The role of the world's rapid urbanization on total fertility rates

I. Introduction

My research question is: how does urbanization affect total fertility rates across the world, after controlling for GDP per capita, education, and food security? For the past century, escalated by the industrial revolution, the world has undergone rapid urbanization. The National Library of Medicine defines urbanization as the "process whereby a society changes from a rural to an urban way of life". But, for the purposes of this paper, we will take a more measurable approach and define it as the "gradual increase in the proportion of people living in urban areas". There are plenty of consequences as a result of urbanization, but one puzzling result is its effects on women's fertility. Macro data from across the world has seen births per woman decrease dramatically in the past few decades; multiple explanations have arisen to explain this, and one explanation among them is urbanization, i.e. the rapid migration from rural to urban areas.

The issue of fertility is interesting because its anticipated decline will have drastic effects on the world; Sanjeev Sanyal, principal economic advisor in India's ministry of finance, in 2011 predicted that the world's population will stabilize by 2050. This is due to replacement rates (total fertility rates to maintain the population) decreasing. These forecasts are important because a declining population in some countries, such as Japan, puts more pressure on institutions as the population grows older and needs financial support. The government needs to plan for these fertility outcomes, and determining what exactly causes these fertility declines is key to accurately plan funding in these areas.

This is where urbanization comes into play. **By running multiple ordinary least squares regressions of urbanization on total fertility rates, while controlling for GDP per capita, education, and food (crop index), I found that urbanization has a negative correlation with fertility. With my more complex counterfactual with added control variables, however, I found there is a less pronounced effect urbanization has on fertility.**

The key variables here are the percentage of urban population to the total population, and total fertility rates, which is measured as births per woman. In order to control for endogeneity of omitted variable biases, I included a counterfactual through meticulous selection of control variables. Controls I decided on include an individual's prosperity, measured as GDP per capita; education, which is measured as primary school enrollment; finally, crop indices to partially measure an individual's food security.

II. Literature Review

Academic literature of the effects of urbanization on fertility is extensive. Papers by Michael J. White (2005) examines case studies in African countries such as Ghana which are undergoing rapid urbanization. In his first paper on coastal Ghana, White uses survey data to find that urbanization not only has a negative relationship with fertility, but there may also be generational effects. This tells me that there will be some endogeneity in my regression that cannot be captured by the macro data. People who migrate to urban areas may maintain cultural or religious ties from their life in rural areas. This suggests that there are cultural or religious values that will also affect my dependent variable of interest, children ever born. It is difficult to measure such values in macro data, but it's important to keep this source of endogeneity into account as to not overstate the effects urbanization alone has on fertility.

Besides papers that use micro data to support their research question, there are also many that use macro data. Yasuhiro Sato and Kazuhiro Yamamoto (2005) analyze the demographic transitions of Europe and less developed countries. One of the main findings of the paper is the "U-shaped" curve of population growth. Urbanization tends to increase population growth up to a point before it rapidly declines again. This is opposite of the common belief that urbanization only drives fertility down. To simulate this, they create a model of rural-urban migration that simulates the effects urbanization has on wage rate, fertility, and overall economic growth. Their idea is the main driving force for having children or not is the cost-savings decision; as their wage rate increases in urban centers, parents have less time to allot to their children, causing them to have less children in the process. Sato and Yamamoto use terms such as "agglomeration economies" and "congestion economies" to describe this answer to the fertility decline. However, there are some effects that are missing from their model. As education and credentials become more important costs when having children, parents must face a quantity-quality tradeoff. In other words, the rising costs of education and thus the "quality" of investment per child drives fertility rates down.

Another paper from Silva and Tenreyo (2017) evaluates a new angle of the fertility decline: population control policies. One type of population control measure is family planning. They run a regression to see how family planning messages affect fertility, controlling for urbanization and mortality rates. The paper's findings demonstrate a strong correlation between family planning messages and declining fertility rates. Urbanization and population control policies are highly correlated. Interestingly they use urbanization and education as control variables here. In my paper, urbanization will be my key independent variable.

