THE IMPACT OF INCOME INEQUALITY ON MORTALITY. A REPLICATION STUDY OF LEIGH AND JENCKS (JOURNAL OF HEALTH ECONOMICS, 2007)

Weilun Wu

Abstract

This study replicates Leigh and Jencks' (2007) analysis of the relationship between income inequality and mortality. Using L&J's preferred specification, I am able to closely reproduce their original findings after reconstructing their data from original sources. When I use multiple imputation instead of their method of linear interpolation, I largely confirm their results. When I extend their data from 2003 to 2018, I again do not find a significant relationship between income inequality and mortality. As a result, I conclude that my replication exercise confirms L&J's results, providing even stronger evidence for the view that income inequality is not adversely related to mortality.

JEL Codes: I12; N30

Keywords: Income Inequality, Health, Mortality, Multiple Imputation

February 6, 2022

1. Introduction

The relationship between income inequality and health has been widely studied. One claim that has attracted particular attention is that greater income equality is associated with better health outcomes in developed economies (Wilkinson, 1998; Wilkinson, 2002; Wilkinson & Pickett, 2006). This is commonly referred to as "Wilkinson's income inequality hypothesis", or just "the income inequality hypothesis". Despite numerous analyses, and numerous systematic reviews, the income inequality hypothesis remains contested (Rodgers, 1979; Fiscella & Franks, 2000; Gravelle et al., 2002; Lynch et al., 2004; Macinko et al., 2003; Subramanian & Kawachi; 2004; Macinko et al., 2004; Kondo et al., 2009; Patel et al., 2018; Kim, 2019). One of the most prominent studies disputing this hypothesis is Leigh & Jencks (2007). L&J has been cited 56 times in the Web of Science and 160 times in Google Scholar.

Using long historical data from 12 OECD countries, L&J estimate the relationship between income inequality, measured by the share of income going to the top 10% of income earners, and (i) life expectancy and (ii) the log of infant mortality. Their key finding is that the estimated effects of income inequality largely disappear when one controls for country and year fixed effects. Not only are the estimated coefficients statistically insignificant, but they are economically negligible, even when evaluated at the endpoints of the associated 95% confidence intervals.

As one might expect when using cross-country data over long time periods, missing data are an issue. L&J's key findings, reported in Columns (7) and (8) of their Table 4, are based on a relatively small subset of complete observations. For example, while the full data allow for a possible 528 country-year observations, the education variable they use as a control variable only has 108 records because the corresponding data are collected at 5-year intervals.

To address the problem of missing data, L&J use linear interpolation to fill in missing values. Unfortunately, interpolation can cause bias in both estimated coefficients and their standard errors (Little, 1992; Allison, 2001; Musil et al, 2002; Enders, 2010). This raises concerns about both the coefficient estimates and the confidence intervals reported in their paper.

Because of its importance in the literature, and the concern about missing data, I have decided to replicate L&J. In the first instance, I am interested in determining whether using proper procedures to address missing values in L&J's dataset might affect their results. I also want to investigate whether their findings hold up when their data are updated to the most recent available. The result from this replication exercise is that I confirm L&J's conclusions, providing even stronger evidence for the view that income inequality is not adversely related to mortality.

2. Replication of the data

The data and code that Leigh and Jencks' (2007) used for their estimation are posted at *Harvard's Dataverse*. Their longitudinal dataset consists of annual observations from 1903 to 2003 for 12 countries. The original, underlying data sources are all publicly accessible. Unfortunately, the only version of the inequality data that L&J provide is post-interpolation. As a result, in replicating their work, I need to go back and reconstruct their dataset from the original sources.

L&J focus on two outcome variables, life expectancy and infant mortality. Both of these are primarily sourced from the Human Mortality Database (HMD).[1] HMD has

[1] According to Leigh and Jencks' (2007) appendix, the mortality data for some countries are differently sourced. Thanks to the authors' notification of these different data sources, I successfully replicated the mortality data for corresponding data sets. These countries are: Australia, Ireland, and the United States.

been updated since L&J collected their mortality data in 2002. My replication uses these updated values.

There are several issues involved in replicating the key independent variable, *"Top Share 10"*, which is the share of total income held by the richest 10% of earners. L&J cite Leigh (2007) as their source for *Top Share 10.* However, there are some differences in these series. One difference is that L&J's *Top Share 10* includes capital gains in its construction of income inequality, while Leigh (2007) does not include capital gains. An explanation for this and other discrepancies is provided in the online supplementary materials that accompany L&J (Leigh & Jencks, 2013).[2] My reconstruction uses Leigh (2007) because pre-interpolation values for *Top Share 10* are not available for L&J.

With respect to the control variables, GDP is sourced from *Maddison Database 2010* (Maddison, 2010). Health expenditure data are taken from *OECD Health Data 2007* (OECD, 2007). Educational attainment is acquired from Barro and Lee (2013). Despite the fact that my reconstructed dataset uses updated values, I am able to closely match L&J's post-interpolation results.

3. The issue of missing data

Missing data are a problem for Leigh and Jencks' (2007) dataset. It is difficult to know how severe a problem because L&J did not provide a copy of their data before they

---

[2] Regarding these discrepancies, Leigh & Jencks (2013) write: "Note that Leigh & Jencks (2007) cites Leigh (2006). That paper (drafted in February 2006) used top incomes series that were as up to date as possible at the time. Where possible, the series in Leigh & Jencks (2007) include capital gains. In a revise and resubmit in December 2006 (after Leigh & Jencks 2007 was in press), referees asked for the series in Leigh (2006) to be updated to exclude capital gains. In addition, several of the top incomes papers had been updated during 2006, so where this had occurred, Leigh also updated the series. The country most affected by this is Ireland, for which Nolan revised upwards his preferred estimates. If Leigh (2006) is published, readers should be aware that there will be some minor differences between the series there and the series used in Leigh & Jencks (2007). These differences are partly a product of compiling top incomes series from working papers, and partly due to the issue of whether to include or exclude capital gains."

applied interpolation. Accordingly, I calculate missingness from the dataset that I reconstructed using their original sources. The results are presented in TABLE 1. The rates of missingness vary widely across the variables.

The main independent variable of interest, *Top Share 10*, has about 14% missing values. *Average years of education of adults aged 15+*, which is used as a control variable in the analysis, has 80% missing values. The latter is due to the fact that the underlying data are reported at 5-year intervals.

The union of multiple variables with missing values increases the missingness rate even further, as all variables must have data to be included in the regression. For example, in the "Complete Case" specifications of TABLE 2, the regressions for life expectancy (LE) and the log of infant mortality (IM) have 64 and 69 observations, respectively. This represents a large loss of information compared to the full set of 528 observations. This explains why L&J turned to linear interpolation to increase the size of their dataset.

