Julianne Zamora
12/15/08
495 Commodity Price Paper – Final

Commodity Prices and the Real Interest Rate

For this paper, I have studied the relationship between real commodity prices and the real interest rate. My initial research into the topic led me to the paper "The Effect of Monetary Policy on Real Commodity Prices" by Jeffrey Frankel. He argues that the real price of a commodity (relative to its long-run equilibrium) is inversely proportional to the real interest rate. This is because as the interest rate rises, the demand for storable commodities is reduced, or the supply is increased. Frankel provides three reasons for this. There is more of an incentive for extraction, there is less desire to carry inventories, and speculators shift out of commodity contracts and into treasury bills (Frankel, 2006). Basically, when interest rates rise there is more to gain by selling what you have today and putting your money in the bank, rather than keeping it invested in storable commodities. I investigated how well this relationship holds for individual food commodities. Specifically, since the relationship exists because of how inventories change, I wanted to see whether the relationship is stronger for more storable commodities (i.e. commodities which can be stored for longer periods). The idea is that the more storable a commodity is, there are more options for when to sell, and thus interest rates will have a stronger effect on the price. If the commodity is not as storable (for example, the food will rot), there will be little option as to when to sell, and interest rates should have little to no effect.

Most of the food commodities that I studied were agricultural, such as corn, oats, soybeans, tomatoes and wheat. I also studied two livestock commodities: live cattle, and live hogs. For a contrast with food commodities, I studied five metals: aluminum, platinum, lead, silver and copper. The idea behind studying the metals is that they will be much more storable over time than any and all of the food commodities.

I begin with a construction of the real interest rate. Using this rate, I confirmed the negative relationship Frankel found between real interest rates and an overall index of commodity prices. With this series, I estimated 3 different models. For each model I first examined the overall relationship between real interest rates and commodity prices, and then looked into storability as a factor. The first model consisted of detrended log real commodity prices regressed on the real interest rate, expected inflation, and world GDP. The second model examines the deviations of the individual commodity prices from overall commodity prices, and how they are related to real interest rates. For this model, I first regressed the log real commodity prices on the CRB Index and subsequently regressed the residuals on the real interest rate. The third model also uses these residuals, but includes the long term interest rate as well. Different levels of storability would imply that for more storable commodities the long term interest rate will be more important, and for less storable commodities the short term interest rate will matter more.

In order to estimate the three models, I needed to construct a credible series for the real interest rate. For this, I used the return on 2 year government notes for nominal interest rates and for each year subtracted the average rate of inflation between the year before and the year after. To ensure my construction of the real interest rate would be useful, I decided to compare a graph of the log real CRB Commodity Price Index versus the real interest rate from Frankel's paper with one using my construction of real interest rates.

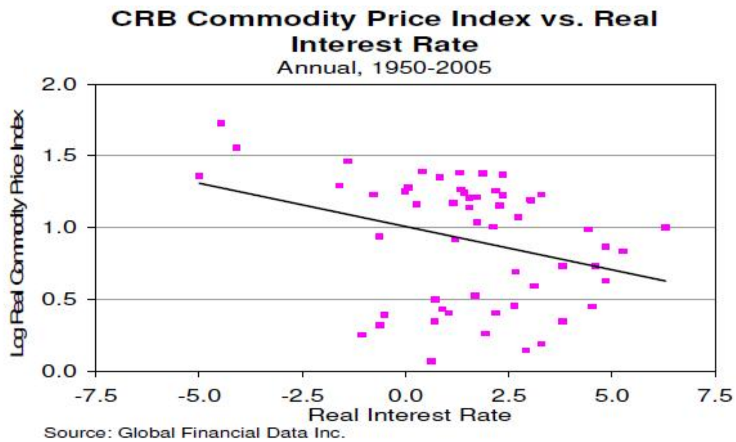


Figure 1: Frankel's Graph

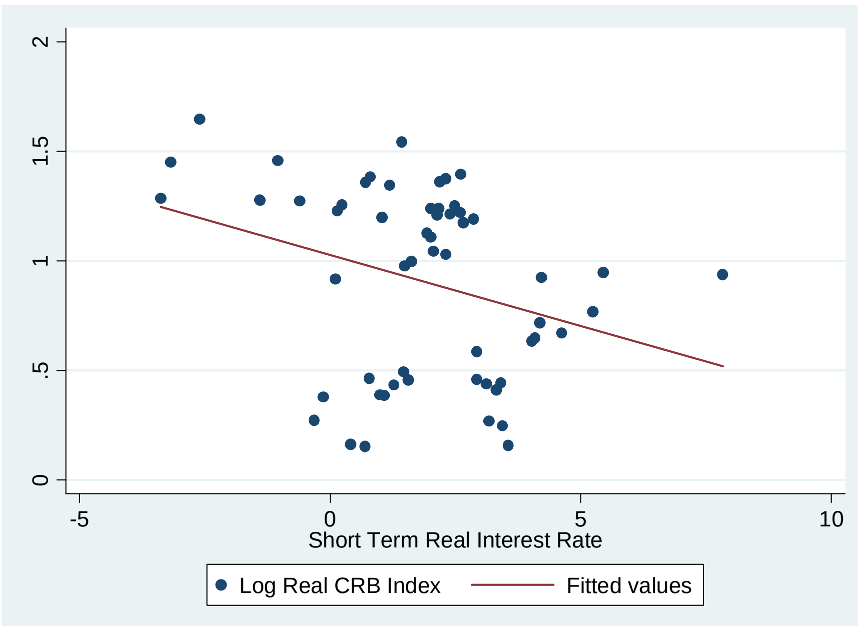


Figure 2: Graph with my construction of the Real Interest Rate

The graphs in Figures 1 and 2 seem close enough to say that my construction of the short term interest rate is reasonably similar to Frankel's. For his analysis, Frankel uses annual data from 1950 to 2005, for the rest of the paper, unless otherwise noted, I use data from 1950 to 2007. Once I had a credible series for the real interest rate, I estimated my three models.

The first model was constructed as follows:

$$c - c^* = \beta_0 + \beta_1 * i + \beta_2 * E(\Delta p) + \beta_3 * w$$

c = real log commodity price
c^* = long run equilibrium commodity price
i = real interest rate
$E(\Delta p)$ = expected inflation

w = world GDP (detrended)

The intuition is that that β_1 will be negative for storable commodities and possibly negative but closer to 0 for less storable commodities. β_2 will be positive for all commodities, since an expectation of an increase in prices should cause prices to rise, even if actual inflation does not turn out to be as high as expected. β_3 should be positive since world GDP is a proxy for changes in demand due to the business cycle. For the measure of c-c* I used the detrended log real commodity price. For i I used the short term real interest rate as calculated above. For E(Δp) I used the average expected inflation for the next 12 months from the Survey of Consumers, this measure was only available for 1960 to 2007. For w I used a measure from Global Financial Data constructed by B. R. Mitchell, which was only available from 1950 to 1998. Therefore, the following regression is for years 1960 to 1998.

The table is a presentation of the results from this regression for the commodities studied:

Commodity (log real spot/cash price, detrended)	Real interest rate (short term) (p-value)	Expected inflation (p-value)	Detrended World GDP (p-value)	R-squared
Corn	-.0247603 (0.145)	.0587638 (0.000)*	.0086846 (0.804)	0.4347
Oats	.0044003 (0.832)	.0605146 (0.003)*	-.0443939 (0.309)	0.2948
Soybeans	-.0225248 (0.229)	.0507444 (0.005)*	-.0086012 (0.825)	0.3336
Tomatoes	-.0229306 (0.024)*	-.0158889 (0.210)	-.0136475 (0.570)	0.1926
Wheat	-.0301836 (0.117)	.0396096 (0.027)*	.0252906 (0.524)	0.2715
Cattle	-.0100138 (0.214)	.018866 (0.013)*	-.0380933 (0.027)*	0.4012
Live hogs	-.0207541 (0.253)	.0440659 (0.011)*	-.1196823 (0.003)*	0.4649
Aluminum	.0405346 (0.002)*	.0461505 (0.000)*	-.0552724 (0.034)*	0.4337
Platinum	.0583871 (0.000)*	.072534 (0.000)*	-.1220782 (0.000)*	0.6580
Lead	-.0394623 (0.051)	.0455704 (0.015)*	.0214307 (0.604)	0.3368
Silver	.0400013 (0.109)	.1730909 (0.000)*	-.189342 (0.001)*	0.7180
Copper	-.0404209 (0.065)	.0164664 (0.401)	-.0857363 (0.061)	0.2787

Table 1: Regressions of Log Real Prices on real interest rates, expected inflation, and detrended world GDP.

With the exception of tomatoes, platinum, aluminum, and lead interest rates were not significant in the regressions. It should also be noted that while interest rates were not significant at a 5% significance level, the coefficient was indeed negative as expected for most of the regressions. With only four commodities showing a significant effect of interest rates, it is difficult to discern whether or not storability is making any difference. However, interest rates were significant at the 10% level for 4 out of the 5 metals studied, and only significant for one food commodity. This is in line with the idea that the more storable the commodity is, the greater the effect of interest rates on the price. Paradoxically, the least storable agricultural commodity studied, tomatoes, was the only one for which interest rates were significant.

It seems as though expected inflation is a much more important explanatory factor. Expected inflation was statistically significant at a 5% level for every regression except for tomatoes and live hogs. I found it interesting in many cases where real interest rates were not significant, expected inflation was, and vice versa. I wondered about the possibility of collinearity between these predictors. Figure 3 shows that the two predictors are somewhat correlated, however, the correlation was only -.3216. The relationship doesn't seem strong enough to suggest that the expected inflation is taking away much explanatory power from the real interest rate. It would be a problem in the measure of the real interest rate if the correlation was very strong, since the nominal component should have been removed.

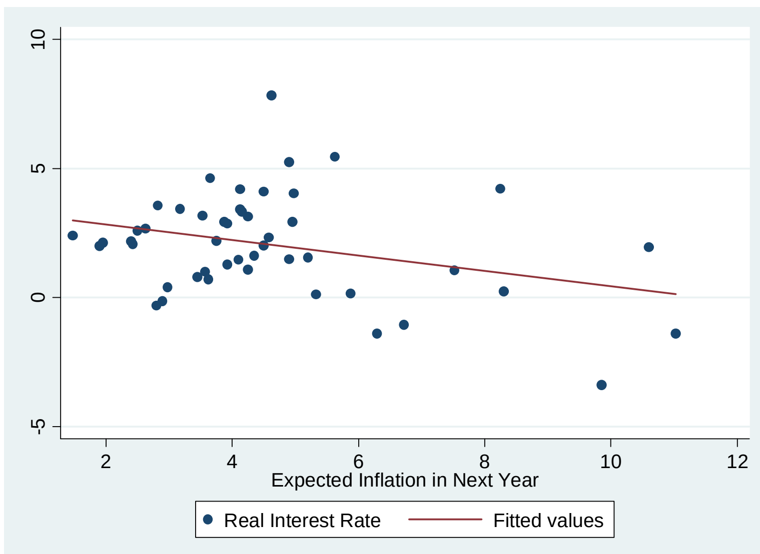


Figure 3: My construction of Real Interest Rates versus Expected Inflation.

It is interesting and unexpected on my part, how significant expected inflation was in these regressions. A story for why the expected inflation would be so important is that investors use commodities to hedge against inflation risk. Since agricultural commodity prices are not as sticky as prices of other goods, such as manufactured goods (Frankel, 2006), they are sure to respond right away to increases in the overall price level of the economy. Higher expected inflation would lead more investors to use these commodities as an inflation hedge which would increase demand and thus raise the real price of the commodities. The coefficients on the expected inflation term tended to be between .04 and .06. The interpretation is that a 1 percentage point increase in expected inflation would increase the real price of the commodity 4 to 6%. The effect of expected inflation was highest for silver, with a 17% increase in the real price of silver for a 1 percentage point increase in the expected inflation. The smallest significant effect was for cattle, with a 2% increase.

World GDP was only significant for cattle, live hogs, platinum, aluminum and silver. This suggests that their prices are more readily affected by the business cycle than the agricultural commodities. It could be the case that the demand for the agricultural commodities is somewhat inelastic to changes in the business cycle; people need to eat whether or not the world economy is doing well.

To explore the relationships between commodity prices, interest rates, and expected inflation further, I will plot the relationships for corn, tomatoes, cattle, and aluminum. I decided to look at these commodities in particular since tomatoes are the least storable commodity studied, cattle are more storable than tomatoes, but still not a very storable commodity, corn is a storable agricultural commodity, and aluminum as a benchmark for a very storable commodity. The plots for the rest of the commodities examined can be found in the Appendix.