These papers analyze a subset of countries. White's papers on Ghana use survey data from a single country, whereas Sato et al. and Silva et al. use United Nations and World Bank data to regress a cross-section of countries. My paper will be more like Silva et al.'s, as I will use macro data on over 170 countries to run my regression, controlling for variables found in this literature to inform my decision.

III. Data

The database I retrieved my data from is The World Bank open database. The World Bank group is a reliable source of development data, with workers in over 170 countries and data on hundreds of developmental variables. To regress the effect of urbanization on fertility, I chose to use panel data of every country available in the World Bank website to create a comprehensive dataset to perform my regression. In order to access the World Bank data, I used the "WDI" package in R Studio, then removed all aggregate values (i.e. North America, East Asia, middle income, etc) since they're already accounted for in the individual country's data. The time period is the maximum amount of time available, 1960-2020. The variables I extracted from the World Bank are total fertility rates (births per woman) to measure fertility, then total urban population (% of total population) to measure urbanization over time. Some control variables I included are GDP per capita to measure income, the crop production index to measure access to food, and finally primary school enrollment (%) in the female population.

GDP per capita is measured in U.S. dollars, and the figures for the GDP are converted from domestic currencies to U.S. dollars using single year official exchange rates. GDP per capita may not be the best measure of an individual's buying power, but it is the best given the macro data and the purposes of this paper. Because GDP per capita takes on larger values compared to my other variables, I took the log of GDP in my OLS regressions otherwise my estimates in my regression will be inaccurate.

Because the dataset is large there are bound to be some N/A values. Since these variables are mostly linear it should be easy for R to fill in the gaps, but it's worth considering there will be some missing values affecting my estimates. I arrived at 13,176 observations of 6 variables. To create a summary statistics table, I used the stargazer package in R.

Table 1 Here

As we can see there is some discrepancy in the number of observations due to missing values. We have the most observations in our dependent variable fertility and independent variable, urban population, but our controls are less complete. But since the controls follow a pattern it shouldn't be too much of a concern for our regression, as R can fill in the blanks.

One of the predicted outcomes is that urbanization decreases total fertility rates in women. If we graph both total fertility and urban population over time, we can see this exact relationship in the data. I took the natural logarithm of urban, otherwise the data was not in a normal distribution. See the graph below to see the key relationship:

Graph 1 Here

As we can see, as the world population becomes more urban over time, fertility rates decrease. Besides regressing over time, we can also graph the linear regression of urbanization on fertility:

Graph 2 Here

Again we can see a negative relationship between urbanization and fertility. As urban population increases as a proportion of the total population, births per woman decreases. To create a counterfactual I will use an OLS regression with control variables. Income (GDP per capita) has an impact on fertility and is also related to urbanization; education also has an impact on fertility, and is correlated with urbanization; finally, access to food also has an impact on the decision to bear children, and may also be correlated to urbanization, so I included the crop index as another control.

IV. Theory

There have been many theories attempting to explain the fertility transition in the world. Urbanization is one of the oldest explanations to the fertility decline in classic demographic transition theory. Notestein (1952) discusses the emergence of the "small family" in urban-industrial society in the past century. As society became more urban, traditional pressures on women to bear as many children began to disappear. Individual achievements, which were separate from raising a family, began to gain more importance in city life. Besides this, Notestein also mentions the growing importance of education, which will also be an important variable in my model determining children ever born.

The primary interaction in my model is the effect urbanization has on fertility. In my literature review we saw that many papers have observed decreasing fertility with rising urbanization. However, there are many sources of endogeneity that may arise with determinants of fertility. The primary source of endogeneity in this regression would be omitted variable bias. One control variable I included to counteract this endogeneity is income, measured as GDP per capita (in U.S. dollars).

As we saw in the paper by Silva and Tenreyo (2017), the decision to bear children today is mostly a cost-savings decision. In their paper, they observe an increasing wage rate over time, due to agglomeration economies. Agglomeration benefits mean as more people enter cities, cost savings tend to rise due to network effects and increasing returns to scale. Costs to production decreasing will mean higher wages (or income) as a result. In the past, more children were necessary to help earn a living, but this becomes unnecessary as individuals can now support themselves in the city. To see how much this rise in income affects fertility, as to not overestimate the other effects of urbanization, I included income as a control variable.

