Abstract:

It's projected that Japan's falling birthrate and dying population will significantly damage its international competitiveness, something an economy stuck in a two-decade long recession definitely cannot afford. Just two years ago Japan's population fell by a record 244,000, with estimates to only accelerate over the near future. Our paper seeks to investigate and study what has caused this negative population growth rate over the past few years, with hopes that identifying the cause can be an initial step in understanding how to boost a lethargic economy. Our methodology will include regressing quantitative and qualitative variables on the population growth rate, to help us determine whether our hypothesis is an accurate estimation for what has caused this downward population trend. However, using our independent variables to regress proved to have certain issues. For one, we only found that the independent variables that were statistically significant in our model, fell into the female empowerment category. Secondly, we found a high level of multicollinearity in our regression. And lastly, we found heteroscedasticity sourced in our female empowerment category. Over the course of our paper we will delve into the importance of female empowerment yielding significant R^2 and F-statistic values, but we will also consider the drawbacks of including certain variables in this category and the positive outcomes of removing the variable in question.

Introduction:

The goal of our research is to investigate what has caused the steady decline in Japan's population. After researching Index Mundi[1] we came across population growth rates for 232 countries. Of these 232 countries, 32 showed population growth rates that were negative (i.e. decreasing population). Japan and Germany were the only two of these 32 countries that have a per capita GDP that is greater than 30,000, so we found it fair to eliminate the other 30 countries as per capita GDP likely contributed significantly to the negative growth rates of these countries. However, Japan's growth rate problem seems to be more significant than Germany's. Germany has more than triple[2] the amount of immigrants than Japan has. These immigrants often have larger families than the native population; they either bring the rest of their family to the country to immigrate or they simply give birth to more children, once they are in the country they immigrate to, than the average person in the native population. This is particularly true in Germany where many of their immigrants come from Middle Eastern countries[3]. It is therefore likely that Germany's population growth rate will increase while Japan's will continue to decrease.

Japan has one of the lowest population growth rates in the world. In fact, on average, there are more people dying in Japan than being born, which would suggest a negative growth rate. 23.9% of Japan's population is 65 and older while only 13.5% is 15 and under and there are higher amounts of adult diapers being sold than baby

---

[1] http://www.indexmundi.com/g/r.aspx?v=24
[2] Source: stats.oecd.org
[3] Source: migrationpolicy.org

diapers. These statistics were shocking and fascinating to us. We thus decided to analyze the population growth rate to determine a regression model that can accurately explain some of the causes of the shocking growth rate percentage. The dependent variable of the model is the population growth rate in Japan, measured in percentage increase or decrease. Our paper hones in on what we believe are contributing factors or independent variables to the population decline evident in Japan today. Our five independent variables are a mixture of both quantitative and qualitative (dummy) variables that we hypothesize will relate to population growth rate in some way. Our quantitative independent variables include, (1) wage and salaried workers (i.e. % of wage earning females employed), (2) ratio of female to male labor force participation rate ('national estimate'), (3) number of marriages **annually** (measured as a % of population) and (4) death rates (crude). Our qualitative independent variable is (5) lifetime employment cut, which will be 0 before **??? and 1 after???** as the trend of providing people with lifetime employment decreasing dramatically beginning in **??** We believe lifetime employment is strongly related to how secure people feel about their financial future, and thereby the willingness to start or expand a family. We collected data for all of the tested independent variables for the range of years 1980-2013. This helped us keep our regression consistent and allows for a more conclusive analysis. Our model is stated as follows:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 D_{5t}$$

Where,

Y_t = population growth rate (%)

X_{1t} = % of employed females earning wages

X_{2t} = ratio of female to male labor force participation rate

X_{3t} = marriages as a % of population

X_{4t} = death rates (crude)

D_{5t} = lifetime employment cut

= 1, During era '______years'

= 0, otherwise

We believe our model is a good predictor and forecaster of population growth rate over the near future. Our regression results exhibit a high R^2 and F-statistic which is a good indication that our independent variables help us understand the trends in Japan's population growth rate. **SOMETHING ABOUT TESTING FOR COLINEARITY IN VARIABLES HERE?** In practical use, if our model shows that a specific variable is statistically significant in relation to the Japanese population growth rate, we will have a better understanding of what is potentially causing this downward population trend in Japan and action can be taken to improve and fix this negative spiral. As we move forward with our paper, we will look towards the literature and articles that sparked our interest in this subject, as well as provided us the groundwork necessary to conduct such a regression. Once completed with our introductory hypothesis and preliminary explanation, we will begin to explain our model, in particular the reason we selected each variable and the impact each variable has on the Japanese population growth rate. Following the model, we will analyze the data obtained for each respective variable as well as our findings from the regression analysis. Lastly, after displaying our model and interpreting its results, we will provide an ample conclusion including an appendix with tables and our final reactions and findings.

Literature Review:

What initially peaked our interest in the growth rate problem that Japan is experiencing was a documentary about the Japanese love industry[4]. The documentary began with some astounding facts. For example, Japan has both the most people over 65 and the least people under 15 in the world. Furthermore, by the end of the century Japan's population was expected to be half of what it is now. The conclusion of the documentary chalked up the growth rate problem to be a product of the Japanese culture's commodification of everything. It claims the Japanese have successfully "distilled each part of a human relationship into a buyable package" which was leading to less marriages, people being single for longer and ultimately, less children being born.

While the sex industry may be a major factor in less children being born, many fear that escapism, an increasing trend in Japan, could be a contributing factor to the problem. One particular concern is an extreme form of escapism called hikikomori[5]. Essentially, what this is is the locking in of oneself into his/her room for an extended amount of time (many cases have been as long as fifteen years), to watch tv, play video games, or simply put - escape life. This is of great concern because estimates have this problem as one which affects somewhere between 500,000 to 2 million people[6]. While

---

[4] http://www.vice.com/video/the-japanese-love-industry
[5] http://theweek.com/articles/453219/everything-need-know-about-japans-population-crisis
[6] http://www.wsj.com/articles/the-fight-to-save-japans-young-shut-ins-1422292138

both the sex industry and the hikikomori trend are likely having an impact on the declining growth rate, there are not a sufficient amount of data to test this.

Although we cannot account for some of the variables stated above, we have found relevant data to growth rate that we will include in our regression. For example The Economist states that “[t]he chief reason for the dearth of births is the decline of marriage[7].” A factor that is attributed to the decline in marriage, and the lower birthrate is the dissolution of the lifetime employment mentality. In the past people knew their financial futures were secure, now however, the thought of not having lifetime employment has led to financial uncertainty and thereby people are less inclined to commit to starting a family. Lastly, a major factor women in the workforce. While in many countries when women's’ participation in the work force increases, birth rates increase, in Japan it is the opposite. Workforce participation has made women more independent. This is undoubtedly good, but the culture of Japan’s work force does not encourage women to have children. Japan’s “old-fashion corporate culture” coupled with a “dire shortage of childcare” has discouraged women from having children and has led them to concentrate on their jobs.