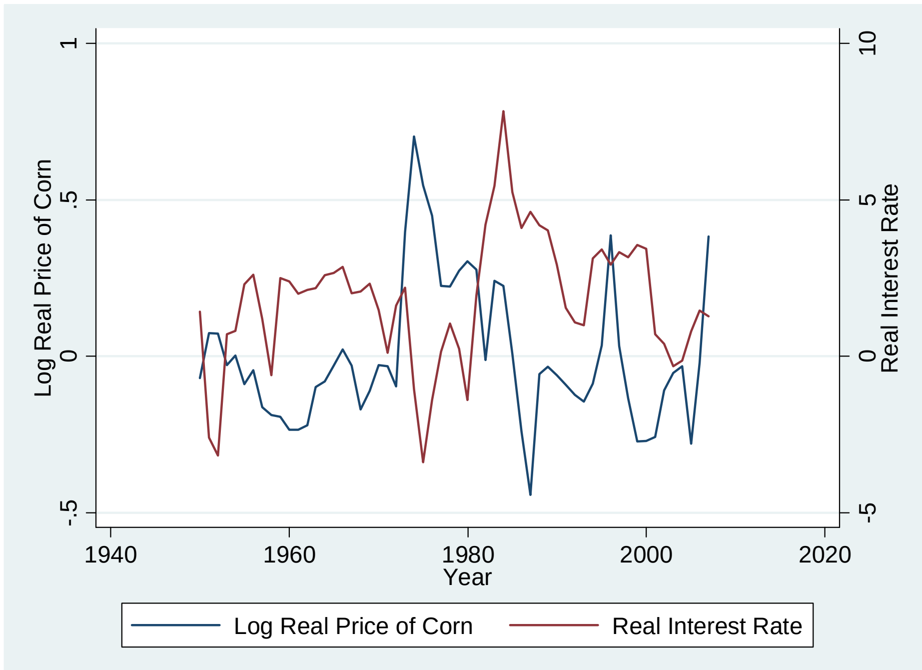


Figure 4: Log Real Price of Corn and Real Interest Rates

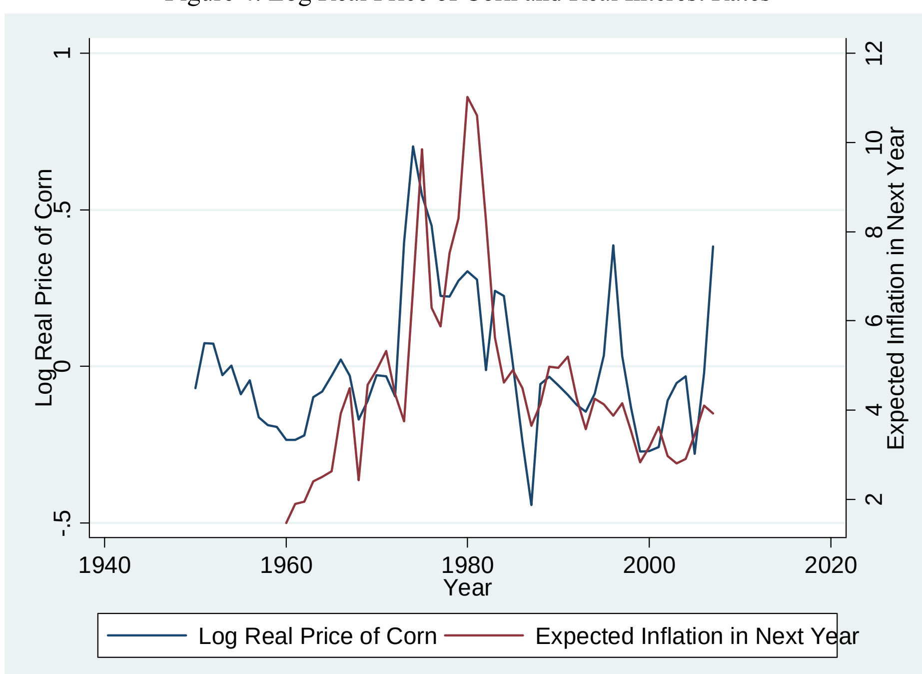


Figure 5: Log Real Price of Corn and Expected Inflation

Figure 4 shows that the negative relationship between the real price of corn and real interest rates can be seen in some time periods but not for others. The relationship seems to be slightly negative in the 1950s, then positive throughout the 1960s until the early 1970s. In the 1970s until the mid-1980s the relationship seems pretty strongly negative, but then it become positive again. However, there could be a lag present here, since a sharp increase in the real interest rate is shown concurrently with an increase in the real price of corn, that is then followed by a sharp decrease in the real price of corn. The relationship then seems to be positive until the 2000s when it becomes negative again. Overall, it seems like the relationship is strong in some periods but that in fact the opposite relationship holds true in other periods.

Figure 5 shows that the positive relationship between real corn prices and expected inflation is pretty strong in most periods. This relationship is much more robust over time, which would be the reason that expected inflation is a much stronger predictor in the regression.

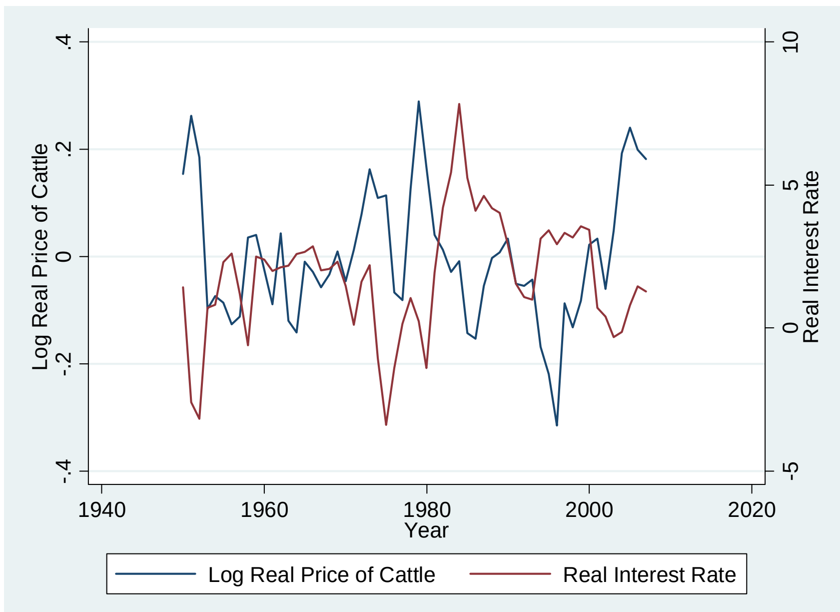


Figure 6: Log Real Price of Cattle and Real Interest Rates

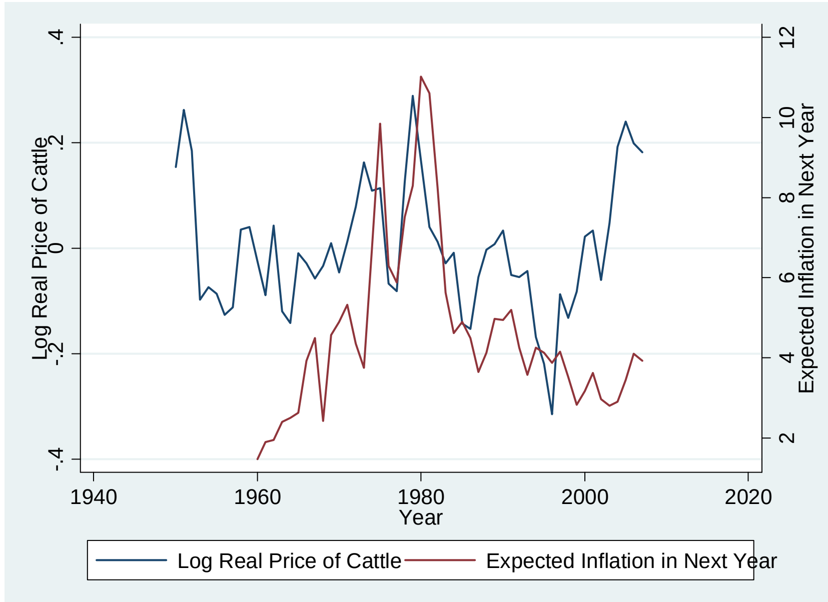


Figure 7: Log Real Price of Cattle and Expected Inflation

Figure 6 shows that for cattle, the real interest rate relationship is not really present before the 1970s. In the 1970s until the 1990s the relationship seems to be negative, but after this period the relationship is not clear.

Figure 7 shows that the relationship between cattle and expected inflation is not as strong as it was with corn, but it seems to be present and fairly robust over time. The only time periods where the relationship isn't there are the early 1960s and perhaps the 1990s, where there is a dip in expected inflation, but there is no visible response in the price of cattle.

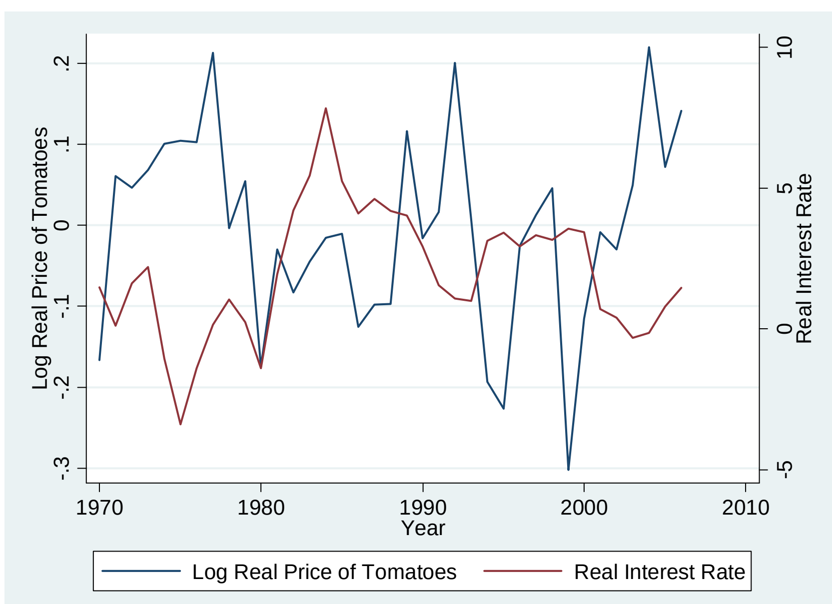


Figure 8: Log Real Price of Tomatoes and Real Interest Rates

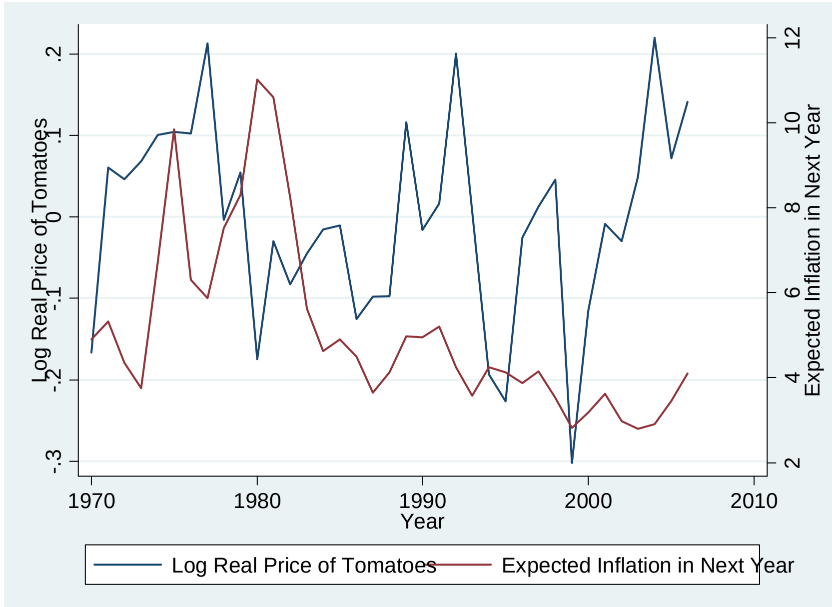


Figure 9: Log Real Price of Tomatoes and Expected Inflation

For tomatoes, data were only available from 1970 to 2007. Recall that tomatoes were the only non-metal commodity for which real interest rates are significant. There is a negative relationship for most of the 1970s and the 2000s. However, most of the graph doesn't show a very strong negative or positive relationship. The explanation for why tomatoes has a significant relationship with real interest rates could be because for the other commodities, such as corn, there were some time periods for which the relationship between prices and real interest rates was positive, and some where it was negative. Since the data set for tomatoes is limited to the period after the 1970s, it may just be the case that the negative relationship "wins out" in this period, where it might not if data was available back to the 1950s. Therefore, the significance probably doesn't tell much about how the lack of storability affects the relationship between tomatoes and real interest rates, compared with other commodities.

Figure 9 shows the relationship between the price of tomatoes and expected inflation is not as strong as it was with corn and cattle. The relationship is somewhat negative from 1970 to the mid-1980s after which it seems to be positive, but again, it is not very strong. It seems that the price of tomatoes is somewhat volatile and expected inflation is not a very good predictor of prices for this commodity.

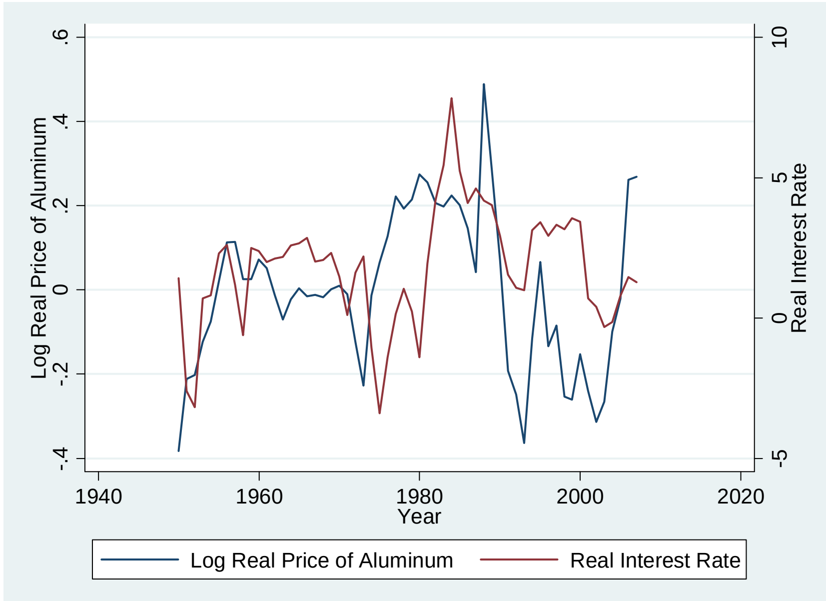


Figure 10: Log Real Price of Aluminum and Real Interest Rates

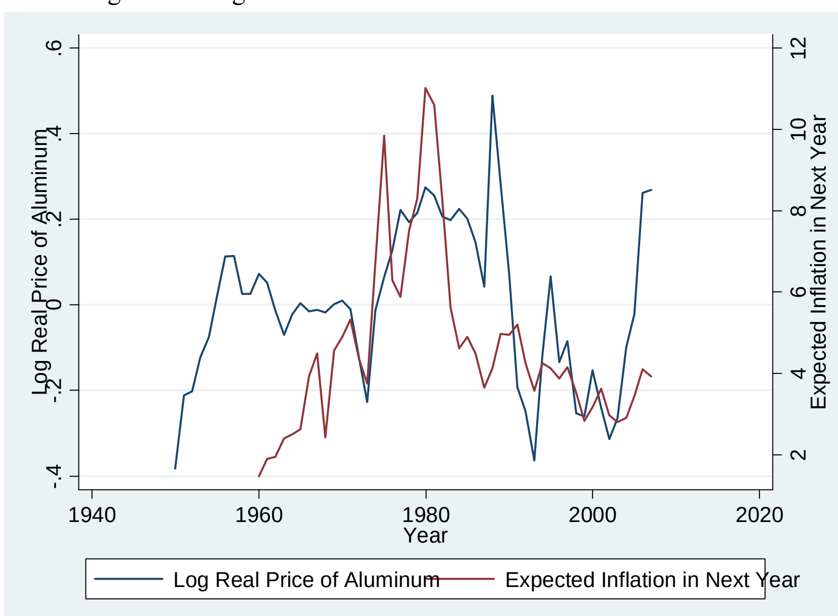


Figure 11: Log Real Price of Aluminum and Expected Inflation

The regression for aluminum showed a significant positive relationship with real interest rates and a significant positive relationship with expected inflation. Figure 10 shows that up until the 1960 as the real interest rate rises, so does the price of aluminum. After the 1960s (which is the period included in the regression) the relationship is negative until around 1985. After this period the relationship is positive.