Education is another variable that has a considerable effect on fertility. It is also correlated with urbanization, as more schools are located in cities. Findings by Kim (2016) conclude that educated women typically have less children than uneducated women. However, they also note that the interpretation is unclear. Two different mechanisms operate here: for one, education can have an impact on child health, through improved childcare, and education can decrease their desire to have children. Second, women's education on contraceptives or other birth control methods work to decrease fertility. To control for these effects, I included primary school enrollment as another variable in my regression.

Finally, another omitted variable could be the impact of food on fertility levels. According to DiClemente et al (2021), women experiencing any form of household hunger are 21.6% less likely to want children after controlling for living children, education and age.

Household hunger here was measured by the Household Hunger Score, a robust indicator that allows them to analyze food insecurity on a household level. In my model, I am using food production indices from World Bank data, to analyze food insecurity on a macro level as opposed to micro data in DiClemente et al.

V. Estimation

As mentioned in the data section, I will be using R studio to perform my multivariable regression. Since my data is macro data, it would be unfitting to set up a quasi-experiment such as differences-in-differences, and instrumental variables would be difficult to use in this setting. To control for endogeneity, I will have to rely on the use of control variables. To see how urbanization affects fertility, I used multiple ordinary least squares regressions, starting with a simple model with urbanization and fertility rates, then progressively adding control variables one at a time to find a model with best fit. My first model is as follows:

$$Fertility = \beta_1 Urban + \varepsilon$$

The coefficients of the regression should be easy to interpret. The main interaction in my estimation is between the percentage of population in urban areas and total fertility rates (children ever born) over time. I expect to find coefficient values that suggest a negative relationship between urbanization and fertility across the world. In other words, as the world becomes more urban, births per woman should decline. This has been a common observation in the literature I reviewed earlier.

As to not overestimate the effects of urbanization on fertility, I will also include some control variables to see how it impacts my coefficients of interest. I expect it won’t change the negative sign of the urban coefficient, but the controls may reveal the magnitude of the effect is overestimated or even underestimated. My second model has these controls added onto the equation:

$$Fertility = \beta_1 Urban + \beta_2 ln(GDP) + \beta_3 Enrollment + \beta_4 Crop\ Index + \varepsilon$$

Since I am taking data from a large number of countries, there are bound to be issues with endogeneity across different cultures, governments, or definitions of what counts as “urban”. Because of this, there may be issues with establishing causality, and I will interpret my results with caution. Furthermore, as I am working with macro data, there will be variables such as cultural and religious differences that will have an effect on the urban variable and total fertility rates. This was seen in some texts covered in the literature review, such as White (2005), which saw significant differences in fertility across religious groups.

VI. Results

To see how urbanization affects total fertility rates, I ran several OLS regressions in R to find results consistent with my theory section. After creating multiple models, starting with the most simple one without any controls, then gradually adding controls, I found a negative relationship between urbanization and fertility. However, with the added controls the effects were less pronounced. This matches with my hypothesis that urbanization will decrease fertility over time. My control variables revealed some omitted variable bias in the original model.

We should begin with my simplest model, only taking Urban and Fertility variables into account. Looking at Model 1, I found the variable Urban has a negative relationship with

Fertility. As the percentage of people living in urban areas to the total population goes up by 1%, total fertility rates decrease by 0.05. The p-value is also less than 0.001; this means the relationship is statistically significant. Besides being backed up by statistical significance, this correlation is backed up by the theory. We expected in our theory section that as people move to cities, fertility will decline as the costs outweigh the benefits of having many children. We should take these results with caution, however, as we haven't ruled out any endogeneity that may arise in the model. For one, there may be omitted variable biases we haven't accounted for, overestimating the negative sign of the coefficient. There are other variables that affect the decision to bear children, one being income. Here, we will use GDP per capita, logged.