(maybe do something with out of wedlock births vs other countries))

(http://stevenscomputersandsociety.blogspot.com/2013/12/the-japanese-love-industry-and.html)

[7] http://www.economist.com/blogs/economist-explains/2014/07/economist-explains-16

http://www.forbes.com/sites/joelkotkin/2014/12/15/can-abe-tackle-the-real-reason-for-japans-decline/#5512f3cf7d78 - Many young Japanese are not only eschewing marriage

http://theweek.com/articles/453219/everything-need-know-about-japans-population-crisis

http://www.bbc.com/news/world-asia-30653825

Model:

The model that we used for our regression is a multivariate linear regression model. Our dependent variable, the population growth rate in Japan, although limited just to one country, requires many independent variables to forecast. In order to decide which specific independent variables to include, we need to discuss certain overall issues that can be a factoring into the declining and currently negative population growth rate.

The first issue that we found in our research that could be factoring into the dwindling population growth rate was financial stability. Aside from unexpected pregnancies, most rational adults are only prepared to have children when they feel financially prepared to support one. A child is the number one financial strain that adults can put on themselves, therefore we thought it would be worthwhile to investigate some specific statistics that have created financial instability and thus had an effect on the population growth rate.

The second issue that we found was that a rise in female empowerment can have a negative effect on the population growth rate. When we look at our domestic history, we can see that there has definitely been an expansion in women's roles in the workplace since the early 20th century. It used to be common for women to be "professional mothers" and housewives but now it is more common for women to take on professional roles in the workplace spending more time away from their children than they would unemployed. Although this is definitely a positive development in terms of gender equality, this has definitely had a negative effect on the population growth rate in

America. The reason being that with no one at home, there is no one to take care of a child and provide for their dependencies. This idea of the rise in female employment having a negative effect on the domestic population growth rate is especially prevalent in Japan where it is frowned upon to have kids if you're a full-time employee. Just like America, Japan has experienced a rise in gender equality in the workplace and due to the fact that there is a negative stigma to juggling a kid and a job, we think this may have a considerable negative impact on our dependant variable.

There's a very big issue going on in Japan and that is the increasing desire for Japanese citizens to remain single. One reason for this is that there are many men in Japan who are satisfied with receiving affection through a virtual girlfriend rather than having a living girlfriend. Another reason is the 'Japanese Love Industry'. There are host and hostess clubs in Tokyo where men and women go to pay the host and hostesses money to sit with them, have drinks, and engage them in nonsexual conversation. These clubs are so popular that some host and hostesses make up to $800,000 a year just for sitting and talking with people! There are also cuddle cafes, where men can go and pay a young girl to do various "relationship-type" things for them. On the "menus" at these cafes, you can pay a girl to cuddle, stare into your eyes, "spoon", or even pick out your ear wax. As extreme as these things are, they seem to have simulated all aspects of a relationship for a price. For many Japanese citizens, there seems to be no need to deal with the "stresses" of being in a relationship and without relationships and marriage there are going to be significantly less children being born. We feel that this can serve as an important factor in our regression model.

Independent Variable D_1: Lifetime Employment Cuts

- Expected sign - (-)

The concept of lifetime employment has existed in Japan for over a century. With lifetime employment, an employee can be confident that they will almost definitely have their job until the day the retire. Lifetime employment became even more common in larger corporations after World War II due to the rising concern of job stability following mass dismissals in the 1940's and 1950's. Now, an employee entering the workforce was assured that even if a depression were to occur, their employer would do their best not to lay off any of their workers. In return, the employees will not quit their jobs until they are prepared to retire.[8] This became customary for years in Japan until 1994 when corporations started to cut down on lifetime employment. What once was customary was now a rarity. In the U.S., we never had lifetime employment therefore our job security has always been the same and no one feels more financially unstable than usual. In Japan, someone entering the job market who once assumed they were going to have an almost guaranteed job for life (if they were able to find a job in the first place) is now not too sure about that. This creates more financial instability for that person and, as we spoke about above, with less financial stability, a person is less likely to get married and have children. This is why we factored lifetime employment cuts as a dummy variable in our regression using 1995 as the year of structural change. Lifetime employment cuts lead to financial instability which leads to less children being born which has a negative impact on population growth rate.

---

[8] Source: http://www.bls.gov/opub/mlr/1984/08/rpt4full.pdf

Independent Variable X_1: Percentage of Employed Females Earning Wages

- Expected sign - (-)

As we mentioned above, the rise in female empowerment could be factoring into our dependant variable having been downward sloping over the years. We also mentioned that there has been a strong rise in female empowerment taking form in the Japanese workplace over the years. In order to include this in our regression, we decided that the percentage of employed females earning wages would serve as a good independant variable to fit our model. By taking the female population and seeing what percentage of that population is paid and employed, we can account for changes in level of female empowerment in Japan. Because female empowerment has been increasing in Japan over the years, we expect that there will be increases in the percentage of females in the workforce. As more females enter the workplace, because of the negative stigma of juggling a job and a kid, we expect that there will be less children being born and thus there will be a negative impact on the population growth rate.

Independent Variable X_2: Ratio of Female to Male Labor Force Participation

- Expected sign - (-)

Due to the importance of female empowerment in our regression model, we found it necessary to include a second independent variable that measures female involvement in the workforce. The Ratio of Female to Male Labor Force Participation compares female involvement in the labor force to male involvement in the labor force. The importance of this is that we are looking at the labor force (those able and looking to work) as opposed to those already employed. An increase in the amount of women actively looking for work, which is reflected in this ratio, will definitely have a negative impact on the population growth rate as well. Looking for a job can be time consuming and if you're hopeful, you may not want to commit to raising a child. Hence why we assume that an increase in this ratio will have a negative impact on the population growth rate.

Independent Variable X_3: Marriages As a Percentage of Population

- Expected sign - (+)

In Japan, the percentage of children born out of wedlock is around 2%. This percentage is extremely low when compared to the percentage of children born out of wedlock in the U.S., which is 40.2%[9]. Theses statistics bring us to our next independent variable, marriages as a percentage of population. Since 98% of children born are to married couples, this variable will be a valuable in our regression model. These statistics lead us to assume that when the number of marriages as a percentage of population in Japan decrease, there will be a negative effect on the population growth rate.