Figure 11 shows that the price of aluminum and expected inflation move together in the same direction, although the magnitude of the movements are smaller for the price of aluminum before 1980 and larger for the price of aluminum after 1980.

After a closer look at these four commodities, as well as the commodities included in the appendix, it appears that the real price is affected by both real interest rates and expected inflation. However, the negative correlation between interest rates and commodity prices seems to be less robust over time than the positive relationship of commodity prices and expected inflation. The first relationship seems to vary from negative, to positive, to non-existent over time, while the other relationship seems mostly positive, with changes in magnitude.

It is likely that there are other macroeconomic factors that need to be controlled for besides world demand and expected inflation in order to study the effects of real interest rates on commodity prices. In order to control for these factors as well as the excess co-movement of commodity prices found by Pindyck (1988), I decided to regress the detrended real log prices against the log CRB Index, a commodity price index constructed by the Commodity Research Bureau. The residuals in this regression would show how the real commodity price moved independently of how commodities moved as a whole. For the second model, I regressed these residuals against the real interest rate as calculated above.

Commodity Residuals (log real spot/cash price, detrended, regressed on logcrb)	Real Interest Rate (p-value)	R-squared
Corn	-.0389636 (0.003)**	0.1428
Oats	-.0304451 (0.041)**	0.0724
Soybeans	-.0280952 (0.065)*	0.0595
Tomatoes	-.0234955 (0.006)**	0.1919
Wheat	-.0407739 (0.003)**	0.1472
Cattle	-.030598 (0.000)**	0.2721
Live Hogs	-.0233729 (0.142)	0.0382
Aluminum	.0204088 (0.077)*	0.0549
Platinum	-.0059068 (0.693)	0.0028
Lead	-.0556345 (0.003)**	0.1426
Silver	-.017738 (0.541)	0.0067
Copper	-.01526 (0.487)	0.0087

Table 2: The residuals for the prices of each commodity regressed on the CRB index, regressed against real interest rates.

Table 2 is a bit more informative about real interest rates and storability, since once the overall price of commodities is controlled for, the real interest rate is significant for every agricultural commodity, as well as cattle, aluminum, and lead. It was not significant for live hogs, platinum, copper, or silver. The significant coefficients were between -.023 and - .055.

Corn and wheat had similar coefficients around -.04. The interpretation is that a 1 percentage point increase in the real interest rate corresponds to a 4% decrease in the residual of the commodity price regressed on the log CRB Index. Oats is also a storable grain, and has a significant coefficient of -.030. However, the p-value was larger for oats, at 0.041, while it was

0.003 for both corn and wheat. For soybeans, which are presumably less storable than corn and wheat since they naturally contain oils, the coefficient is slightly less negative at -0.028. The coefficient on soybeans was significant at the 10% level but not the 5% level, since the p-value was 0.065. The least storable agricultural commodity, tomatoes, has a smaller, yet still significant coefficient of -.023.

For live hogs, real interest rates were not significant. However, the log real price of hogs did not have a strong linear trend like the other commodities (Figure 12). For the regressions, I still detrended the price of hogs linearly. This could be part of the reason that the negative coefficient on live hogs is not significant

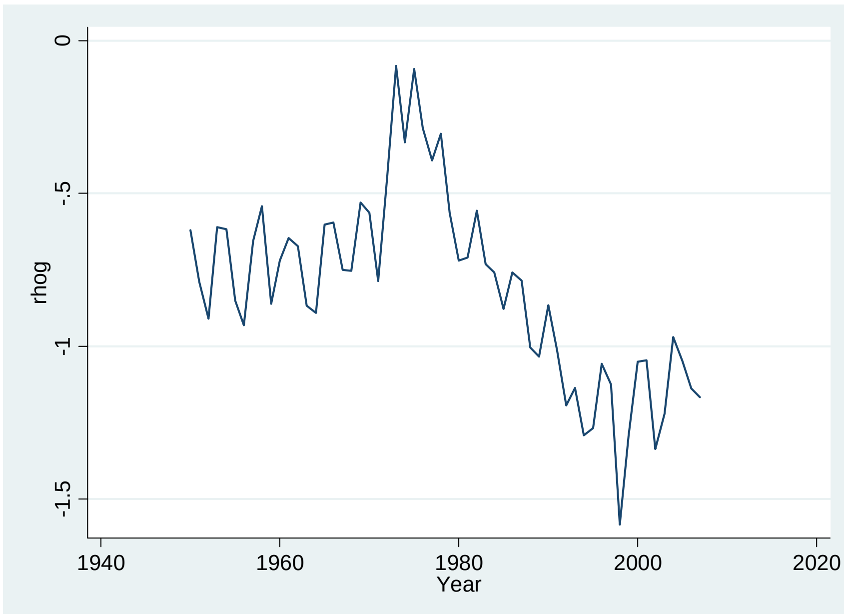


Figure 12: Real Log Hog Prices

Cattle had a significant coefficient of -.0305. Interestingly, although one would assume that cattle are not a very storable commodity, the coefficient is not remarkably different from the coefficients of the grains. In addition, the p-value of the coefficient is the smallest for all commodities.

As for the metals, one would expect that they are the most storable commodities. Since they do not need to be harvested at a certain time and cannot spoil, there are many more opportunities to sell them. However, lead was the only metal which was significant at the 5% level. It has a coefficient of -.056, which is a bit higher than the coefficients for the agricultural commodities. Aluminum was significant at the 10% level, however the coefficient was positive. The regressions on the rest of the metals were negative, although not significant, which is consistent with the findings in the first regression, interest rates do not seem to have a strong impact on the price of metals.

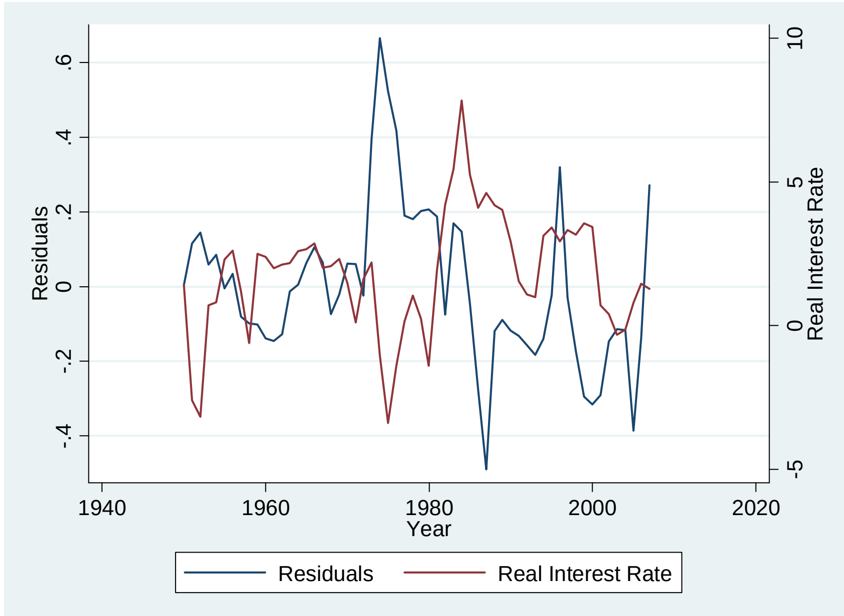


Figure 13: Residuals of Corn Prices and Real Interest Rates

A look at the graph of the corn residuals and real interest rates shows the same pattern that was in the original graph of corn prices and interest rates. The negative relationship holds true for most of the 1950s until both values start to increase after 1960. The strong negative relationship is clear in the 1970s and the 1980s, although there appears to be a bit of a lag in the 1980s. In the 1990s the relationship appears to be positive for the most part, and returns to being negative after 2000. It could be the case that since the regression in Table 1 only included data from 1960 until 1998, and the regression corresponding to this graph is for data from 1950 to 2007, the extra data may have contributed enough to make the real interest rate significant. This goes to show that the relationship is not robust over time, at least for corn.

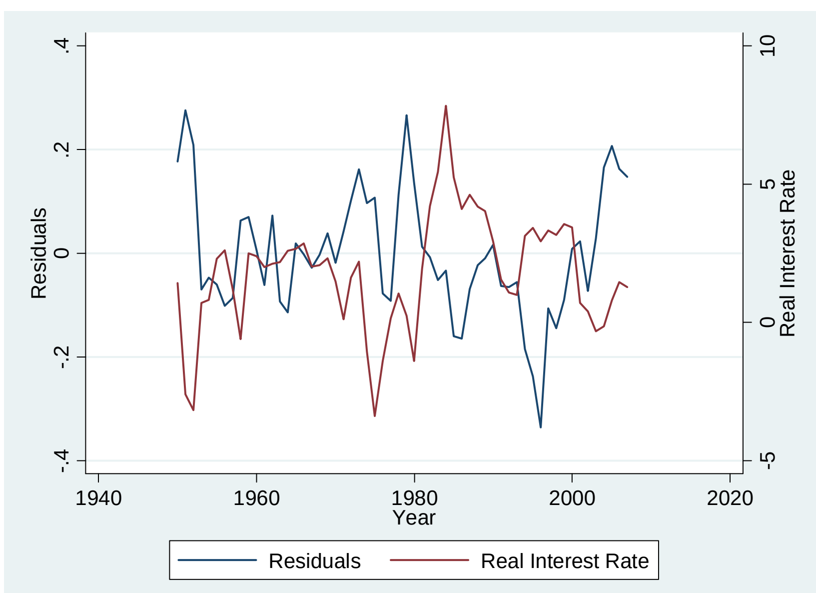


Figure 14: Residuals of Cattle Prices and Real Interest Rates

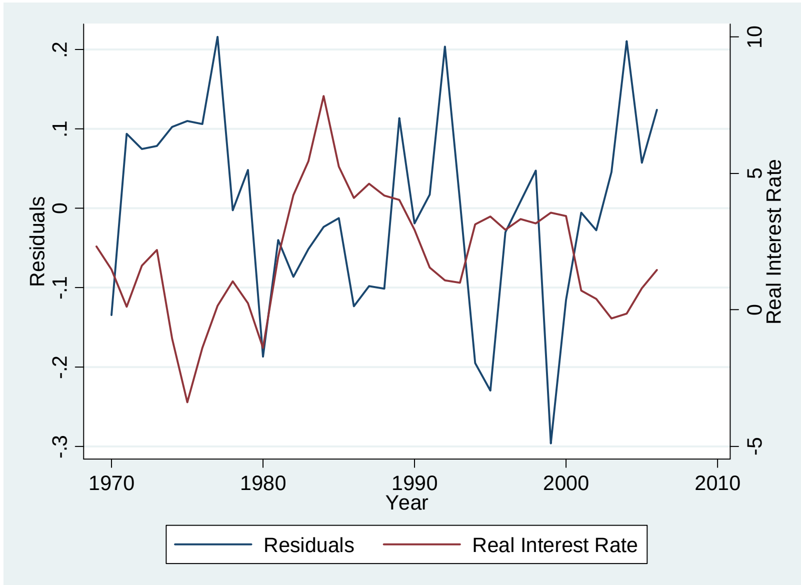


Figure 15: Residuals of Tomato Prices and Real Interest Rates

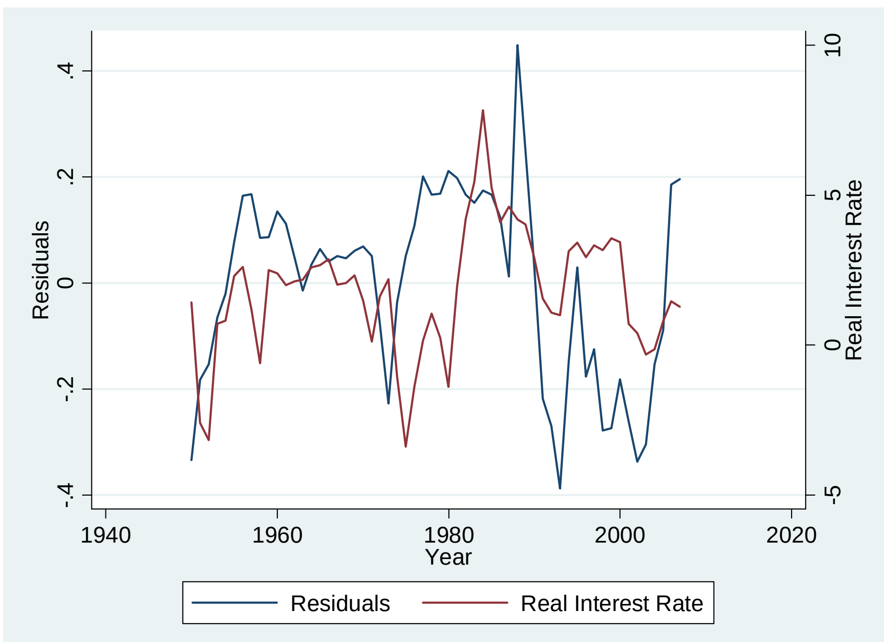


Figure 16: Residuals of Aluminum Prices and Real Interest Rates

Figures 14, 15, and 16 show the relationships for cattle, tomatoes, and aluminum. The relationships between the residuals and the real interest rates are not remarkably different. This is evidence that the additional time periods included could have been a contributing factor to the significance of interest rates in these regressions. It should not be the case that the interest rates are significant because expected inflation is no longer included in the regressions. This is because the expected inflation had such a strong effect on the commodity prices, it has an effect on the commodity price index as well. Since the residuals of the regression on the index are what is being studied here, the expected inflation in included implicitly. Indeed, a regression of the log CRB index on expected inflation gives a significant positive coefficient of 0.0716.

So far there has not been conclusive evidence that storability has an effect on the negative correlation of interest rates and commodity prices. For the third model, I decided to look at short term versus long term interest rates to see if differences in storability had a noticeable impact there. I constructed a measure of the real long term interest rate using returns on 20 year bonds minus the average inflation over the past 10 years and the next 10 years. The correlation between the short term and long term real interest rates as constructed is 0.5236.

