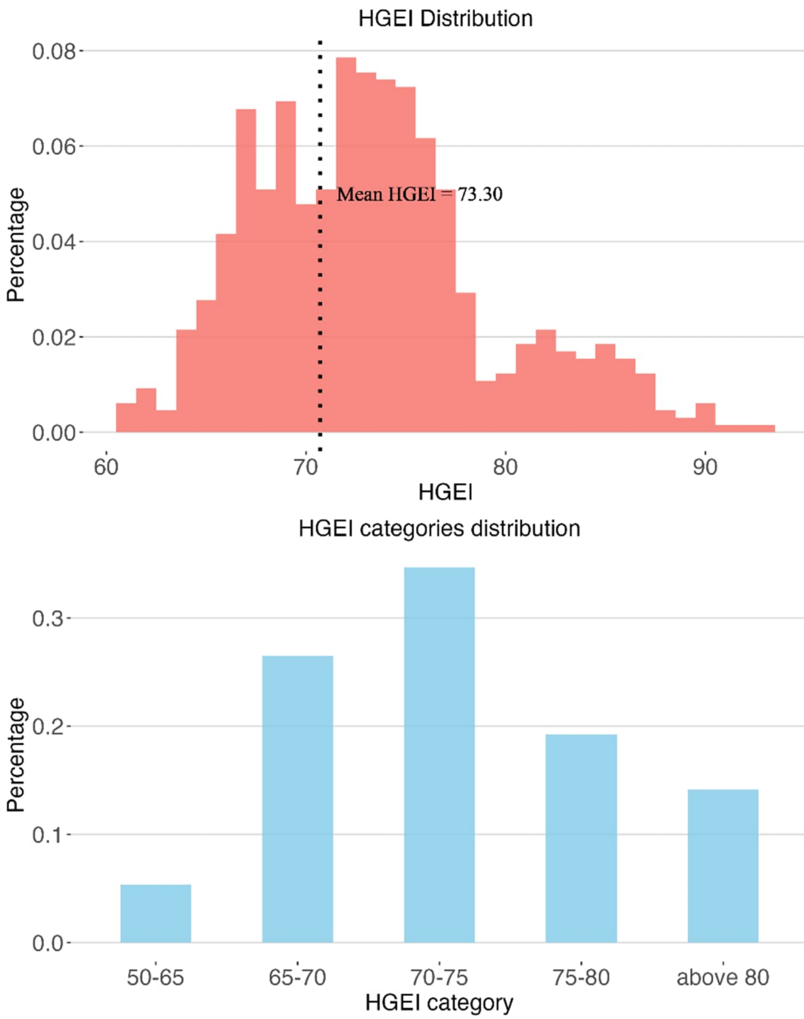
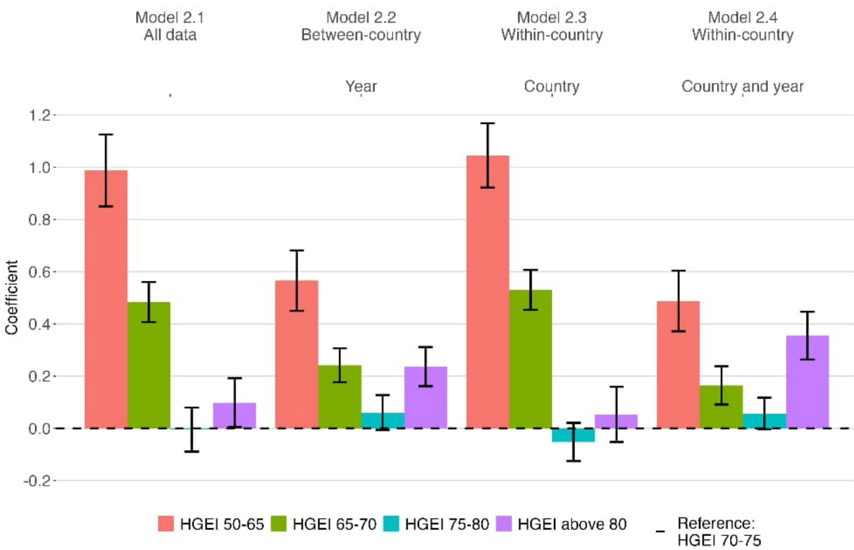


Figure 6 shows the relationship between the HGEI and the tempo-adjusted TFR. According to Models 2.1 to 2.4, after adjusting for tempo distortion, the results of each model remain relatively stable. One small change is that in Model 2.2, an HGEI between 75 and 80 is significantly correlated with a higher tempo-adjusted TFR at the 5% level. In Model 2.4, countries with an HGEI above 80 exhibit a strong and statistically significant correlation at the 5% level with the tempo-adjusted TFR, controlling for year and country fixed effects, providing conditional support for the U-shaped relationship theory. Overall, the U-shaped relationship between gender equality and fertility remains robust after adjusting for the tempo distortion in period fertility.