Independent Variable X_4: Death Rates (/population)

- Expected sign - (-)

Our last independent variable is death rates in Japan. Death rates' effect on population growth rate are pretty self-explanatory. The two factors that make up the change in the Japanese population are the number of births and the number of deaths

[9] Source: http://www.cdc.gov/

that occur domestically. An increase in deaths leads to a decrease in the population, holding births constant. When dealing with the growth rates, an increase in death rates leads to a decrease in the population growth rates.

Data:

As important as our model is too the regression analysis, there would be no where to go without substantive data. This section focuses on outlining and discussing the data for each variable we are testing in our model. For each quantitative independent variable we will provide a scatterplot displaying the relationship between the respective variable and the dependent variable, population growth rates. To begin, below is a table with the mean and standard deviation for each respective independent variable:

Table 1:

Variable	Mean	Standard Deviation	Sources
Y: Population Growth Rate (%)	0.2775%	0.2666%	*http://data.worldbank.org/country/japan*
X_1: Salaried Female Workers (%)	77.7617%	8.3794%	*http://data.worldbank.org/country/japan*
X_2: Females/Males in Labor Force[8]	64.6164%[10]	2.2816%	*http://data.worldbank.org/country/japan*
X_3: Marriages as % of Population	0.5949%	0.0397%	*http://www.mhlw.go.jp/toukei/list/dl/81-1a2.pdf*

[10] Ratio in a percentage form is interpreted as 64.6164 females in the workforce per 100 males.

X_4: Deaths as % of Population	0.76%	0.12%	*http://data.worldbank.org/country/japan*
D_5: Lifetime Employment Cuts	N/A (dummy)	N/A (dummy)	*http://www.japantimes.co.jp*

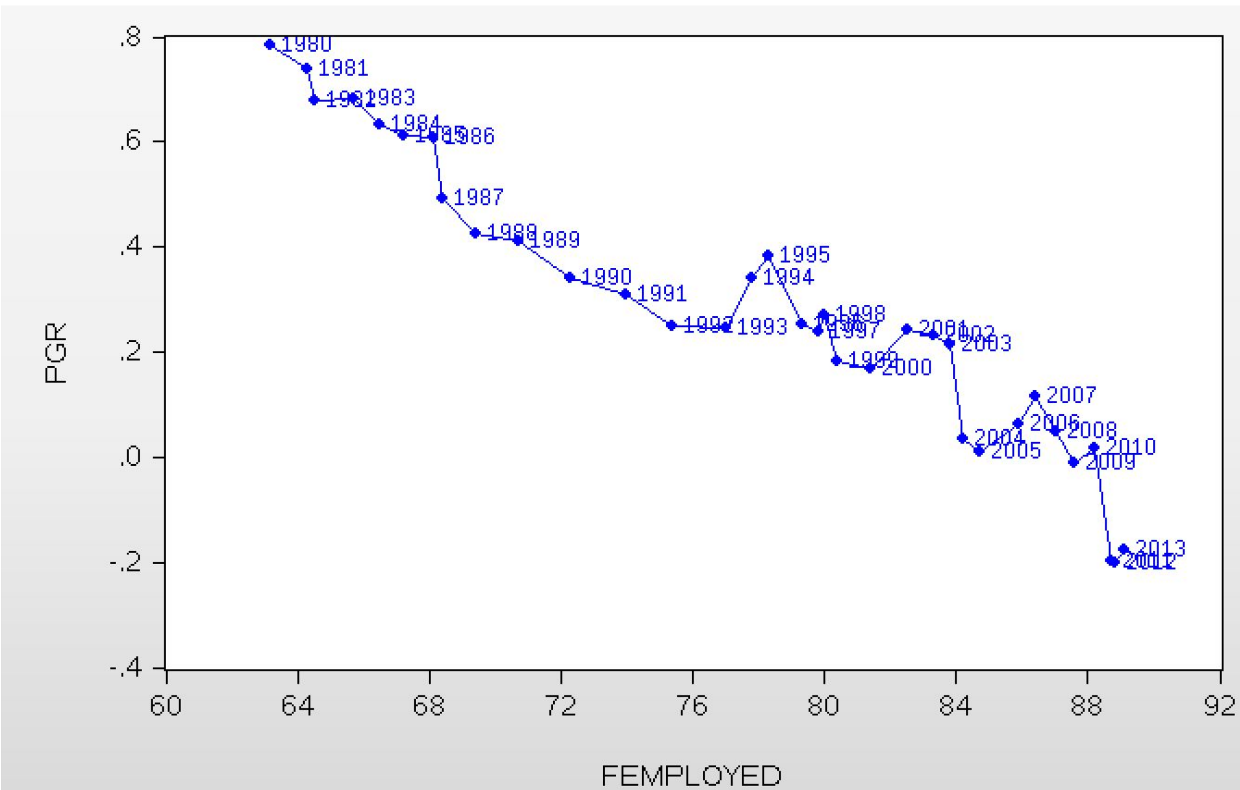
The data that we used is all time-series data from the years 1980-2013, yielding n=33 observations per variable. The data we obtained for our quantitative variables (both dependent and independent) came from various sources. *Data.Worldbank for Japan,* provided the data for our dependent variable, population growth rates in Japan, as well as three independent variables, X_1 salaried female workers, X_2 the female-to-male ratio in the labor force, and X_4 deaths as a percentage of population. We found the data for our last quantitative variable X_3 from the Japanese *Ministry of Health, Labour and Welfare.* The values we show for the qualitative dummy variable was obtained from an article published by the *Japan Times*. All of our sources are provided in the table including the mean and standard deviation for each respective variable. The table on the next page labeled *Table 2* lists the data for each of the five variables we are including in our regression function. To obtain a consistent form we uploaded all the data we obtained to EViews to better organize the data for regression analysis. After introducing the data in *Table 2*, we will further discuss the sources for the data of each variable as well as the methodologies we took in computing the data and preparing it for regression analysis

Table 2:

LEAVE THIS PAGE FOR THE DATA ON THE WORD DOCUMENT THAT U BUILT OF ABE'S INCLUDING ALL THE INFORMATION

The data for our dependent variable, population growth rates in Japan was directly imported from the *Data.Worldbank for Japan* sourced above. We exported the data from Excel into EViews and the values are measured by percentage increase and decrease as shown in the Table. The Japanese population has shown mean growth rate of 0.2775% and standard deviation of 0.2666% over the range of years we are sampling.