Next I added my control variables to control for omitted variable biases in my initial model. As I said in my data section, I took the natural logarithm of GDP to properly estimate the coefficients in my model. This second model with GDP per capita added to the model as a control variable is a level-log regression. In this second model (Model 2 in the appendix) I found the effect of urbanization on fertility was less pronounced, with 1% increase in urbanization only decreasing births by 0.02 per woman, all else held constant. Urban and GDP was also significant at $p<0.001$, meaning a good fit for the model.

The economics behind this was described in the theory section: as income increases, women typically have less children. Children in developing countries are often a source of income in rural areas, relying on these children to work jobs or through subsistence agriculture, as helping hands on the farm. As GDP per capita in the country increases, the typical family moves away from subsistence agriculture and the need for help on the farm. This is translated into our model with a 1% increase in GDP per capita decreasing births per woman by $(\frac{\beta_2}{100})$or $\frac{0.4822}{100} = 0.00482$ births, since this is a level-log model. Of course, moving away from rural areas into urban areas is the main interaction of our model, and adding income as a control variable helps see that the initial coefficient was overestimated.

My second control variable to account for endogeneity is education, measured as primary school enrollment. In our more complicated model with education, it further reduces the coefficient of urbanization, showing its possible overestimation in the first model. Education is also statistically significant at the 0.001 significance level, with a 1% increase in enrollment decreasing births per woman by 0.02. This was predicted in our theory, where we see increased education on contraceptives and other birth control, as well as increased income with a degree, leads to less children desired.

Finally, I included the crop index of each country in the model, which also turned out to be significant at $p<0.001$. What is important to note here is that increasing crop output actually has a negative impact on fertility. In our theory section I established the literature was divided on how food security impacted fertility levels. Here I can estimate that increasing food supplies actually decreases fertility. This runs counter to the intuition that women would only have children if they can provide nourishment, but in this case higher crop yields is correlated with less children. This result should be treated cautiously, however, as higher crop yield does not necessarily mean better food security for the general populace.

Overall, we can see the model accurately predicts lower fertility rates with urbanization over time. The adjusted R^2 value is particularly strong, with a value of 0.6477 in my second model, meaning 64.7% of the variance in my dependent variable, fertility, can be explained by my model. This is an improvement over the simple model without a counterfactual, only giving us an R^2 of 0.4134. However, the model has many missing values, meaning we should accept these results cautiously.

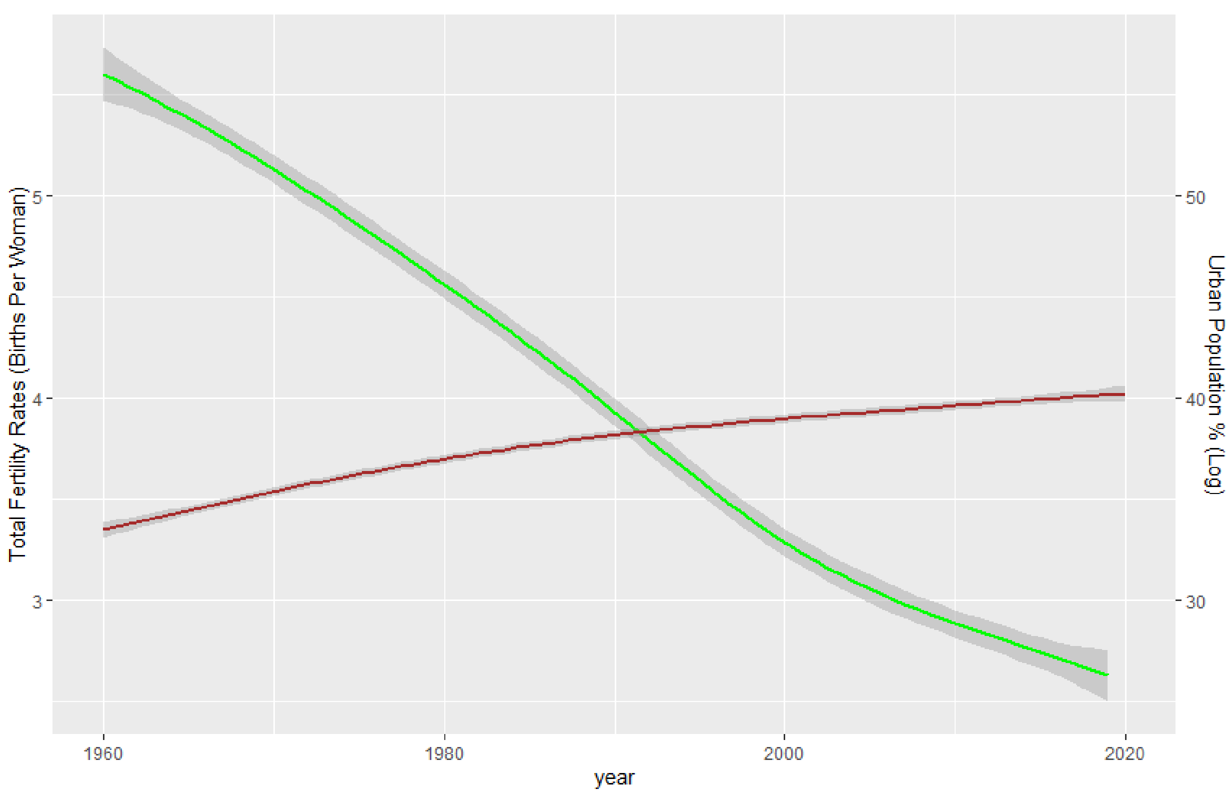
As mentioned earlier, I didn't take religious or cultural differences into account, a source of endogeneity immeasurable in the macro data available. Additional avenues of research could be to look individually at all the control variables here. For one, the effects of food security on fertility is interesting, as different data suggests different results. The fact that the world is rapidly urbanizing, and this suggests decreasing fertility, means governments should inform their policies based on this. Whether governments want to increase fertility or decrease it, investments into urban planning should be made with this effect on fertility in mind.

VII. Appendix

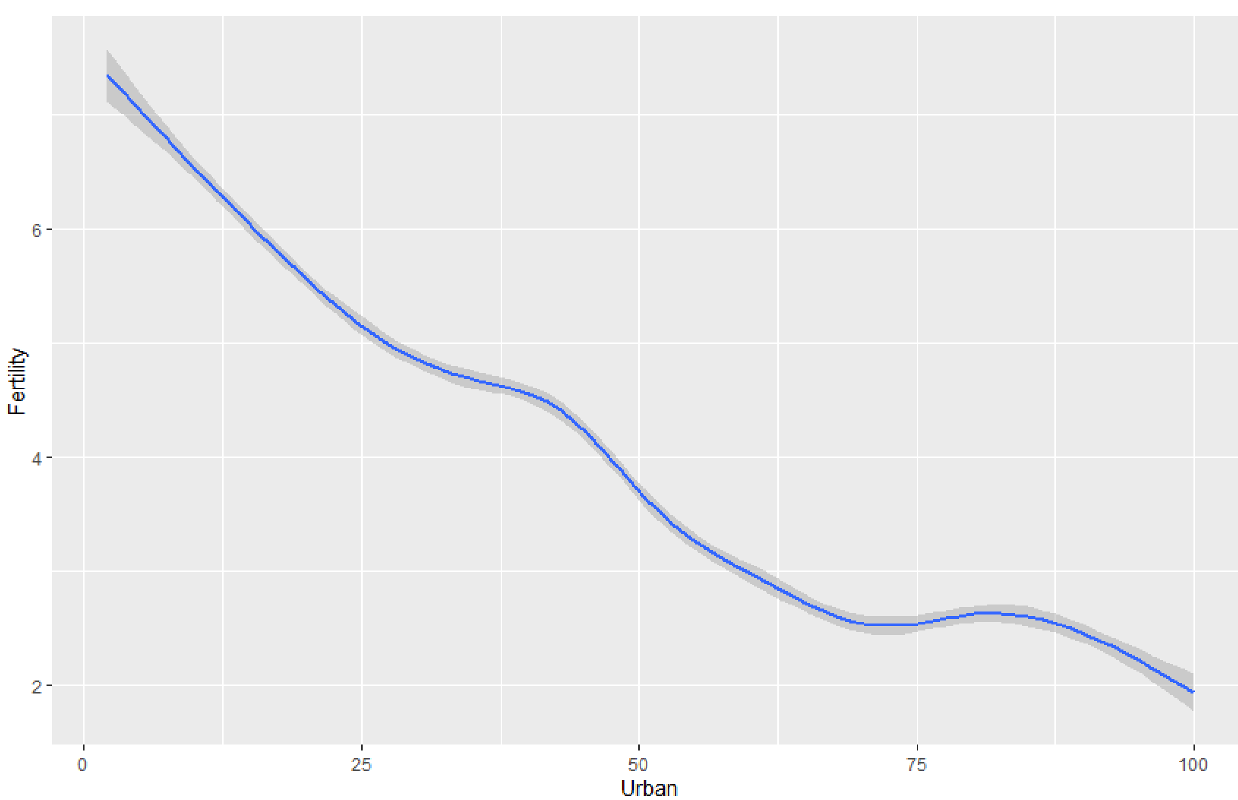
Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Year	13,176	1,990	17.607	1,960	1,975	2,005	2,020
GDP	9,505	11,686.180	18,521.810	124.276	1,341.290	14,069.510	183,244.600
Fertility	11,688	3.985	2.018	0.860	2.107	5.902	8.864
Urban	13,045	51.276	25.769	2.077	30.246	72.096	100.000
Enrollment	7,234	94.335	24.826	0.000	90.600	107.184	223.566
Crop Index	10,562	81.490	66.187	0.070	46.893	99.048	916.620

Graph 1: Total Fertility Rates and Urban (Log) Over Time



Graph 2: The Effect of Urbanization on Fertility



Regression Equations:

Model 1

$$Fertility = \beta_1 Urban + \varepsilon$$

Model 2 with counterfactual

$$Fertility = \beta_1 Urban + \beta_2 ln(GDP) + \beta_3 Enrollment + \beta_4 Crop\ Index + \varepsilon$$

Model 1: Urbanization and Fertility

Predictors	*Estimates*	*Std. Error*
(Intercept)	6.56 ***	0.031
Urban	-0.05 ***	0.0005
Observations	11603	
R^2 / R^2 adj	0.413 / 0.413	

* *p<0.05* ** *p<0.01* *** *p<0.001*

Model 2: Added Control Variables

Predictors	*Estimates*	*Std. Error*
(Intercept)	11.35 ***	0.112
Urban	-0.02 ***	0.001
GDP [log]	-0.48 ***	0.018
Enrollment	-0.02 ***	0.000
Crop Index	-0.01 ***	0.000
Observations	5952	
R^2 / R^2 adj	0.648 / 0.648	

* *p<0.05* ** *p<0.01* *** *p<0.001*

References

DiClemente K, Grace K, Kershaw T, Bosco E, Humphries D. "Investigating the Relationship between Food Insecurity and Fertility Preferences in Tanzania" Matern Child Health J. Feb. 2021. doi: 10.1007/s10995-020-03022-1

Kim, Jungho. "Female education and its impact on fertility" Institute of Labor Economics, Feb. 2016. https://wol.iza.org/articles/female-education-and-its-impact-on-fertility

National Library of Medicine. "Urbanization" Retrieved 15 Nov. 2021.

Notestein, Frank W. "Economics of Population and Food Supplies" Michigan State College, 1952. doi: 10.22004/ag.econ.209548

Sanyal, Sanjeev. "The End of Population Growth" Project Syndicate, 30 Oct. 2011. https://www.project-syndicate.org/commentary/the-end-of-population-growth-2011-10

Sato, Yasuhiro and Yamamato, Kazuhiro. "Population concentration, urbanization, and demographic transition" Journal of Urban Economics, 17 Mar. 2005. doi: 10.1016/j.jue.2005.01.004

Silva, Tiloka and Tenreyo, Silvana. "Population control policies and fertility convergence" Journal of Economic Perspectives, 2017. doi: 10.1257/jep.31.4.205

The World Bank, World Development Indicators (2020). Retrieved from http://data.worldbank.org/