Figure 6: OLS regression between HGEI and tempo-adjusted TFR between and within countries



5. Conclusion and discussion

The main finding of this study is that in contrast to Kolk’s (2019) conclusion of weak support, conditional support for a U-shaped relationship between gender equality and fertility in developed countries emerges when a more comprehensive gender equality indicator is employed in the analysis. Although Kolk’s (2019) study demonstrates a high level of verifiability and low inaccuracy (as shown in the Appendix), his results appear to be sensitive to the choice of gender-equality indicator.

Kolk's (2019) analysis focuses on public-sphere gender equality, particularly women's political empowerment. By contrast, incorporating a multidimensional indicator reveals a sharp and significant curvilinear relationship between gender equality and fertility once year and country fixed effects are included. The U-shaped relationship between gender equality and fertility is not, as Kolk (2019) claims, simply the result of a cross-sectional pattern where countries with high gender equality are more likely to have high fertility levels. My findings align with most gender–fertility reversal theories in social demography regarding the U-shaped relationship between gender equality and fertility, but highlight that this pattern is conditional on the secular downward trend of fertility being controlled for. This rebound is insufficient to offset the overall long-term decline in fertility. This result remains robust after adjusting for tempo distortion in period fertility.

The empirical finding that improvements in gender equality are associated with a conditional reversal in fertility, but are not associated with an absolute fertility reversal, underscores the need to integrate gender–fertility reversal theories within a broader theoretical framework that addresses the universal decline in fertility. While gender-equality reversal mechanisms may generate a modest U-shaped increase in fertility at very high levels of gender equality, the overall decline in fertility, described by frameworks such as the demographic transition theory (Kirk 1996) and the second demographic transition theory (Lesthaeghe 2014), is substantially stronger than any reversal effect attributable to gender equality alone. This aligns with recent findings from the Nordic countries, which show that in highly gender-equal contexts, fertility intentions are low among individuals who hold strongly egalitarian attitudes. These attitudes, characterised by the prioritisation of non-familial life goals such as career advancement and personal fulfilment, as predicted by second demographic transition theory, diminish the potential fertility-reversing effects of gender equality improvement (Begall and Hiekel 2025).

This study does not establish a causal effect of gender equality on fertility. Although the country and year fixed effects specification accounts for all time-invariant heterogeneity and a shared linear fertility trend, it cannot adjust for country-specific time-varying characteristics, such as development (Myrskylä, Kohler, and Billari 2009), family policy (Zhang 2017), or labour-market regulations (Bennett 2021), which may simultaneously influence both gender equality and childbearing. More robust empirical evidence based on quasi-experiments is required to identify the actual causal relationship between gender equality and fertility.

This replication and extension have certain limitations. Although the HGEI aims to capture various aspects of gender equality both within and outside the household, it still lacks sufficient information on private-sphere gender equality, an element that most gender–fertility theories emphasise as central to fertility reversal. The HGEI indicator attempts to use the gender gap in age at marriage to capture gender equality in the private sphere. Although various studies indicate that the gender gap in age at marriage is a strong predictor of the unequal position of spouses and various domestic violence outcomes (Caldwell, Reddy, and Caldwell 1983; Kishor and Johnson 2005), it is still not a direct measure of gender equality in the private sphere. Better indicators, such as time-use data regarding unpaid housework, may provide a more accurate measure of private-sphere gender equality; however, there is insufficient longitudinal time-use data covering decades for such a large number of countries. Another potential option is to use attitudes toward gender equality inside the household, as captured by the World Values Survey, as a measure of private-sphere gender equality. However, there is still insufficient longitudinal data for this approach: it would only allow for examining the relationship between gender equality and fertility at a few time points, as in Arpino, Esping-Andersen, and Pessin's (2015) study. This makes it difficult to determine whether the U-shaped pattern observed at these time points reflects an actual longitudinal trend or is simply the result of fluctuations in period TFR. While the HGEI does not directly measure private-sphere gender equality, it currently provides the most suitable available indicator for longitudinal cross-national analysis.

Additionally, as the HGEI does not provide gender equality data beyond 2003, recent developments in gender equality and fertility relationships cannot be analysed here. Male involvement in childcare, such as the male share of parental leave, is a critical aspect emphasised by gender-equality reversal theories (Goldscheider, Bernhardt, and Lappegård 2015), which remained limited even in highly egalitarian countries around 2003 (Duvander and Cedstrand 2022). Detecting a conditional U-shaped relationship during this period thus suggests that a stronger fertility rebound might emerge if more recent data were available. As Kolk's (2019) original study suggests, the curvilinear trend is particularly sharp in the post-2000 period, and it is reasonable to believe that the U-shaped pattern still exists, or has perhaps become even sharper in the last two decades. Future research could benefit from re-evaluating this relationship when harmonised longitudinal indicators covering more recent periods become available.