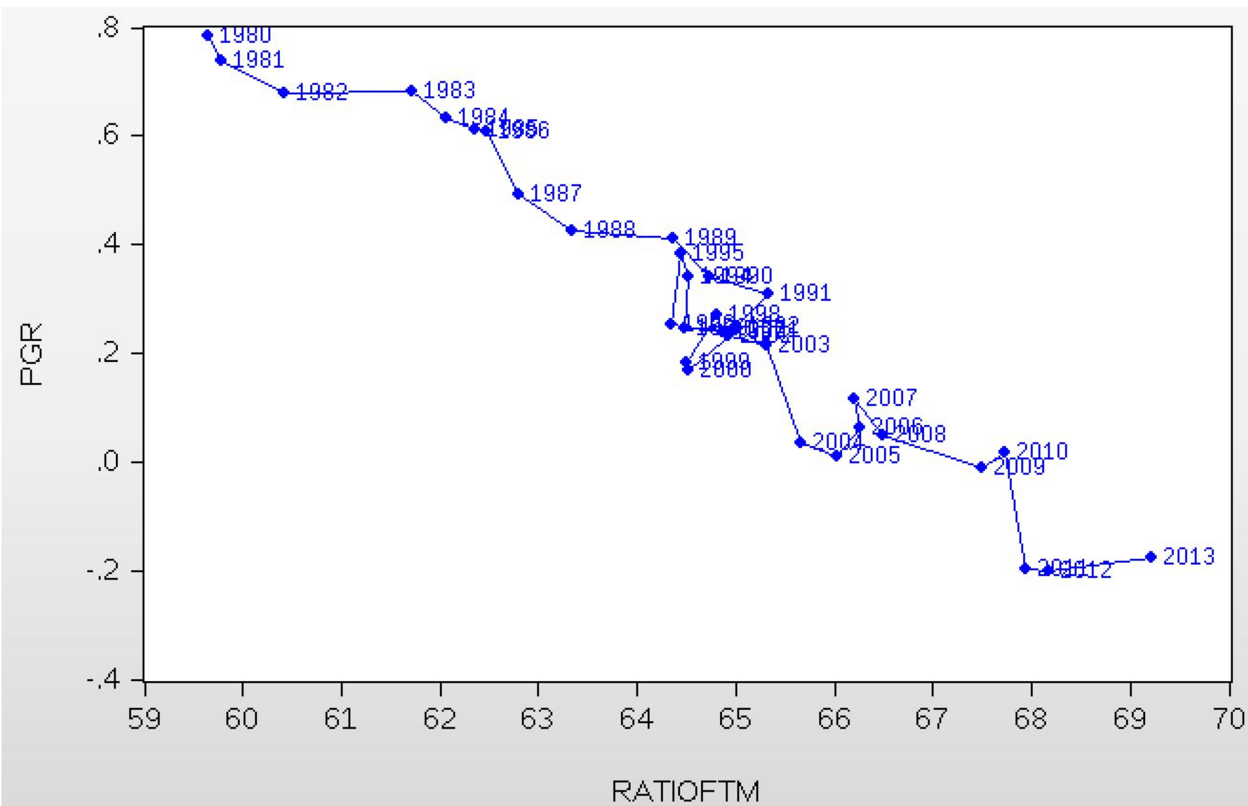
Graph 1:



Graph 1 above details the relationship between Y and X_1, our population growth rate and the percentage of salaried females in Japan. There appears to be a negative relationship between the two, which displays how, as the percentage of women earning

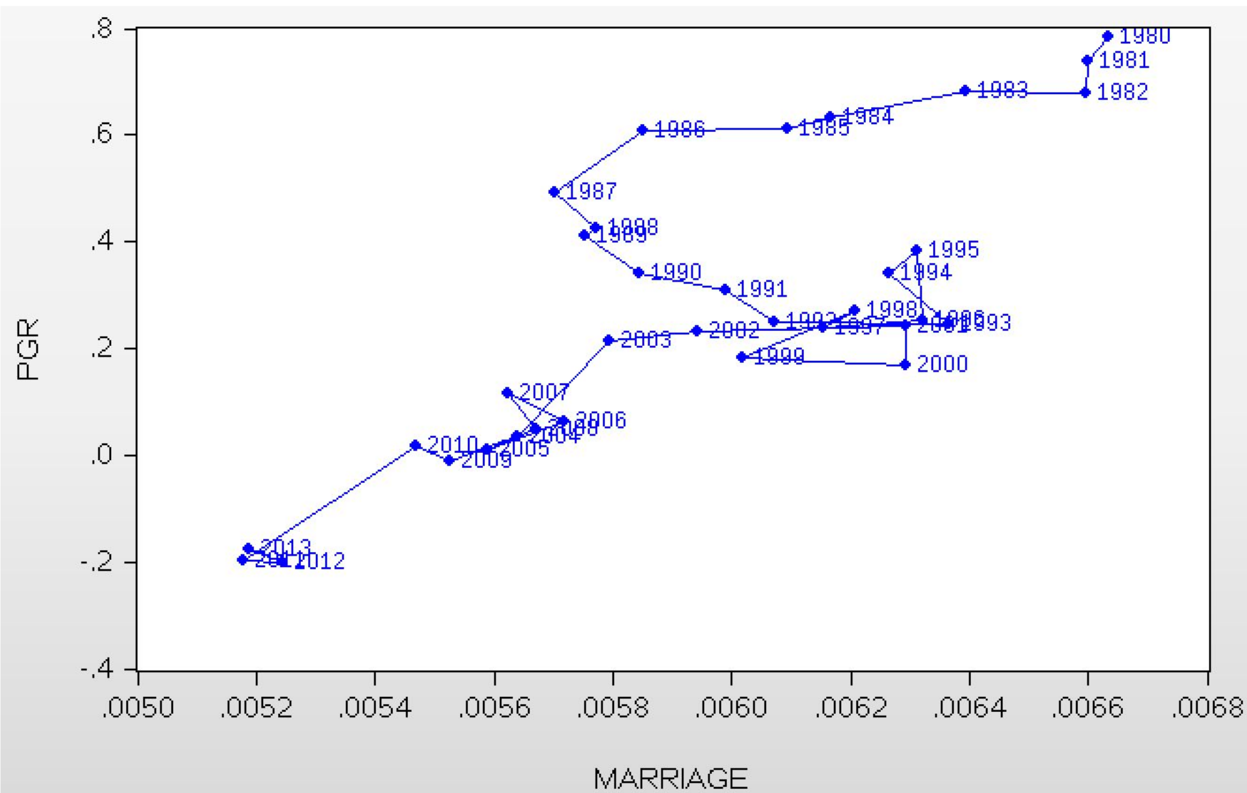
wages increases the population growth rate appears to decrease. Overtime the percentage of females employed and earning wages has increased while the Japanese population has decreased, sometimes with a negative growth rate (i.e. actual decrease in Japan’s population). This can be a result of various reasons, but for purposes of our regression, we find it safe to assume their is a negative correlation between the two variables.