Linear interpolation assumes that variables are a function of time and uses the closest data points to "predict" the missing values. This approach has several shortcomings. It ignores the association between missing values and other variables. It only relies on time to "predict" missing values and ignores the predictive power of other variables. Further, interpolation artificially reduces variation in the data. Interpolated values all lie on a straight time line, ignoring the randomness that occurs in naturally occurring data.

In fact, the linear interpolation method that L&J employed did not fill in all of the missing values in their dataset. For example, it did not interpolate *Top Share 10* data when there was a gap in the data of more than four years. The post-interpolation datasets that they used to estimate the specifications in Columns 7 and 8 in their

Table 4 only used 430 of the 528 observations. Thus, they did not exploit all the information that was available in their dataset.

To address these shortcomings, this study uses multiple imputation (MI). MI is superior to linear interpolation for several reasons. Firstly, MI uses correlations with all other variables to impute missing values. Secondly, MI uses information from any observations that have non-missing data. Thirdly, MI produces estimates with attractive properties. If the data are “missing at random” (MAR) -- that is, the probability that a variable is missing is independent of that variable -- then MI will produce estimates that are (i) consistent, (ii) asymptotically efficient, and (iii) asymptotically normal (Rubin, 1987; Rubin, 1996; Paul, 2001; Enders, 2010; Pedersen et al., 2017; Little and Rubin, 2019).

MI entails two phases, an imputation phase and an analysis phase. In the imputation phase, MI iteratively produces multiple versions of a completed dataset without any missing values. Each version uses stochastically generated values to replace missing values. The analysis phase runs regressions on each of the individual datasets, resulting in multiple sets of estimates. The estimates are then combined into a single, final set of estimates via Rubin’s rule (Rubin, 1987).

The imputation process uses non-missing values of all the other variables plus the country and year dummy variables.[3] My exploratory analysis sets the “burnin” and “burnbetween” parameters sufficient to produce datasets that demonstrated “stationarity” and “independence” (Ender, 2010) (see Appendix B). To determine the appropriate number of imputations, I used von Hippel’s (2020)

[3] The preferred approach to imputing panel data is to convert the data into wide form before imputing. This allows one to predict missing values of a given variable in a given year with observed values of that variable in other years. This was not possible with my data because the number of years was greater than the number of countries. As a result, I imputed the data in long form and relied on the year and country fixed effects, along with other observed variables, to predict missing values.

"how_many_imputations", user-written Stata program. One of the disadvantages of MI is that it stochastically creates different datasets. Differing attempts to implement MI using the same data and code produce different results with each attempt. To address this issue, von Hippel's approach calculates the number of imputed datasets sufficient to cause the standard errors of the respective parameter estimates to vary less than 5%.

I proceed by re-estimating L&J's key regression using the dataset I reconstructed and applying MI to fill in values for missing data.

4. Replication of Leigh and Jencks' (2007) results

Leigh and Jencks (2007) were mindful that rapid advances in epidemiology contributed to dramatic increases in life expectancy in the period leading up to 1960. Recognizing that controlling for these advances was nigh impossible, their main analysis focused on the period 1960 to 2003.

L&J's key findings are reported in Columns (7) and (8) of their Table 4. The corresponding two-way fixed effects model is given by Eq. (1).

$$m_{ij} = \alpha + \beta(Top\ Share\ 10)_{ij} + \gamma Z_{ij} + \delta_i + \rho_j + \varepsilon_{ij} \quad (1)$$

where m_{ij} is an indicator of mortality, measured by either life expectancy or the log of infant mortality. The subscript i stands for country, and the subscript j for year. *Top Share 10* is the income share of the richest 10% of the population. This variable serves as the measure of income inequality. An implication of the income inequality hypothesis is that greater inequality is positively related to mortality. Z is a vector of covariates including GDP, educational attainment, and health expenditures. ε is an error term. All variables are continuous. Note that Eq. (1) includes both country and time fixed effects, indicated by δ_i and ρ_j, respectively.

I use Eq. (1) to re-estimate the relationship between inequality and health. I first attempt to reproduce Leigh and Jencks' (2007) results using my reconstruction of their dataset. In replicating their results, I rely on their programming code, which provided details about the linear interpolation approach they used. After replicating their results, I then employ MI to reexamine the relationship between inequality and health.

The results of my analysis are presented in TABLE 2. The regression estimates are presented in pairs, with the left-hand of the pair reporting the results for life expectancy (LE), and the right-hand of the pair reporting results for the log of infant mortality (IM). As L&J drew attention to the 95% confidence intervals implied by their estimated standard errors, TABLE 2 reports these below the estimated coefficients.

The first pair of results, designated as "Original", are copied directly from L&J's paper and are the results they report in Columns (7) and (8) of their Table 4. The next pair of results is identified as "Complete Case". It only includes observations for which there are no missing values. "Linear Interpolation" consists of the "Complete Case" observations plus additional observations for which missing values were replaced by their linearly interpolated values. The final pair of regression results ("Multiple Imputation") are based on the full sample of 528 observations where all missing values were replaced with simulated values following the MI procedure described above.

The fact that the "Linear Interpolation" results are similar to the "Original" indicates that my reproduction of L&J's results was largely successful. For example, the linear interpolation estimate of the effect of *Top Share 10* on life expectancy is 0.040 with a 95% confidence interval of [-0.091,0.170]. This compares with L&J's original estimates of 0.033 and [-0.100,0.166], respectively. In words, this specification implies that a 10-percentage point increase in *Top Share 10* is associated with an increase in life expectancy of between 0.3 and 0.4 years. Note that the sign is counter to what the income inequality hypothesis predicts. Further, the estimate is statistically insignificant at the 5 percent level. This is indicated by the fact that both confidence intervals include 0. On the basis of these and similar results, L&J conclude that income inequality is not associated with greater mortality.

As noted above, in the absence of any efforts to fill in missing values, the number of usable observations shrinks from 528 observations to 64 (LE) and 69 observations (IM) (cf. "Complete Case" regressions). This represents a substantial loss in information. It should be obvious that it is non-optimal to not utilize the information in these discarded observations, especially since many of the lost observations are the result of missing values in control variables, particularly *Average years of education of adults 15+*.

When I apply Leigh and Jencks' (2007) linear interpolation approach, I am able to increase the size of the dataset to 427 observations. Note that this is still less than the full size of the dataset, which is 528 observations. The reason L&J's method of linear interpolation did not impute all missing values was because there were instances where they determined that linear interpolation would lead to misleading values. One such instance occurred whenever there were more than four missing values in a "row". The effect of not using all available observations is that L&J do not exploit all the available information in the dataset. An additional problem with this approach is that linear interpolation treats the imputed values as "true" values and does not account for the sampling randomness in the real data. As a result, the linear imputed estimates of standard errors are expected to be biased.

The last pair of estimates reports results from my MI analysis. By employing MI, I am able to use all 528 observations. While there are differences, they are not dramatic. The size of the estimated effect of *Top Share 10* increased from 0.040 to 0.054 for life expectancy, but remained the same for infant mortality. Note again the signs of both coefficients run counter to the income inequality hypothesis.

The 95% confidence intervals have also become narrower under MI estimation. For example, the width of the 95% confidence interval for *Top Share 10* in the Linear Interpolation/LE regression is 0.091 + 0.170 = 0.261. The width of the corresponding 95% confidence interval in the Multiple Imputation/LE regression is 0.038 + 0.146 = 0.184. Turning to the other variables, the MI confidence intervals are generally narrower than the "Linear Interpolation" confidence intervals. While MI has allowed me to take full advantage of the information in the dataset and increase the number of observations, the final results are still consistent with L&J's conclusion. There does not appear to be a statistically and economically significant relationship between income inequality and mortality.

5. Reanalysis with the updated data

In this section, I extend L&J's analysis by using more recent data. A major concern in all studies involving income inequality is the reliability of the inequality data. In recent years, academic studies of income inequality have increasingly turned to the World Inequality Dataset (WID) as the authoritative source of world income inequality (Atkinson et al., 2018; Alstadsæter et al., 2019; Blanchet et al., 2019).

WID covers 115 countries and provides a variety of indicators describing income, wealth and inequality levels for the top and bottom groups. It has earned wide acceptance because it (i) integrates micro data sources with national accounts, ii) provides consistency in economic indicator measurements, and iii) strives to create comparability across countries and years (Piketty and Atkinson, 2010; Alvaredo et al., 2013; Alvaredo et al., 2016; Alvaredo et al., 2017). WID data are free to use and easily accessible.[4]

Accordingly, I proceed by replacing Leigh and Jencks' (2007) income inequality data with WID data. The "top 10% share" indicator in WID is calculated as the pre-tax national income share held by the richest 10% of the population. Though it employs the same general methodological approach as Leigh and Jencks' (2007) *Top Share* 10, WID is different in its income concepts, population age cut-offs and income units.[5] Other than *Top Share* 10, all other variables remain the same.

Updating the other variables with more recent data was relatively straightforward given all were available online. For example, the measures of life expectancy and infant mortality were updated through 2018 and available at the HMD website. Similarly, updated GDP data were available via the Maddison project website.[6] In this manner, I was able to extend L&J's data from 2003 to 2018.

The MI estimates using the updated data are reported in TABLE 3. As a point of comparison, the first two columns use data from 1960-2003. The only difference with the previously reported MI results in TABLE 2 is that I have replaced L&J's measure of inequality with the WID measure. The last two columns report results from estimating the same specifications using 1960-2018 data.

The first thing to note is that the respective estimates change signs. They are now consistent with the income inequality hypothesis, though they remain statistically

[4] See https://wid.world/wid-world/ for details.

[5] As noted, the norm pre-tax national income in WID covers all pre-tax personal income flows (including capital gains), after taking into account the operation of pension system. And the population is estimated upon individuals over age 20. The income unit is individual-based (rather than the household), with resources split equally within couples.

[6] See https://www.rug.nl/ggdc/historicaldevelopment/maddison/ for more details.

insignificant. A 10-percentage point increase in *Top Share 10* is estimated to lower life expectancy by 0.2 years and increase the infant mortality rate by 5%.

If we take a "best/worst" case scenario as L&J did and use the bounds of the confidence intervals as a measure of largest "possible" effect, the corresponding effects from a 10-percentage point increase in income inequality are a decrease in life expectancy of 1 year and an increase in the infant mortality rate of about 20 percent. These effects are not negligible. Of course, using values from the other end of the confidence intervals reverses these effects. Evaluated at these alternative values, a 10-percentage point increase in income inequality is estimated to increase life expectancy by 0.5 years and reduce the infant mortality rate by 9 percent.

As a point of reference, it should be noted that the *Top Share 10* estimates of -0.021 and 0.005 in the 1960-2018 regressions of TABLE 3 fit firmly within the confidence intervals estimated by L&J using their 1960-2003 data. As reported in TABLE 2, the 95% confidence intervals for *Top Share 10* in the Original/LE regression and Original/IM regressions are [-0.100,0.166] and [-0.034,0.013], respectively. Seen from that perspective, using MI and updating the data to 2018 has resulted in estimates consistent with what L&J report in their paper.

<u>6. Conclusion</u>

This study replicates Leigh and Jencks' (2007) analysis of the relationship between income inequality and mortality. Using L&J's preferred specification, I am able to closely reproduce their original findings after reconstructing their data from original sources. When I use multiple imputation instead of their method of linear interpolation, I largely confirm their results. When I extend their data from 2003 to 2018, I again do not find a significant relationship between income inequality and mortality. As a result, I conclude that my replication exercise confirms L&J's results, providing even stronger evidence for the view that income inequality is not adversely related to mortality.

TABLE 1

Missing Data in Reconstructed Leigh and Jencks' (2007) Dataset

Variable	Total observations	Number missing	Missing proportion (%)
Income share of richest 10% (*Top Share 10*)	528	72	13.6
Average life expectancy at birth	528	43	8.1
Infant mortality rate	528	38	7.2
Real GDP per capita	528	24	4.5
Average years of education of adults aged 15+	528	420	79.5
Real public health spending per capita	528	125	23.7
Real private health spending per capita	528	87	16.5

Note: The sample period is 1960 to 2003.

TABLE 2

Replication of Leigh and Jencks (2007), Columns 7 and 8, Table 4

	Original		Complete Case		Linear Interpolation		Multiple Imputation	
	Reg. (1)	Reg. (2)	Reg. (1)	Reg. (2)	Reg. (1)	Reg. (2)	Reg. (1)	Reg. (2)
	LE	IM	LE	IM	LE	IM	LE	IM
Income share of richest	0.033	-0.010	0.003	-0.001	0.040	-0.006	0.054	-0.006
10% (Top Share 10)	[-0.100,0.166]	[-0.034,0.013]	[-0.123,0.130]	[-0.024,0.022]	[-0.091,0.170]	[-0.026,0.013]	[-0.038,0.146]	[-0.021,0.009]
Real GDP per capita	0.170	-0.075 *	0.218	-0.061	0.125	-0.068 *	0.031	-0.087 ***
($ 1,000s)	[-0.106,0.446]	[-0.158,0.009]	[-0.432,0.867]	[-0.171,0.050]	[-0.158,0.407]	[-0.144,0.009]	[-0.212,0.275]	[-0.135, -0.039]
Real GDP per capita	-0.005	0.002 **	-0.009	0.002**	-0.003	0.002 **	-0.005 *	0.003 ***
squared ($ 1,000s)	[-0.014,0.005]	[0.000,0.004]	[-0.022,0.005]	[0.000,0.004]	[-0.011,0.005]	[0.001,0.004]	[-0.011,0.001]	[0.002,0.005]
Average years of	-0.332	-0.050	-0.178	-0.034	-0.114	-0.024	-0.170	0.002
education of adults	[-1.002,0.338]	[-0.115,0.016]	[-0.683,0.326]	[-0.129,0.061]	[-0.517,0.288]	[-0.098,0.051]	[-0.592,0.251]	[-0.060,0.063]
aged 15+								
Log real public health	0.105	-0.049	1.224 *	-0.174 *	0.041	-0.041	1.880 ***	-0.388 ***
spending per capita	[-0.440,0.651]	[-0.152,0.055]	[-0.226,2.673]	[-0.363,0.015]	[-0.427,0.510]	[-0.155,0.073]	[1.158,2.602]	[-0.515, -0.261]
Log real private health	0.295	-0.050	0.156	0.047	0.328	-0.045	-0.031	0.024
spending per capita	[-0.368,0.958]	[-0.243,0.142]	[-0.522,0.835]	[-0.126,0.220]	[-0.328,0.985]	[-0.227,0.138]	[-0.616,0.554]	[-0.127,0.175]
Country and year FE	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
N	*430*	*430*	*64*	*69*	*427*	*427*	*528*	*528*

*Note: "Original" labels the results published in Leigh and Jencks (2007); "Complete Case" labels the replication results without missing data treatment); "Linear Interpolation" labels the replication results with linear interpolation; "Multiple Imputation" labels the replication results with multiple imputation. Numbers in brackets are robust standard errors, clustered at Numbers in brackets are robust standard errors, clustered at the country level. Significance level is labeled with stars, * for 10% ** for 5% and *** for 1%. "LE" stands for "Life expectancy at birth", "IM" is the "Log of the infant mortality rate (per 1,000 live birth)".*

TABLE 3

Reanalysis of Leigh and Jencks (2007) Using Updated Data

	Multiple Imputation *(period 1960-2003)*		Multiple Imputation *(period 1960-2018)*	
	Reg. (1)	Reg. (2)	Reg. (1)	Reg. (2)
	LE	IM	LE	IM
Income share of richest 10%	0.054	-0.013	-0.021	0.005
(Top Share 10)	[-0.055, 0.162]	[-0.031,0.006]	[-0.098,0.055]	[-0.009,0.019]
Real GDP per capita ($ 1,000s)	0.179 *	-0.086 ***	0.104	-0.075 ***
	[0.001, 0.359]	[-0.124, -0.048]	[-0.065,0.274]	[-0.112, -0.037]
Real GDP per capita squared ($ 1,000s)	-0.004 **	0.001 ***	-0.001	0.001 ***
	[-0.007, -0.001]	[0.001,0.002]	[-0.003,0.001]	[0.000,0.001]
Average years of education of adults aged	-0.160	-0.000	-0.074	-0.035 *
15+	[-0.447, 0.128]	[-0.065,0.064]	[-0.291,0.143]	[-0.080,0.010]
Log real public health spending per capita	1.050 *	-0.349 ***	-0.641	0.039
	[-0.197, 2.296]	[-0.568,-0.131]	[-1.614,0.332]	[-0.143,0.220]
Log real private health spending per capita	-0.441	0.140 *	0.529 **	-0.049
	[-1.075, 0.193]	[-0.008,0.287]	[0.029,1.030]	[-0.183,0.086]
Country and year FE	*YES*	*YES*	*YES*	*YES*
N	*528*	*528*	*708*	*708*

*Note: "Multiple Imputation" labels the replication results with multiple imputation. Different sample periods are noted. Numbers in brackets are robust standard errors, clustered at Numbers in brackets are robust standard errors, clustered at the country level. Significance level is labeled with stars, * for 10% ** for 5% and *** for 1%. "LE" stands for "Life expectancy at birth", "IM" is the "Log of the infant mortality rate (per 1,000 live birth)".*

References

Allison, P. D. (2001). *Missing Data*. Sage publications.

Alstadsæter, A., Johannesen, N., & Zucman, G. (2019). Tax evasion and inequality. *American Economic Review*, *109*(6), 2073-2103.

Alvaredo, F., Atkinson, A. B., Piketty, T., & Saez, E. (2013). The top 1 percent in international and historical perspective. *Journal of Economic Perspectives*, *27*(3), 3-20.

Alvaredo, F., Atkinson, A., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2016). Distributional National Accounts (DINA) guidelines: Concepts and methods used in WID. world.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2017). Global inequality dynamics: New findings from WID. world. *American Economic Review*, *107*(5), 404-09.

Atkinson, A. B., Casarico, A., & Voitchovsky, S. (2018). Top incomes and the gender divide. *The Journal of Economic Inequality*, *16*(2), 225-256.

Barro, R. J., & Lee, J. W. (2013). A new dataset of educational attainment in the world, 1950–2010. *Journal of Development Economics*, *104*, 184-198.

Blanchet, T., Chancel, L., & Gethin, A. (2019). How unequal is Europe? Evidence from distributional national accounts, 1980-2017. *WID. world working paper*, *6*.

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford press.

Fiscella, K., & Franks, P. (2000). Individual income, income inequality, health, and mortality: what are the relationships? *Health Services Research,* 35(1 Pt 2), 307.

Gravelle, H., Wildman, J., & Sutton, M. (2002). Income, income inequality and health: what can we learn from aggregate data? *Social Science & Medicine*, 54(4), 577-589.

Kim, K. T. (2019). Income inequality, welfare regimes and aggregate health: Review of reviews. *International Journal of Social Welfare*, *28*(1), 31-43.

Kondo, N., Sembajwe, G., Kawachi, I., Van Dam, R. M., Subramanian, S. V., & Yamagata, Z. (2009). Income inequality, mortality, and self-rated health: Meta-analysis of multilevel studies. *BMJ*, 339.

Leigh A. & Jencks C. (2007). Inequality and mortality: Long-run evidence from a panel of countries. *Journal of Health Economics*, 26(1):1-24.

Leigh A. & Jencks C. (2013). healthtopinc_readme_web.pdf, Inequality and Mortality: Long-Run Evidence from a Panel of Countries, https://doi.org/10.7910/DVN/QB76A1/VJVTSS, Harvard Dataverse, V1.

Leigh, A. (2007). How closely do top income shares track other measures of inequality? *The Economic Journal*, *117*(524), F619-F633.

Little, R. J. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association*, 87(420), 1227-1237.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Lynch, J., Smith, G. D., Harper, S., & Hillemeier, M. (2004). Is income inequality a determinant of population health? Part 2. US national and regional trends in Income inequality and age- and cause-specific mortality. *The Milbank Quarterly*, 82(2), 355-400.

Macinko, J. A., Shi, L., Starfield, B., & Wulu Jr, J. T. (2003). Income inequality and health: a critical review of the literature. *Medical Care Research and Review*, 60(4), 407-452.

Macinko, J. A., Shi, L., & Starfield, B. (2004). Wage inequality, the health system, and infant mortality in wealthy industrialized countries, 1970–1996. *Social Science & Medicine*, 58(2), 279-292.

Maddison, A. (2010). Statistics on world population, GDP and per capita GDP, 1-2008 AD. Historical Statistics, 3, 1-36.

Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, *24*(7), 815-829.

OECD (2007). *OECD Health Data 2007: Statistics and Indicators for 30 Countries*. Organisation for Economic Co-operation and Development, Institute for Research and Information in Health Economics (IRDES): Paris.

Patel, V., Burns, J. K., Dhingra, M., Tarver, L., Kohrt, B. A., & Lund, C. (2018). Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms. *World Psychiatry*, 17(1), 76-89.

Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology*, *9*, 157.

Piketty, T., & Atkinson, A. B. (2010). *Top Incomes: A Global Perspective*. Oxford University Press.

Rodgers, G. B. (1979). Income and inequality as determinants of mortality: an international cross-section analysis. *Population Studies*, 33(2), 343-351.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473-489.

StataCorp, L. (2017). Stata survival analysis reference manual. Texas, College Station.

Subramanian, S. V., & Kawachi, I. (2004). Income inequality and health: what have we learned so far? *Epidemiologic Reviews,* 26(1), 78-91.

Von Hippel, P. T. (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods & Research*, *49*(3), 699-718.

Wilkinson, R. G. (1998). Mortality and distribution of income. Low relative income affects mortality. *BMJ*, 316(7144), 1611-1612.

Wilkinson, R. G. (2002). *Unhealthy Societies: The Afflictions of Inequality*. Routledge.

Wilkinson, R. G., & Pickett, K. E. (2006). Income inequality and population health: a review and explanation of the evidence. *Social Science & Medicine*, 62(7), 1768-1784.

APPENDIX A:
Summary Statistics

	Mean	Std. dev	Min	Max	N
Income share of richest 10%	31.901	4.129	22.3	43.108	407
Average life expectancy at birth (years)	74.916	2.833	66.6	80.55	473
Log infant mortality rate (per 1,000 live births)	2.330	0.567	1.147	3.865	497
Real GDP per capita ($ 1,000s)	14.577	4.781	3.072	29.074	528
Average years of education of adults aged 15+	9.137	1.967	4.18	12.67	108
Log real public health spending per capita	6.278	1.144	1.792	7.835	420
Log real private health spending per capita	5.278	1.231	1.946	8.057	420

Note: Data period ranges from 1960 to 2003. All variables are summarized with missing values.

APPENDIX B:
Option Settings and Diagnostics in Multiple Imputation

In multiple imputation, estimates are sensitive to the options used in the imputation model algorithm. An imputation model is mostly determined by three options. In Stata, these are (i) "*burnin*", (ii) "*burnbetween*", and (iii) "*add*". The first two options determine which set of imputed data should be retained, while the last option determines the total number of retained datasets prior to analysis. This appendix explains how I set the values for these three options.

Like maximum likelihood, multiple imputation employs a concept of "convergence". I follow the approach of focusing on the "worst linear function" (WLF) to diagnose the convergence of my imputation models (Enders, 2010; StataCorp, 2017). "Convergence" in imputation models refers to the properties of "stationarity" and "independence", which are mediated by the options *burnin* and *burnbetween*, respectively.

My study includes three pairs of MI regressions, according to whether the dependent variable is life expectancy (LE) or log of infant mortality (IM) (cf. Columns 7 and 8 in TABLE2, Columns 1 and 2 in TABLE 3, and Columns 3 and 4 in TABLE 3). Although there are two different specifications of regression models, one for LE and one for IM, the imputation model is the same for both for a given dataset because each imputation algorithm uses the other dependent variable as an auxiliary variable in imputing missing values.

In an imputation model, the *burnin* and *burnbetween* values are set first and then diagnostics are used to assess them using worst linear function (WLF) plots (Enders, 2010; StataCrop, 2017). After specifying *burnin* and *burnbetween*, I set the total number of imputations via the *add* option following von Hippel's (2020) approach.

To produce the Multiple Imputation results in TABLE 2, I set *burnin* equal to 1000 and kept *burnbetween* at its default value of 100. To obtain the diagnostic plots, I first ran a pilot imputation with 5 datasets. FIGURE A.1 plots a "time series" graph of the mean of WLF for each of the 1000 iterations, where iterations substitute as a measure of “time”. For *burnin* = 1000, no trend is apparent, indicating the multiple imputation algorithm is “stationary”.

FIGURE A.2 plots a correlogram of mean WLF values over iterations. The imputation algorithm builds in a certain degree of dependence from one imputation to another. One wants sufficient “spacing” between imputations so that that the datasets are independent. According to FIGURE A.2, anything beyond 15 to 20 iterations should be sufficient to ensure independence. Thus I keep the default value for *burnbetween* of 100.

I then applied the user-written (by von Hippel) Stata command “*how_many_imputations*” to determine the total number of imputations. This command selects a total number of imputations such that the standard errors of the respective estimates will vary less than 5%. For the estimates in TABLE 2, the associated total numbers of imputations were 181 for the life expectancy regression and 191 for the infant mortality regression.

TABLE 3 uses two imputation models, one to generate data for the regressions in Columns 1 and 2, and another for the regressions in Columns 3 and 4. I continue to set the *burnin* and and *burnbetween* options at 1000 and 100, respectively. The diagnostic trend figures for WLF are presented in FIGURES A.3 and A.5., while the diagnostic correlograms are reported in FIGURES A.4 and A.6. All the plots are consistent with “stationarity” and “independence”, confirming the selection of my *burnin* and *burnbetween* option values. The corresponding numbers of total imputations are (i) 142 and 178 for the regressions in Columns 1 and 2 of TABLE 3, and (ii) 142 and 133 for the regressions in Columns 3 and 4 of TABLE 3.