I also undertook some robustness checks based on the main analysis. In the regression analysis, the gender equality indicator could be treated as a continuous variable instead of categorical variables. By including a squared transformation of the gender equality indicator in the regression, it would be possible to test whether there is a significant curvilinear relationship between gender equality and fertility. This approach would eliminate concerns about potential issues arising from the choice of a gender equality threshold. The robustness check results (Table A-5, Appendix) show that, after accounting for country and year fixed effects, HGEI is negatively associated with TFR (β = –0.454; 95% CI [–0.710, –0.199]), while the squared term of HGEI is positively associated with TFR (β = 0.00313; 95% CI [0.00138, 0.00487]), indicating a U-shaped relationship. The estimated turning point of the U-shape occurs when HGEI is around 72.7. These findings provide evidence that the conditional U-shaped relationship identified in the main analysis is not sensitive to the choice of HGEI thresholds.

In addition, future research could further investigate the association between gender equality and parity-specific fertility rates. For example, it would be valuable to investigate whether the reversal in fertility associated with increased gender equality is driven by an increase in the number of first births, second births, or perhaps both. By breaking down the fertility rate into parity-specific fertility rates, it would be possible to identify the specific subgroups affected by gender equality. This could provide a deeper understanding of the underlying mechanisms behind the relationship between gender equality and fertility.

Overall, the weak support for the U-shaped relationship identified in Kolk’s (2019) research may stem from the choice of gender equality measures. By employing a more comprehensive indicator of gender equality, this analysis reveals a conditional U-shaped relationship between gender equality and fertility once year and country fixed effects are applied. Although gender-equality reversal mechanisms may contribute to a modest increase in fertility at very high levels of gender equality, this effect appears to be outweighed by broader, universal forces that continue to depress fertility across contexts.

6. Acknowledgements

Haohao Lei acknowledges funding from the Clarendon Fund and Nuffield College, University of Oxford, and from the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS), Max Planck Institute for Demographic Research. Support from the Leverhulme Trust (Grant RC-2018-003) through the Leverhulme Centre for Demographic Science is also acknowledged. Haohao is grateful to Dr Dirk Witteveen for his supervision of the MPhil replication project and to Dr Andrea Tilstra for her valuable guidance. The author declares no competing interests. All data used are publicly available. Replication and extension code in R is available at: https://github.com/Haohao-Lei/Replication-and-Extension-of-Kolk-2019-Demographic-Research.

References

Anderson, T. and Kohler, H.-P. (2015). Low fertility, socioeconomic development, and gender equity. *Population and Development Review* 41(3): 381–407. doi:10.1111/j.1728-4457.2015.00065.x.

Arpino, B., Esping-Andersen, G., and Pessin, L. (2015). How do changes in gender role attitudes towards female employment influence fertility? A macro-level analysis. *European Sociological Review* 31(3): 370–382. doi:10.1093/esr/jcv002.

Begall, K. and Hiekel, N. (2025). Examining the gender equality–fertility paradox in three Nordic countries. *Population and Development Review* 51(2): 858–888. doi:10.1111/padr.12721.

Bennett, F. (2021). *Gendered economic inequalities: A social policy perspective*. London: Institute for Fiscal Studies (IFS Deaton Review of Inequalities).

Bongaarts, J. and Feeney, G. (1998). On the quantum and tempo of fertility. *Population and Development Review* 24(2): 271–291. doi:10.31899/pgy6.1010.

Caldwell, J.C., Reddy, P.H., and Caldwell, P. (1983). The causes of marriage change in South India. *Population Studies* 37(3): 343–361. doi:10.1080/00324728.1983.10408866.

Coppedge, M., Gerring, J., Knutsen, C.H., Lindberg, S.I., Teorell, J., Altman, D., Bernhard, M., Borella, C., Cornell, A., Fish, M.S., Fox, L., Gastaldi, L., Gjerløw, H., Glynn, A., God, A.G., Grahn, S., Hicken, A., Kinzelbach, K., Krusell, J., Marquardt, K.L., McMann, K., Mechkova, V., Medzihorsky, J., Natsika, N., Neundorf, A., Paxton, P., Pemstein, D., Pernes, J., Rydén, O., von Römer, J., Seim, B., Sigman, R., Skaaning, S.-E., Staton, J., Sundström, A., Tzelgov, E., Wang, Y.-t., Wig, T., Wilson, S., and Ziblatt, D. (2024). V-Dem dataset v14. *Varieties of Democracy (V-Dem) Project*.

Dilli, S., Carmichael, S.G., and Rijpma, A. (2019). Introducing the historical gender equality index. *Feminist Economics* 25(1): 31–57. doi:10.1080/13545701.2018.1442582.

Duvander, A.-Z. and **C**edstrand, S. (2022). Gender equal parental leave use in Sweden: The success of the reserved months. In: de la Porte, C., Eydal, G.B., Kauko, J., Nohrstedt, D., 't Hart, P., and Tranøy, B.S. (eds), *Successful public policy in the Nordic Countries: Cases, lessons, challenges*. Oxford: Oxford Academic: 263–283. doi:10.1093/oso/9780192856296.003.0013.

Esping-Andersen, G. and Billari, F.C. (2015). Re-theorising family demographics. *Population and Development Review* 41(1): 1–31. doi:10.1111/j.1728-4457.2015.00024.x.

Folbre, N. (1983). Of patriarchy born: The political economy of fertility decisions. *Feminist Studies* 9(2): 261–284. doi:10.2307/3177490.

Fraser, N. (1994). After the family wage: Gender equity and the welfare state. *Political Theory* 22(4): 591–618. doi:10.1177/0090591794022004003.

Goldscheider, F., Bernhardt, E., and Lappegård, T. (2015). The gender revolution: A framework for understanding changing family and demographic behaviour. *Population and Development Review* 41(2): 207–239. doi:10.1111/j.1728-4457.2015.00045.x.

Goldstein, J.R., Sobotka, T., and Jasilioniene, A. (2009). The end of "lowest-low" fertility? *Population and Development Review* 35(4): 663–699. doi:10.1111/j.1728-4457.2009.00304.x.

Kishor, S. and Johnson, K. (2005). Profiling domestic violence: A multi-country study. *Studies in Family Planning* 36(3): 259–261. doi:10.1111/j.1728-4465.2005.00068.x.

Kirk, D. (1996). Demographic transition theory. *Population Studies* 50(3): 361–387. doi:10.1080/0032472031000149536.

Kolk, M. (2019). Weak support for a U-shaped pattern between societal gender equality and fertility when comparing societies across time. *Demographic Research* 40(2): 27–48. doi:10.4054/DemRes.2019.40.2.

Lesthaeghe, R. (2014). The second demographic transition: A concise overview of its development. *Proceedings of the National Academy of Sciences* 111(51): 18112–18115. doi:10.1073/pnas.1420441111.

McDonald, P. (2000a). Gender equity in theories of fertility transition. *Population and Development Review* 26(3): 427–439. doi:10.1111/j.1728-4457.2000.00427.x.

McDonald, P. (2000b). Gender equity, social institutions and the future of fertility. *Journal of the Australian Population Association* 17(1): 1–16. doi:10.1007/BF03029445.

McDonald, P. (2013). Societal foundations for explaining low fertility: Gender equity. *Demographic Research* 28(34): 981–994. doi:10.4054/DemRes.2013.28.34.

Myrskylä, M., Kohler, H.-P., and Billari, F.C. (2009). Advances in development reverse fertility declines. *Nature* 460(7256): 741–743. doi:10.1038/nature08230.

Philipov, D. and Kohler, H.-P. (2001). Tempo effects in the fertility decline in Eastern Europe: Evidence from Bulgaria, the Czech Republic, Hungary, Poland, and Russia. *European Journal of Population/Revue européenne de démographie* 17(1): 37–60. doi:10.1023/A:1010740213453.

Ryder, N.B. (1956). Problems of trend determination during a transition in fertility. *The Milbank Memorial Fund Quarterly* 34(1): 5–21. doi:10.2307/3348329.

Ryder, N.B. (1959). *An appraisal of fertility trends in the United States*. Indianapolis: Bobbs-Merrill, College Division.

Ryder, N.B. (1964). The process of demographic translation. *Demography* 1(1): 74–82. doi:10.1007/BF03208446.

Ryder, N.B. (1983). Cohort and period measures of changing fertility. In: Bulatao, R.A. and Lee, R.D. (eds.). *Determinants of fertility in developing countries*, Vol. 2. New York: Academic Press: 737–756.

Trivers, R.L. (2017). Parental investment and sexual selection. In: Campbell, B.G. (ed.). *Sexual selection and the descent of man*. London: Routledge: 136–179. doi:10.4324/9781315129266-7.

United Nations (1995). Programme of action of the 1994 International Conference on Population and Development (Chapters I–VIII). *Population and Development Review* 21(1): 187–213. doi:10.2307/2137429.

United Nations Development Programme (2010). *Human Development Report 2010: The Real Wealth of Nations – Pathways to Human Development* [electronic resource]. New York: United Nations Development Programme. http://hdr.undp.org/en/reports/global/hdr2010.

Zhang, J. (2017). The evolution of China's one-child policy and its effects on family outcomes. *Journal of Economic Perspectives* 31(1): 141–160. doi:10.1257/jep.31.1.141.

Appendix

Replication of Kolk's (2019) study

Kolk's (2019) original analysis was based on Stata software; in this study, I try to replicate Kolk's (2019) finding in R version 4.1.3. Specifically, I replicate Figures 1, 2, and 3 in Kolk's original paper. My visualisation is based on the ggplot2 package, version 3.4.0. The result of my replication is shown in the Appendix Table A-1 and Figures A-1 to A-3.

For the replication, I accessed the same data source as Kolk (2019). For the WPEI, the latest version is version 12, and data from different versions is not comparable. Because the version of data used in the original research was not clearly stated, I tried several versions of the dataset and contacted the V-Dem Institute to gain access to the version 7.1 data used in the original study. However, I found it difficult to replicate the specifications of Kolk's (2019) sample size exactly. Instead of including data from 1,993 cumulative years, my replication includes 2,054 cumulative years, 61 more years of data than in the original sample, because after Kolk (2019) published his original paper, the latest TFR data became available. Kolk did not specify the exact period he used for each country. I have provided summative statistics regarding the differences between my replication and Kolk's (2019) original paper in the Appendix Figure A-4. The difference between most countries captured by the newest data compared with the data Kolk (2019) used is just one to two years of data. As I discuss later, additional years would not have any great impact on Kolk's (2019) original conclusion except for a small coefficient fluctuation in the regression.

For countries with vague definitions or oversea territories, I followed Kolk's classification methods: the United Kingdom refers to England and Wales, France refers to France excluding overseas territories, and Germany refers to Western Germany until reunification (1990), then all of Germany. Finally, in my replication I removed the samples missing the WPEI or the TFR. Most of the data excluded due to missing values are prior to 1960. For most countries, a continuous longitudinal observation of the WPEI and the TFR is available since 1960.

My replication is almost identical to Kolk's original study. The overall negative association and U-shaped relationship between the TFR and the WPEI after 2000 is noticeable in both visualisations (shown in the Appendix). However, there are also some small differences between my replication and Kolk's (2019) results. For example, in Spain, the TFR and the WPEI are 2.86 and 0.348 in 1937, 2.55 and 0.333 in 1938, and 2.12 and 0.335 in 1939, respectively. These data points are not shown in Kolk's (2019) original visualisation but exist in mine, as I highlight in the Appendix. There are also some missing data points around the lowest WPEI level, as in Portugal before 1960 in

Kolk's (2019) visualisation. Although these data points are missing in the original paper, adding them does not lead to any dramatic change in the conclusion.

Regarding the country-specific trends over time, consistent with Kolk's (2019) findings, my replication indicates that in most countries, no U-shaped relationship exists between the TFR and the WPEI. A minor reversal pattern only occurs in 4 of the 35 countries: Belgium, Denmark, France, and the Netherlands. In Figure A-2, the previous missing data points in Figure B.2, such as the data points for Spain from 1937 to 1939, are present in Kolk's (2019) visualisation. Their presence makes the missing data points in Figure A-1 more puzzling.

Finally, as I discussed above, my replication of the regression model confirms Kolk's (2019) original argument, which suggested only weak support for a U-shaped pattern between gender equality and fertility in a society over time. Fertility and the WPEI exhibit a negative between-country correlation; however, a U-shaped pattern emerges after controlling for period trends. Focusing on only the post-1990 period and controlling for period trends make the U-shaped pattern between the TFR and the WPEI appear particularly significant. However, controlling for the time-invariant country characteristics renders the relationship between the TFR and the WPEI negative and provides no evidence in support of the U-shaped relationship. In general, Kolk's (2019) study had a high level of verifiability with minimal inaccuracy.

Table A-1: Country year difference

Country	Kolk's original paper	My replication	Extension with HGEI	Extension with tempo-adjusted TFR
Australia	56	57	44	0
Austria	64	66	53	19
Belarus	24	26	0	0
Belgium	56	57	44	0
Bulgaria	63	63	54	54
Canada	91	96	54	54
Chile	14	14	12	11
Czech Republic	65	67	11	11
Denmark	56	57	44	35
Estonia	24	26	12	12
Finland	77	78	54	21
France	70	71	54	0
Germany	58	61	48	0
Hungary	65	67	54	51
Iceland	53	57	0	0
Ireland	56	57	44	0
Italy	61	63	50	0
Japan	68	70	54	35
Lithuania	25	27	13	13
Netherlands	65	67	54	53
New Zealand	56	57	44	0
Norway	48	50	37	36
Poland	44	46	33	32
Portugal	75	76	53	43
Russia	56	58	9	9
Slovakia	17	22	11	11
Slovenia	25	27	13	13
South Korea	56	57	44	3
Spain	83	85	54	28
Sweden	117	117	54	33
Switzerland	83	85	54	5
Taiwan	39	41	0	0
Ukraine	23	23	13	13
United Kingdom	77	79	54	4
United States	83	84	54	54

Table A-2: Results of the regression analysis

Model	HGEI Category	Estimate	Std. Error	t value	95% CI
Model 1.1	50–65	1.18	0.05	23.55	[1.08, 1.28]
	65–70	0.47	0.03	12.02	[0.40, 0.55]
	70–75 (ref)	–	–	–	–
	75–80	–0.09	0.05	–1.72	[–0.19, 0.01]
	≥ 80	–0.02	0.06	–0.24	[–0.14, 0.11]
Model 1.2	50–65	0.61	0.04	12.67	[0.52, 0.71]
	65–70	0.15	0.03	4.17	[0.08, 0.22]
	70–75 (ref)	–	–	–	–
	75–80	0.04	0.04	0.80	[–0.05, 0.12]
	≥ 80	0.20	0.05	3.85	[0.10, 0.31]
Model 1.3	50–65	1.35	0.05	26.85	[1.25, 1.45]
	65–70	0.53	0.03	14.06	[0.46, 0.61]
	70–75 (ref)	-	-	-	-
	75–80	–0.13	0.04	–2.76	[–0.22, –0.04]
	≥ 80	–0.13	0.06	–2.18	[–0.26, –0.01]
Model 1.4	50–65	0.29	0.05	5.71	[0.19, 0.39]
	65–70	–0.03	0.03	–1.15	[–0.11, 0.03]
	70–75 (ref)	–	–	–	–
	75–80	0.11	0.03	3.20	[0.04, 0.18]
	≥ 80	0.46	0.05	9.25	[0.36, 0.56]
Model 2.1	50–65	0.99	0.07	14.05	[0.85, 1.12]
	65–70	0.48	0.04	12.35	[0.41, 0.56]
	70–75 (ref)	–	–	–	–
	75–80	-0.00	0.04	–0.11	[–0.09, 0.08]
	≥ 80	0.10	0.05	2.05	[0.00, 0.19]
Model 2.2	50–65	0.59	0.06	10.09	[0.48, 0.70]
	65–70	0.29	0.03	8.90	[0.22, 0.35]
	70–75 (ref)	–	–	–	–
	75–80	0.13	0.03	3.81	[0.06, 0.20]
	≥ 80	0.25	0.04	6.91	[0.18, 0.32]
Model 2.3	50–65	1.04	0.06	16.67	[0.92, 1.17]
	65–70	0.53	0.04	13.61	[0.45, 0.61]
	70–75 (ref)	–	–	–	–
	75–80	–0.05	0.04	–1.39	[–0.13, 0.02]
	≥ 80	0.05	0.05	1.00	[–0.05, 0.16]
Model 2.4	50–65	0.55	0.06	9.53	[0.44, 0.66]
	65–70	0.23	0.04	6.37	[0.16, 0.30]
	70–75 (ref)	–	–	–	–
	75–80	0.10	0.03	3.39	[0.04, 0.16]
	≥ 80	0.26	0.04	5.83	[0.17, 0.34]

Note: All models include country and year fixed effects; standard errors are clustered by country. The 95% Wald confidence interval is computed as $\beta \pm 1.96 \times SE$.

Figure A-1: Association between TFR and female political empowerment in 35 countries, by period, for all periods (3.1 a) and post-2000 (3.1 b); my replication for all periods (3.2 a) and post-2000 (3.2 b)

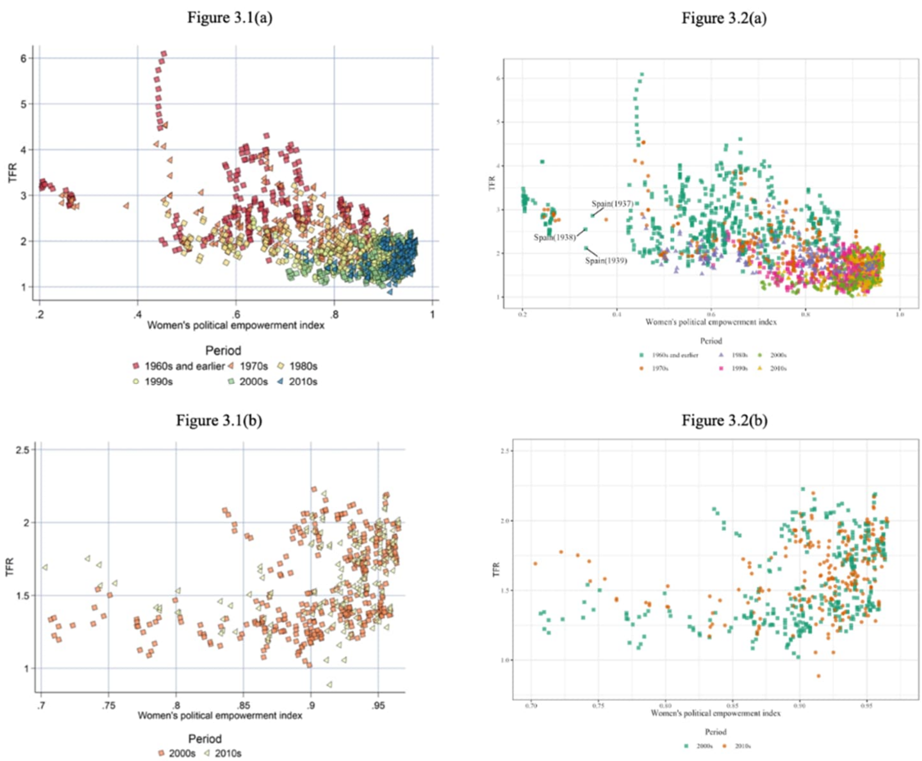


Figure A-2: Association between TFR and female political empowerment in 35 countries, by country and period (only TFR below 4 shown); Kolk's result (4.1) vs. my replication (4.2)

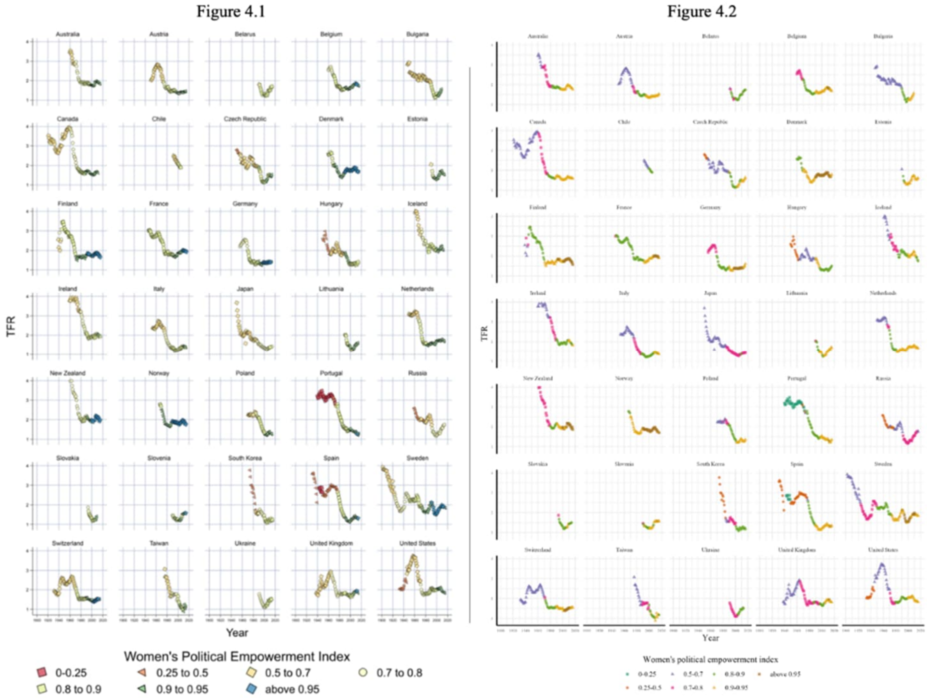


Figure A-3: OLS Regression between TFR and WPEI (female empowerment index) in six models, 35 countries and 1,993 cumulative years; Kolk's result (5.1) vs. my replication (5.2)

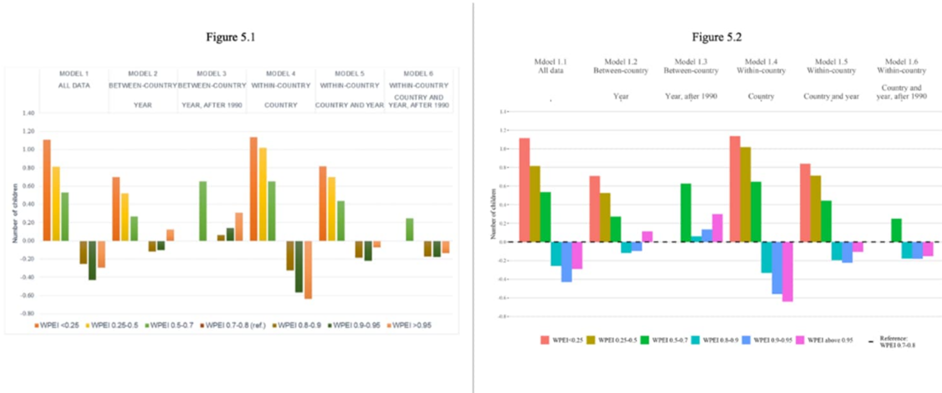


Figure A-4: Country-year comparison between Kolk (2019) and my replication

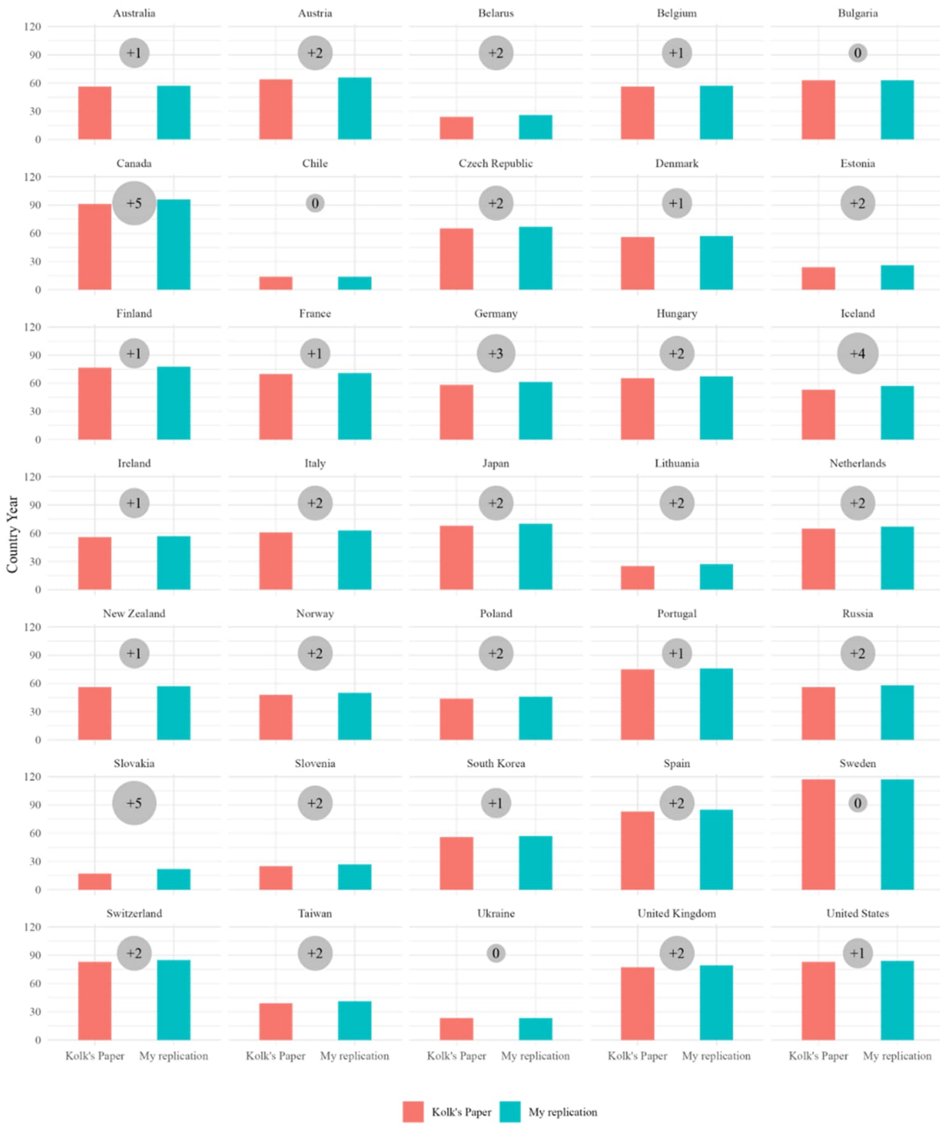


Table A-3: OLS model specifications: TFR and tempo-adjusted TFR regressions

Model	Outcome	Predictors	Year term?	Country FE?	Interpretation
1.1	Period TFR	HGEI category (ref = 70–75)	–	–	Between-country association of TFR with HGEI levels
1.2	Period TFR	HGEI category + year	✓	–	Between-country, controls for common period trend
1.3	Period TFR	HGEI category	–	✓	Within-country association (absorbs time-invariant factors)
1.4	Period TFR	HGEI category + year	✓	✓	Within-country net of period trend
2.1	Tempo-adj. TFR	HGEI category (ref = 70–75)	–	–	Between-country association using tempo-adjusted fertility
2.2	Tempo-adj. TFR	HGEI category + year	✓	–	Between-country, controls for tempo-adjusted period trend
2.3	Tempo-adj. TFR	HGEI category	–	✓	Within-country, absorbs all time-invariant country traits
2.4	Tempo-adj. TFR	HGEI category + year	✓	✓	Within-country net of both period trend and country FE

Table A-4: Coverage window for TFR and tempo-adjusted TFR

Country	First Year (TFR)	Last Year (TFR)	N Years (TFR)	First Year (adjTFR)	Last Year (adjTFR)	N Years (adjTFR)
Australia	1960	2003	44	-	-	-
Austria	1951	2003	53	1985	2003	19
Belgium	1960	2003	44	-	-	-
Bulgaria	1950	2003	54	1950	2003	54
Canada	1950	2003	54	1950	2003	54
Chile	1992	2003	12	1993	2003	11
Czech Republic	1993	2003	11	1993	2003	11
Denmark	1960	2003	44	1969	2003	35
Estonia	1992	2003	12	1992	2003	12
Finland	1950	2003	54	1983	2003	21
France	1950	2003	54	-	-	-
Germany	1956	2003	48	-	-	-
Hungary	1950	2003	54	1953	2003	51
Ireland	1960	2003	44	-	-	-
Italy	1954	2003	50	-	-	-
Japan	1950	2003	54	1969	2003	35
Lithuania	1991	2003	13	1991	2003	13
Netherlands	1950	2003	54	1951	2003	53
New Zealand	1960	2003	44	-	-	-
Norway	1967	2003	37	1968	2003	36
Poland	1971	2003	33	1972	2003	32
Portugal	1950	2003	53	1960	2003	43
Russia	1992	2000	9	1992	2000	9
Slovakia	1993	2003	11	1993	2003	11
Slovenia	1991	2003	13	1991	2003	13
South Korea	1960	2003	44	2001	2003	3
Spain	1950	2003	54	1976	2003	28
Sweden	1950	2003	54	1971	2003	33
Switzerland	1950	2003	54	1999	2003	5
Ukraine	1991	2003	13	1991	2003	13
UK	1950	2003	54	-	-	-
United States	1950	2003	54	1950	2003	54

Table A-5: Regression results (robustness check)

$$TFR_{c,t} = \beta_1 HGEI_{c,t} + \beta 2\, \mathrm{HGEI}_{c,t}^2 + \mu_c + \delta_t + \varepsilon_{c,t}$$

	Estimate	Std. Error	t value	95% CI
Intercept	19.4766	4.7588	4.0927	[10.1513, 28.8019]
HGEI	-0.4544	0.1305	-3.4824	[-0.7101, -0.1987]
$HGEI^2$	0.0031	0.0009	3.5045	[0.0014, 0.0049]