Graph 2:



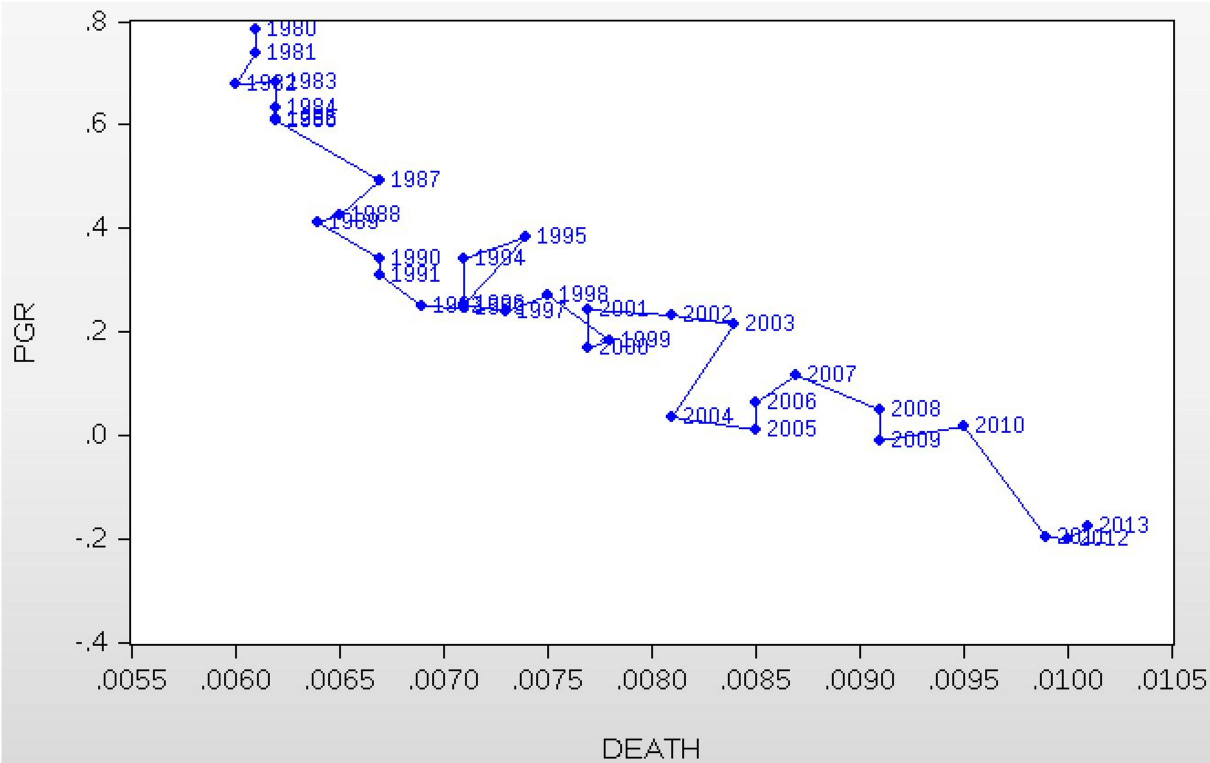
Our second graph looks at the relationship between our dependent variable and the ratio of females to males in the labor force. Similar to X_1 in *Graph 1*, our X_2 exhibits the same negative relationship where an increase in the ratio of females to males in the labor force, contributes to a decreasing and sometimes negative growth rate in the Japanese population. The graph labels the different years for which the data corresponds too, so over the progression of our data set, as more females join the labor force in proportion to men, the growth rate in Japan's population has been decreasing. More recently, the population growth rate hasn't only been decreasing, but has become negative, where the actual population has been decreasing in number. We will continue to look at the remaining two quantitative variables X_3 and X_4, to better determine the relationships these variables have with the dependent variable

Graph 3:



Graph 3 looks at our relationship between the dependent variable and our third independent variable X_3. X_3 is represented by the marriages as a percentage of Japan's population. In other words, the independent variable is a measure of how many marriages occur annually in Japan, as a percentage of that year's population. Unlike the previous two independent variables, *Graph 3* exhibits a positive relationship between marriage rates and population growth rates. This means that as more marriages occur annually as a percentage of population, the population growth will increase as well. However, our data points in *Graph 3* show that over time the marriage rate has decreased and therefore the population growth rate has decreased as well. *Graph 3* highlights the importance of not focusing on just the relationship between the dependent and independent variables, but rather looking towards how the variables respect over time. Our model is looking to identify how each of our independent variables can help explain why there has been a decreasing population growth rate in Japan.

Graph 4:



Our last quantitative independent variable is X_4, a measure of the death rates as a percentage of population observed in Japan over our experimental range of years (n=33). The relationship between Y and X_4 is very similar to those observed in *Graph 1* & 2. *Graph 4* shows us that since 1980 the death rates in Japan have increased, which appears to have a negative relationship with the population growth rate in Japan. This seems to suggest that holding immigration constant, there are more people dying in Japan than children being born over the 33 years our data is analyzing. Our independent variables, X_1, X_2, X_3, and X_4 each have unique and important relationships with the empirical population growth rate in Japan. The data section of our experiment is coming to display how each of these independent variables relates to our Y variable on hand.

Before proceeding with our regression it is important that we point out a few potential points of error in our data and assumptions. In *The Model* section of our paper, and now in *The Data* section when we showed the relationship between each variable over the time of our analysis we made assumptions on what we expect the "sign" of the coefficient to be for each respective independent variable. For variables X_1, X_2, and X_4 we showed and assume a negative coefficient. For X_3 we believe there is a positive relationship between the dependent variable and marriage rates taken as a percentage of Japanese population. There is a possibility that this positive or negative assumption could be wrong when we conduct our regression output. Chart 1 below identifies the expectation we have for each coefficient in our regression. In addition, multicollinearity is a problem that we must identify prior to doing any regression output and analysis. There is a chance that two or more of our independent variables relate with each other in a positive or negative fashion. Any correlation between our independent variables must be identified for the purpose of a correct regression analysis, and we will test for potential multicollinearity in the next section of our paper. Multicollinearity is a problem for a number of reasons. For one, it can increase the variance in our betas, i.e. the coefficients of our variables. Secondly, it makes assessing the statistical significance of the model difficult and may even change the signs of the coefficients from one regression to the next. Lastly, it is important that we point out the need for a test of Heteroscedasticity in our residuals prior to proceeding with our regression. The problem with heteroscedastic in the residuals is that when performing a regression using OLS it will treat all the residuals with the same weight (as it is an unbiased estimator), however the residuals with greater disturbances will have a larger effect and in effect render the coefficients inefficient.