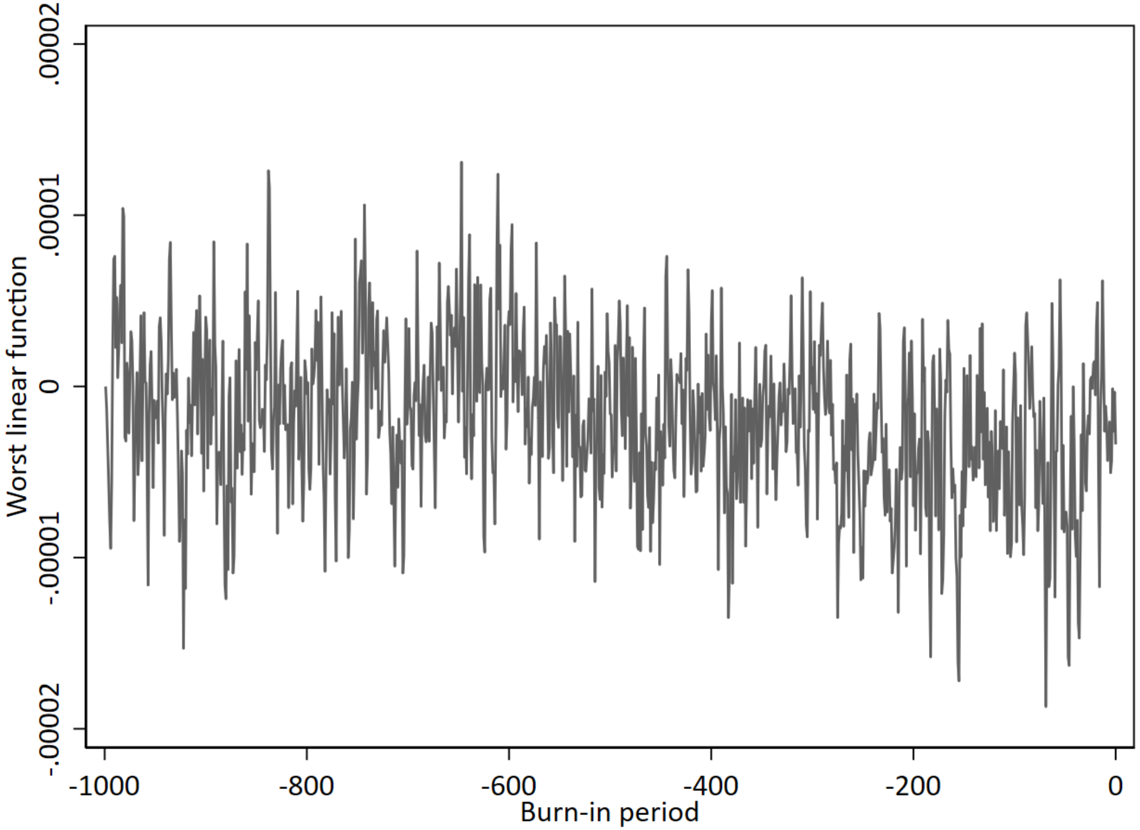

Figure A.1: Time-series plot of WLF (TABLE 2/Columns 7 and 8)

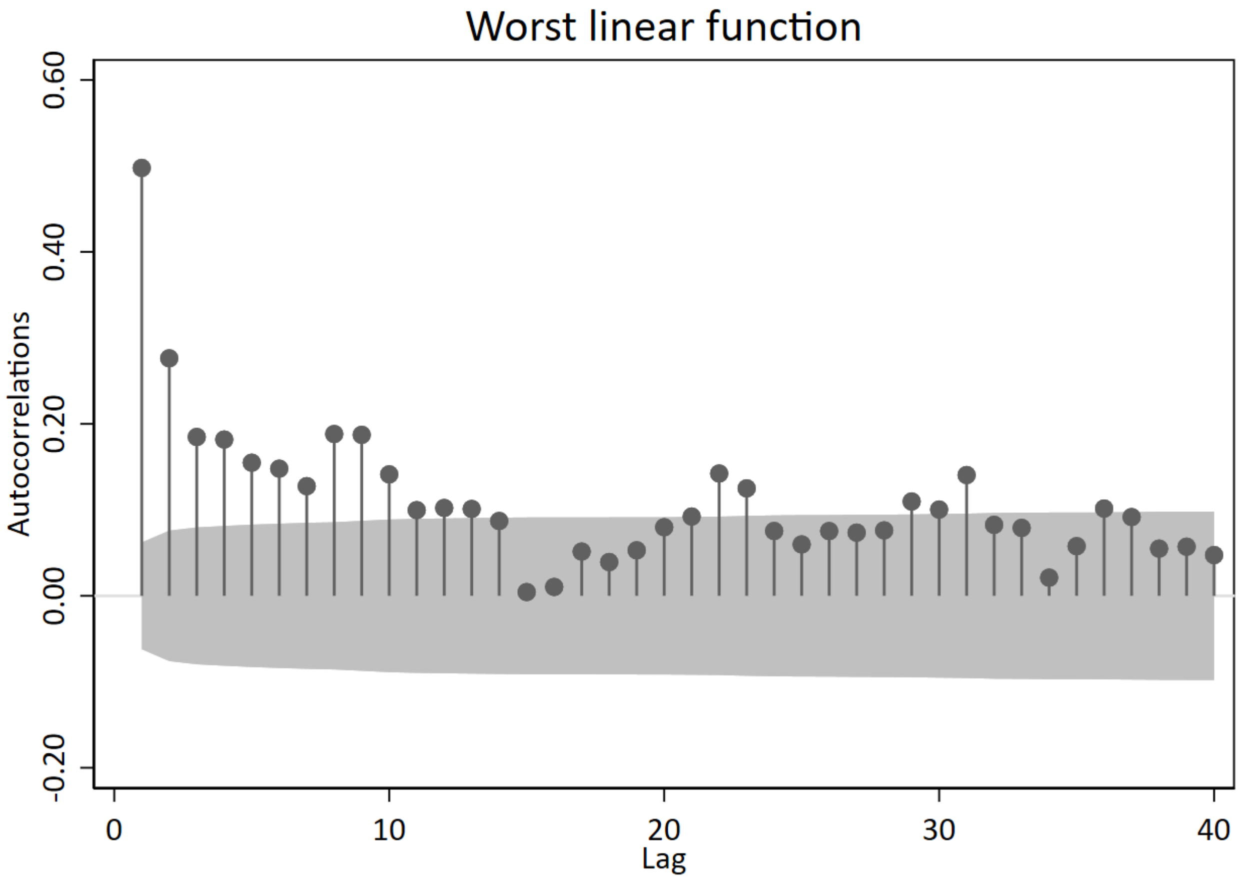

Figure A.2: Autocorrelation plot of WLF (TABLE 2, Columns 7 and 8)

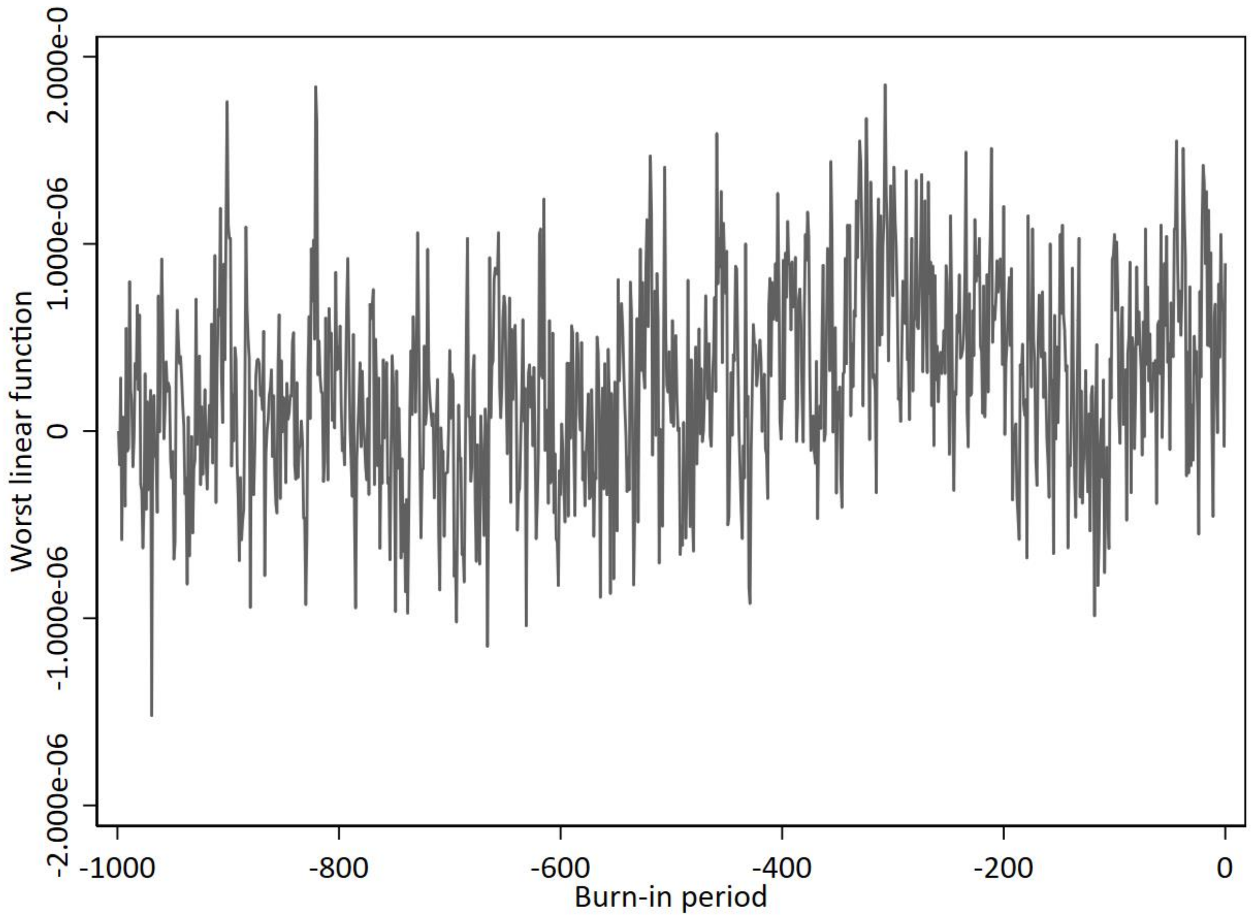

Figure A.3: Time-series plot of WLF (TABLE 3/Columns 1 and 2)

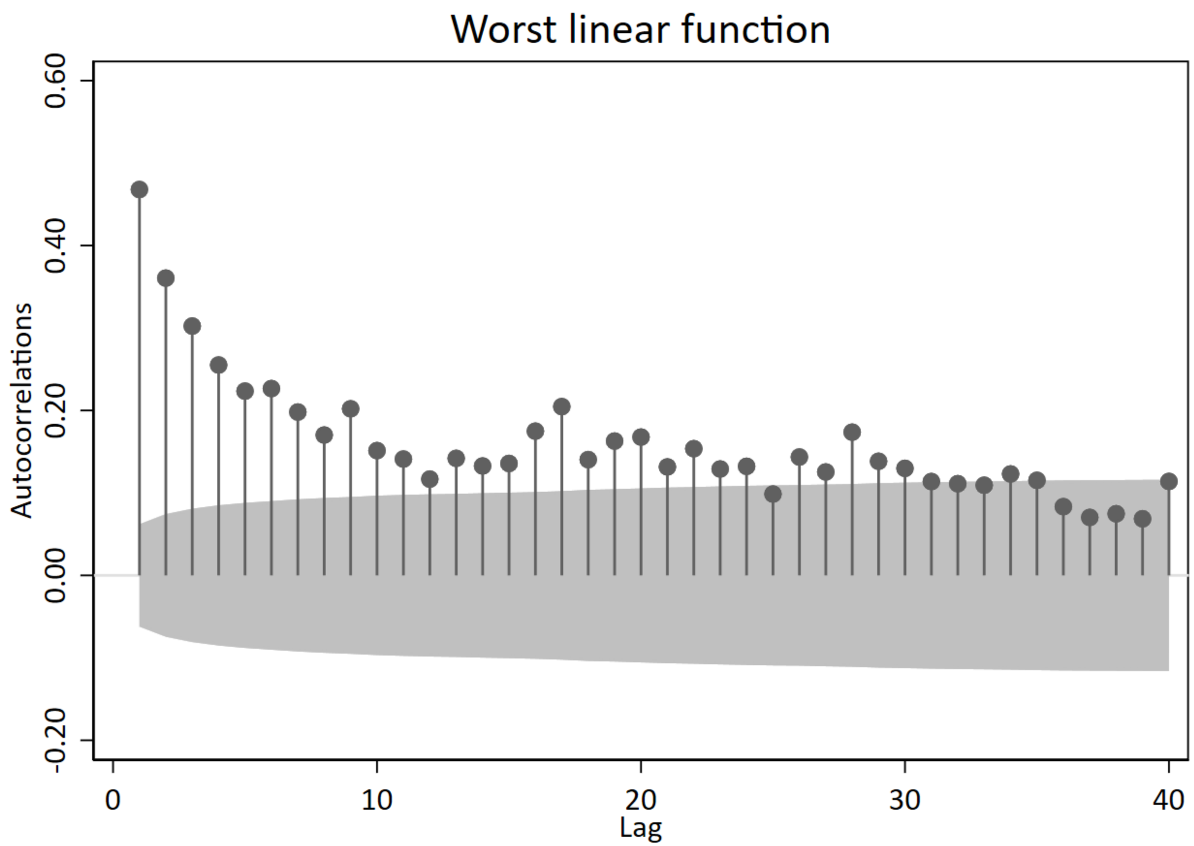

Figure A.4: Autocorrelation plot of WLF (TABLE 3, Columns 1 and 2)

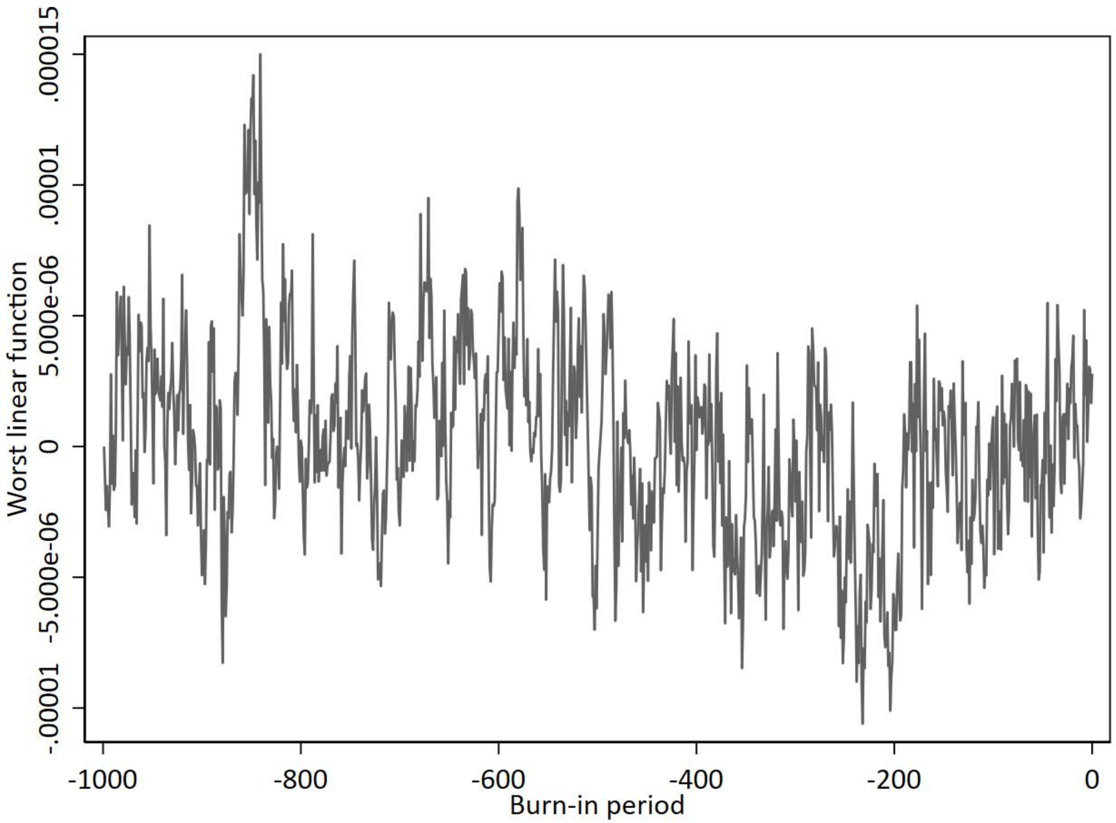

Figure A.5: Time-series plot of WLF (TABLE 3/Columns 3 and 4)

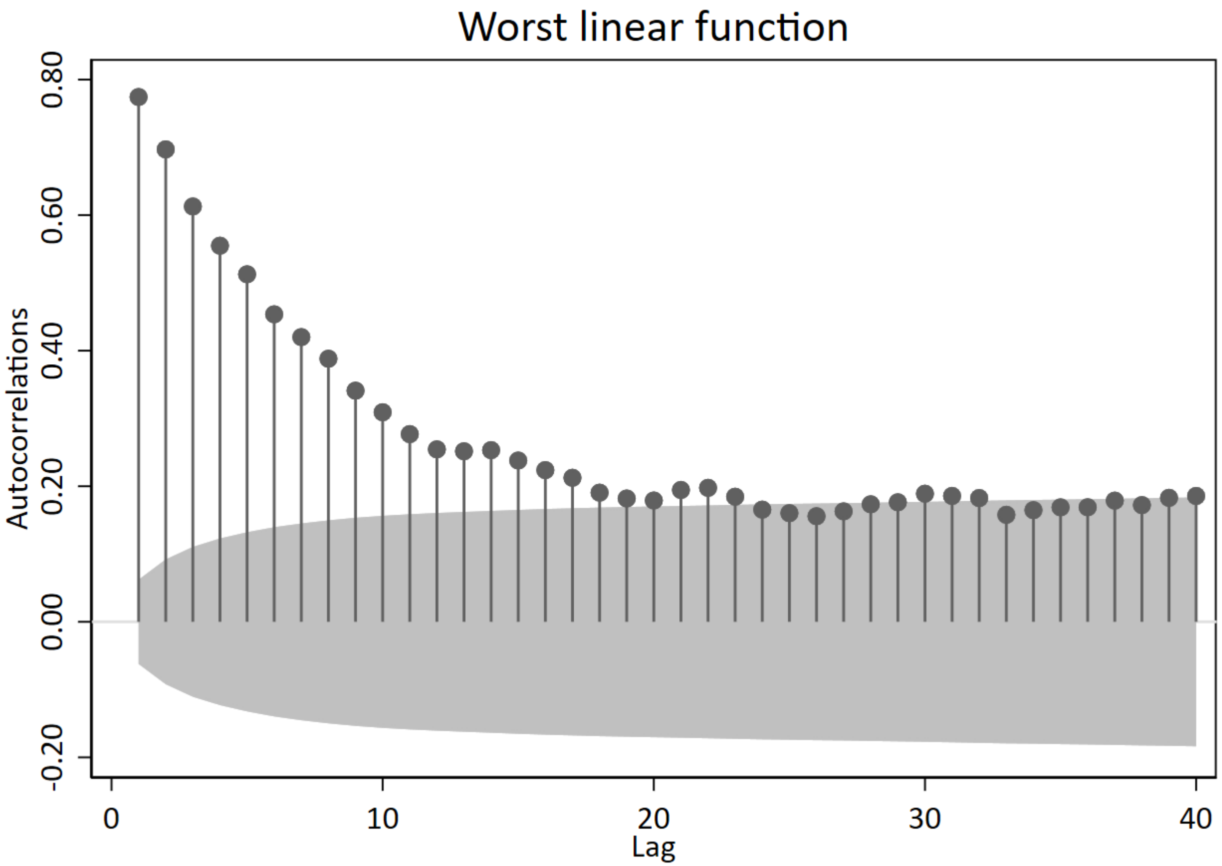

Figure A.6: Autocorrelation plot of WLF (TABLE 3, Columns 3 and 4)