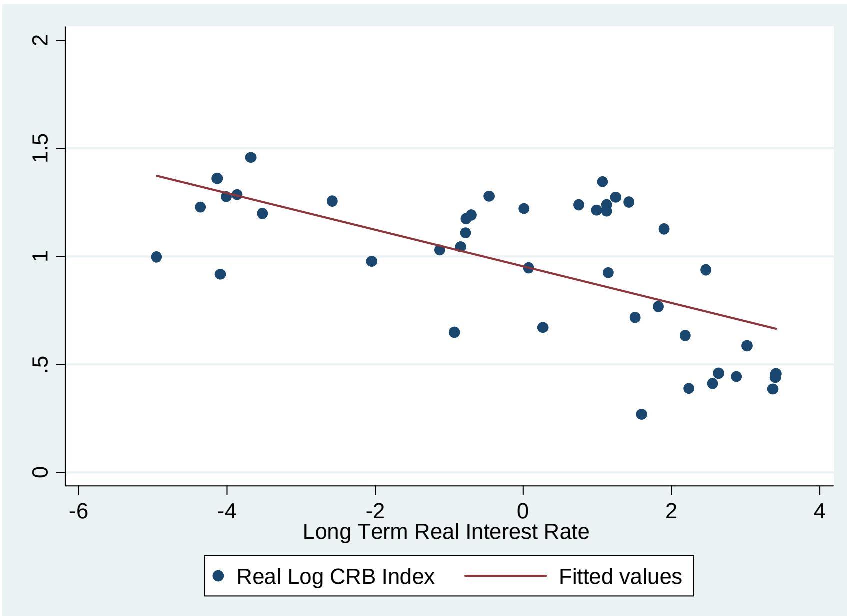


Figure 17: My construction of Long Term Interest Rates and the Real Log CRB Index

The coefficient on the real log CRB index regressed on the long term interest rate was -.0847 with a p-value of 0.000. The R-squared of the regression was 0.3796. In comparison the coefficient was -.0647 for the short term interest rate with a p-value of 0.017 and an R-squared of 0.0981. For the overall index, the long term interest rate seems to have a stronger effect on the real price of commodities. The correlation between the short term interest rate and long term interest rate I used was .5236. The long term rate could only be constructed for 1957 to 1998.

Commodity Residuals (log real spot/cash price, detrended regressed on logcrb)	Real Interest Rate (p-value)	Long Term Interest Rate (p-value)	R-squared
Corn	-.0266943 (0.102)	-.0379166 (0.007)*	0.3619
Oats	.0083287 (0.656)	-.0601582 (0.000)*	0.3264
Soybeans	-.0000506 (0.997)	-.0693279 (0.000)*	0.5446

Wheat	-.0389248 (0.027)*	-.0174373 (0.221)	0.2572
Tomatoes	-.011336 (0.267)	-.0099175 (0.240)	0.1993
Cattle	-.0115582 (0.182)	-.0186204 (0.012)*	0.3082
Live Hogs	-.0009405 (0.949)	-.0825338 (0.000)*	0.6203
Aluminum	.0219899 (0.131)	-.0203499 (0.0926)	0.0833
Platinum	.0395234 (0.043)*	-.0449582 (0.006)*	0.1834
Lead	-.0442541 (0.030)*	-.0250386 (0.131)	0.2819
Silver	.0093261 (0.783)	-.1094137 (0.000)*	0.3368
Copper	-.0006654 (0.974)	-.0724866 (0.000)*	0.3988

Table 3: The residuals for the prices of each commodity regressed on the CRB index, regressed against real interest rates, short and long term.

For most of the commodities the long term rate was more significant. An explanation for this is that investors are interested in the long term as well as the short term. That is to say, investors are more interested about how much their money can make in the bank over multiple periods rather than just the short term. For instance, it might not be worthwhile to take your money out of the commodity if you expect interest rates to rise, but then shortly return to where they were. This is especially true since there are other factors influencing the price of the commodity. In short, a higher interest rate over the long term will give a higher return and thus is more incentive to put money in the bank rather than commodities.

Short term interest rates should have more of an impact on the less storable commodities, such as the grains and livestock, while long term rates should be more important for very storable commodities such as metals. However, this is not the case in this regression. For all of the agricultural and livestock commodities except for wheat, long term interest rates were more important. Another exception is tomatoes, for which neither rate was significant. This is probably due to collinearity of the predictors, since the short term interest rate was significant for the tomatoes in both of the previous regressions.

As for the metals, long term rates had a very strong significant effect on copper and silver. This is interesting since the short term interest rates were not significant for either of the previous regressions. It lends some credence to the idea that storability matters. However, for lead the short term interest rate was more significant, and for platinum, both rates were statistically significant. The aluminum regression most likely suffers from the same problem that the regression on tomatoes does, namely collinearity of the predictors.

Overall, it is clear that there is a significant negative relationship between real interest rates and commodity prices. From the regression in the first model, it is apparent that this relationship is not robust over time. For many commodities, it may hold true in some time frames, but it other time frames the relationship is positive, or even non-existent. This is in contrast with expected inflation, which was a significant predictor of the log real commodity price and was robust over time. As far as storability is concerned, for the most part there was no discernable difference in the slopes of the coefficients for more or less storable commodities. Even when the movement of the commodity index was taken into account in the second model, the prices of metals as the more storable commodities were not significantly affected by real interest rates. Meanwhile the less storable food commodities were significantly affected by real interest rates. In the third model, most of the commodity prices had a more significant relationship with the long term rate, even though they cannot be stored for a long time. All together, there is not conclusive evidence that the storability of commodities determines the strength of the negative relationship between the real commodity price and the real interest rate.

References:

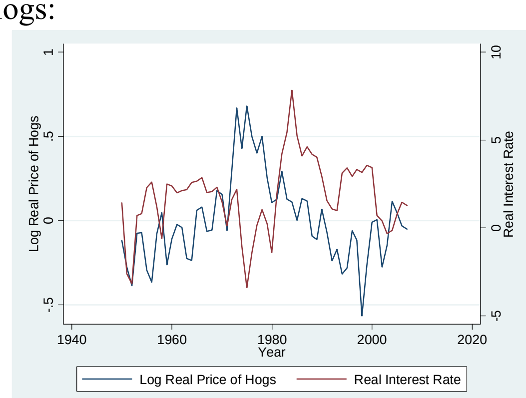
Frankel, Jeffrey A. (2006). The effect of monetary policy on real commodity prices. National Bureau of Economics Working Paper 12713.

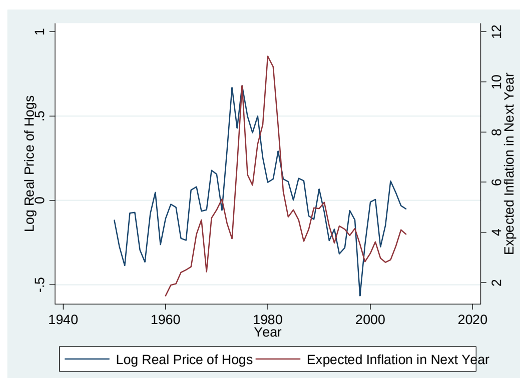
Pindyck, Robert S. & Rotemberg, Julio J.(1988). The excess co-movement of commodity prices. National Bureau of Economics Working Paper 2671

Appendix:

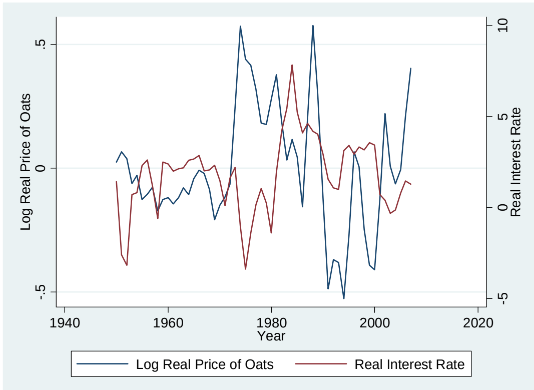
Graphs of Prices, Real Interest Rates, and Expected Inflation:

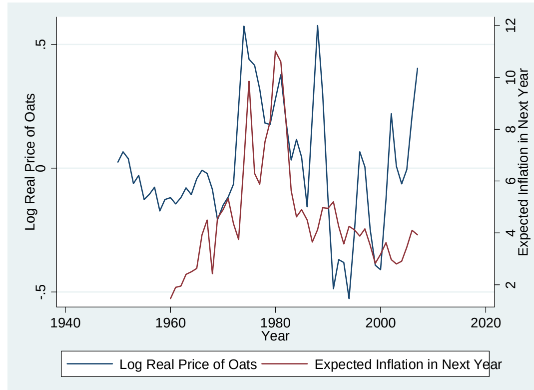
Live Hogs:



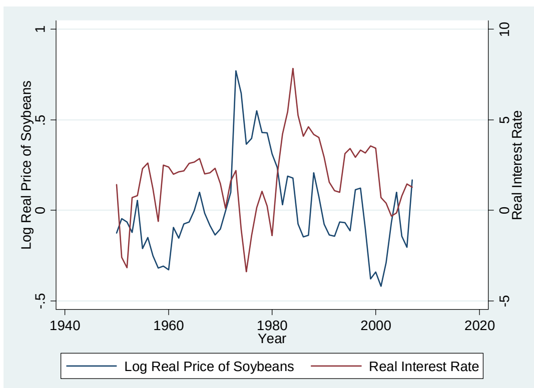


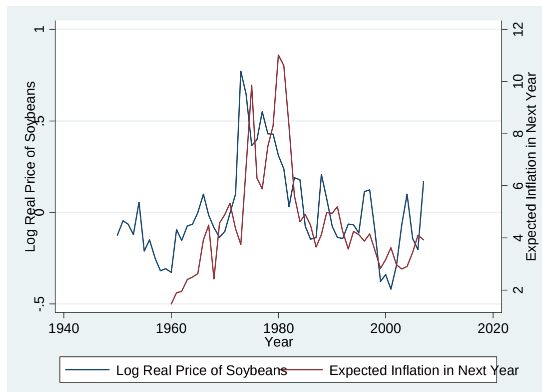
Oats:



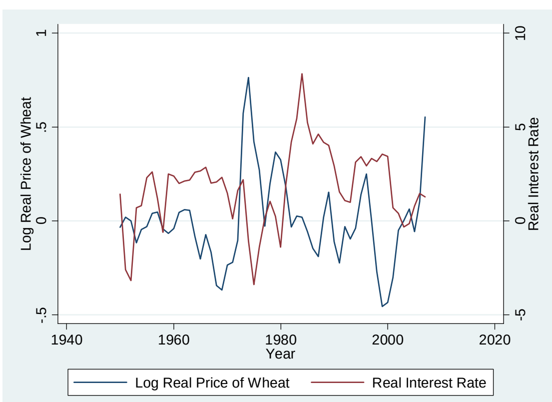


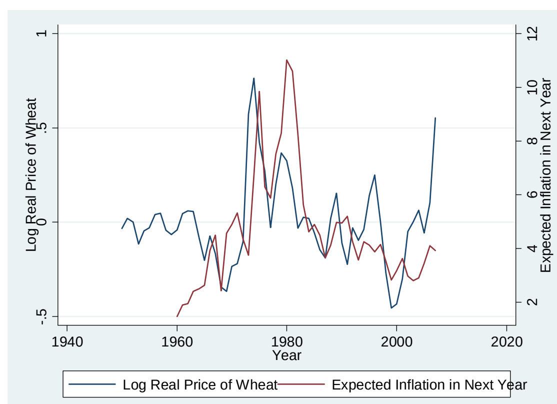
Soybeans:



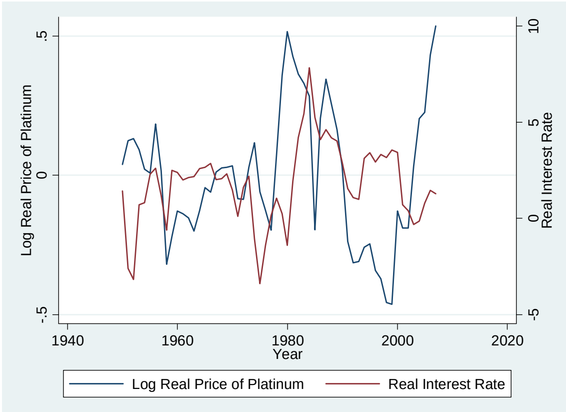


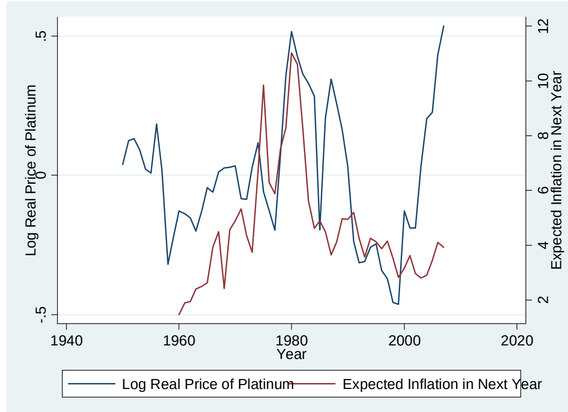
Wheat:



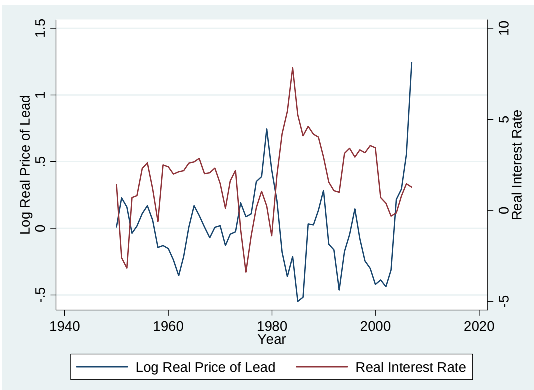


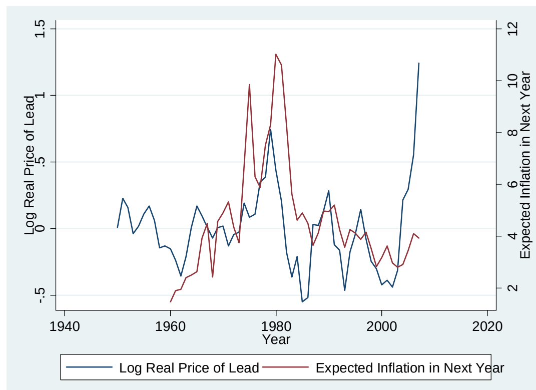
Platinum:



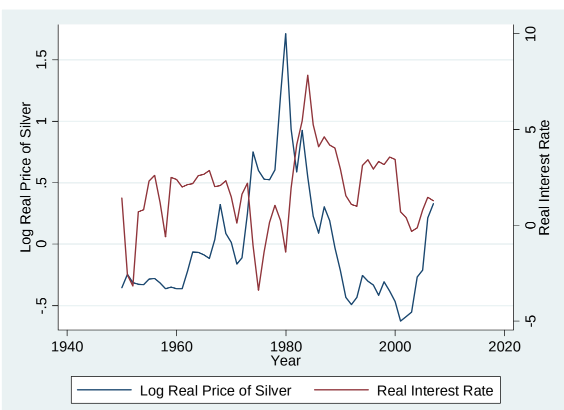


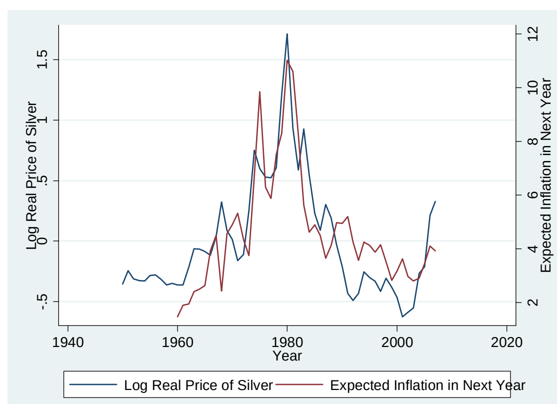
Lead:



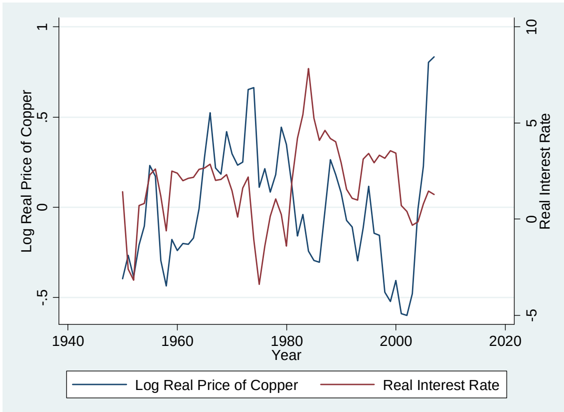


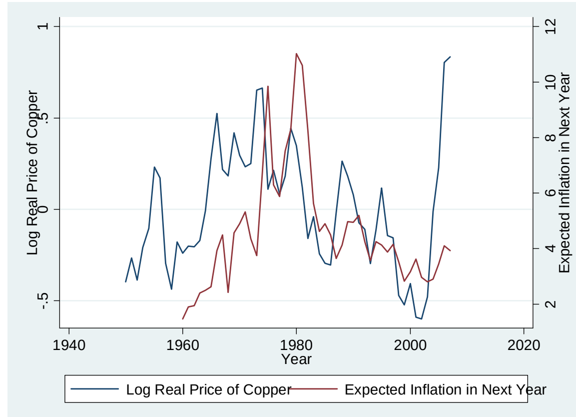
Silver:





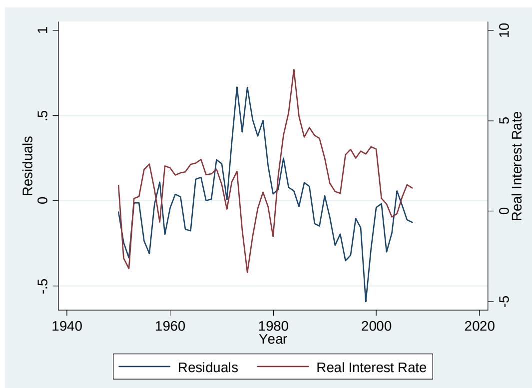
Copper:



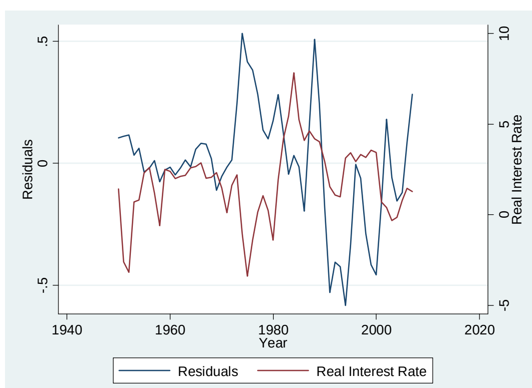


Graphs of Residuals and Real Interest Rates:

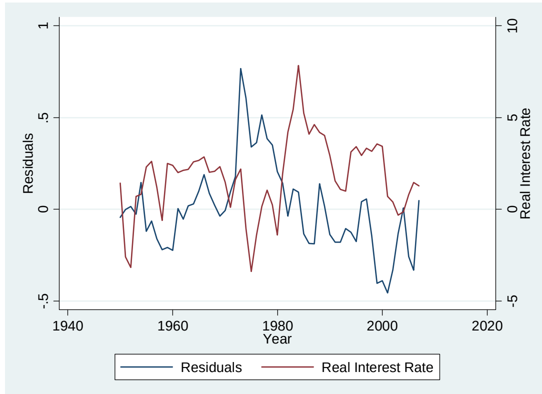
Live Hogs:



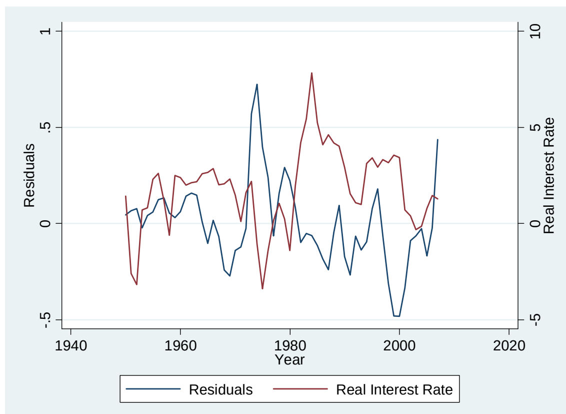
Oats:



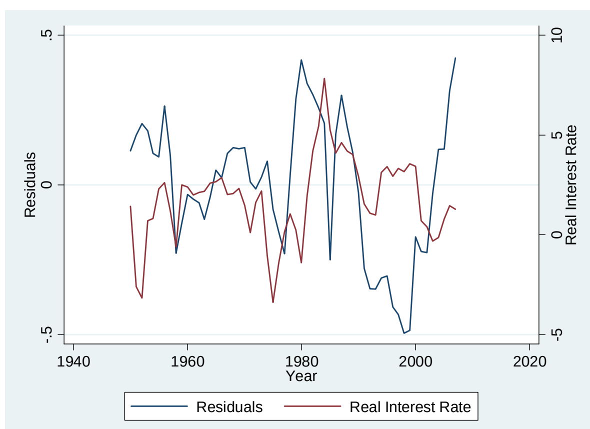
Soybeans:



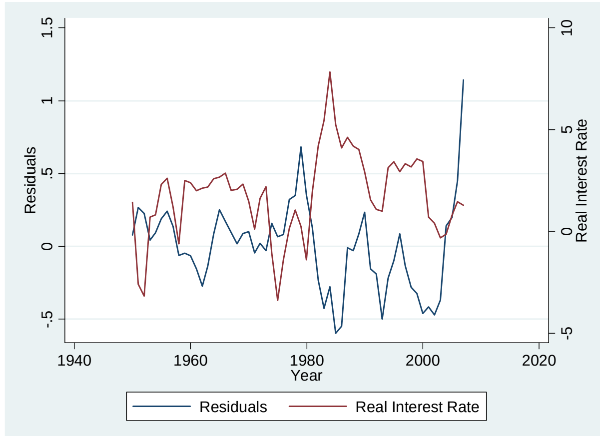
Wheat:



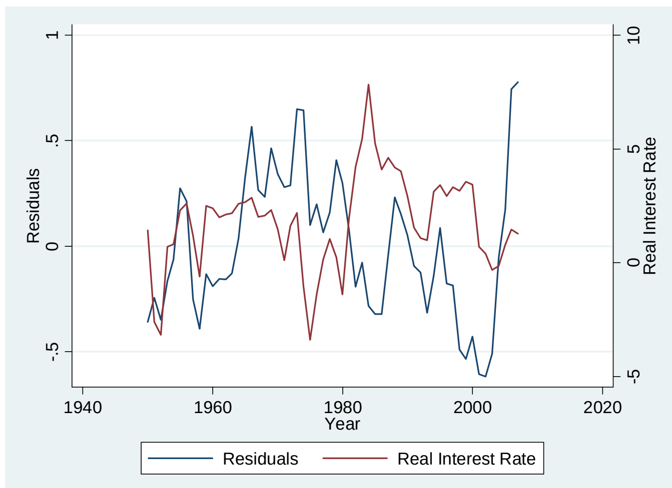
Platinum:



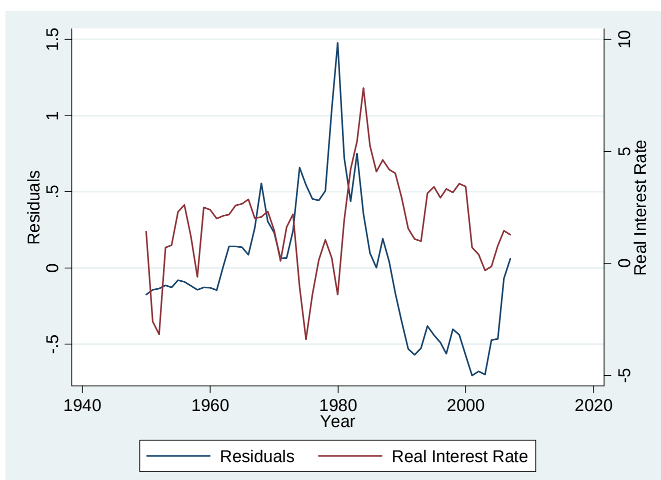
Lead:



Copper:



Silver:



Julianne Zamora
Project 3
4/18/08

Which Has a Larger Impact on Mental and Physical Health: Work Life or Family Life?

Introduction:

For those of us just beginning our adult lives, there is a lot of uncertainty about the road ahead. One way to go is to plant the focus solely on one's future career. Will this make us healthier both mentally and physically in our adult lives, or will it just add unnecessary stress? The alternative is to focus on finding your special someone and starting a family together, and letting work just be something you have to do to be able to enjoy the real time investment in your life, your family. Many people chose is to focus on both family and work life but at different levels of intensity throughout their lives. The path you chose really is a question of whether you carefully plan your work life and leave your family life to luck, or if you really plan your family life and leave your work life to luck.

What you decide to spend your time on when you're young may determine what you spend most of your time on when you're older. For those who find a spouse and have children a significant amount of time is spent caring for children in addition to what you would normally do. This is especially significant for those who spend a lot of time multitasking while caring for their children. This leads to higher levels of reported tension, especially for women. In addition, those who do not have children, and primarily spend time working are less likely to engage in multitasking while at work, and thus multitasking will not increase their tension levels (Michelson, 2005). How we spend our time does not only help determine our mental state, but it also determines where tension is coming from (i.e. multitasking as opposed to working long hours).

In the end, we want to choose the path that will improve our wellbeing. One way to measure wellbeing on a very personal level is to study mental and physical health. For many of us who want to have it all, the real question is which will have more of an impact on my health, my family life, or my work life?

In order to answer this question I regressed 3 different health related variables from the Panel Study of Income Dynamics (PSID). The first variable I used is the K-6 Non-specific Psychological Distress Scale, which is a measure of emotional distress. The second variable I used is self-reported health status. The third variable I used is a variable that measures how often the individual feels rushed or pressed for time. I used this variable as an indication of how "stressed out" the individual is. First, I ran regressions with independent variables dealing with family life, then with independent variables concerning work life, and then I combined the two.

Data:

The data used was taken from the PSID data for 2003. Only heads of households as well as wives living in the household at the time of the study were used. Age was controlled for, as only adults between the ages of 18 and 59 were studied. I chose these ages because most people 60 and older will be retired and the children they have will be living outside the home. For that reason, most of these people will be out of the workforce and living without children at home, but that is because they have finished their career or have already raised their children, as opposed to younger singles who cannot find work, or have yet to have children. These are important distinctions to make when the question being asked is concerned with lifestyle factors.

The independent variables I used can be split into three separate categories. First is what I call the "Family Life Variables". These variables include births to the individual, which for the data studied ranges from 0 to 13. Another lifelong variable is the individual's number of marriages, which ranged from 0 to 6. The other numerical values studied were the number of children living in the house (ranging from 0 to 8) and the number of others supported by the individual (ranging from 0 to 6). The others supported could include children living outside the house, the ex-spouses, parents, etc. In addition, the marital status of the individual was studied along with the age of their youngest child.

The next category of independent variables was "Work Life Variables". These included the family's income, the number of years the individual spent in school, their employment status, as well as the number of hours they worked in 2002 for all jobs and including overtime.

The third category of independent variables used were variables were responses to the question "Has a doctor ever told you that you have or had the following…".The conditions studied were a learning disorder, permanent loss of memory or mental ability, any emotional, nervous, or psychiatric problems, asthma, arthritis, heart disease, heart attack, lung disease, cancer, diabetes, hypertension, and stroke .

For each regression, those individuals with inappropriate values (a refusal to answer, missing value, or otherwise) for any of the independent variables studied or for the dependent variable, were not included. I could not find a variable to determine the ages of the children in the family, aside from the variable for the age of the youngest child. I wanted to study what effect the age of the children would have on the parent's health. In order to do this I made a dummy variable based of the age of the youngest child to determine whether there was a child or children in the house who were 6 or younger. For the marital status variable, I made 2 different dummy variables. The first dummy variable determines whether the individual was married at the time of their interview in 2003. The next dummy variable determines whether they were divorced and not remarried in 2003.

The work life variables were modified for the regressions as well. The log of income was studied in order to deal with the skewness of the income data. Dummy variables were made to study the effect of employment status. The first dummy was for the unemployed and included those who were temporary laid off or looking for work. Next were dummy variables for those who were out of the workforce, which included the retired, the permanently disabled, housewives and students.

Another work life variable that was made into a categorical variable was the annual work hours. I divided this variable by 52 to get the average time each person spent at work weekly. Since many people were unemployed or out of the work force, in order to keep them from being counted twice by the regression, I used categorical dummy variables for the weekly work hours. Most people worked between 30 and 50 hours a week, so one variable was made for those people who worked, but worked less than 30 hours a week, and those who worked more than 50 hours a week.

Dummy variables were made to study the effect of education on the various regressions. The years of education were used to determine who had graduated high school or college. Those with 12 or more years of education were considered high school graduates, while those who had 16 or more years of education were considered college graduates. The education variable used to make these dummies distinguishes high school graduates from those who receive their GED, if he/she received a GED the education variable had the number of years they actually spent in school. It should be noted that high school and college were not mutually exclusive categories, so those who had a 1 for college also had a 1 for high school.

Dummy variables were created for each health condition as well. To make the regressions a bit less cumbersome, I grouped the medical conditions into serious conditions and "non-serious" conditions. The conditions I considered serious were heart attack, stroke, heart disease, and cancer. Those considered "non-serious" were asthma, arthritis, lung disease, diabetes, and hypertension. The dummy variables for learning disorders, memory loss, and emotional problems were kept separate.

Results:

K6 Distress Scale:

The K-6 Non-Specific Distress Scale (K6) was developed by Dr. Ronald Kessler, Professor of Healthcare Policy at Harvard Medical School. It is called K-6 because it involves 6 questions. Respondents of the PSID were asked the following questions: In the past 30 days how often have you felt a) so sad nothing could cheer you up? b) nervous? c) restless or fidgety? d) hopeless? e) that everything was an effort? f) worthless? For each question, there were five possible choices for the respondent. Each response would garner a certain number of points. The possible responses were: all of the time (4), most of the time (3), some of the time (2), a little of the time (1), and none of the time (0). After these six questions were asked, (each is a separate variable in the PSID), the points were tallied and that sum is the K6 variable I used to do my analysis. The point values range from 1-24. A rating of 13 or more on the distress scale indicates "sensitivity around the threshold for the clinically significant range of the distribution of nonspecific distress."

An important thing to note about this distress scale is that it is not a measure of happiness or satisfaction. A lack of emotional distress does not indicate that someone is always in a good mood, he/she could be indifferent, or maybe he/she had a relatively good past 30 days. A high rating on the distress scale does not necessarily mean that someone is never happy either. Another thing to note is that the questions are asked about the past 30 days, but I imagine for many people it is hard to remember all of your emotions for the past 30 days. If I had to answer the same questions, I would probably think about the past 10 days or so, and extrapolate that for an answer about the past 30 days.

In the previous project, I looked at how marital status and the number of children affected the K6 variable. The first thing I did was to use the Individual Weight variable to come up with the weighted averages for different every category of adult based on sex, marital status, and number of children. I graphed these averages on one 3-D bar graph.

Graph 1: Average Distress Levels

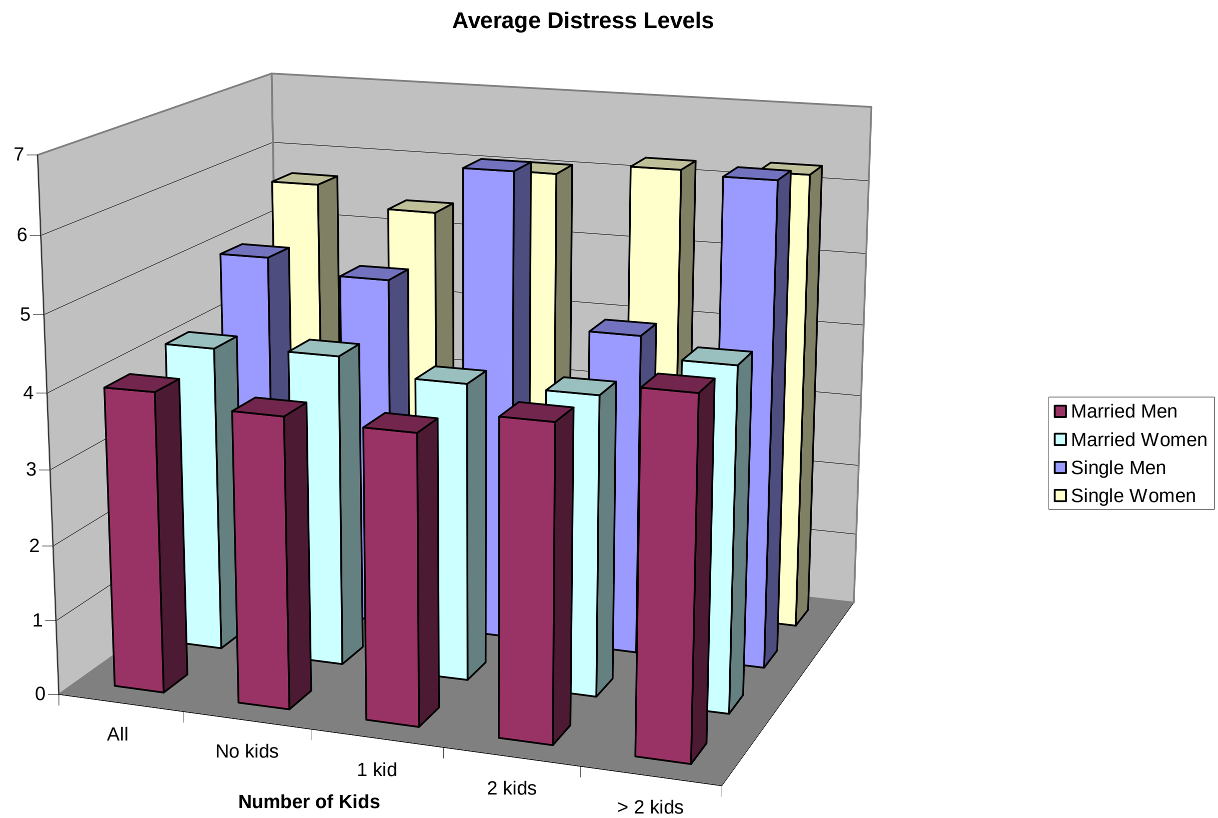


Table 1: Average Distress Levels (number of observations)

	Married Men	**Married Women**	**Single Men**	**Single Women**
All	3.9871 (916)	4.1375(1051)	4.970 (626)	5.6715 (792)
No kids	3.8411 (375)	4.1843 (464)	4.8131 (522)	5.3939 (454)
1 kid	3.8000 (210)	3.9778 (221)	6.3941 (57)	6.0533 (157)
2 kids	4.1038 (214)	3.9900 (239)	4.3453 (30)	6.2176 (107)
More than 2 kids	4.6205 (117)	4.5282 (157)	6.5010 (17)	6.2598 (74)
All Men	4.3550 (1542)	**All Women**	4.7258 (1843)	

Of course, as was expected women score higher than men overall. This was my suspicion going in since I had learned in a psychology class in high school that women are more likely to suffer from mental distress. In addition, women have a higher reported mean tension level than men (Michelson, 2005). Something to consider is that this question was only asked for the respondent, and men may be less likely to report feelings of sadness or nervousness because of social stigma that is attached to mental/emotional problems.

In the overall distress averages, regardless of the number of children, married men show less distress than any other category. Next came married women, then single men and single women. This supports the idea that marriage leads to lower distress levels. As one might expect, overall, the adults with more children in the household had higher average distress. There were some deviations from these overall trends when the number of children and marital status were combined to make categories. For married men and women with 2 children or more, men showed slightly more distress than women, but the difference is minimal. Another deviation is that single men with 1 child, or more than 2 children, seem to have higher average distress than single women in the same categories, but with no children involved, single women have higher average distress. This suggests that single mothers may be better off (distress-wise) than their single father counterparts whereas single, childless women are worse off than their male counterparts.

When examining the difference between men and women where children are involved, as mentioned above multitasking is an important factor. Women spend more time caring for children, and when they care for children they are likely to combine it with an activity such as housework or some form of leisure. This takes away from "pure leisure" and increases tension levels with each additional activity done at once. This increased tension over time could cause more distress for married women who have children than for married men. The graph shows that for married people with children, men do indeed have less distress than married women, but as the number of children increase, married men have more distress than married women. This is supported by the literature on multitasking. When there are more children, both parents will have to spend more time on child care. Additional activities done at one time add more to men's tension levels than they do for women. This could be causing the additional children more distress for married men then for married women. In addition, this could be an explanation for why single fathers are worse off than their single mother counterparts.

An anomaly in the weighted averages of single men is that they report more distress with 1 child, and more than 2 children, than they do with 2 children. I think could be caused by the lack of observations for single men. The data used to make these graphs did not include those that had 0 individual weights. There were only 30 observations for single men with 2 children, 57 observations for single men with 1 child, and only 17 for single men with more than 2 children.

The next thing I decided to do was look at those people who had K6 scores of 13 or above, which is past the threshold for clinical significance. First, I pooled both sexes together, and tested whether being single or married was a predictor of having a distress level of 13 or above (i.e. having some sort of psychological disorder). Then I investigated whether or not being a parent or being childless was a predictor of having the distress level of 13 or above. One should note that parental status was determined by the number of births to the individual as opposed to the number of children living in the household. After pooling the sexes together, I studied these factors for males and females separately. Table 2 illustrates each contingency table with row percents, and associated chi-square statistics.

Table 2: Contingency tables and associated chi-square tests (row percents)

All	**Distress ≥13**	**Distress <13**	**Total**	**Pearson Chi-Square**
Single	159(7.32%)	2013(92.68%)	2172	36.0895
Married	152(3.81%)	3835(96.19%)	3987	**p-value**
Total	311(5.05%)	4068(94.95%)	6159	0.000

All	**Distress ≥13**	**Distress <13**	**Total**	**Pearson Chi-Square**
Parent	260(5.23%)	4712(94.77%)	4972	1.7388
Childless	51(4.30%)	1136(95.70%)	1187	**p-value**
Total	311(5.05%)	5848(94.95%)	6159	0.187

Men	**Distress ≥13**	**Distress <13**	**Total**	**Pearson Chi-Square**
Single	46(4.95%)	884(95.05%)	930	2.4538
Married	73(3.71%)	1895(96.29%)	1968	**p-value**
Total	119(4.11%)	2779(95.89%)	2898	0.117

Men	**Distress ≥13**	**Distress <13**	**Total**	**Pearson Chi-Square**
Parent	92(4.12%)	2143(95.88%)	2235	0.0025
Childless	27(4.07%)	636(95.93%)	663	**p-value**
Total	119(4.11%)	2779(95.89%)	2898	0.960

Women	**Distress ≥13**	**Distress<13**	**Total**	**Pearson Chi-Square**
Single	113(9.10%)	1129(90.90%)	1242	37.3143
Married	79(3.91%)	1940(96.09%)	2019	**p-value**
Total	192(5.89%)	3069(94.11%)	3261	0.000

Women	**Distress ≥13**	**Distress<13**	**Total**	**Pearson Chi-Square**
Parent	168(6.14%)	2569(93.86%)	2737	1.9265
Childless	24(4.58%)	500(95.42%)	524	**p-value**
Total	192(5.89%)	3069(94.11%)	3261	0.165

Only 2 of the tables showed a p-value that was statistically significant, when $\alpha = 0.05$. All in all, none of the tables which compared parenting status and distress levels showed any statistical significance. It appears that having children is not a factor in determining if someone will be over the clinically significant threshold. However, marital status has a p-value very close to 0, which means that there is some relationship between marital status and being above the threshold.

Looking at this relationship for men and women separately, the relationship is only significant for women. This could be because there is a smaller amount of data for men, or as mentioned above, there is some stigma for men in being “emotional”. Another reason the relationship might prevail for men rather than women is that it is less socially acceptable for women to remain single long into adulthood than it is for men to do the same. This is a reason that being single could cause increased emotional distress for women. This is also important to keep in mind when considering why single women with children may actually be less distressed than single men with children. If the woman has a child, she may be divorced, widowed, or has already had a serious relationship, and would not be considered a “spinster”. If this is a cause of distress for single women outside of parenting status, it may be lowered for single women with children, and might not be something that is different for single men with children. This could explain why distress levels are higher for single women than single men *without* children, but the same is not true when children are involved.

Age is another factor which could be at play here. The women who are unmarried are also likely to be younger than those who are married (by about 3 years on average for this data). Those between the ages of 18-24 have a higher risk of mental distress, and this could be a contributing factor to marital status to be a significant predictor of mental distress above the threshold.

It could also be the case that women are less likely to take on a commitment such as marriage when they are emotionally unstable. On the other hand, it could be that their emotional instability has ended many of their relationships or caused a divorce. Another way to look at it would be that being in a committed relationship such as the relationship between a married man and women lowers distress enough for certain women to keep them away from the clinically significant level.

The next thing I did was to run regressions using with the K6 variable as the dependent variable and the family life and work life variables mentioned above as the independent variables. The base group in these regressions is single childless males currently employed working between 30 and 50 hours a week, without a high school education and who have none of the medical conditions listed. The regressions are displayed in Table 3.

Table 3: K-6 Variable Regressions

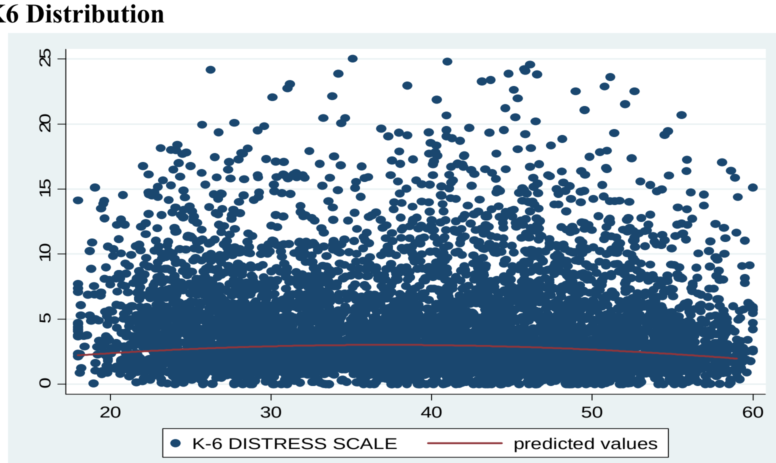
	Family Life Variables	Work Life Variables	Combined
Constant Variables			
Age	0.1155 (0.002)**	0.1922 (0.000)**	0.1632 (0.000)**
Age Squared	-0.0017 (0.000)**	-0.0025 (0.000)**	-0.0022 (0.000)**
Female	0.0485 (0.600)	0.1728 (0.068)*	0.1422 (0.136)
Selected Variables			
Births	0.1819 (0.000)**	___	0.0208 (0.627)
Children 6 or under	-0.0378 (0.744)	___	-0.1181 (0.295)
Children	0.0386 (0.469)	___	0.0532 (0.309)
Number of Others Supported	0.0169 (0.798)	___	0.1016 (0.116)
Marriages	-0.0496 (0.561)	___	-0.0233 (0.779)
Married	-1.1548 (0.000)**	___	-0.3343 (0.034)**
Divorced	-0.0879 (0.637)	___	0.0555 (0.760)
Log Income	___	-0.6058 (0.000)**	-0.5267 (0.000)**
Work Hours Less than 30	___	-0.2565 (0.024)**	-0.2489 (0.029)**
Work Hours Greater than 50	___	0.2214 (0.112)	0.1953 (0.162)
Unemployed	___	1.5457 (0.000)**	1.5123 (0.000)**
Student	___	0.1067 (0.798)	0.1294 (0.757)
Housewife	___	-0.1869 (0.287)	-0.1359 (0.447)

Disabled		1.4820 (0.000)**	1.5379 (0.000)**
Retired		0.7785 (0.047)**	0.8133 (0.038)*
High School		-1.1773 (0.000)**	-1.1422 (0.000)**
College		-0.2144 (0.053)*	-0.1999 (0.075)*
Learning Disorder	1.3676 (0.000)**	1.1092 (0.000)**	1.0955 (0.000)**
Mental Loss	3.8217 (0.000)**	3.0204 (0.000)**	3.0211 (0.000)**
Emotional Problem	3.1818 (0.000)**	2.9753 (0.000)**	2.9606 (0.000)**
Serious Health Problem	0.4626 (0.008)**	0.2848 (0.093)*	0.2718 (0.110)
Non-serious Health Problem	0.6894 (0.000)**	0.4768 (0.000)**	0.4656 (0.000)**
Constant	2.7783	8.1254	7.9808
R-squared	0.1437	0.1931	0.1955
Adjusted R-squared	0.1416	0.1907	0.1922

Number of observations: 6129
Each cell contains the beta coefficient and its significance in parentheses
** Significant at 5% level
*Significant at 10% level

One of the key variables I held constant in the K6 regressions was age. As mentioned above, age has a significant effect on the K6 distress scale. The scale has a U- shaped distribution with a maximum at age 37 for the combined regression. Graph 2 shows this distribution graphically.

Graph 2: Age and K6 Distribution



Interestingly, the differences between males and females I noticed in my earlier analysis were not statistically significant (at a 5% level) in the combined K6 regression. However, the coefficients for all regressions do show a positive effect for being female, as would be expected. In addition, with family life variables the female variable is clearly not at all significant, but for the work life variable regression this variable has a significant positive coefficient (0.17) at the 10% significance level. The difference between the distress levels of men and women become apparent when work life variables are held constant, and disappear when the family life variables are held constant. Clearly women are more significantly affected by the family life variables than men. This is supported by the earlier chi-square tests which found that being single is a significant predictor for being above the threshold for clinical significance in distress. It is also supported by the fact that women are more likely to multitask, which most of the time includes taking care of children or doing housework. The increased amount of multitasking leads to more tension for women which over time could be what is causing the significant difference in mental distress levels (Michelson, 2005).

To get a better idea of what's going on with respect to the differences in gender, one should examine the significant family life variables. The family life variables that are statistically significant are the number of births and current marital status. The number of births to the individual has a positive coefficient of .18 when the regression is run not holding constant the work life variables. When the family life and work life variables are combined the number of births is not statistically significant. This suggests that perhaps there is a work life variable that will negate the additional distress caused by an additional birth. Upon further runs of the family life regression, with each work life variable added by itself, it appears the education dummies high school and college are what causes the number of births to become insignificant. This seems to suggest that those who are well educated are better able to handle the added distress that is caused by additional children. This might mean that they are better prepared to have children and/or they had planned ahead if and when they would have them.

The most interesting part about this is that the number of births is significant, while the number of children under 18 is not significant, and neither is whether or not there are children living in the household under age 7. The difference between the number of children birthed and the number of children living in the household could be the result of children who are living in another household as a result of divorce, children who have died, or children who are 18 or older. This might suggest that it is not the age of the children, or whether or not they live in the household that has an effect on distress, but actually being a parent to the child. This would mean that multitasking for women (with regard to child care specifically) alone cannot account for the significant difference between men and women in distress levels.

As far as marital status goes, the negative coefficient in these regressions shows that being married decreases ones reported distress. This is in line with the previous analysis which showed that married people on average have less distress and are less likely to be above the threshold for clinical significance. When the work life variables are included the negative coefficient drops from 1.2 to .3. The income variable is measured on a family level, so perhaps additional income earned by spouses is what is diminishing the coefficient; however this is pure speculation at this point.

Overall, the R-squared value for the regression with work life variables is higher at .19 than it is for the regression with family life variables at .14. This suggests that the significant variables for work life could be more important determinants of distress levels. These variables include the log of income, weekly work hours, employment status, and level of education. The coefficients for the combined regression are not very different from the coefficients for the regression that included the work life variables alone, so in my discussion I'll refer to the coefficients from the combined regression.

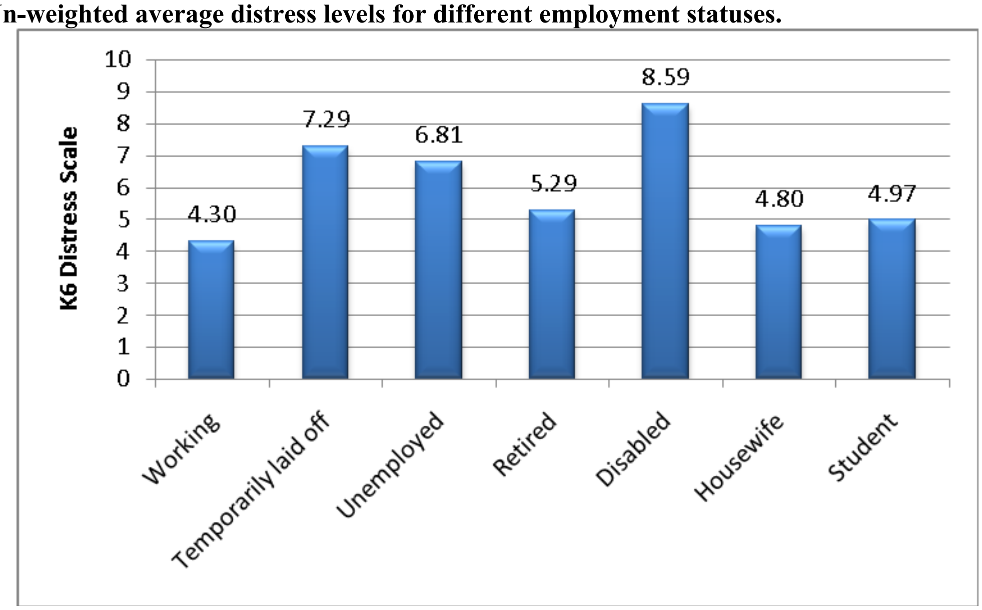
The log of income was significant, for each 1% increase in family income there is a -.53 point decrease on the K6 distress scale. It should be noted that this does not imply that more income makes one happier, but it does tend to decrease emotional distress, which includes anxiety or depression.

The weekly work hours were only significant for those who worked less than 30 hours a week, and not for those who worked more than 50 hours a week. Those who worked less than 30 hours a week had a .25 lower score on average, than those working between 30 and 50 hours a week. One would expect that working more than 50 hours a week would have a significant positive coefficient on the K6 scale. However, according to the study on multitasking done by Michelson, paid work is the least likely activity to be paired with other activities. This means that those who are busy working, are just that, busy working. They are not doing many different things at once and are thus less susceptible to the stress/distress that could be brought on by constant multitasking.

The level of education attained for these adults is also significant. Table 3 shows that high school graduates score on average 1.2 points lower on the K6 distress scale, while graduating college as well decreases the scale by .2 points more. This makes sense considering anxieties about getting or keeping a good job will be decreased when one has a high school diploma or college degree to fall back on. Many things can happen in life, but an education unlike a job, can never be lost.

Employment status is also important to look at. Being able to work, but not having a job adds on average 1.5 points to the K6 scale. This is interesting when compared with those who are out of the work force. Being a student or a housewife was not significant; while being permanently disabled had a coefficient of 1.5 and being retired added a coefficient of .8. So while being disabled or unemployed had similar coefficients, the other out of the work force variables did not. This may suggest that the frustration and worries of not being able to work and put food on the table cause this distress rather than a lack of things to do or something similar. However, boredom or a lack of satisfaction with the way one spends their time could be contributing to the positive coefficient for those who are disabled or retired, since housewives and students who are busy with keeping the home and getting an education. Graph 3 gives the un-weighted average distress levels for each different employment status studied.

Graph 3: Un-weighted average distress levels for different employment statuses.



Clearly the disabled and the unemployed have it worst. I think what could be happening with disabled people is that they may have the same concerns about getting food on the table that the unemployed do, combined with the boredom or restlessness that the retired or temporarily laid off people may have to contend with. Another thing to consider is that disabled people we well as the unemployed are probably not out of work voluntarily, while the other categories of people are. Over time these things would definitely lead to anxiety and depression which are measured by the K6 scale. The contributions of these factors also explain how much lower the distress rate is for those who are employed.

I added the health conditions to this regression to see if they had significant effects on the K6 scale, which indeed they did. Having a learning disorder added on average 1.1 points to the scale, while both having memory loss or having been previously diagnosed with an emotional problem added approximately 3 points to the scale. These are quite large coefficients when one considers the median point value on the scale is a 4. However, it is to be expected that prior mental and emotional health diagnoses will be a good predictor of current conditions. All of the health condition variables have positive coefficients and thus contribute to mental distress, with the exception of serious medical conditions for the combined regression. In the combined regression (as well as the work-life regression), serious medical conditions have a positive coefficient of around .3 while the non-serious medical conditions contribute about half a point to the distress scale. This makes sense since the non-serious conditions tend to be the more chronic conditions. It would be these conditions and their symptoms that would be causing anxiety or distress on a daily basis rather than a heart attack that has already happened. The problems in these categories are all physical, so there is a relationship between physical and mental health states in these regressions.

Health Status Scale:

The health status variable was collected by asking the head and wife of the household to rate their health in general on a scale from 1 to 5. The points were defined as follows: 1 is Excellent, 2 is Very Good, 3 is Good, 4 is Fair, and 5 is Poor. I ran logistic regressions by first creating 2 dummy variables. The first is an indicator of good health, which I took to mean a score of 1 or 2 on the health status scale. The second is an indicator of poor health with a score of 5 on the health status variable. The base group for both regressions is those who scored a 3 for Good or a 4 for Fair. Table 4 shows the results of the regression including the odds ratios and the significance level for each cell.

Table 4: Logistic Regression of Health Status Variable

Dependent Variable	Good Health			Poor Health		
	Family Life Variables	Work Life Variables	Combined	Family Life Variables	Work Life Variables	Combined
Constant Variables						
Age	0.9882 (0.000)**	0.9767 (0.000)**	0.9777 (0.000)**	1.0345 (0.001)**	1.0375 (0.000)**	1.0355 (0.001)**
Female	0.8028 (0.000)**	0.8392 (0.001)**	0.8543 (0.002)**	0.7552 (0.081)*	0.6528 (0.011)**	0.6715 (0.020)**
Learning Disorder	0.5934 (0.003)**	0.6507 (0.019)**	0.6542 (0.021)**	1.2648 (0.495)	0.7854 (0.510)	0.7987 (0.545)
Mental Loss	0.4266 (0.007)**	0.5668 (0.093)*	0.5689 (0.095)*	3.2977 (0.000)**	2.0131 (0.025)**	2.0697 (0.021)**
Emotional Problem	0.5402 (0.000)**	0.5639 (0.000)**	0.5623 (0.000)**	2.9540 (0.000)**	2.5869 (0.000)**	2.4818 (0.000)**
Serious Health Problem	0.3929 (0.000)**	0.3843 (0.000)**	0.3824 (0.000)**	1.6514 (0.007)**	1.4004 (0.081)*	1.4246 (0.069)*
Non-Serious Health Problem	0.3405 (0.000)**	0.3572 (0.000)**	0.3551 (0.000)**	2.8879 (0.000)**	2.4827 (0.000)**	2.5138 (0.000)**
Selected Variables						
Births	0.8688 (0.000)**	______	0.9635 (0.102)	1.1050 (0.062)*	______	0.9897 (0.861)
Children 6 or under	1.1913 (0.002)**	______	1.2391 (0.000)**	1.4434 (0.063)*	______	1.2130 (0.347)
Children	1.0081 (0.762)	______	0.9993 (0.981)	1.0442 (0.589)	______	1.0522 (0.534)
Number of Others Supported	0.9661 (0.320)	______	0.9099 (0.008)**	0.7816 (0.091)*	______	0.9332 (0.631)
Marriages	1.0172 (0.697)	______	1.0388 (0.397)	1.2158 (0.079)*	______	1.2268 (0.085)*
Married	1.8008 (0.000)**	______	1.1659 (0.080)*	0.5675 (0.020)**	______	1.0753 (0.790)
Divorced	1.1898 (0.072)*	______	1.0438 (0.670)	0.5352 (0.026)**	______	0.6733 (0.179)
Log Income	______	1.3792 (0.000)**	1.3395 (0.000)**	______	0.7611 (0.001)**	0.6999 (0.000)**
Work Hours Less than 30	______	1.0626 (0.339)	1.0565 (0.389)	______	0.6881 (0.094)*	0.6976 (0.108)
Work Hours More than 50	______	1.2015 (0.017)**	1.2233 (0.009)**	______	1.0056 (0.988)	1.0365 (0.923)
Unemployed	______	0.7422	0.7314	______	4.0390	3.8427

		(0.000)**	(0.000)**		(0.000)**	(0.000)**
Student	____	1.4844 (0.105)	1.5332 (0.081)*	____	2.8255 (0.180)	2.8517 (0.176)
Housewife	____	1.6654 (0.071)*	1.6429 (0.080)*	____	____	____
Disabled	____	1.7431 (0.000)**	1.7325 (0.000)**	____	____	____
Retired	____	1.6318 (0.000)**	1.6325 (0.000)**	____	0.7718 (0.726)	0.7210 (0.661)
High School	____	1.7109 (0.000)**	1.6585 (0.000)**	____	0.5136 (0.000)**	0.5099 (0.000)**
College	____	1.9240 (0.000)**	1.8958 (0.000)**	____	1.0043 (0.989)	1.0269 (0.931)
Pseudo R-squared	0.0980	0.1371	0.1400	0.1382	0.2199	0.2258
Number of Observations			8779			3573

Each cell contains the odds ratio and its significance in parentheses
**Significant at the 5% level
*Significant at the 10% level

The variables which were controlled for in these regressions were age, sex, and having any of the already mentioned medical conditions. All of these variables were statistically significant in each regression. As is expected, age has a negative effect on reported health status with each year decreasing the odds of reporting a good health status, and increasing the odds of reporting a poor health status. Interestingly, while sex was not significant for the K6 distress scale, it is significant for the health status scale. Women have about 15% lower odds of reporting good health, and 33% lower odds of reporting poor health (taken from the combined regressions). Indeed for this data, about 40% of women studied reported Good or Fair health, while this was true only for 34% of men. Mental health problems were significant in all of the regressions, at least at the 10% level. Each of these variables increased the odds of reporting poor health and decreased the odds of reporting good health. Both serious and non-serious health conditions decreased the odds of reporting a good health status by around 62-65% while increasing the odds of reporting poor health by 42% for serious conditions and 250% for non-serious conditions.

As with the K6 scale, the number of births to the individual was significant. Each additional birth leads to 13% lower odds of reporting good health for the regression with only family life variable. In addition, for reporting poor health, this variable is only significant (with 11% higher odds for each birth) when work life variables are not controlled for, so it could be the income and amount of work necessary that contributes to more births making one more likely to report poor health. In contrast to the K6 variable, having a child 6 or under is significant for reporting a good health status. It gives one 19% higher odds of reporting a good health status when work life variables are not held constant and 24% higher odds when work life variables are accounted for. It should be kept in mind that age has been controlled for so this is not an effect that is due to people with younger children being younger and having better health. An interesting anomaly is that having at least 1 child age 6 or younger also increases the odds of reporting poor health by 44%, for the family life variable regression. However this is not significant at the 5% level, so it may be a coincidence.

Marital status is significant for both reporting good health and reporting poor health, but only in the family life regressions and not in the combined regressions. This is a confusing effect since both being married or divorced(not remarried) have lead to higher odds of reporting good health and lower odds of reporting poor health. This would seem to suggest having never been married would lead to worse self reported health, but this is difficult to tell since there is not a variable for those who are widowed, and the number of marriages variable is insignificant. Since these variables (particularly being married) are not significant once work life variables are accounted for this may have something to do with income. An interesting thing is that for the good health regression, while being divorced becomes insignificant when income is accounted for, the number of others becomes significant with 9% smaller odds of reporting good health.

Looking at the work life variables, comparing the pseudo R-squared values implies that for both reporting good health and reporting bad health, work life variables are more important than family life variables. Indeed, for poor self reported health, once work life variables are accounted for all of the family life variables become insignificant.

Income, just as it has with K6 has a positive effect on health status (a negative coefficient for K6), as does employment. This makes sense since health insurance coverage is not controlled for in this regression, and intuitively being employed and having a higher income will give one access to better health related opportunities such as expensive produce, multivitamins, and gym memberships. Weekly work hours were only significant in the good health regression, with those working more than 50 hours a week being 22% more likely to report good health. This is also probably due to having better insurance coverage than those who are unemployed or working part time.

Education has a positive effect as well. Those who have graduated from high school report good health status statistically more often than those who have not, and are significantly less likely to report poor health. Those who have graduated college are even more likely to report good health, and are not significantly different from those who graduated high school in reporting poor health. Since employment and income are held constant, this may imply that those who have graduated high school are more likely to have learned how to take care of themselves through education, either in school itself, or through being able to understand the bodily processes that lead to poor health and avoiding them.

Interestingly, in every regression the unemployed had significantly greater odds (around 400%) of reporting poor health, and smaller odds (around 27%) of reporting good health. This could stem from a lack of insurance, since an increased likelihood of reporting poor health is not seen for retired people or students who are likely covered through other means (i.e. Medicare or parents). In addition, being a housewife or a student have no effect on good health, while being retired or disabled have 63% and 74% greater odds of reporting good health respectively. For the poor health regression, none of the housewives or disabled people studied reported bad health. This is interesting; housewives most likely have health care coverage through their husbands, while those who are disabled may be insured through government programs. Those who are unemployed may not have insurance that provides enough from them to maintain their health, or they may be sick enough to not be able to work, but still are not considered permanently disabled.

Feeling Rushed Variable:

The next variable I regressed was the feeling rushed variable mentioned above. The question asked in the interview was how often to do you feel rushed or pressed for time. The answers were given on a scale from 1 to 5 with possible answers of almost always (1), often (2), sometimes (3), rarely (4), and almost never (5). To do the regressions I split up those who were rushed often which included scores of 1 or 2, and those who were not rushed often, which included scores of 4 or 5 and compared both groups with the people who reported feeling rushed sometimes(3).

Table 5: Logistic Regression of Feeling Rushed Variable

Dependent Variable	Rushed Often			Not Rushed Often		
	Family Life Variables	Work Life Variables	Combined	Family Life Variables	Work Life Variables	Combined
Constant Variables						
Age	1.0397 (0.120)	1.0322 (0.177)	1.0203 (0.433)	0.9807 (0.505)	0.9560 (0.105)	1.0013 (0.966)
Age Squared	0.9992 (0.012)**	0.9992 (0.011)**	0.9994 (0.057)*	1.0004 (0.316)	1.0007 (0.036)**	1.0001 (0.696)
Female	1.2229 (0.002)**	1.4148 (0.000)**	1.4051 (0.000)**	0.8815 (0.106)	0.7648 (0.001)**	0.8133 (0.012)**

Selected Variables						
Births	0.9831 (0.553)	________	1.0379 (0.221)	1.0598 (0.068)*	________	0.9948 (0.877)
Children 6 or under	1.2174 (0.010)**	________	1.2356 (0.007)**	1.1993 (0.069)*	________	1.1461 (0.182)
Children	1.0917 (0.017)**	________	1.0878 (0.024)**	0.8893 (0.012)**	________	0.8721 (0.004)**
Number of Others Supported	1.1156 (0.019)**	________	1.0763 (0.122)	0.9815 (0.749)	________	1.0096 (0.873)
Marriages	1.1336 (0.025)**	________	1.1220 (0.044)**	0.9429 (0.382)	________	0.9179 (0.217)
Married	1.1339 (0.204)	________	0.9200 (0.438)	0.7127 (0.004)**	________	0.9786 (0.869)
Divorced	1.0716 (0.547)	________	0.9975 (0.983)	0.8790 (0.336)	________	0.9825 (0.898)
Log Income	________	1.1702 (0.000)**	1.1608 (0.001)**	________	0.8333 (0.000)**	0.8554 (0.003)**
Work Hours Less than 30	________	1.0292 (0.718)	1.0216 (0.790)	________	1.1031 (0.313)	1.1042 (0.311)
Work Hours More than 50	________	1.6937 (0.000)**	1.6678 (0.000)**	________	1.1918 (0.198)	1.1893 (0.206)
Unemployed	________	0.6632 (0.000)**	0.6325 (0.000)**	________	1.4934 (0.000)**	1.5444 (0.000)**
Student	________	0.7881 (0.381)	0.7760 (0.353)	________	0.5363 (0.100)*	0.5557 (0.124)
Housewife	________	1.1000 (0.747)	1.1130 (0.718)	________	0.5741 (0.246)	0.5644 (0.233)
Disabled	________	0.9904 (0.950)	0.9877 (0.937)	________	0.5682 (0.018)**	0.5608 (0.016)**
Retired	________	0.9854 (0.916)	0.9762 (0.863)	________	0.8617 (0.435)	0.8719 (0.475)
High School	________	1.0178 (0.850)	1.0691 (0.483)	________	0.8961 (0.290)	0.8466 (0.116)
College	________	1.0736 (0.361)	1.1131 (0.175)	________	0.7787 (0.017)**	0.7606 (0.010)**
Pseudo R-squared	0.0178	0.0294	0.0339	0.0140	0.0276	0.0337
Number of Observations			4594			2960

Each cell contains the odds ratio and its significance in parentheses
**significant at the 5% level
*significant at the 10% level

Both age and sex were controlled for, and this was beneficial since both variables had significant effects on which category of the rushed variable one fell under. Women were far more likely to feel rushed with around 41% greater odds of feeling rushed often, and around 19% lesser odds of not feeling rushed often for the combined regressions. Interestingly, when the family life variables are regressed, without accounting for work life variables the odds for women feeling rushed decrease to 22% greater odds and being female becomes insignificant for the poor health variables. So once work life variables are accounted for women become even more likely to feel rushed often than men. This would support the idea that family life variables are more important for women.

Age is a quadratic function with respect to the rushed variable, with a minimum at the age of 31. A typical explanation would be that people at this age have better jobs than their counterparts in their 20s, but are still not quite at the top of the career ladder where they can relax more often. However, these results are seen in all of the regressions for feeling rushed as well as the work life regression for not feeling rushed.

The number of births was significant for not feeling rushed, but was not significant for feeling rushed. In addition, it was only significant for the regression in which work life variables were not held constant, so it would seem that the 6% greater odds of not feeling rushed with each additional birth actually has to do with work life somehow. If the odds were lower odds, this would make sense since the work life variables might account for the extra work needed to support more children. However, since the odds are greater of not feeling rushed with each additional birth and 11% lower with another child(under 18) living in the house, perhaps this is being caused by children above the age of 18 living in the household and contributing to income or caring for the younger children and thus lowering the stress for the parents.

The number of children under 18 living in the household was significant in the regression for feeling rushed. Each additional child led to around 9% higher odds of being rushed often for both the regression with only family life variables and the combined regression. This is in line with what one would intuitively expect since children need to be fed, have their clothes washed, live in a clean house, be taken to school, etc. These duties most often fall on parents and thus increase how often they feel rushed. This also has to do with multitasking, since childcare is one of the activities with which the most multitasking occurs (Michelson, 2005). More children will lead to more things happening at once for most people and thus make them more likely to feel rushed. As was mentioned above, children have a parallel effect on the not feeling rushed variable, in that each additional child leads to 11% smaller odds of not feeling rushed.

The dummy variable for having children under 7 is significant with 24% greater odds of feeling rushed often.. This makes sense when we consider what was previously said about multitasking; the younger the child the more time will be spent caring for them, which means more multitasking and more feeling rushed.

The number of others supported is significant in the regression for feeling rushed often, both when only family life variables are considered (12% greater odds). An explanation could be that supporting others costs more money and those who support more people have work more often and thus the significance disappears when weekly work hours are controlled for.

The number of marriages is significant for the regression of feeling rushed often. Each additional marriage leads to 12% greater odds of being rushed often. This is not showing up in the variable for divorce, but the variable for divorce does not count those who have divorced and are remarried, so it could be these people who are making the difference, by perhaps having different family members in different locations.

Being married is significant for the not rushed often variable. It decreases the odds of not feeling rushed by 29% in the family life regression, but this significance disappears when work life variables are controlled. This leads me to believe that the rushing around caused by being married could actually be caused by having to work more often and a more stressful job in order to make the income to support another person. This makes sense to me even if both spouses are working since married people might consider themselves "grown up" and are likely to move out of bachelor pads and apartments and into houses that require mortgage payments and upkeep costs.

Work life variables are more significant in these regressions as well since the relative R-squared values are higher for the regressions with work life variables alone than for those with family life variables alone. The log income has a big effect with each 1% increase in income leading to 17% greater odds of feeling rushed often (17% smaller odds of not being rushed often)when family life variables aren't accounted for and 16% greater odds of being rushed often(14% smaller odds of not) when they are. This is most likely because those who make more money work more often and have more stressful jobs. The significance in a college education (22% lesser odds of not feeling rushed often) is also captured here since those with college educations probably have more stressful jobs even if income is controlled for since their jobs require an education and thus probably more mental skill.

Employment status is interesting because it shows that being unemployed significantly reduces the odds of feeling rushed by 37% and increases the odds of not feeling rushed by about 54% (taken from the combined regressions). However, the same is not true for those who are out of the work force, which did not significantly reduce the odds of being rushed, or increase the odds of not being rushed. This is most likely caused by housewives caring for children and students going to school or retirees keeping busy doing enough to feel rushed at times. The exception is those who are disabled had 44% lesser odds of not feeling rushed. Perhaps the disabled feel more inclined to be rushed at least some of the time since they may take longer to do everyday tasks.

Conclusions:

The analysis of each of these variables gave an interesting picture for work life and family life in relation to physical and mental health. One can conclude that the work life variables are more important in predicting each of these physical and mental health variables, but family life variables should not be discounted. For mental distress, the work life variables of log income, employment status, and weekly work hours dominate, but family life variables such as marital status and number of births take on significant importance, especially for women. It its very clear a big part of the difference that can be observed in the average distress levels of men and women has to do with the family life variables and not work life. So to take care of our mental health it is important for everyone to devote some time to planning their work life. For women, planning out family life is also very important. Women should be aware that more children will likely lead to more multitasking in their lives and increased tension, while finding a partner and getting married will decrease their distress levels during their adult lives.

Overall, it is clear that income and employment status are the most important indicators of the health that one reports. Weekly work hours are also important, however income has a greater effect on the odds of reporting good or bad health. In further analysis health insurance coverage will need to be accounted for get a clearer picture of what is going on with these variables. Family life variables did not do much to explain the differences in health status, and one should take care to find a good career if they are worried about their future health status.

The feeling rushed variable showed that work life variables were more important when measuring stress, if one considers feeling rushed being stressed. However, the family life variables had a similar R-squared, and the combined regression had a much larger R-squared value than either of the other two. As one might expect, the number of children as well as income and the number of hours worked during the week were all important when considering stress levels. More children and more work hours lead to more stress, while a higher level of income leads to less.

Since there were many significant variables in these regressions, I think there is still much to uncover regarding the relationship between lifestyles both personal and professional and health both physical and mental.

References

Michelson, William (2005). “Patterns beneath the Surface: The Texture of Multitasking,” Time Use: Expanding the Exploratory Power of the Social Sciences. Boulder, CO: Paradigm Publishers, pp. 122-136.