Table 3:

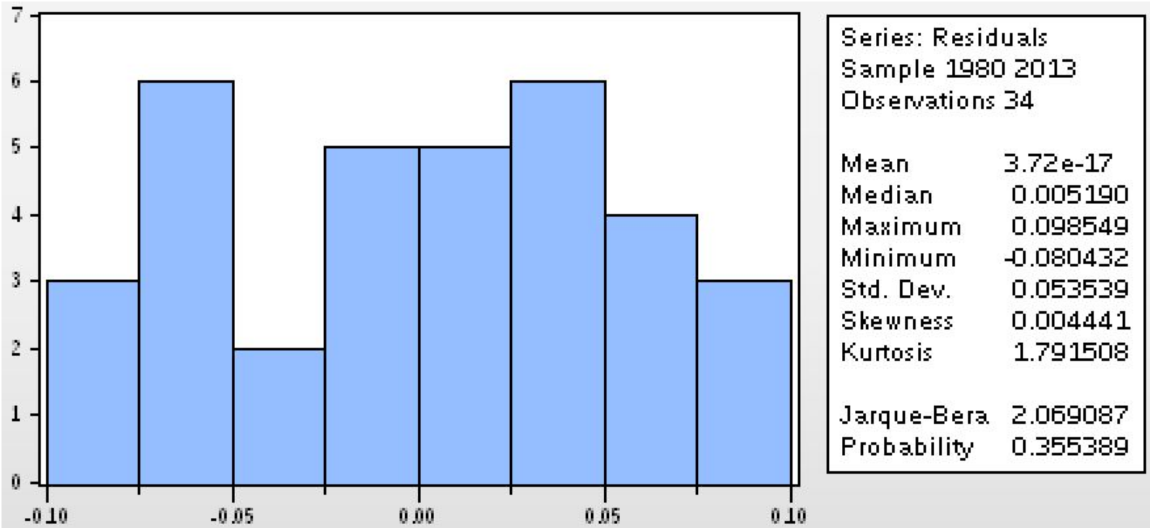
Variable	Expected Sign	Actual Sign
X_1: Salaried Female Workers (%)	Negative "-"	Negative "-"
X_2: Females/Males in Labor Force (%)	Negative "-"	Negative "-"
X_3: Marriages as % of Population	Positive "+"	Positive "+"
X_4: Deaths as % of Population	Negative "-"	Negative "-"
D_5: Lifetime Employment Cuts[11]	**Negative "-"**	**Positive "+"**

[11] **Bold -** symbolizes different expected and actual sign

Regression Results:

As we can see from the *Output 1* table in the Appendix section, all of our expected signs were correct with the exception lifetime employment cuts (LEC). In order to determine whether the coefficient of each independent variable is statistically significant, the absolute value of the t-statistic needs to be greater than 2 and the p-value needs to be close to 0. Our X_1 variable, Salaried Female Workers (%) (FEMPLOYED), and X_2 variable, Females/Males in Labor Force (%) (RATIOFTM) have significant t-statistics and p-values. Our X_3 variable, Marriages as % of Population (MARRIAGE), X_4 variable, Deaths as % of Population, and D_5 dummy variable, Lifetime Employment Cuts, have insignificant t-statistics and p-values. The issue that D_5 has an unexpected sign is minimized because we cannot tell if this sign is accurate due to its insignificant t-statistic and p-value. Additionally, the independent variables of the only two coefficients that are statistically significant fall into the female empowerment category. Unfortunately, the coefficient's of the independent variables of the financial stability, marriage, and death categories are statistically insignificant. *Output 1* helps us determine how each variable regresses together with each other and the dependent variable, for our overall regression analysis.

Graph 5:



In order to test whether our residuals are normally distributed we analyze the histogram above, labeled *Graph 5*. At first glance we can see that our residuals certainly do not qualify as perfectly normally distributed but we can definitely see that it looks normally distributed. In order to test this we look at skewness and kurtosis. If there were a perfect normal distribution, the skewness would be 0 and kurtosis would be 3. In our histogram we find that skewness is 0.004441 and kurtosis is 1.791508, which is close enough to their optimal values to conclude that our residuals are indeed normally distributed. Our R^2 value of .959691 is extremely high which is great in testing our regression's goodness of fit. Additionally, our F-statistic of 133.3265, with a p-value of 0, is extremely significant. So far, the results of our regression model seem to support our model's accuracy.

To better show the difference between our group of statistically significant independent variables (X_1 and X_2) and the remaining statistically insignificant variables (X_3, X_4 and D_5) we provided *Output 2 and Output 3* respectively (Appendix). *Output 2* regresses the female empowerment group that shows to be statistically significant from *Output 1*. As you can see on the table in *Output 2*, the t-statistic for both FEMPLOYED (X_1) and RATIOFTM (X_2) are greater than 2 and have a p-value that is low enough for us to again conclude their relevant significance to the regression. The reason we split up the group of significant variables and the insignificant ones was to show to difference in R^2. For *Output 2* we have a R^2 value of 0.956136 which is minimally smaller than the R^2 in *Output 1*, 0.959691. We know that as you incrementally add independent variables to a regression the R^2 will increase as well, because of the larger range of potential explanation the group of explanatory variables has on the regressand. However, in our case, the addition of the variables shown in *Output 3* have a minimal impact on the overall R^2. From *Output 3*, MARRIAGE (X_3), DEATH (X_4) and LEC (D_5), the variables that weren't significant to the overall regression in *Output 1* the R^2 value is 0.878959. Therefore, by differentiating the female empowerment group of variables in *Output 2* (X_1 and X_2), to the death rates, marriage rates, and lifetime employment rates (X_3, X_4 and D_5) our split regressions exhibit what it means to say that certain variables are statistically significant and certain variable are not statistically significant. *Output 3's* variables do not add much, if any value to our R^2 level, which is a clear violation of the benefit of adding new variables to a model. We are not saying that there is no additional explanation from adding the second group of variables in *Output 3*, we are just saying there is no statistical significance from doing so.

To test for multicollinearity we observed the variance inflation factor (VIF), as seen in appendix table 1. According Gujarati a VIF of 10 or higher may indicate multicollinearity. Three variables possess this characteristic; FEMPLOYED, RATIOFTM and DEATH. We therefore do further testing, namely create a correlation matrix (see appendix , to see if multicollinearity may be legitimately occurring here, and if so, between which variables.

