Theory and Evidence…..

An Evaluation of Investor Sentiment Measures

by

Zhou Jiang

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science

Undergraduate College

Leonard N. Stern School of Business

New York University

May 2011

Professor Marti G. Subrahmanyam	Professor Jeffrey Wurgler
Faculty Adviser	Thesis Advisor

Abstract: This paper evaluates investment sentiment measures by testing their predictability for future returns. Different measures are combined to form sentiment indices to also test their predictive power. Correlations between future returns of different portfolios and the sentiment measures and indices are calculated and tested for significance. This study finds that combining different sentiment measures create better proxies for sentiment and contain significant predictive power for monthly returns, especially for small stock portfolios.

I especially thank professor Jeffrey Wurgler for his help in making this paper possible. I also thank professor Gregory Brown and the Conference Board for providing their data for this paper. I appreciate the Stern School of Business for organizing the Honors Program.

Table of Contents

I. INTRODUCTION

Investor sentiment has been a myth to measure until two decades ago with the emergence of behavioral finance theories. The reason is the classical finance theory posits that the effects of sentiment on stock prices could be easily arbitraged away by rational traders and investors. However, numerous research have concluded that the effects of sentiment could persist in stocks given the nature of uncertainty in some stocks.

DeLong, Shleifer, Summers, and Waldmann (1990) suggest that some assets are more heavily traded by individual investors and have such high transaction costs so that the stock prices of these assets are more influenced by sentiment and these influences are not easily arbitraged away. Lee, Shleifer, and Thaler (1991) went on to prove that closed-end fund discount could be used as a negative predictor of sentiment.

It is not until the last decade, however, when an abundance of individual proxies for sentiment has been studied and researchers have started to think of combining these proxies to form sentiment indices that might better measure sentiment. Baker and Wurgler (2006) combined six previously studied sentiment proxies to form a sentiment index and tested it against various portfolios of stock returns. In particular, they found that this index correlate with next period returns of smaller and younger stocks particularly significantly. Brown and Cliff (2004) also combined twelve direct and indirect measures of sentiment to form an index and tested it on both contemporaneous and near-term stock returns. What they found is that their index correlates strongly with contemporaneous returns but not with future returns. Despite the results, both papers point to the robustness of combining different sentiment proxies together to generate a better measure of investor sentiment.

This study seeks to draw on these two papers and makes use of some of their variables to test again a more expansive set of data and extend the data set by at least ten years. This paper also tests two survey-based measures of sentiment that are not widely tested as sentiment proxies. This study offers two primary findings:

1) **Return Predictability:** All the variables are tested individually against future stock returns to find significant correlations. They are also combined in different ways to form sentiment indices. It is found that combining different variables may create better sentiment measures as they correlate more significantly and highly with future returns, especially those of small stocks and the spread between small and large cap stocks.

2) **Long Term Reversal Predictability:** This study does not find sentiment indices to be predictive of long term reversal in stock returns as in Baker and Wurgler (2006).

II. LITERATURE REVIEW AND HYPOTHESIS

A. Literature Review

Numerous recent studies have considered the measurability of investor sentiment and its predictive power for stock returns. Although many of them use very distinctive proxies to measure investor sentiment, their common conclusion is that sentiment can be measured and correlates highly with stock returns in both in general and cross-sectionally.

Baker and Wurgler (2006) constructed a sentiment index to test its predictive power on returns of portfolios of stocks differentiated by characteristics such as size, profitability, volatility and others. They formed the index based on the first principle component of six variables, namely the closed-end fund discount, NYSE turnover, the number and average first-day returns on IPOs, the equity share in new issues and the dividend premium. They found that the returns of different portfolios could be predicted by the beginning of the year sentiment index number. Specifically, stocks that are younger, smaller, unprofitable, non-dividend paying, highly volatile, extremely high growth and distressed tend to have lower subsequent returns when the sentiment is estimated to be high. The reasoning is that these stocks are more likely to attract the attention of optimists and speculators who buy on the hype of stocks and leave after the hype is over. This research is valuable in confirming the sentiment components of the variables used in their sentiment index and serves as a basis for proxies of investor sentiment in this research.

Brown and Cliff (2004) used very similar techniques to analyze the relationship of investor sentiment and stock returns. They also formed a sentiment index using rather different proxies, however. These include: survey data from Investor Intelligence; technical indicators such as the ratio of the number of advancing issues to declining issues; variables that relate to particular types of trading activity such as margin borrowing, the percent change in short interest and the ratio of specialist's short sales to total short sales; and many others. They found that there is strong evidence of co-movement of the sentiment measures with the market but little evidence of short-run predictability in returns. Although this result is in contradiction with Baker and Wurgler's analysis, this research tries to combine with the sentiment proxies used in both papers in an effort to try to extract the sentiment components of these proxies more effectively.

Other research studying the predictive power of investor sentiment generally focuses on specific proxies. For example, Fisher and Statman (2000) used Wall Street strategists' mean allocation to stocks in their recommended portfolios as a proxy for sentiment of large investors and found it to have a negative relationship with S&P 500 returns. Kaniel, Saar and Titman (2004) constructed a daily measure of investor sentiment with the imbalances in the orders of individuals in NYSE and found that it has a strong predicative power of future returns. However, many of these papers have proprietary data that this paper is not able to obtain so they are not included in the analysis. Nonetheless, this paper represents the most comprehensive analysis of investor sentiment yet performed.

B. Hypothesis

This study examines whether investor sentiment can predict future stock returns, especially those that are theoretically more subject to sentiment, such as smaller, younger and high growth stocks. The aggregate investor sentiment could be measured by extracting common components of different sentiment measures in previous research. Also, we want to test whether investment sentiment measures could predict long term reversal in stock returns, which was suggested by Baker and Wurgler (2006).

III. DATA

Several variables are taken from Baker and Wurgler (2006) and Brown and Cliff (2004) and combined together. No one of them is uncontroversially predictive of returns so this paper forms a composite sentiment index comprising the following variables: the closed-end discount, NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, the dividend premium, the Conference Board Consumer Confidence Index, the University of Michigan Consumer Sentiment Index, advance decline ratio, margin debt, short interest and specialist short sales.

The closed-end fund discount, *CEFD*, is the average difference between the net asset values of closed-end stock fund shares and their market prices. Numerous works cited in Baker and Wurgler (2006) suggests that CEFD is inversely related to sentiment. This paper takes the value-weighted average discount on closed-end stock funds for 1962 through 1993 from Neal and Wheatley (1998), for 1994 through 1998 from CDA/Wiesenberger, for 1999 through 2001 from turn-of-the-year issues of the Wall Street Journal, and for 2002 through 2010 from the ETF Connect website.

NYSE share turnover is based on the ratio of reported share volume to average shares listed from the *NYSE Fact Book*. Jones (2001) finds that high turnover forecasts low market returns.

The number of IPOs, *NIPO*, and the average first-day returns, *RIPO*, are taken from Jay Ritter's website, which updates the sample in Ibbotson, Sindelar, and Ritter (1994). Both of these series were found to be positively correlated with sentiment in Stigler (1964) and Ritter (1991).

The share of equity issues in total equity and debt issues, *S,* is another measure of financing activity that may capture sentiment. Baker and Wurgler (2000) find that high values of the equity share predict low market returns. The equity share is defined as gross equity issuance divided by gross equity plus gross long-term debt issuance using data from the *Federal Reserve Bulletin*.

Dividend premium, *PDND*, is the log difference of the average market-to-book ratios of payers and nonpayers. Baker and Wurgler (2004) use this variable to proxy for relative investor demand for dividend-paying stocks. Given that payers are generally larger, more profitable firms with weaker growth opportunities (Fama and French (2001)), the dividend premium may proxy for the relative demand for this correlated bundle of characteristics.

The Conference Board Consumer Confidence Index, *CCI* and the University of Michigan Consumer Sentiment Index, *UMCSI*, are two direct measures of consumer sentiment that use surveys to find out about the public's view on the economy and business environment. By their construction, they are likely to be positively correlated with sentiment, at least consumer sentiment. These could be obtained from the publishers' websites.

The ratio of the number of advancing issues to declining issues, *ADVDEC*, is obtained from Alex Matulich / Unicorn Research Corporation that contains the daily historical data for shares on NYSE. *ARMS* is a modification of *ADVDEC* which standardize the number of advances and declines by their volumes.

The percent change in margin borrowing, *MD*, is obtained from the *NYSE Factbook*. This measure is frequently cited as a positive sentiment indicator since it represents investors using borrowed money to invest. It was tested in Brown and Cliff (2004).

The percent change in short interest, *SI*, is also obtained from the *NYSE Factbook*. It is generally understood to be a bearish indicator, as proposed by Brown and Cliff (2004), because more sophisticated investors tend to take on the short side more often.

The ratio of specialists' short sales to total short sales, *SPSS*, is obtained from the *NYSE Factbook* as well. Brown and Cliff (2004) see it as a bearish indicator because specialists are supposedly well-informed and relatively savvy investors, so when their short-selling becomes relatively large, the market is likely to decline.

For all the variables used, the data is detrended by a moving. For each variable, I regress up to 12 lags and define the optimal detrending length as the regression with the highest adjusted R^2, (marked by a "dt" behind the name). This is to eliminate some of the correlation of the variables with time rather than with other variables. Each variable is also standardized (marked by an "s" in front of the name). For example, *SARMSD11* denotes *ARMS* detrended by a moving average of 11 months, and then standardized.

Outliers that are more than four standard deviations away from the mean are censored at 4 or -4. Both the original and detrended data are examined to find the more suitable candidate to form a sentiment index. (Refer to Figure 1.)

IV. EMPIRICAL RESULTS

A. Data Summary

Most of the data are available since 1960s and end not before 2007. Table 2 shows the pairwise correlation of the variables. Also, Figure 1 shows the comparison of variables and their detrended, standardized and censored versions to explore whether detrending is more suitable for each variable.

From Figure 1, we can see that PDND, TURN, S and SPSS show rather apparent trends as with time. Others show no significant trend.

TURN's time trend could be easily explained by the development in technology of electronic trading and global integration of financial markets which increased trading activity drastically, while the number of shares listed increased at a lower rate.

PDND, S and SPSS have no theoretical basis for having a time trend.

Since some variables show little correlation with time, we would also test them together with the detrended ones to see if they form a more predictive sentiment index.

As expected, the most noticeable trend is the strong correlation between the standardized data and their detrended, standardized and censored counterparts, shown in Table 2. Also, except for *SS*, the predictors used in Baker and Wurgler's sentiment index all show relatively strong correlation with *SPDND*. There is also strong correlation between the two consumer confidence survey indices, *SUMCSI* and *SCCI*. There are some other rather strong correlations but there are no straightforward theoretical explanations behind them.

B. Factor Analysis

From the various versions of variables, we first wanted to get a general combination of all variables to form sentiment indices that include as many variables that we included here as possible. However, one variable, *UMCSI*, has many fewer data points compared to other variables so it is dropped out in the factor analysis to reduce the amount of data lost when forming sentiment indices. All of the different principal components (PCs) listed are the first principal component from factor analysis on different sets of variables. All these are displayed in Table 3 and plotted in Figure 2.

PC1 is simply a combination of all the variables in their standardized forms while PC2 is the detrended, standardized and censored version of PC1. PC3 is a combination of PC1 and PC2. In Figure 2, we can see the graph of these three PCs. It is interesting to note that PC3 varies highly with PC1. This is logical since PC2 which consists of the detrended variables that should show less variation.

There is a fair mix of expected and unexpected signs of the correlation of the different predictors and PC1. Four out of the thirteen predictors turn out to have their expected signs. They are namely, *SNIPO*, *SRIPO*, *STURN*, and *SMD*. The rest have opposite signs from their theoretically suggested signs. However, six out of the thirteen predictors have the expected correlation signs with PC2 which is a good improvement. They are namely, *SPDNDD2*, *SNIPOD12*, *SRIPOD12*, *SCCID9*, *SMDD6*, and *SADVDECD12*. This shows that detrending, standardizing and censoring does improve the quality of the predictors. Surprisingly, PC3 shows more expected correlations with predictors. 18 out of 26 predictors show the correct signs of correlation with PC3. This suggests that the combination of more predictors make PCs more reliable and in line with theoretical reasoning.

PC4 is a combination of only two variables, which are the two surveys of consumer confidence. Both predictors correlate with PC4 with the correct sign. This is reasonable as only two predictors which correlate highly with each other are used here. This again confirms the expectation that consumer sentiment should correlate positively with investor sentiment.

PC5 and PC6 are formed from two versions of variables, standardized and detrended/standardized, taken from Brown and Cliff (2004). Only one predictor in PC5, *SMD*, and two predictors in PC6, *SMDD6* and *SADVDECD12*, have the correct signs. This might be the reason why they did not find their sentiment index to predict near term stock returns as the make-up of their principal component is not consistent with theoretical reasoning.

PC7 is a combination of variables that show the most significant correlations with forward equal-weighted returns and small cap returns. The variables are *SPDNDD2, SNIPOD12, SRIPOD12, SS, SCCID9, SMDD6* and *SADVDECD12*. This could be seen in Table 4 where single variables are correlated with the different portfolios of returns.

Correlations of the predictors with PC7 all show the expected signs. This makes PC7 the most robust PC so far.

Not included in table 3 is another sentiment index, PC8, which we also want to test. It is formed following the original Baker and Wurgler (2006) method. (It includes raw variables and lagged variables so it could not be shown in Table 3.) It includes six variables which are *S*, *CEFD*, *PDND_LAG12*, *RIPO_LAG12*, *NIPO*, and *STURND12_LAG12*. It was suggested in their paper to have significant negative correlation with future returns and we tested it with the benefit of additional and more recent data. The signs of the correlations are found to be negative for *S*, *RIPO_LAG12*, *NIPO* and *STURND12_LAG12* and positive for *CEFD* and *PDND_LAG12*. Interestingly, only two variables, *S* and *STURND12_LAG12*, show the expected signs.

C. Test of Return Predictability

To test whether any sentiment measure is predictive of future returns, we examine its correlation with the following month's returns of six different portfolios. The significance (p-value) of the correlation is also reported.

The six portfolios all come from the combination of stocks in the New York Stock Exchange, American Stock Exchange (currently merged with NYSE) and Nasdaq Exchange. The portfolios of returns include: value-weighted market returns, *FVWR*; equal-weighted market returns, *FEWR*; small cap returns, *FCAP1R* (first decile of portfolios categorized by size); large cap returns, *FCAP10R* (tenth decile of portfolios categorized by size); and the difference between small cap and big cap returns, *FCAP1CAP10R* (sometimes abbreviated with a "~" in tables).

The results of single variable's correlation with returns can be found in Table 4. The returns that are in general the most predictable are *FEWR*, *FCAP1R* and *FCAP1CAP10R*. This is consistent with our hypothesis that these sentiment proxies are better at predicting small stocks that are theoretically more likely to be influenced by investor sentiment.

In addition, 16 out of the 26 variables are showing expected signs of correlations with *FCAP1R*. Out of the 17 significant correlations (p-values of at most 0.01) with *FCAP1R*, 11 predictors are showing the expected sign. These predictors are namely *SPDNDD2*, *SRIPO*, *SRIPOD12*, *SCEFDD9*, *SS*, *SSD10*, *SSI*, *SMD*, *SMDD6*, *SADVDEC*, and *SADVDECD12*. Detrending turns the sign of one variable, *SNIPO*, to the expected sign and another variable, *STURN*, to the unexpected sign.

It is interesting to note that *SPDNDD2* and *SADVDECD12* are showing very high correlation with small cap returns of greater than 20% and both have p-values that are close to 0. This means they are relatively good predictors of small stock returns by themselves.

In general, detrending seems to make the variables from Brown and Cliff (2004) more significantly correlated with small cap returns whereas it does the opposite for the confidence surveys. It only makes some Baker and Wurgler (2006) variables more significantly correlated with returns.

We then performed the same test with the PCs. The PCs that are most significantly correlated with small cap returns are PC2, PC3, PC6, PC7 and PC8, as shown in Table

5.

It is surprising that merely combining all the detrended versions of variables in PC2 could produce significant and high correlations of 24.17% with small cap returns and 26% with small minus big cap returns. This shows that combining the variables using principal component analysis is useful in extracting the sentiment-measuring properties from each proxy. This again is confirmed by PC3 which includes all versions of variables.

For PC6 and PC8, this test confirms previous results of Baker and Wurgler (2006) and Brown and Cliff (2004) that the sentiment indices they create hold significant predictive power for small cap returns.

It is reasonable for PC7 to have the highest correlation since it is composed of all the variables that by themselves have significant correlations (p-values below 0.05) with small cap returns. This again confirms that combining the variables is a better way to measure sentiment than using any variable alone.

D. Test of Long Term Reversal

PCs are lagged by 12 months to create lagged PCs labeled by "_LAG12" behind the names of PCs. Their correlations with different returns are then tested again and the results are shown in table 6. However, except for PC4_LAG12 and PC8_LAG12, no significant correlations are found between the returns and the lagged PCs.

For PC4_LAG12, it is interesting that it does seem like it predicts long term reversal of stock returns. However, it's short term correlation with small cap returns is not significant. This suggests that consumer sentiment surveys might be better long term reversal of returns predictors.

For PC8_lagged12, its correlation with small cap returns is significant but positive. This is inconsistent with our hypothesis that it should have a negative correlation with returns to predict reversal. Also, this contradicts the conclusion of Baker and Wurgler (2006) which suggests that investor sentiment predicts short term high returns but long term reversal in returns for smaller stocks.

In addition, the PCs are also lagged by 24 months and tested. No significant correlations are found as well. It seems to suggest that investor sentiment could not predict long term reversal of stock returns as Baker and Wurgler (2006) suggested. One possible explanation is that this study did not performance as many types of tests as Baker and Wurgler (2006) did in their paper. It is possible that simple correlation analysis cannot reveal such a property of investor sentiment.

V. CONCLUSION

This study finds that aggregate investor sentiment measures can predict small stock returns in this study. Predictors *SPDNDD2*, *SNIPOD12*, *SRIPOD12, SS, SCCID9, SMDD6* and *SADVDECD12* are found to have the most significant correlations with small stock returns. Combining several sentiment measures creates even better proxies for investor sentiment as they are more highly correlated with future returns. Long term reversal predictability of investor sentiment is generally not found in this study.

VI. REFERENCES

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance 4*, 1645-1680.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance 11*, 1-27.

De Long, J., Shleifer, A., Summers, L. H., & Waldmann, R. (1990). Noise trader risk in financial markets. *Journal of Political Economy 98*, 703-738.

FisherL.Kenneth, & StatmanMeir. (2000). Investor sentiment and stock returns. Financial Analyst Journal 56, 16-23.

Ibbotson, R. G., Sindelar, J. L., & Ritter, J. R. (1994). The market's problems with the pricing of initial public offerings. *Journal of Applied Corporate Finance*, 66-74.

Kaniel, R., Saar, G., & Titman, S. (2004). *Individual investor sentiment and stock returns*. Salomon Center for the Study of Financial Institutions.

Lee, C., Shleifer, A., & Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *Journal of Finance 46*, 75-109.

Matulich, A. (2002, 1 1). *Advances and Declines*. Retrieved 4 1, 2011, from http://unicorn.us.com/advdec/

Nuveen Closed-end Funds. (2001, 1 1). *CEF Connect*. Retrieved 4 01, 2011, from http://www.cefconnect.com/Default.aspx

NYSE Euronext, Inc. (2009, 1 1). *NYSE Technologies*. Retrieved 4 1, 2011, from NYSE Facts & Figures: http://www.nyxdata.com/nysedata/asp/factbook/main.asp

Stigler, G. J. (1964). Public regulation of the securities markets. *Journal of Business 37*, 117-142.

Survey Research Center: University of Michigan. (2010, 10 1). University of Michigan Consumer Sentiment Index. *Surveys of Consumers*.

The Conference Board. (2011, 2 22). Consumer Confidence Survey.

The Federal Reserve Board. (2009, 2 6). *Board of Governors of the Federal Reserve System*. Retrieved 4 1, 2011, from http://www.federalreserve.gov/econresdata/releases/corpsecure/corpsecure20090131.htm

The Wharton School of the University of Pennsylvania. (1994, 1 1). *The Center for Research in Security Prices*. Retrieved 4 1, 2011, from Wharton Research Data Services: http://wrds-web.wharton.upenn.edu/wrds/ds/crsp/index.cfm

Table 1: Summary of Variables and Future Monthly Returns

Description: The variables are defined as follows: fvwr is forward value-weighted returns; fewr is forward equal-weighted returns; fcap1r is forward returns of the first decile of stocks differentiated by size (i.e. small caps); fcap10r is forward returns of the tenth decile of stocks differentiated by size (i.e. large caps); fspr is forward returns of the Standard & Poor's index; fcap1cap10r is the difference between fcap1r and fcap10r.

Variable	Mean	Std. Dev.	Min	Max	Years Covered
fvwr	0.009	0.04	-0.23	0.17	1958 - 2010
fewr	0.012	0.06	-0.27	0.30	1958 - 2010
fcap1r	0.018	0.08	-0.29	0.54	1958 - 2010
fcap10r	0.009	0.04	-0.21	0.18	1958 - 2010
fspr	0.006	0.04	-0.22	0.16	1958 - 2010
fcap1cap10r	0.009	0.07	-0.15	0.51	1958 - 2010
pdnd	-4.271	17.73	-60.13	32.90	1961 - 2007
spdndd2	-0.016	0.95	-4.00	4.00	1961 - 2007
nipo	26.250	23.61	0.00	122.00	1960 - 2010
snipod12	-0.001	1.00	-3.33	4.00	1960 - 2010
ripo	0.145	0.15	-0.34	0.78	1960 - 2010
sripod12	-0.007	0.97	-3.11	4.00	1960 - 2010
turn	0.375	0.24	0.10	1.33	1958 - 2008
sturnd12	0.043	0.66	-4.00	2.27	1958 - 2008
cefd	9.561	7.01	-10.91	25.28	1965 - 2010
scefdd9	0.001	1.00	-4.00	3.07	1965 - 2010
s	0.181	0.11	0.01	0.63	1958 - 2010
ssd10	-0.006	0.97	-3.26	4.00	1958 - 2010
umcsi	86.431	13.04	51.70	112.00	1978 - 2010
sumcsid12	0	1	-3.61	3.19	1978 - 2010
cci	94.631	24.72	25.30	144.71	1967 - 2010
sccid9	0	1	-3.49	2.23	1967 - 2010
si	0.016	0.07	-0.23	0.39	1960 - 2010
ssid12	-0.003	0.99	-3.50	4.00	1960 - 2010
md	0.008	0.04	-0.22	0.39	1960 - 2010
smdd6	-0.006	0.94	-4.00	4.00	1960 - 2010
advdec	1.233	0.28	0.65	2.63	1965 - 2010
sadvdecd12	-0.004	0.98	-2.12	4.00	1965 - 2010
arms	1.051	0.31	0.69	3.00	1965 - 2010
sarmsd11	-0.017	0.88	-4.00	4.00	1965 - 2010
spss	0.410	0.13	0.04	0.69	1965 - 2009
sspssd11	0	1	-3.49	3.39	1965 – 2009

Table 2: Pairwise Correlation of Predictors

Description: Correlation between the standardized versions of each variable and the detrended, standardized and censored versions is shown here.

	spdnd	snipo	sripo	sturn	scefd	ss	sumcsi	scci	ssi	smd	sadvdec	sarms	sspss	spdndd2	snipod12	sripod12	sturnd12	scefdd9	ssd10	sumcsi 12	sccid9	ssid12	smdd6	sadvdec 12	sarmsd1 1	sspssd11
spdnd	1.000																									
snipo	-0.442	1.000																								
sripo	-0.452	0.107	1.000																							
sturn	-0.528	0.076	0.093	1.000																						
scefd	0.347	-0.384	0.035	-0.296	1.000																					
ss	-0.039	0.270	0.067	-0.390	0.157	1.000																				
sumcsi	-0.010	0.237	0.100	0.300	-0.324	-0.449	1.000																			
scci	-0.229	0.188	0.282	0.098	-0.079	-0.174	0.791	1.000																		
ssi	-0.005	0.025	-0.055	0.036	0.046	-0.065	0.022	-0.010	1.000																	
smd	-0.107	0.165	0.238	0.036	0.038	0.036	0.000	0.044	0.095	1.000																
sadvdec	0.072	-0.247	-0.082	0.192	0.069	-0.173	-0.310	-0.366	-0.019	0.020	1.000															
sarms	0.169	-0.229	-0.184	0.328	-0.034	0.007	-0.308	-0.333	-0.129	-0.247	0.276	1.000														
sspss	0.245	0.081	0.204	-0.779	0.137	0.291	0.123	0.266	-0.039	0.083	-0.272	-0.414	1.000													
spdndd2	0.112	0.087	-0.348	0.014	-0.161	0.045	0.067	0.036	-0.028	-0.374	-0.135	0.312	-0.068	1.000												
snipod12	-0.192	0.527	0.223	-0.052	-0.019	0.233	-0.197	-0.094	0.002	0.256	-0.067	-0.122	0.078	-0.108	1.000											
sripod12	0.014	-0.144	0.668	-0.003	0.126	-0.089	-0.061	-0.013	-0.002	0.293	0.118	-0.085	0.096	-0.429	0.029	1.000										
sturnd12	-0.173	0.066	0.162	0.456	-0.063	-0.054	0.225	0.225	0.136	0.164	-0.082	-0.385	0.109	-0.112	-0.007	0.127	1.000									
scefdd9	0.195	0.019	-0.075	0.054	0.296	-0.024	0.095	-0.029	-0.008	0.027	0.063	0.014	-0.076	-0.043	-0.046	0.088	-0.028	1.000								
ssd10	-0.108	0.122	0.200	0.037	0.056	0.601	-0.171	0.014	-0.118	0.123	-0.163	0.032	0.029	-0.060	0.331	0.083	0.051	0.043	1.000							
sumcsid12	-0.236	0.334	-0.055	-0.089	-0.166	0.063	0.369	0.166	0.026	0.116	-0.098	-0.214	0.108	-0.018	0.025	-0.133	0.184	-0.072	-0.204	1.000						
sccid9	-0.148	0.337	0.082	-0.098	0.041	0.146	0.066	0.334	-0.069	0.308	-0.079	-0.129	0.144	-0.038	0.223	0.072	0.063	0.052	0.124	0.207	1.000					
ssid12	0.022	0.019	-0.072	0.030	0.026	-0.067	0.048	0.007	0.967	0.081	-0.007	-0.096	-0.042	0.002	0.017	-0.028	0.100	0.009	-0.123	0.017	-0.079	1.000				
smdd6	-0.006	-0.034	0.149	-0.019	0.006	-0.017	-0.086	-0.049	0.146	0.814	0.118	-0.120	0.061	-0.390	0.073	0.265	-0.021	0.001	0.022	-0.094	0.039	0.143	1.000			
sadvdecd 12	0.047	-0.173	0.009	0.031	-0.008	-0.170	-0.083	-0.131	-0.003	-0.078	0.843	0.096	-0.067	-0.102	-0.200	0.089	-0.023	0.050	-0.238	-0.122	-0.171	0.004	0.082	1.000		
sarmsd11	0.006	-0.016	-0.051	0.073	-0.037	0.089	-0.018	0.057	-0.108	-0.279	-0.102	0.660	-0.136	0.317	-0.114	-0.110	-0.142	-0.053	0.076	-0.136	-0.079	-0.120	-0.223	0.016	1.000	
sspssd11	-0.109	0.061	0.280	-0.033	0.099	0.019	-0.048	0.017	-0.085	0.240	0.104	-0.137	0.357	-0.198	0.071	0.301	0.105	0.170	0.053	-0.154	0.039	-0.124	0.225	0.080	-0.166	1.000

Table 3: Principal Component Comparison

Description: The numbers shown here are the correlation between each variable used in the principal component and the principal component itself. It shows how much each variable influences each principal component.

Principal Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7
spdnd	0.022		0.187				
spdndd2		-0.592	-0.046				-0.395
snipo	0.184		0.203				
snipod12		0.189	0.214				0.042
sripo	0.231		0.154				
sripod12		0.565	0.277				0.296
sturn	-0.837		-0.155				
sturnd12		0.193	-0.279				
scefd	0.336		-0.064				
scefdd9		0.202	0.018				
ss	0.619		-0.139				-0.025
ssd10		0.188	-0.096				
sumcsi							
sumcsid12				0.353			
scci	-0.049		0.081				
sccid9		0.092	0.183	0.353			0.016
ssi	0.051		0.005		0.109		
ssid12		0.067	0.036			0.200	
smd	0.148		-0.051		0.334		
smdd6		0.573	-0.125			0.530	0.267
sadvdec	-0.243		0.165		-0.319		
sadvdecd12		0.201	0.142			0.143	0.072
sarms	-0.409		0.105		-0.631		
sarmsd11		-0.533	0.081			-0.457	
sspss	0.756		-0.037		0.537		
sspssd11		0.376	-0.243			0.324	

Table 4: Single Variable Return Predictability

Description: For each pair of correlation, the first number shown is the correlation, followed by its significance level and lastly by the number of observations used for the correlation analysis. Correlations of high significance are marked with stars behind them. More stars represents higher significance.

Returns	fvwr	fewr	fcap1r	fcap10r	fspr	fcap1c~r
spdnd	-0.0035	0.0126	-0.0144	-0.01	-0.0157	-0.0103
	0.9347	0.7656	0.7323	0.8121	0.7095	0.8063
	564	564	564	564	564	564
spdndd2	-0.0352	-0.1229***	-0.2058***	-0.0249	-0.0273	-0.222***
	0.4049	0.0035	0	0.5552	0.5177	0
	563	563	563	563	563	563
snipo	-0.0373	-0.1317***	-0.1414***	-0.0127	-0.0039	-0.1576***
	0.3576	0.0011	0.0005	0.7549	0.9232	0.0001
	611	611	611	611	611	611
snipod12	-0.0206	-0.0143	0.0179	-0.0178	-0.0153	0.0329
	0.6119	0.7238	0.6587	0.66	0.7053	0.4178
	610	610	610	610	610	610
sripo	0.02	0.0655	0.1424***	0.016	0.0142	0.1567***
	0.6278	0.1124	0.0005	0.6993	0.7317	0.0001
	588	588	588	588	588	588
sripod12	0.0842**	0.1151***	0.1772***	0.0827**	0.0766*	0.1539***
	0.0415	0.0053	0	0.0452	0.0638	0.0002
	587	587	587	587	587	587
sturn	-0.0211	-0.033	-0.0309	-0.0148	-0.0097	-0.0263
	0.6047	0.4172	0.4467	0.715	0.8115	0.5181
	607	607	607	607	607	607
sturnd12	0.0756*	0.0484	0.0192	0.0786**	0.0793**	-0.0298
	0.057	0.2233	0.6289	0.0479	0.046	0.4544
	634	634	634	634	634	634
scefd	0.0125	0.0451	0.0391	0.003	-0.0031	0.0437
	0.771	0.2932	0.3623	0.9445	0.9422	0.3087
	545	545	545	545	545	545
scefdd9	-0.0245	-0.059	-0.0982**	-0.0178	-0.0225	-0.103**
	0.5693	0.1696	0.022	0.6794	0.601	0.0162
	544	544	544	544	544	544
ss	-0.0892**	-0.1245***	-0.1108***	-0.0809**	-0.0842**	-0.0761*
	0.0247	0.0017	0.0052	0.0415	0.0339	0.0552
	635	635	635	635	635	635
ssd10	-0.0565	-0.0884**	-0.0671*	-0.0494	-0.0465	-0.0458

	0.1556	0.026	0.0916	0.2142	0.2427	0.2493
	634	634	634	634	634	634
sumcsi	-0.0167	-0.0883*	-0.1316***	-0.0017	0.0133	-0.1522***
	0.7423	0.0825	0.0094	0.9732	0.7932	0.0027
	388	388	388	388	388	388
sumcsid12	0.0792	0.0275	-0.0231	0.0846*	0.0852*	-0.0871*
	0.1198	0.5896	0.6506	0.0964	0.0943	0.0871
	387	387	387	387	387	387
scci	-0.0741	-0.1452***	-0.1433***	-0.0577	-0.049	-0.1283***
	0.1109	0.0017	0.002	0.2147	0.2927	0.0056
	464	464	464	464	464	464
sccid9	-0.0558	-0.1071**	-0.0887*	-0.0413	-0.0398	-0.0757
	0.2305	0.0212	0.0564	0.375	0.393	0.1037
	463	463	463	463	463	463
ssi	0.056	0.0816**	0.0794*	0.0497	0.0583	0.0605
	0.1695	0.0452	0.0513	0.2233	0.1528	0.1378
	603	603	603	603	603	603
ssid12	0.0376	0.0648	0.0659	0.0316	0.0398	0.0565
	0.3573	0.1123	0.1064	0.4389	0.33	0.1662
	602	602	602	602	602	602
smd	-0.0099	0.0151	0.0719*	-0.0058	0	0.0883**
	0.8076	0.7096	0.0759	0.8857	0.9997	0.0291
	610	610	610	610	610	610
smdd6	-0.0149	0.0474	0.1172***	-0.0176	-0.0151	0.1493***
	0.7136	0.2433	0.0038	0.6642	0.7105	0.0002
	609	609	609	609	609	609
sadvdec	0.0776*	0.2223***	0.2643***	0.049	0.0393	0.2769***
	0.0692	0	0	0.2514	0.3584	0
	549	549	549	549	549	549
sadvdecd12	0.0584	0.2133***	0.268***	0.0289	0.0192	0.2943***
	0.1721	0	0	0.5	0.6539	0
	548	548	548	548	548	548
sarms	-0.0157	-0.0316	-0.0424	-0.0125	-0.0105	-0.0414
	0.7138	0.4597	0.3211	0.7696	0.8053	0.3331
	549	549	549	549	549	549
sarmsd11	-0.0186	-0.0895**	-0.1196***	-0.0055	-0.0057	-0.1361***
	0.6635	0.0361	0.0051	0.8976	0.8936	0.0014
	548	548	548	548	548	548
sspss	0.0088	0.0296	0.0554	0.0056	0.0094	0.061
	0.8395	0.496	0.2027	0.8981	0.8289	0.1608
	530	530	530	530	530	530
sspssd11	0.0449	0.0569	0.0858**	0.0475	0.0459	0.0692
	0.3025	0.1913	0.0487	0.2753	0.2925	0.1118

	529	529	529	529	529	529

**p-value < 0.1*

***p-value < 0.05*

****p-value < 0.01*

Table 5: Return Predictability of Principal Components

Description: This table is the same as table 4. The only difference is the left column where principal components are listed instead of single variables. PC1 is made up of all 13 standardized predictors. PC2 is made up of all 13 detrended, standardized and censored versions of the predictors. PC3 is made up of all the components from PC1 and PC2. PC4 is made up the detrended, standardized and censored versions of the two survey-based measures. PC5 is made up of 5 standardized predictors used in Brown and Cliff (2004). PC6 is detrended, standardized and censored version of PC5. PC7 is made up of predictors that have the most significant correlations with small stock returns by themselves. PC8 is made up of the predictors used in Baker and Wurgler (2006).

Returns	**fvwr**	**fewr**	**fcap1r**	**fcap10r**	**fspr**	**fcap1c~r**
PC1	-0.0168	0.0139	0.0473	-0.0218	-0.0272	0.0676
	0.7325	0.7771	0.335	0.6579	0.5795	0.1683
	417	417	417	417	417	417
PC2	0.0306	0.1366***	0.2417***	0.0215	0.0203	0.26***
	0.5332	0.0053	0	0.6618	0.6791	0
	416	416	416	416	416	416
PC3	-0.0051	0.0748	0.1822***	-0.0094	-0.0084	0.2123***
	0.9179	0.1279	0.0002	0.8489	0.8651	0
	416	416	416	416	416	416
PC4	0.0241	-0.0344	-0.0473	0.0356	0.037	-0.0804
	0.6361	0.4993	0.3536	0.4852	0.4678	0.1144
	387	387	387	387	387	387
PC5	0.0275	0.025	0.046	0.0297	0.0364	0.0344
	0.5273	0.5654	0.2906	0.495	0.4027	0.4295
	530	530	530	530	530	530
PC6	0.055	0.159***	0.2283***	0.0419	0.0423	0.2389***
	0.2066	0.0002	0	0.3357	0.3312	0
	529	529	529	529	529	529
PC7	0.0421	0.1443***	0.2527***	0.0327	0.0317	0.2652***
	0.3915	0.0032	0	0.5056	0.5185	0
	416	416	416	416	416	416
PC8	0.0022	0.0721	0.0908**	-0.0157	-0.0263	0.1137**
	0.9601	0.1065	0.0417	0.726	0.5563	0.0107
	503	503	503	503	503	503

**p-value < 0.1*

***p-value < 0.05*

****p-value < 0.01*

Table 6: Test of Long Term Reversal

Description: Principal components that have significant correlation with small cap returns are listed here together with their lagged versions. "_lag12" represents that the PC is lagged by 12 months.

Returns	**fvwr**	**fewr**	**fcap1r**	**fcap10r**	**fspr**	**fcap1c~r**
PC1_lag12	0.0722	0.0853*	0.068	0.0672	0.0627	0.0329
	0.1409	0.0819	0.1655	0.1708	0.2011	0.503
	417	417	417	417	417	417
PC2_lag12	-0.041	-0.0631	-0.0421	-0.0365	-0.0294	-0.0236
	0.4037	0.199	0.3923	0.4576	0.5496	0.6309
	416	416	416	416	416	416
PC3_lag12	-0.014	-0.0431	-0.044	-0.01	-0.0021	-0.0432
	0.7765	0.3804	0.3709	0.8382	0.9665	0.379
	416	416	416	416	416	416
PC4_lag12	-0.0131	-0.0992*	-0.145***	-0.0013	-0.0001	-0.1666***
	0.8008	0.0546	0.0049	0.9799	0.9983	0.0012
	376	376	376	376	376	376
PC5_lag12	-0.0162	-0.0377	-0.0346	-0.013	-0.0123	-0.032
	0.7106	0.3866	0.4265	0.7651	0.7768	0.4622
	530	530	530	530	530	530
PC6_lag12	-0.0117	-0.0145	0.0084	-0.0119	-0.0087	0.0176
	0.7875	0.7387	0.848	0.7854	0.8419	0.6871
	529	529	529	529	529	529
PC7_lag12	-0.0511	-0.049	-0.0021	-0.0497	-0.0434	0.0304
	0.2986	0.3191	0.9664	0.3119	0.3769	0.5367
	416	416	416	416	416	416
PC8_lag12	-0.0012	0.0483	0.0737*	-0.0125	-0.0195	0.0942**
	0.9788	0.2797	0.0985	0.7801	0.6634	0.0346
	503	503	503	503	503	503

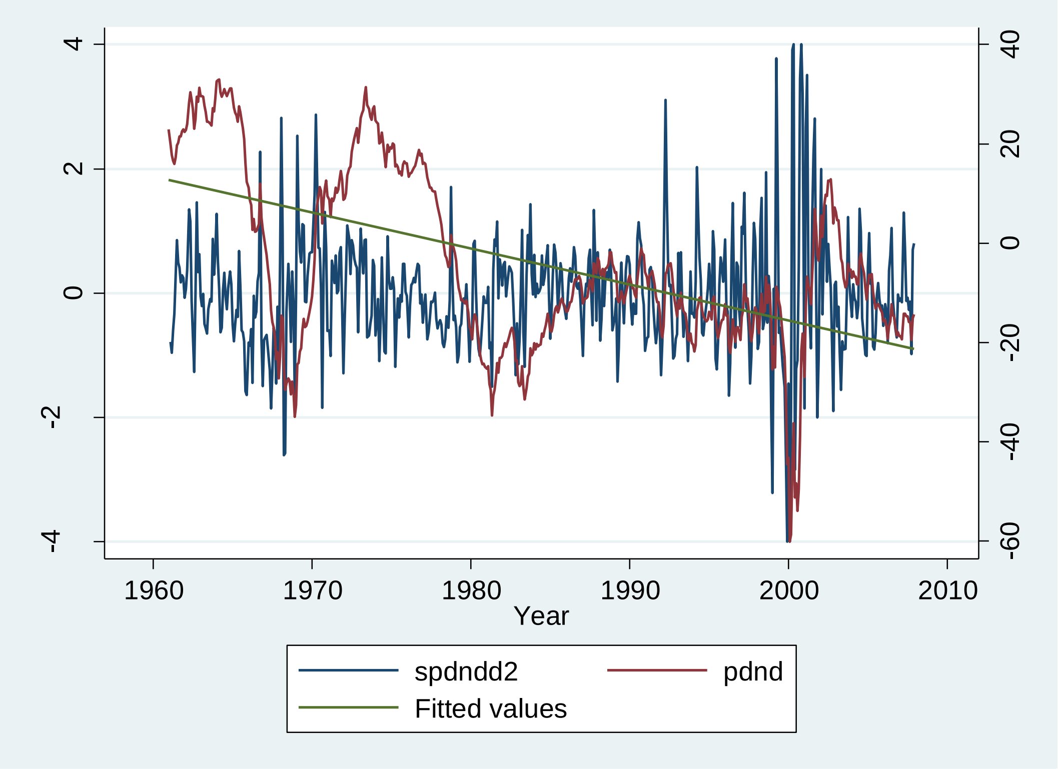
**p-value < 0.1*

***p-value < 0.05*

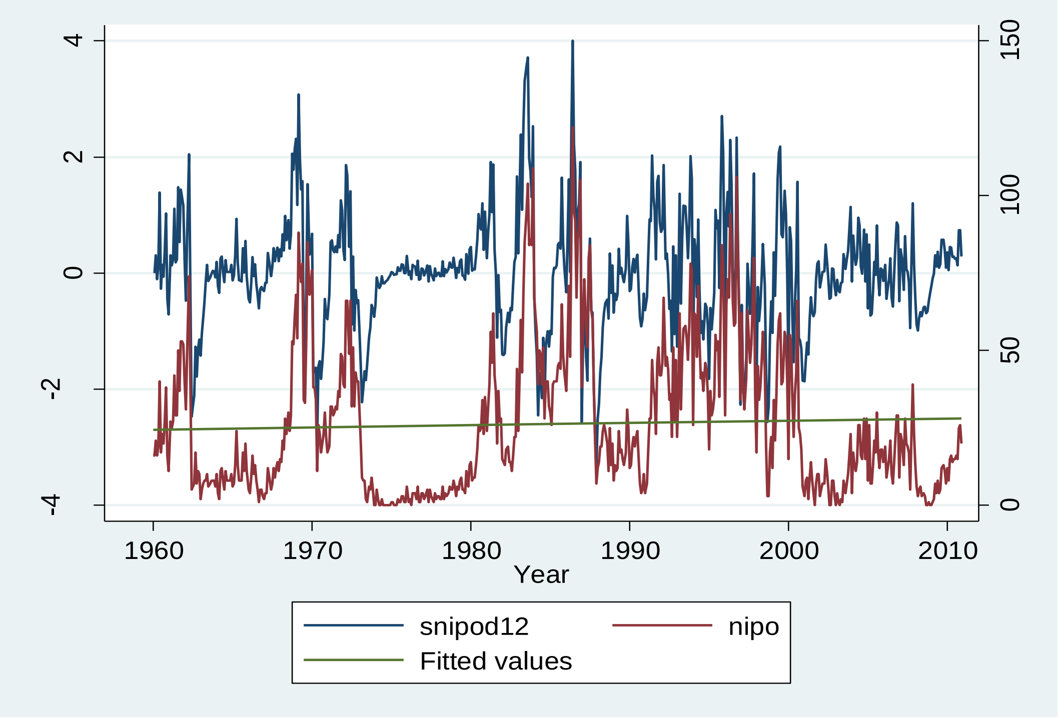
****p-value < 0.01*

Figure 1: Comparison of Raw Values and Detrended, Standardized and Censored Values of Predictors

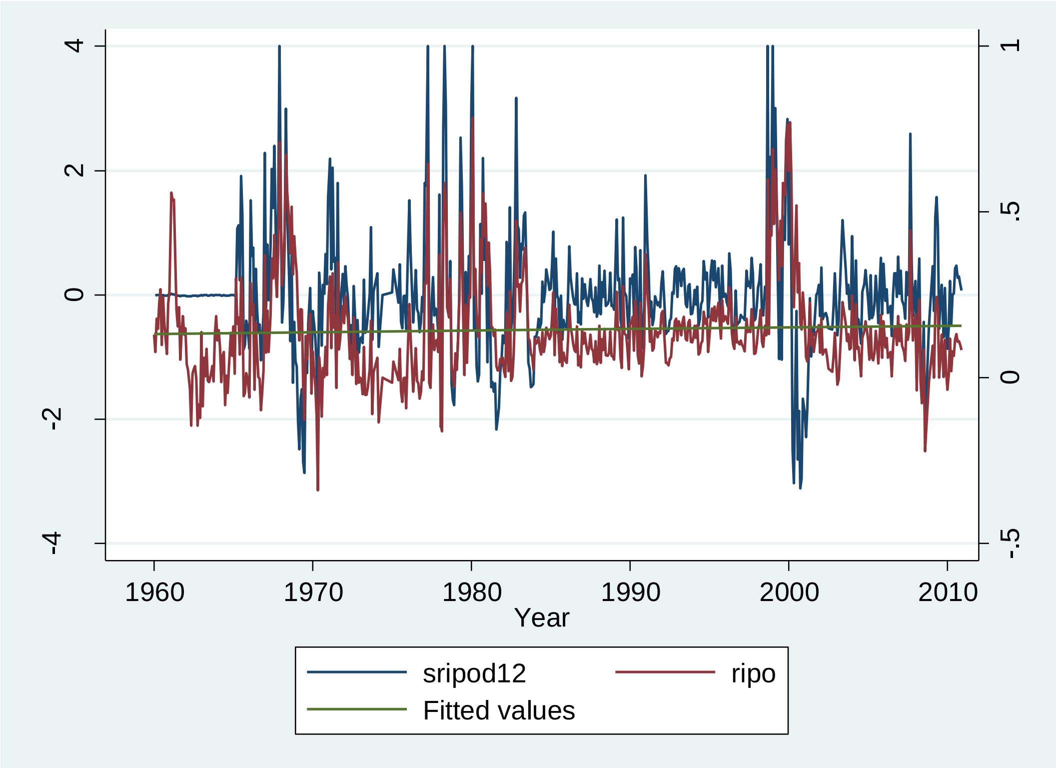
PDND



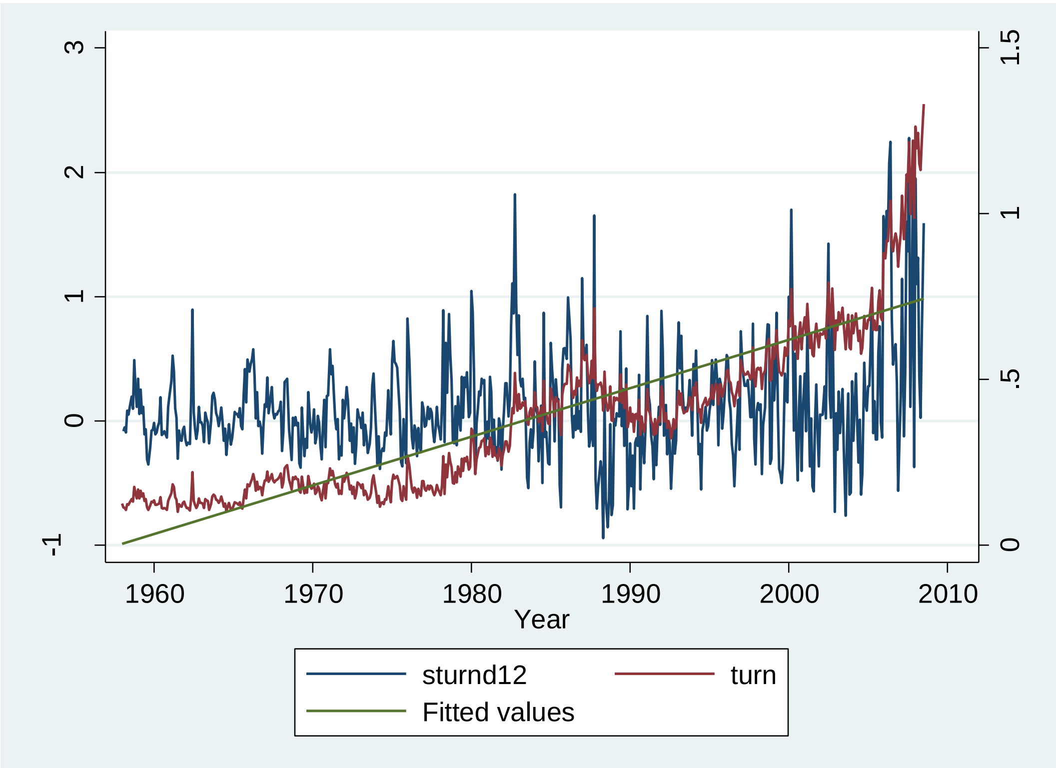
NIPO



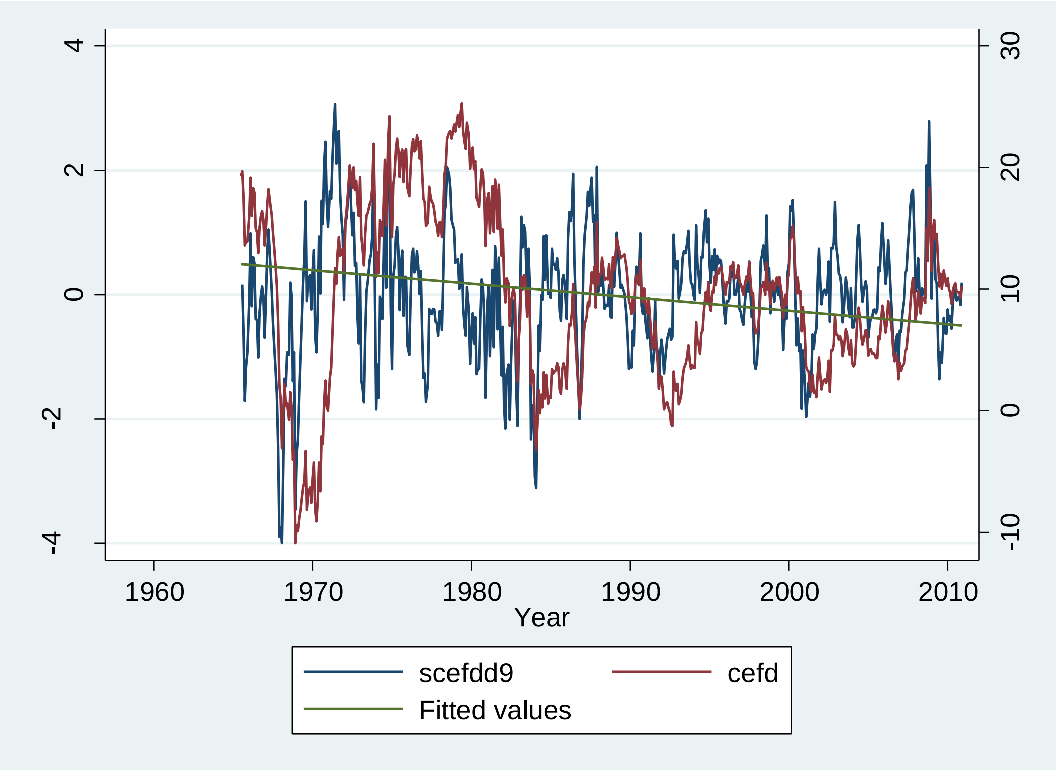
RIPO


4
2
0
-2
-4
1
.5
0
-.5
1960
1970
1980
1990
2000
2010
Year
sripod12
ripo
Fitted values

TURN

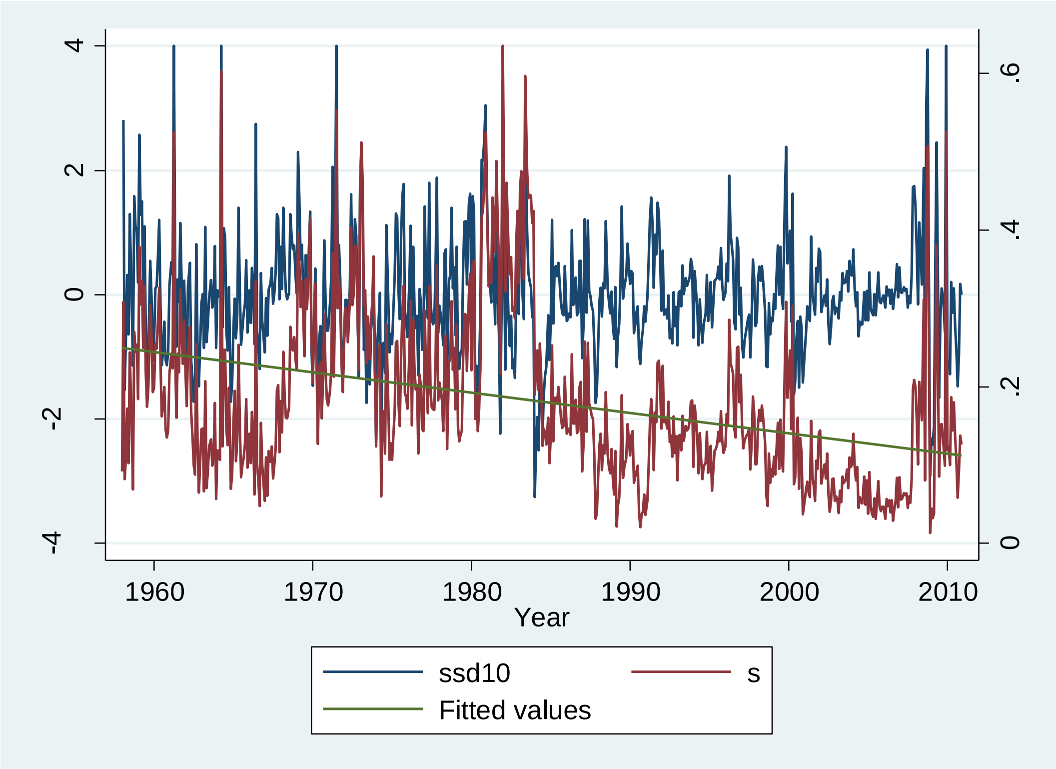

3
2
1
0
-1
1.5
1
.5
0
1960
1970
1980
1990
2000
2010
Year
sturnd12
turn
Fitted values

CEFD



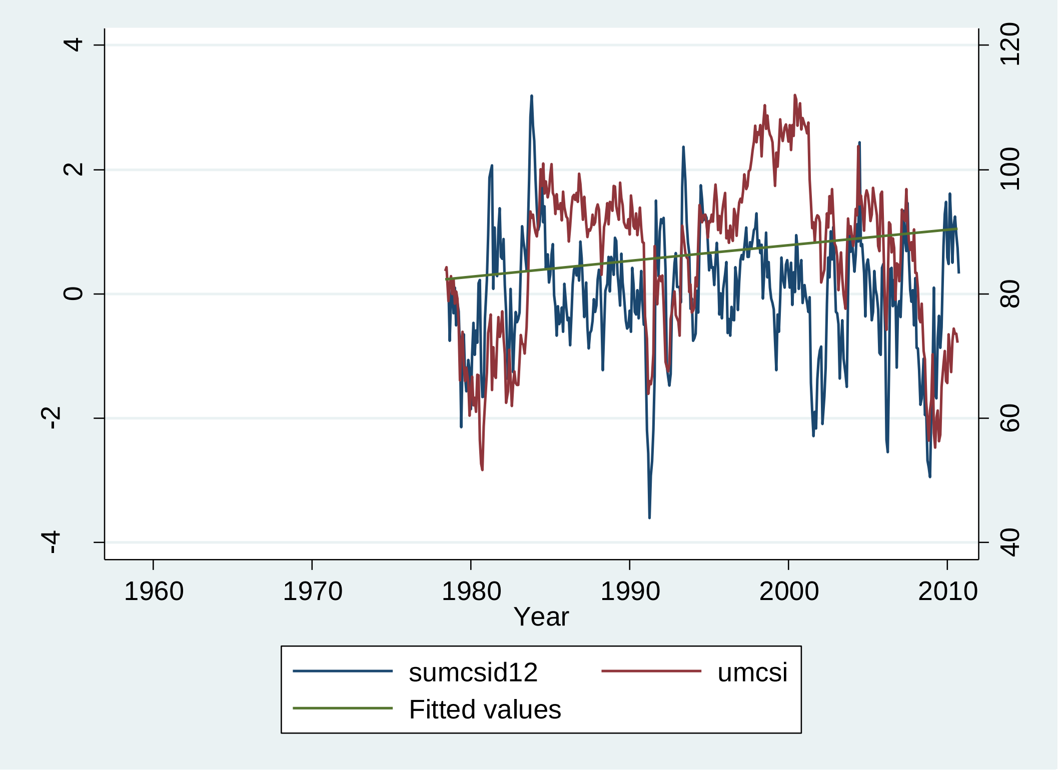
4
2
0
-2
-4
30
20
10
0
-10
1960
1970
1980
1990
2000
2010
Year
scefdd9
cefd
Fitted values

S



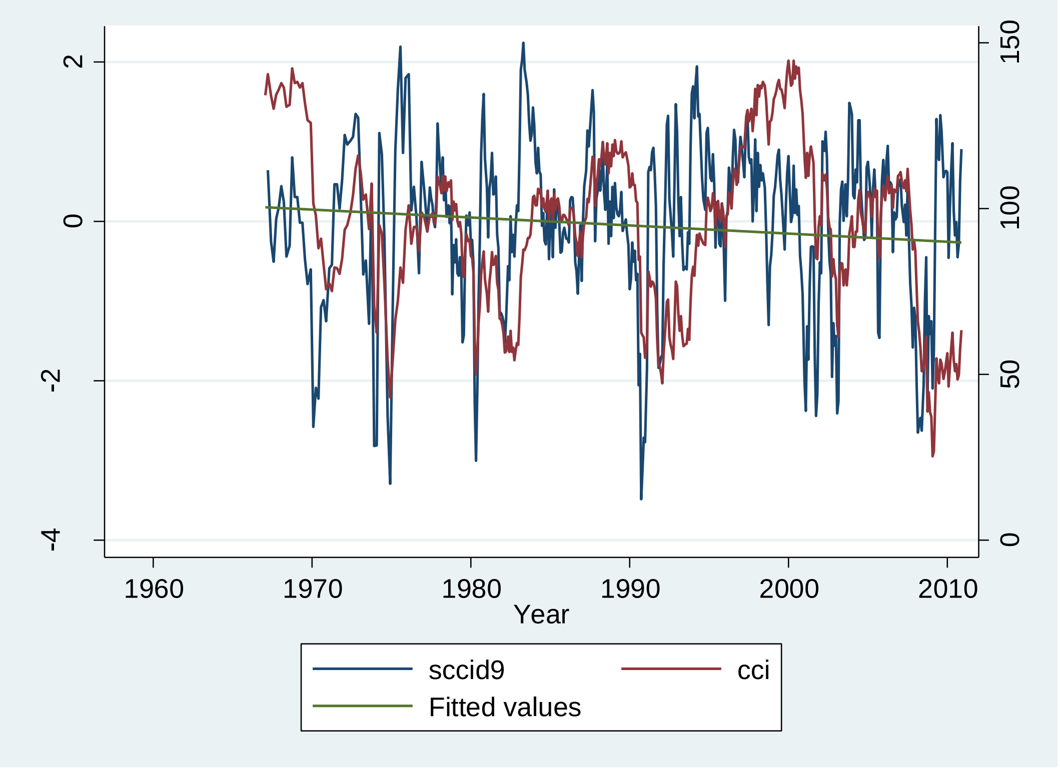
4
2
0
-2
-4
.6
.4
.2
0
1960
1970
1980
1990
2000
2010
Year
ssd10
s
Fitted values

UMCSI



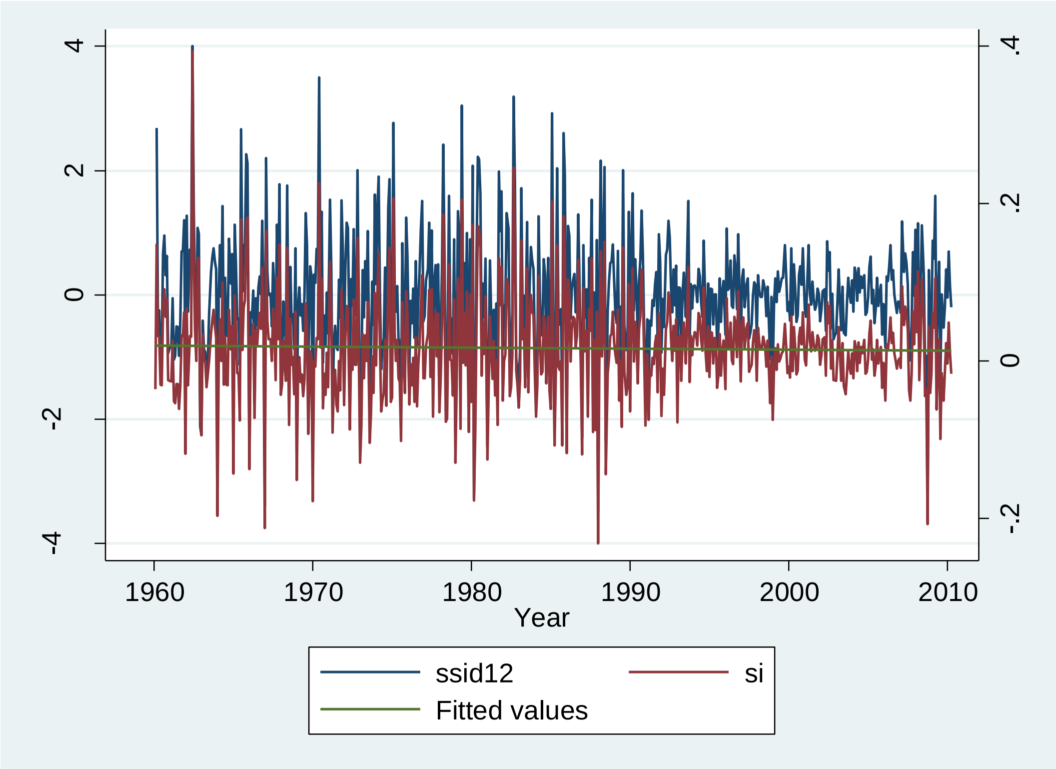
4
2
0
-2
-4
120
100
80
60
40
1960
1970
1980
1990
2000
2010
Year
sumcsid12
umcsi
Fitted values

CCI



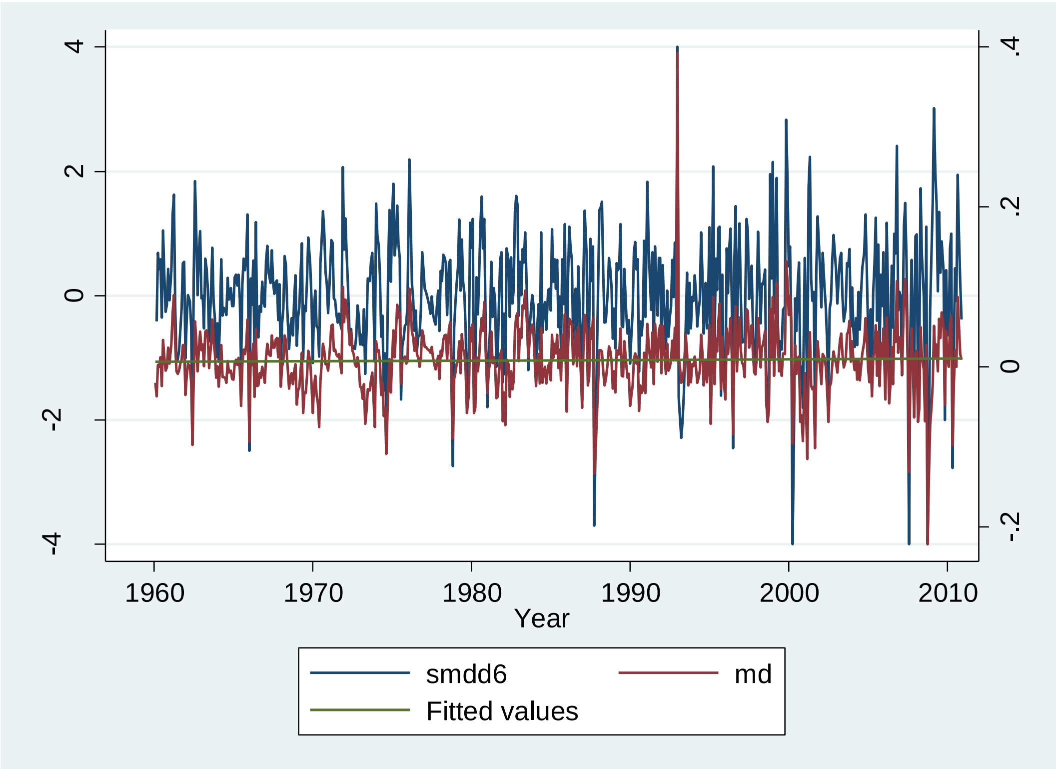
2
0
-2
-4
150
100
50
0
1960
1970
1980
1990
2000
2010
Year
sccid9
cci
Fitted values

SI



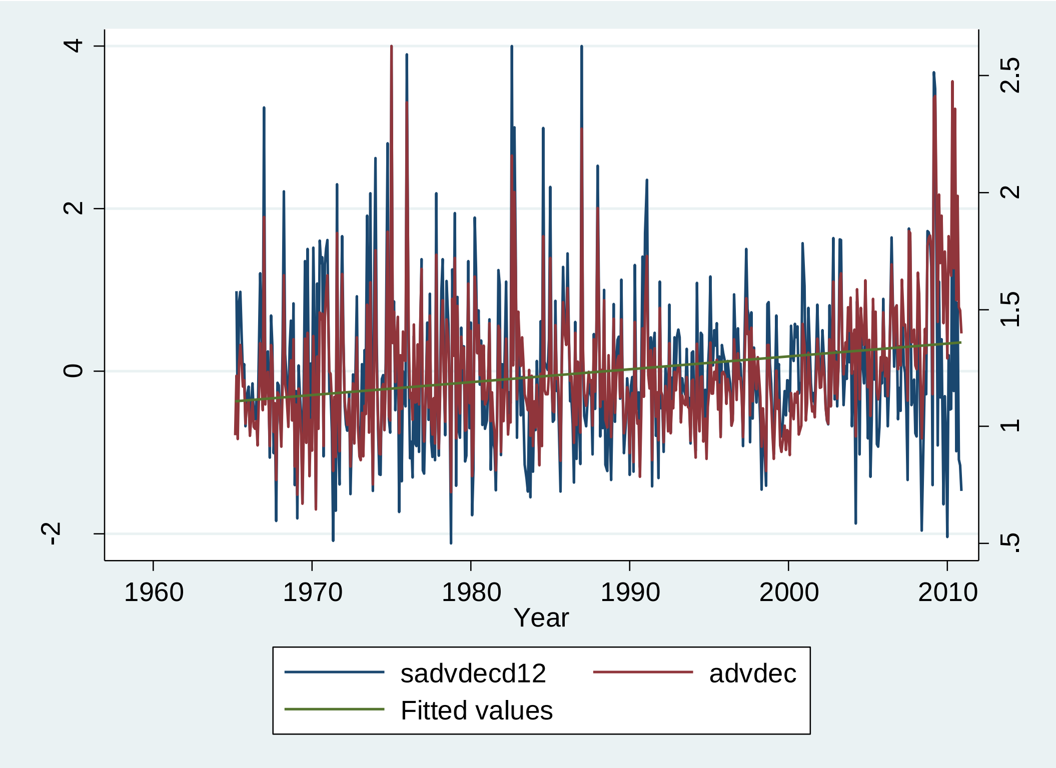
4
2
0
-2
-4
.4
.2
0
-.2
1960
1970
1980
1990
2000
2010
Year
ssid12
si
Fitted values

MD

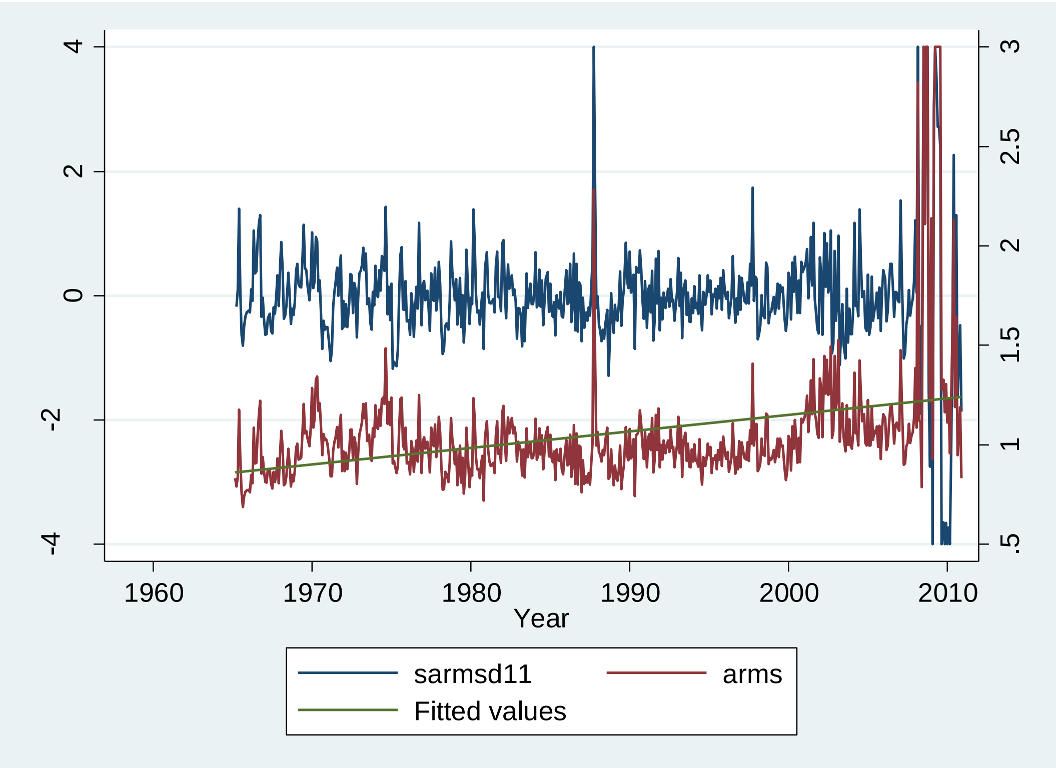


4
2
0
-2
-4
.4
.2
0
-.2
1960
1970
1980
1990
2000
2010
Year
smdd6
md
Fitted values

ADVDEC


4
2
0
-2
2.5
2
1.5
1
.5
1960
1970
1980
1990
2000
2010
Year
sadvdecd12
advdec
Fitted values

ARMS


4
2
0
-2
-4
3
2.5
2
1.5
1
.5
1960
1970
1980
1990
2000
2010
Year
sarmsd11
arms
Fitted values

SPSS

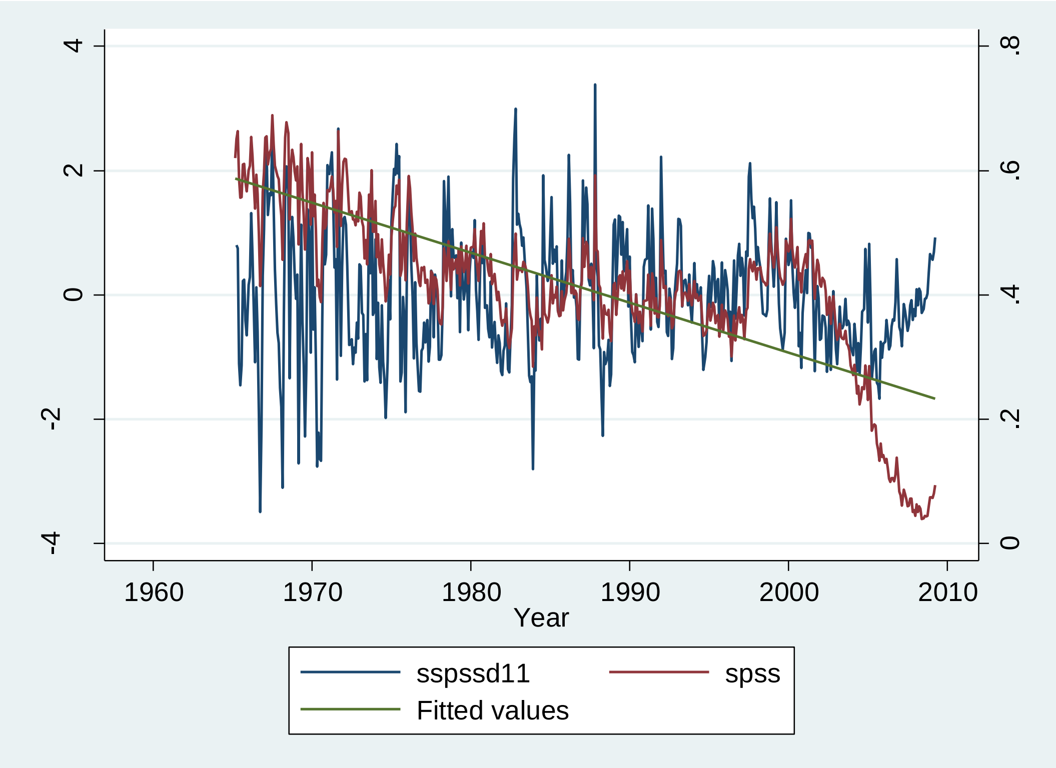
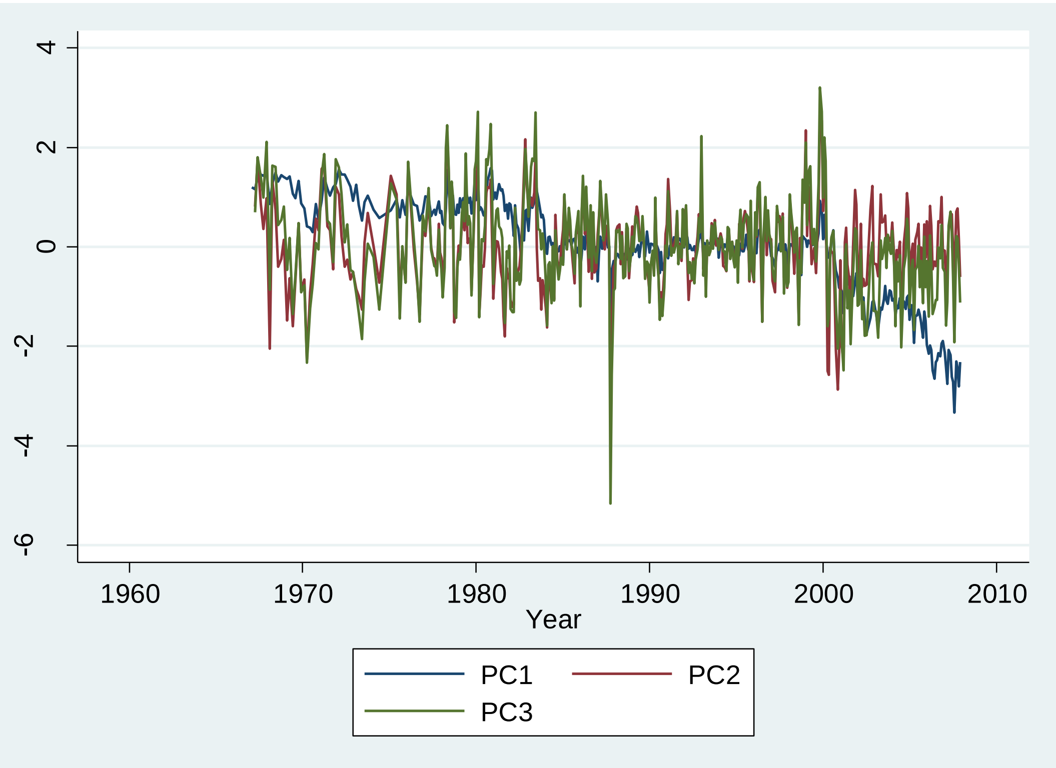
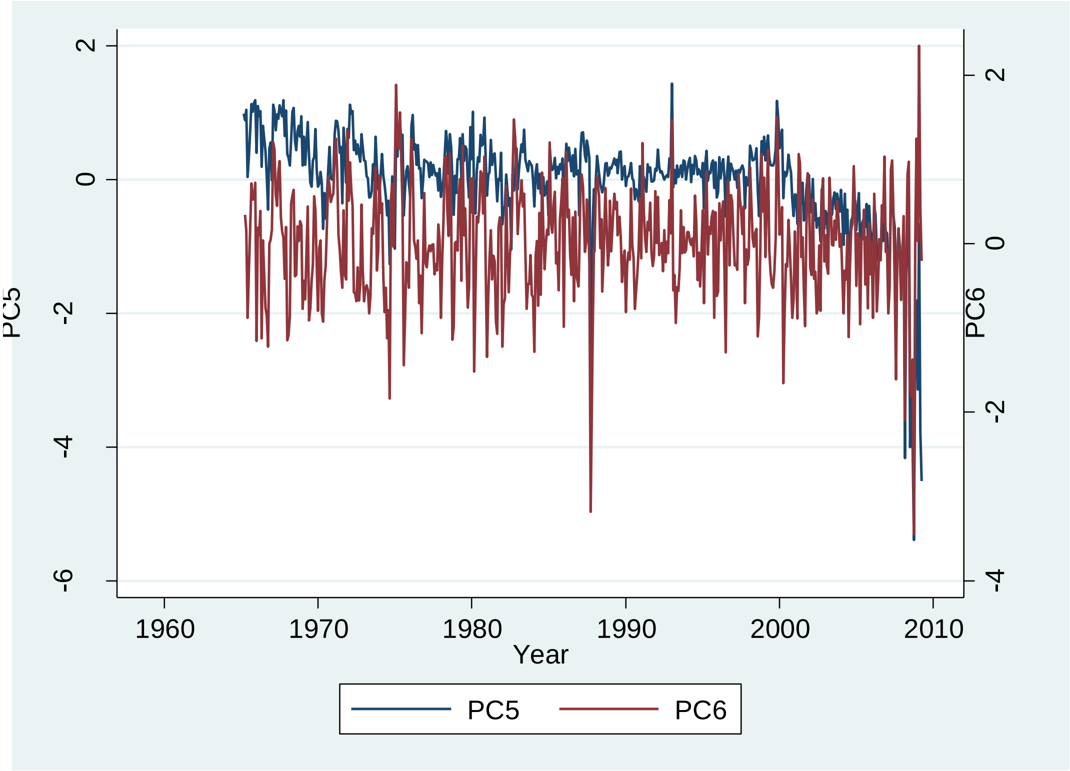
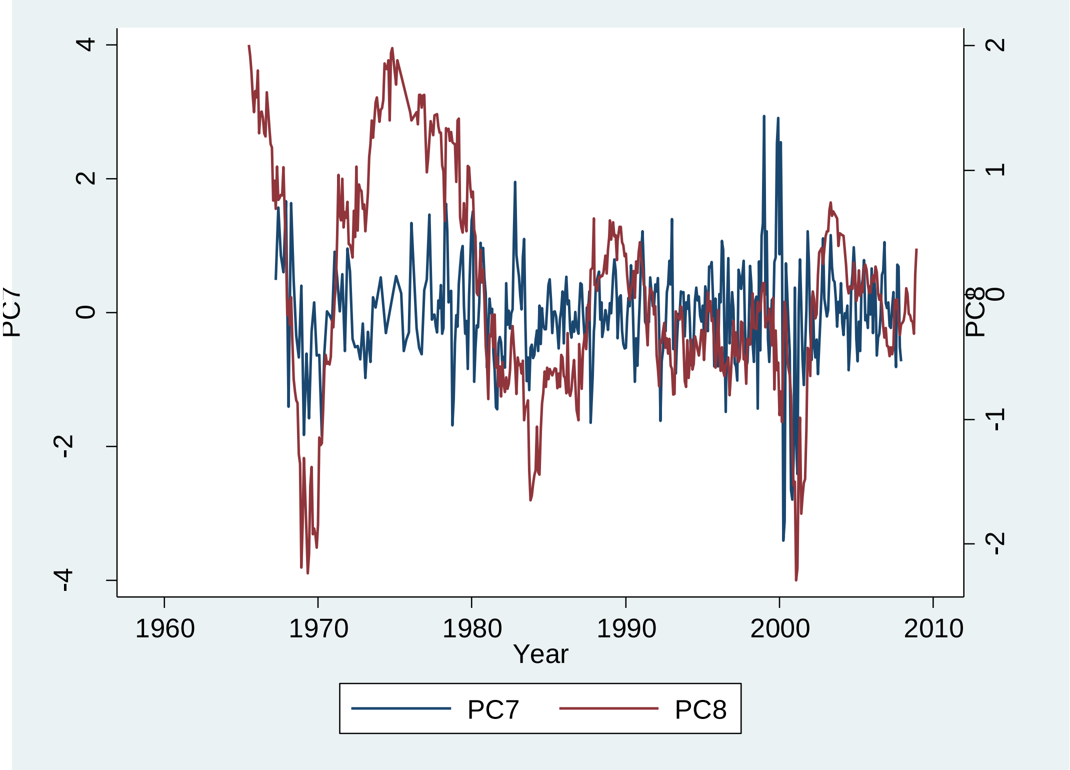


Figure 2: Monthly Graph of Principal Components




PC5
PC6
Year
PC5
PC6
1960
1970
1980
1990
2000
2010


PC7
PC8
Year
PC7
PC8
1960
1970
1980
1990
2000
2010