	Correlation				
	FEMPLOYED	RATIOFTM	MARRIAGE	DEATH	LEC
FEMPLOYED	1.000000	0.925827	-0.661641	0.940517	0.874868
RATIOFTM	0.925827	1.000000	-0.807642	0.901845	0.691127
MARRIAGE	-0.661641	-0.807642	1.000000	-0.771692	-0.432835
DEATH	0.940517	0.901845	-0.771692	1.000000	0.802022
LEC	0.874868	0.691127	-0.432835	0.802022	1.000000

We observe here that, with the exception of LEC and MARRIAGE, all of our variables have a high correlation number and therefore likely exhibit multicollinearity. In order to correct for multicollinearity we remove FEMPLOYED as it has the highest VIF of 42.48916 (more than double the VIF of the next closet variable) as well as the highest correlation with each variable.

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.820405	7326.557	NA
RATIOFTM	0.000143	5356.071	6.473693
MARRIAGE	2830.764	898.6572	3.869746
DEATH	734.1937	386.1640	9.792788
LEC	0.001690	8.432884	3.720390

After removing FEMPLOYED we can see significant improvements in our model. What is most important is that the VIF for all our variables is now under ten. This indicates that we have successfully accounted for and removed multicollinearity. When analyzing our regression without FEMPLOYED (see appendix) we can see an increase in significance in three of our four variables. RATIOFTM's t-stat increases in significance, as it goes from -2.23 to -6.75, and its p-value decreases, as it goes from .04 to .00. LEC's t-stat increases in significance, as it goes from -.61 to -1.22, and its p-stat decreases, as it goes from .55 to .23. Death's t-stat not only increases in significance, as it goes from -.56 to -2.02, but it also increases to become significant in our model. Death's p-value also decreases, as it goes from .58 to .05. Marriage's t-stat decreases, as it goes from to -.776 to -.37, and it's p-value increase from .44 to .71. It is also important to note that the signs of the coefficients of marriage and LEC went from positive to negative. While marriage's expected sign and actual sign were consistent in our old regression and not in our new one, LEC's expected sign and actual sign were at odds with each other in our old regression, and now are consistent in our new one. However, this is not of particular importance as both marriage and LEC are not statistically significant . Lastly, although the R-squared decreased in our new regression, it only did so by .00641.

Dependent Variable: PGR
Method: Least Squares
Date: 05/06/16 Time: 15:10
Sample: 1980 2013
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.064452	0.905762	6.695415	0.0000
RATIOFTM	-0.080889	0.011978	-6.753199	0.0000
MARRIAGE	-19.87915	53.20492	-0.373634	0.7114
DEATH	-54.60735	27.09601	-2.015328	0.0532
LEC	-0.050251	0.041107	-1.222442	0.2314

R-squared	0.952950	Mean dependent var	0.277591
Adjusted R-squared	0.946460	S.D. dependent var	0.266665
S.E. of regression	0.061703	Akaike info criterion	-2.597928
Sum squared resid	0.110409	Schwarz criterion	-2.373463
Log likelihood	49.16478	Hannan-Quinn criter.	-2.521379
F-statistic	146.8414	Durbin-Watson stat	1.740214
Prob(F-statistic)	0.000000		

APPENDIX THIS ^

In order to test for heteroscedasticity we used the Glejser test which regresses the absolute value of the error term against the independent variable. A high t-stat, or a low p-value, for the estimated coefficient of our independent variables would indicate the presence of heteroskedasticity.

Heteroskedasticity Test: Glejser

F-statistic	2.199101	Prob. F(5,28)	0.0827
Obs*R-squared	9.586930	Prob. Chi-Square(5)	0.0878
Scaled explained SS	5.485453	Prob. Chi-Square(5)	0.3595

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 05/06/16 Time: 16:07
Sample: 1980 2013
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.860793	0.496525	1.733635	0.0940
FEMPLOYED	0.006604	0.003351	1.970792	0.0587
RATIOFTM	-0.017639	0.008558	-2.061150	0.0487
MARRIAGE	-25.04993	24.91139	-1.005561	0.3232
LEC	-0.021348	0.023280	-0.917020	0.3670
DEATH	-3.699693	13.12670	-0.281845	0.7801

Our results show that we may have a problem with heteroskedasticity as Femployed and Ratioftm have a high t-stats (1.97 and -2.06 respectively). We therefore attempted to fix this problem by removing Ratioftm which had the highest t stat. After doing so the t-stats of all our variables dropped with Femployed being highest at 1.66 (SEE APPENDIX). Before concluding that we had successfully adjusted our model to become homoscedastic, we decided to run the Glejser test with Femployed removed. Our motivation for doing so was the results obtained from our regression that we previously ran without Femployment; results that indicated more significance in our model without the Femployment variable. After running the Glejser test without Femployemnt, we found that the t-stats of each variable dropped significantly with the highest variable's t-stat now being Death at .51 (SEE APPENDIX). With these results the likelihood of our model having any heteroskedasticity in its residuals is minimal. With the results of our multicollinearity and heteroskedasticity tests we concluded that our model is significantly improved with the omission of Femployment.

Heteroskedasticity Test: Glejser

F-statistic	0.937463	Prob. F(4,29)	0.4562
Obs*R-squared	3.892994	Prob. Chi-Square(4)	0.4207
Scaled explained SS	2.635399	Prob. Chi-Square(4)	0.6206

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 05/06/16 Time: 16:09
Sample: 1980 2013
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.041624	0.457732	0.090936	0.9282
RATIOFTM	-0.001113	0.006053	-0.183827	0.8554
MARRIAGE	3.207431	26.88739	0.119291	0.9059
LEC	0.010521	0.020774	0.506441	0.6164
DEATH	7.046158	13.69311	0.514577	0.6107

R-squared	0.114500	Mean dependent var	0.048071
Adjusted R-squared	-0.007638	S.D. dependent var	0.031063
S.E. of regression	0.031182	Akaike info criterion	-3.962915
Sum squared resid	0.028197	Schwarz criterion	-3.738451
Log likelihood	72.36956	Hannan-Quinn criter.	-3.886366
F-statistic	0.937463	Durbin-Watson stat	1.838619
Prob(F-statistic)	0.456202		

Heteroskedasticity Test: Glejser

F-statistic	2.569283	Prob. F(4,29)	0.0589
Obs*R-squared	8.896334	Prob. Chi-Square(4)	0.0637
Scaled explained SS	5.075007	Prob. Chi-Square(4)	0.2797

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 05/06/16 Time: 16:08
Sample: 1980 2013
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.189038	0.199531	-0.947411	0.3513
FEMPLOYED	0.003312	0.001996	1.659093	0.1079
MARRIAGE	3.775624	20.76067	0.181864	0.8570
LEC	-0.018265	0.020678	-0.883319	0.3843
DEATH	-4.084124	13.53947	-0.301646	0.7651

R-squared	0.261657	Mean dependent var	0.049812
Adjusted R-squared	0.159816	S.D. dependent var	0.028649
S.E. of regression	0.026260	Akaike info criterion	-4.306482
Sum squared resid	0.019998	Schwarz criterion	-4.082017
Log likelihood	78.21019	Hannan-Quinn criter.	-4.229933
F-statistic	2.569283	Durbin-Watson stat	2.514822
Prob(F-statistic)	0.058945		

Summary and Conclusions:

As mentioned at the start of our paper, we found our topic, the Japanese population growth rate to be equally exciting and important. There are many reasons why an entire country's population could be experiencing such a prolific and persistent decrease. Yet, after careful consideration we chose our group of independent variables for good reason. Female empowerment, deaths, marriages, and financial stability all contribute uniquely to Japan's recent empirical population decline. After compiling and experimenting with the relevant data we have come to the following realizations and conclusions.

Our complete regression output (incorporating all quantitative and qualitative variables) gave us an R^2 value of 0.959691. About 96% of our model was able to explain the variation in population growth rates experienced in Japan over our n=33 years of observations. Additionally, our F-statistic of 133.3265 and its p-value of virtually zero led us to believe that our model was highly accurate. However, it is important to note that not all of our variables played an equal role in this explanation of the variation of Japanese population growth rates. Two of our variables, the percentage of salaried female workers and the ratio of females to male in the labor force ($X_1$and X_2 respectively), produced statistically significant results. Unfortunately, these results show that the only statistically significant independent variables in our model come from what we categorize as female empowerment. The other three variables, marriage rates as a percentage of population, death rates as a percentage of population and lifetime employment cuts (X_3, X_4 and D_5 respectively), proved to be statistically insignificant which was initially upsetting. We were hoping in our regression experiment, to find more significant determinants for the downward population trend evident in Japan.

To better understand our independent variables we grouped the statistically significant variables and the statistically insignificant variables into their own regressions (*Output 2* and *Output 3* respectively) on population growth rate. What we found was that the group of insignificant variables from *Output 3* contributed little to the overall R^2 value from *Output 1*. This helped us solidify our understanding of the significance and contribution each group of variables incrementally had on the regressand.

Next we tested for multicollinearity: In order to test for this, we created a collinearity matrix and found that there was definitely a high level of correlation between variables. To determine whether this multicollinearity was significant to our model we created a VIF chart and found that most VIF's were above 10 and therefore significant. To fix this, we decided to remove the variable (FEMPLOYED, X_1) that had the highest VIF value and run a new VIF test. Doing so, we found that the VIF's for each variable were now below 10, thus showing no issue with multicollinearity.

Running a heteroscedasticity test, we found that our our female empowerment variables were significantly heteroscedastic. In removing FEMPLOYED (X_1) from our model, we fixed our multicollinearity issue, so we decided to see if it would help our heteroscedasticity issue as well. To our delight, removing FEMPLOYED almost guaranteed our model to be homoscedastic.

Although our results seemed convincing at first and seemed to rely heavily on the female empowerment variables, we found in our various tests that there were substantial issues with multicollinearity and heteroscedasticity. When we removed FEMPLOYED from our model, not only did death rates; coefficient become significant, we also fixed our issues with multicollinearity and heteroscedasticity. It seemed counterproductive to remove an independent variable with a statistically significant coefficient, but when we did we found that our problems were solved and a statistically significant coefficient for death rates took its place.

Appendix:

Output 1: Original Regression

Dependent Variable: PGR
Method: Least Squares
Date: 05/06/16 Time: 13:34
Sample: 1980 2013
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.344206	1.166170	3.725192	0.0009
DEATH	-17.18804	30.83020	-0.557507	0.5816
FEMPLOYED	-0.017031	0.007871	-2.163899	0.0392
LEC	0.033278	0.054676	0.608648	0.5477
MARRIAGE	45.44871	58.50846	0.776789	0.4438
RATIOFTM	-0.044896	0.020099	-2.233674	0.0337

R-squared	0.959691	Mean dependent var	0.277591
Adjusted R-squared	0.952493	S.D. dependent var	0.266665
S.E. of regression	0.058123	Akaike info criterion	-2.693738
Sum squared resid	0.094591	Schwarz criterion	-2.424381
Log likelihood	51.79355	Hannan-Quinn criter.	-2.601880
F-statistic	133.3265	Durbin-Watson stat	1.630780
Prob(F-statistic)	0.000000		

Output 2: Regression with Respect to X_1 and X_2

Dependent Variable: PGR
Method: Least Squares
Date: 05/06/16 Time: 13:33
Sample: 1980 2013
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.516877	0.531912	10.37179	0.0000
FEMPLOYED	-0.014321	0.003167	-4.521379	0.0001
RATIOFTM	-0.063849	0.011632	-5.488942	0.0000

R-squared	0.956136	Mean dependent var	0.277591
Adjusted R-squared	0.953306	S.D. dependent var	0.266665
S.E. of regression	0.057623	Akaike info criterion	-2.785695
Sum squared resid	0.102932	Schwarz criterion	-2.651016
Log likelihood	50.35681	Hannan-Quinn criter.	-2.739765
F-statistic	337.8659	Durbin-Watson stat	1.665272
Prob(F-statistic)	0.000000		

Output 3: Regression with Respect to X_3, X_4 and D_5

Dependent Variable: PGR
Method: Least Squares
Date: 05/06/16 Time: 13:31
Sample: 1980 2013
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.682560	0.678818	1.005512	0.3227
DEATH	-144.8829	37.16768	-3.898090	0.0005
LEC	-0.076194	0.064541	-1.180552	0.2471
MARRIAGE	123.5945	76.92353	1.606720	0.1186

R-squared	0.878959	Mean dependent var	0.277591
Adjusted R-squared	0.866855	S.D. dependent var	0.266665
S.E. of regression	0.097304	Akaike info criterion	-1.711830
Sum squared resid	0.284040	Schwarz criterion	-1.532259
Log likelihood	33.10112	Hannan-Quinn criter.	-1.650591
F-statistic	72.61642	Durbin-Watson stat	0.708206
Prob(F-statistic)	0.000000		

Table 1 : VIF

Variance Inflation Factors
Date: 05/06/16 Time: 13:56
Sample: 1980 2013
Included observations: 34

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	1.359952	13687.12	NA
FEMPLOYED	6.19E-05	3812.516	42.48916
RATIOFTM	0.000404	16996.80	20.54342
MARRIAGE	3423.240	1224.742	5.273912
DEATH	950.5013	563.4177	14.28779
LEC	0.002989	16.81329	7.417626

Bibliography: