Do Financial Market Variables Predict Unemployment Rate Fluctuations?

Jingyi Chen
Department of Economics
College of Arts & Science
East Carolina University
M.S Research Paper

June, 27, 2002

<u>Abstract</u>

This paper examines empirically the Granger-causal relationship between financial market variables and real economic activity as measured by the unemployment rate. We find in our paper that the in-sample measures of fit are largely affected by one particular influential observation: 1974:12. This observation accounts for superior performance of the paper-bill spread in explaining the unemployment rate. We then show that none of the commonly employed measures of monetary policy contain incremental information useful in forecasting the unemployment rate. A simple pure autoregressive model performs better than three-variable models that contain the paper-bill spread, the federal funds rate or M2 in out-of-sample forecasting. The different data vintage matters in evaluating a model. The fact that the results are sensitive to the different data vintage make us suspect the robustness of the Granger Causality between financial market variables and the unemployment rate concluded in the earlier studies.

I. Introduction:

Monetary policy is a central bank's actions to influence the availability and cost of money and credit, as a means of helping to promote national economic goals. Broadly, the monetary policy affects the real economy through three channels: 1) through the cost of borrowing in the market, 2) through the exchange rate and 3) through the prices of financial assets, especially equities. The primary tools of monetary policy include open market operations, discount policy, and reserve requirements. Financial market developments and asset prices thus provide useful information for a monetary policy that focuses on price stability, and form an integral part of the overall assessment of economic developments required for the successful conduct of monetary policy.

Over the past three decades, the researchers and economists have reached the consensus that the monetary policy has effects on the real activity, such as the real output growth rate and the unemployment rate. This paper focuses on the predictive power of the financial market variables on real activity as measured by unemployment rate.

Since the 1970s', there have been extensive studies and researching papers on the Granger causality of the financial market variables on the real economic activities. These studies have broadened the money-income causality literature spawned by Christopher Sims and encompassed a diverse set of measures of monetary policy, more comprehensive and sophisticated empirical assessment strategies and a richer array of explanations for observed correlations between financial market variables and economic activity. The common evidence presented in the previous literature is that particular interest rates and spreads not only dominate monetary aggregates as predictors of economic performance, but also are remarkably powerful predictors. The ensuing papers

also found that the predictive power of the paper bill spread weakened during the second half of the 1980s and the early 1990s.

Thoma and Gray (1998) examined the predictive power of financial market variables on real economic activity as measured by industrial production. They reviewed the important methodological pitfalls in the earlier studies that rely primarily on in-sample measures of fit. Since test statistics are sensitive to some influential observations during the sample period, in-sample measures of fit will become misleading indicators of out-of-sample measures (forecast errors) with the presence of the extreme outliers. Thoma and Gray raised their concern that the previous conclusions of the Granger-causality between financial market variables and the real activity need reassessment because these conclusions are potentially incorrect when relying on in-sample measures of fit. They made a striking illustration of the drawbacks of in-sample measures of fit. The technique employed in their paper is the rolling (recursive) regression with the goal of addressing the extreme sensitivity across sample periods of the causality statistics to assess the explanatory power of financial market variables. As the empirical studies have suggested that the in-sample measures of fit are sometimes heavily influenced by individual observations, Thoma and Gray found that the observation of 1974:12 "accounts for the uniformly superior performance of the paper-bill spread reported in many studies". The outliers present in the data in 1974, they argued, could lead one to conclude incorrectly that the paper-bill spread contains information generally useful in forecasting real activity. In-sample measures of fit do not provide reliable indicators of out-of-sample fit. To make this point clearer, they evaluate out-of-sample forecasting

ability of paper-bill spread, federal funds rate, and M2. The comparison of the RMSEs shows that the forecasting ability of three models deteriorates dramatically in late-1974.

To address whether the financial market variables contain useful information to predict the industrial production, they departed from the common practice of framing empirical exercises as horse races among those competing financial market variables. Rather, they address this question by comparing the out-of-sample forecasting power of a simple autoregressive model of industrial production to the predictive power of models that include the paper-bill spread, the federal funds rate, and M2. What they found is that none of the financial market variables considered aid systematically in forecasting industrial production, whether the variables are considered alone or in combination. They also concluded that "either monetary policy innovations have no significant real effects, or we (collectively) have failed in our efforts to measure monetary policy".

Besides the fact that the individual observation might influence the in-sample measures, the particular data vintage used might matter for evaluating forecasts. The economic data such as real output, money stock, the consumption spending, etc, that are publicized by the Fed are subject to revisions and redefinitions according to the information available. The information set available at particulate date is called a 'vintage'. The real-time data specifically referred to the information set available to economic researchers who were making forecast in real time. Throughout our paper, for the sample period 1960:02 to 1995:04 we studied, we use real-time data dated May 1995 as well as the most recent data vintage that is updated in February 2002. Data revisions are important because they may affect the test statistics and the results may be sensitive

to the revisions and redefinitions. The more sensitive the results to data vintages, the less robust the facts are.

Croushore and Stark (2000) plotted differences in the data across vintages for the same date. By examining the features of the plots, they found revisions do affect data. They suggested that in analyzing forecasts, one should be very careful about what vintages of the data one uses as actual, since redefinitions, changes in methodology, and changes in relative prices seem to have dramatic effects on economic data. They testes a simple empirical ARIMA model of real output growth and compare forecasts generated from models estimated on latest available data to those generated from models estimated on real-time data. However, they found RMSEs of forecasts from models estimated on two data sets are not quite different. This finding is quite surprising because it says that having today's vintage gives no better forecast performance than having available just real-time data, when the goal is to forecast the data as they appear today. Or this finding just simply means that the variable they choose to forecast is not very productive in the sense that the forecast errors are large relative to the revisions today. They concluded that "Forecasts based on real-time data are certainly correlated positively with forecasts based on final data, but data revisions to real output may cause forecasts based on current-vintage data to be considerably different from forecasts based on real-time data over selected sample periods".

One more example to illustrate that the data vintage matters is the paper by Swanson and Amato (2000) where they reassessed the evidence on the marginal predictive content of M1 and M2 for real and nominal output, taking the imperfection from data revision into account. They used the latest version of the data that is available

and the sequences of historical time series that would have been available to forecasters in real time. What they concluded is that the generally significant marginal predictive content of M1 or M2 for output that is found using a recently revised data set is not duplicated in a real time setting.

The main task of our paper is to study the Granger-causality relationship between the financial market variables and the unemployment rate. We use the real-time data over the selected sample period as well as the recent vintage data dated February 2002 that are taken from the Federal Reserve Bank of Philadelphia. Our findings are similar with what Thoma and Gray concluded. We find that the results are sensitive to the observation of 1974:12, which occurred in the midst of the first OPEC oil-price shock and brought on the second deepest post-World War II recession in the U.S. After we exclude this observation, none of the commonly employed indicators of monetary policy contain useful information in forecasting the unemployment rate. Another observation, 1982: 12, even though it has the highest unemployment rate in the post-Word War II period, however, does not influence the in-sample measures. It would be another interesting topic to explore why the 1982:12 observation appears to have no such influence as the 1974:12 observation has.

The out-of-sample forecasting errors exhibit very similar patterns for the paper-bill spread, the federal funds rate and M2. It is not easy at the first glance to determine whether one of these three variables dominates the other in the clear advantage owned by one of the three variables. The only apparent discrepancy in which the paper-bill spread dominates the other two in forecasting performance happens during the period immediately after 1974:12 till the end of 1979. It illustrates that the short-lived

advantages of the paper-bill spread might caused by presence of the influential observation of 1974:12.

To be apart from go beyond the conventional practice of the 'horse racing' amongst alternative models, we compare the forecasting errors of models including financial variables to the pure simple autoregressive model with lag order six. The comparison yields the similar conclusion made by Thoma and Gray (1998) that the financial market variables do not help predict the industrial production. We find that adding the financial variables to explain the unemployment rate does not improve the forecast ability. Therefore, none of the financial market variables studied in our paper provides systematic aid to forecasting the unemployment rate.

We test our model using both the recent revised data and the real-time data to examine whether the results are considerably different from each other. We then show that the results are sensitive to the data vintages. The real-time data and the February 2002 vintage data do differ in the in-sample model evaluations. The in-sample Granger-causality test statistics using the February 2002 revised data vintage tends to be lower than those using the real-time data, especially for M2 and the federal funds rate. If we use the February 2002 vintage data as "actual" values, then the attenuation of in-sample Granger-causality relationship would actually reflect the weak relationship between the financial market variables and the unemployment rate. However, the results of forecasting errors using two data sets do not seem to be different from each other, which is similar to Croushore and Starks results. . The sensitivity of the results to data revisions lead us to question about the robustness of the earlier claims about the Granger-causality relationships.

The second part of our paper describes the real-time data and the February 2002 data vintage we used throughout this paper. The methodology is also illustrated in the second part. The empirical results are presented in the third part. It has three sections of the results: the first section is about the in-sample measure of fit and sample sensitivity, the second section is about evaluation of out-of-sample forecasting performance and the third section concentrates on the comparison between the simple autoregressive model and the three-variable models that contain either the federal funds rate, the paper-bill spread or M2. The conclusion is made in the final part of our paper.

II. Methodology and Data

In our paper, we evaluate the explanatory power of three commonly employed financial market variables that also appeared in the paper by Thoma and Gray (1998): FF--the federal funds rate, SP--the paper-bill spread (more specifically, the difference between the six-month commercial and treasury bill rates), and M2.

The data we use has two parts: one is the monthly real-time data dated May 1995 and the other one is the recent vintage monthly data updated in February 2002 by the Federal Reserve Bank of Philadelphia. We select our sample period as 1960:02-1995:04. We use the above two data sets to explore whether our results are sensitive to the particular data vintage used. As Croushore and Stark (2002) stressed, the data vintage matters. Because the analysis of new forecasts is often based on the final, revised data, rather than the data that were available to economic agents who were making forecasts in real time, "the results of such exercises may be misleading". Therefore, to avoid such problems in creating forecasting models, Federal Reserve Bank of Philadelphia has

developed a data set that gives a modeler a "snapshot" of the macroeconomic data available at any given date in the past. As we mention before, the information set available at a particular date is called a 'vintage', and the collection of such vintages a 'real-time data set'. Developing a real-time data set is not a simple process of entering old data into spreadsheet. It actually requires a substantial amount of effort, including digging through old source data and figuring out what data were available at what time. Given the lack of documentation of much of the historical data, such procedure is definitely not a trivial one. In the real-time data we used, the observations are identical to those one would have seen in published sources during the period of 1960:02 to 1995:04. The revisions are made later on to the real-time data incorporating new source data. The most recent revision available to this paper can be seen in the February 2002 data vintage.

Previous studies have found that the paper-bill spread is highly significant in explaining real activity. However, it appears that this significant effect results from one particular individual observation. To illustrate the sensitivity of test-statistics to the individual observation in sample period, we employ rolling (recursive) regressions. To be specific, we estimate first over the period 1960:02 through 1965:01 and a test of the hypothesis that the financial variable does not Granger-cause the unemployment rate increase is conducted. We then conduct an F-test to examine whether the individual financial variable: the paper-bill spread or the federal funds rate or M2 is statistically significant in explaining the unemployment rate. The F-statistics are calculated based on two equations: one is unrestricted model by adding one of three financial variables and the other is restricted to only the unemployment rate itself and the inflation rate:

Equation (1): $ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + \mu_t$ -----Unrestricted Model

Equation (2): $ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \mu_t$ -----Restricted Model

Here ur is the unemployment rate; p is the inflation rate (using CPI less shelter). f is one of three financial variable noted above: FF, SP and M2. μ is a well-behaved disturbance term.

$$F_{(q,\ n-k)} \sim \frac{(SSRr - SSRur)/q}{SSRur\ /n-k}$$

Here q equals 6, which is the number of restricted parameter. In other words, it equals to autoregressive lag order six of itself. The degree of freedom of the unrestricted model equals the total number of observation minus the number of the coefficients, which include the intercept.

One month then is added to the dataset so that the sample covers 1960:02 through 1965:02 and the estimation and the hypothesis test are repeated. The process of adding one month to the data and repeating the causality test continues until the entire data set 1960:02 through 1995:04 is exhausted in performing the test. We apply this recursive regression for each of three financial variables: federal funds rate, paper-bill spread, and M2. Therefore, for each financial variable, the result is a set of 364 F-statistics with different degree of freedom under the null hypothesis of granger-causality

Equation (1) is used to estimate the predictive power of three variables separately and test the hypothesis of granger causality on the unemployment rate. We construct one more models to estimate the relative predictive power of one of the above three financial variables by controlling the rest two.

Equation (3): $ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i M2_{t-i} + \sum_{i=1}^{6} \eta_i SP_{t-i} +$

$$\sum_{i=1}^{6} \varphi_i FF_{t-i} + \mu_t$$

The only difference between Equation (1) and Equation (3) is that now Equation (1) is expanded to include all three financial variables simultaneously and allows them to compete against each other in the same model.

As we mentioned before, the recursive regression we introduced here would be helpful in detecting the sample sensitivity to an individual observation. The monthly observation standing out in our sample turns out to be 1974:12, which produces large increase in the test statistics. Thoma and Gray (1998) used the recursive regressions and found that that the observation of 1974:12 has exerted large influence on test statistics. This data appears to be particularly important in evaluating the paper-bill spread. The clear-cut dominance of the paper-bill spread over the federal funds rate and M2 in explaining the unemployment rate does not appear until the sample is extended to include 1974:12.The time series for the growth rate of industrial production exhibit by far its largest negative value in 1974:12, while the paper bill spread reaches a value almost double any other post-war high in 1974:07. Therefore, a record high in interest rates and spreads in mid-1974 preceded a record low growth rate in late 1974, raising the question of the extent to which the dominance of the paper-bill spread in explaining output growth can be attributed to a single observation. Our suspect is that our data might also be subject to the influence of the 1974:12 observation. Therefore, we re-estimate the Equation (1) and Equation (3) by excluding this observation. The interesting thing is that the value of 1974:12 (7.2%) is not the largest positive value in the time series but it has exerted large influence on the in-sample measures. However, in contrast, the highest unemployment rate for the observation of 1982:12 (10.8%) seems to have no special

influence on the in-sample measures. It would be interesting to explore further why the 1982:12 observation, in contrast to the 1974:12 observation, appears to be relatively non-influential.

The out-of-sample measures of fit are generally viewed as the acid test of an economic model. We aware that in-sample measures of fit may be misleading indicators of out-of-sample measures with the presence of some influential observation in our data. In other words, the in-sample performance of a model, both absolute and relative, is not necessarily a good indicator of the out-of-sample performance of the model.

Based on Equation (1), we make forecasts of the unemployment rate using three financial variables and compare the relative forecasting power. We use the Root Mean Square Error (RMSE) of 36 consecutive forecasts. There are three forecasting horizons: three-month horizon, six-month horizon and nine-month horizon.

For example, the three-month horizon forecasts using the recursive in-sample model is calculated as follows: first we estimated the six-lag vector autoregressive model over the period 1960:02 through 1965:02 and obtain the estimated model:

$$u\hat{r}_t = a + \sum_{i=1}^{6} \hat{\lambda}_i ur_{t-i} + \sum_{i=1}^{6} \hat{\gamma}_i p_{t-i} + \sum_{i=1}^{6} \hat{\beta}_i f_{t-i}$$

Using these estimated coefficients, we generate a forecast of unemployment rate in 1965:05. The forecast error for 1965:05 therefore is calculated as:

$$E_t = ur_t - u\hat{r}_t$$

We then update the sample by adding one observation, 1965:03, re-estimate the in-sample model, generate a forecast, and calculate the forecast error for the unemployment rate in 1965:06. We continue updating and generating the 3-step-ahead forecasts until the model is estimated over 1960:02 through 1968:01 and used to generate

a forecast for 1968:04. We have 36 consecutive forecast errors by now and we calculate the RMSE by the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{36} Et}{36}}$$

A low RMSE indicates relative good out-of-sample forecasting performance, and a high RMSE indicates the poor forecasting ability. We calculate the RMSE of 36 consecutive forecasts till reaching the last observation 1995:04. The methods of calculating 6-step-ahead and 9-step-ahead follow the same way as the 3-step-ahead. We save our effort to illustrate the methods of calculating 6-step-ahead and 9-step-ahead RMSEs.

Thoma and Gray argued in their paper that "the focus on 'horse races' has distracted most researchers from the more fundamental question of whether any of these financial variables contain incremental information useful in forecasting economic activity" and they compared the forecast errors generated by the models utilizing three financial variables to the forecast errors generated by the simple autoregressive model of the industrial production. We follow in the same way and compare the forecast errors of the models to the forecast errors of a pure autoregressive model of lag order six of the unemployment rate, shown by Equation (4):

Equation (4): $ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \mu_t$

Our purpose is to try to verify whether the models using the financial market variables reduce or increase forecasting power. The RMSE is calculated in the same way we describe above. If adding the financial market variables do not help predict the

unemployment rate, like Thoma and Gray's result for the industrial production, we would expect the RMSEs of the pure autoregressive model would be not very different, if not better, from the RMSEs of the models containing one of the financial market variables.

III. Empirical Results

Sample Sensitivity of Granger Causality

Figure 1 reports F-statistics for tests of the hypothesis that the lagged values of Federal Funds Rate (FF), the Paper-Bill Spread (SP), and M2 do not Granger cause the unemployment rate in Equation (1). The upper figure (a) uses the real time data and lower figure (b) uses the vintage data of February 2002.

By examining two figures, we find that before 1975, the test statistics of three financial variables are very close and it is difficult to tell which dominates which, though the test statistics of M2 is always slightly higher than the other two and most of the time exceeds the critical value. However, since 1975, it is obvious that the test statistics of paper-bill spread increases dramatically and is higher above the other two. The paper-bill spread outperforms federal funds rate and M2 in explaining the unemployment rate since then.

The federal funds rate, in both figures, becomes statistically significant in explaining the unemployment rate after 1974:12 and then plummets at the end of 1979 when the Federal Reserve shifted its operating procedure from targeting the funds rate to targeting non-borrowed reserves. Since then, the federal funds rate, in general, does not provide any explanatory power either in the real-time data or in the revised vintage data.

The explanatory power of M2 differs in the real-time data and the revised vintage data since 1979:12. In real-time data, what we observe is that it remains predictive for the unemployment rate through 1979:12 to the last observation of our sample. Yet, in the vintage data of February 2002, the explanatory power weakened to a large extent and since the end of 1987, it stays below the 5 percent critical value and remains statistically insignificant onwards. The revisions do have impact on M2 data and makes the test statistics considerably different from the results based on real-time data.

Figure 2 makes a further step. It reports F-statistics for Granger causality tests based on Equation (3), which is expanded to include all three financial variables simultaneously. Therefore we are able to control the effect of the other two variables when we examine whether the third one Granger causes the unemployment rate. FF, M2 and SP are competing against each other in the same model. The upper figure (a) uses the real time data and lower figure (b) uses the vintage data of February 2002.

For all three financial variables, the Granger-causality statistics are below the 5 percent critical value and these three financial variables remains statistically insignificant before 1975. They differ in their relative explanatory powers after that and we are going to describe separately.

The results for SP are consistent in both real-time data and the February 2002 vintage data. It outperformed M2 and the federal funds rate with its F-statistics high above the other two. Yet its explanatory power weakened in the 1990s and only remains slightly above the 5% critical value.

The federal funds rate consistently has no explanatory power for the unemployment rate in both figures. Though the federal funds rate is the main tool of

Feed's monetary policy, its effect on real economy is not so direct as other measures. It might explain the weak explanatory power in the presence of other variables in the model.

For M2, the real-time data show that during the first half of the 1980s, it Granger-causes the unemployment rate. Note that this time period is when the worst post-WWII recession occurred and the unemployment rate reached its highest value in this time period. Since then, the explanatory power weakened until 1992, when the F-statistics rise slightly above the critical value. It is possible that the Granger-causality relationship between M2 and the unemployment rate in the early 1980s is due to the shift in Federal Reserve operating procedures in the late 1979, which resulted in the decline in the explanatory power of the federal funds rate. However, in the February vintage data, M2 is consistently insignificant in explaining the unemployment rate.

Figure 1 and 2 convey useful information regarding the sample sensitivity of the test statistics to the individual observations in our data. We notice that there are two dates standing out. The first one is 1974:12. The sharp increases of the explanatory power of the paper-bill spread in both models occurred following this observation. This drives us to suspect that the previous findings of the importance of the paper-bill spread in explaining the unemployment rate are actually the effect of this single observation. Our finding is similar to what Thoma and Gray (1998) concluded. They showed that the 1974:12 observation could explain the uniformly superior performance of the paper-bill spread reported in many earlier studies. As we mentioned earlier, 1974:12 is right in the midst of the first OPEC oil-price shock. This supply shock, thought did not cause the unemployment rate to rise immediately to the highest point, has a long adverse impact on

the unemployment rate. Another important macroeconomic indicator, the growth rate of the industrial production, has its largest negative value in 1974:12. Also the mid 1970s has seen that the paper-bill spread and interest rate reached the highest point after World War II. Therefore, this observation is particularly important in the whole data. As we will see later, the exclusion of 1974:12 decreases dramatically the explanatory power of the paper-bill spread and drives it to statistically insignificant.

We re-test the model using Equation (1) by excluding the observation 1974:12 and report the test results in Figure 3. We use the same lag order six for all right-hand side variables. Therefore, eliminating the effects of 1974:12 would result in deletion of its effects in the subsequent six months. Compared with Figure 1, the change is dramatic. In Figure 3, after the 1974:12 was removed, even though the explanatory power of the paper-bill spread still dominates the other two, it has been weakened to a large extent and the magnitude of the increase after 1974:12 is much smaller. Yet the exclusion has little to do with the pattern of the explanatory power of M2 either in the real-time data or in the February 2002 vintage data. For the federal funds rate, the changes due to the exclusion are reflected in two main periods. First, during the period of 1975 to 1980, eliminating the effects of 1974:12 made federal funds rate statistically insignificant in explaining the unemployment rate, while before the exclusion it is statistically significant. Second, though it exhibits the same pattern of dramatic plummeting in its of explanatory power in the later 1979 when the Federal Reserve shift the operating procedure, the federal funds rate gains explanatory power after1982 in the real-time data, which is in contrast with its performance in Figure 1 where we did not exclude the observation of 1974:12.

Figure 4 reports F-statistics under the null of no granger causality between the financial variables and the unemployment rates. We re-test the model using Equation (3), where we include three financial variables simultaneously and allow them to compete against each other. Compared with Figure 2, eliminating the effects of 1974:12 has little impact on the performance of the federal funds rate. It remains consistently a poor explanatory variable for the unemployment rate. However, the paper-bill spread, in both the real-time data and the February 2002 vintage data, is now statistically insignificant in explaining the unemployment rate. This drives us to conclude that the seemly important explanatory power of the paper-bill spread is a post-1974 phenomenon. Its dominance beginning in late 1974 depends to a very large extent on the observation of 1974:12. The exclusion, on the other side, has no obvious effect on M2, which still remains statistically significant in explaining the unemployment rate since late 1979 in the real time data. The February 2002 vintage data shows that none of the three financial market variables is statistically significant in explaining the unemployment rate. The different data vintage makes different claims of the Granger-causality between monetary policy and real output possible.

The above analysis of the sample sensitivity of the test statistics to the individual observations mainly focuses on 1974:12, which appears to be the most influential one. The exclusion of 1974:12 results in the sharp drop of the explanatory power of the paper-bill spread from a superior performance to an insignificant and poor one. The performance of M2 seems not to be affected by this observation and remains significant in explaining the unemployment rate after 1979:12. The results also suggest that other individual observations, such as 1982:12, when the unemployment reached its historical

highest level, give some notable increase in the test statistics. However, the 1982:12 observation appears to be of no special influence as it results in the similar pattern when we exclude this observation. What we are concerned with and try to study here is that the in-sample measures are easy to be influenced by the individual observation. If we omit the potential problem posed by this influential observation, the in-sample measure of fit will turn to be a misleading indicator of out-of-sample measures of fit because the in-sample model fit cannot guarantee out-of-sample forecasting performance. We examine the out-of sample forecasting performance in the proceeding sections.

Evaluating Out-of-sample Forecasting

This section provides further evaluation of the performances of the different models. As we argued earlier, the in-sample measures of fit might be misleading indicators of out-of-sample measures, especially in our data where we have found that the in-sample test statistics are sensitive to some individual observation. The in-sample performance of a model is not necessary a good indicator of the out-of-sample performance of the model. As one of the most important functions of a time series model is to perform the task of forecasting, we argue that out-of-sample measures of fit are the correct metric. The out-of-sample measures of fit are generally regarded as the ultimate test of a model. Forecast error is one of the most important out-of-sample metrics when we evaluate the in-sample model. A good forecasting model therefore, will have the smallest forecast errors among all the alternative models, i.e., the forecast values are as close to the actual path as possible.

Figure 5 through 7 report measures of the forecasting power of models that include the paper-bill spread, the federal funds rate, and M2. We use the method we described in Section II to calculate RMSE on three different horizons: three-month (Figure 5), six-month (Figure 6), and nine-month (Figure 7). In these three figures, we show the results from the real-time data by the left side and the results from the February 2002 vintage data by the right side.

The forecasting power varies considerably across sample period. There are two periods that marked with dramatic deterioration in the forecasting ability of three financial variables. The first period is after 1974:12, where the RMSEs of the paper-bill spread, the federal funds rate, and M2 jump upwards. To be more detailed, the RMSE of the paper-bill spread is relatively lower than the RMSEs of the other two variables, yet for the federal funds rate and M2, it is difficult to distinguish the relative advantage. The second period is after 1981. The forecasting ability for the federal funds rate deteriorates the most, which is consistent with the dramatic plummeting of F-statistics under the null hypothesis of no Granger-causality depicted by Figure 1. The paper-bill spread and M2 have the similar forecasting ability in this period. We also notice that after 1988, the RMSE of the paper-bill spread increases above the RMSEs of the other two variables. The deterioration of the forecasting ability for the paper-bill spread during these periods contrasts sharply with the large jump in causality statistics shown in Figure 1. For M2, this contrast also can be visually told, even though it is not so obvious as for the paper-bill spread. Furthermore, Figure 5 through 7 show that the paper-bill spread does not has the relative advantage in forecasting over the federal funds rate and M2. The only period where it has consistently lowest RMSE for three forecasting horizons is the period from

1974:12 through 1977:12. But this is far from sufficient to support the claim that the in-sample measures of fit provide reliable indicators of out-of-sample fit as such a claim based on the in-sample measures of fit, which have been greatly affected by the 1974:12 observation. Meanwhile Figure 5 though 7 address again the problem of the sample sensitivity of the test statistics to the observation of 1974:12 as well as the observation of early 1980s.

Simple Autoregressive Model vs. Vector Autoregressive model

Up to now we have shown and analyzed the results of the in-sample measures of fit and out-of-sample explanatory power of three financial market variables of interest. We have reached conclusion that in our sample, the test statistics are sensitive to the individual observation in the sample period and the in-sample measures of fit is a misleading indicator of a forecasting model. These results are limited to the procedure of so called 'horse race' where we only focus on the question of whether any of these variables contain incremental information useful in forecasting the unemployment rate. We argue that the lag values of the unemployment rate itself contain more useful incremental information than the financial market variables do. We proceed to compare the forecast errors of the three-variable models to the forecast errors of a simple autoregressive model of the unemployment rate. The results are shown in Figure 8 through 10. We present the results on forecast horizons of three-, six- and nine-month from both the real-time data and the February 2002 vintage data.

By examining these three figures, we find that none of the financial variables is systematically useful in forecasting the unemployment rate. To make it clearer, we can visually divide the whole sample period into four sub-periods.

The first period is before 1974:12. The results for the paper-bill spread, the federal funds rate and M2 are mixed. In general, for three forecast horizons the RMSEs of the paper-bill spread model and the federal funds rate model are higher than the RMSEs from the simple autoregressive model. For M2 model, the RMSEs are higher than the RMSEs from the autoregressive model except the period of mid 1970 though 1973.

The second period starts from 1974:12 and ends at about 1979. The RMSEs for all three financial market variables are all considerably lower than the RMSEs for the autoregressive model. This appears to be the most systematic and prominent difference during the whole sample period. By the end of 1978, the RMSEs for M2 and the federal funds rates are rising again above the RMSEs for the AR model, except that the RMSEs for the paper-bill spread still stand slightly lower.

The third period appears to be the first half of 1980s, where the unemployment rate keeps rising to its historical height. During this period, the RMSEs for M2 and the paper-bill spread models are consistently lower than the RMSEs for the AR model, but for the federal funds rate, the 3-month horizon displays a pattern that clearly favors the AR model as the RMSEs for the AR model is much lower.

The last period is from the second half of 1980s until the end of the sample period. The RMSEs of models that include a financial variable are difficult to distinguish from the RMSEs of the AR model. Still, we can find that the RMSEs of the paper-bill

spread and the federal funds rate models are consistently higher than the RMSEs of the AR model.

In general, the enhancement of the importance of three financial variables in explaining the unemployment rate occurs immediately following 1974:12. Over most of the sample periods, the RMSEs of models include the paper-bill spread, or the federal funds rate, or M2 perform no better than the AR model. The RMSEs are not different between the real-time data and the February 2002 vintage data. These findings drive us to conclude that the unemployment rate itself has higher forecasting ability than any model that includes the financial market variables.

IV. Conclusion

Our paper, following Thoma and Gray (1998), studies the Granger-causality relationship between financial market variables and the unemployment rate. We use recursive regressions to address the sample sensitivity of the test statistics to the individual observations. We start with constructing the in-sample models and testing the null hypothesis that the financial variables, either alone or in combination, do not Granger-cause the unemployment rate. Through this we compare the relative explanatory power of three financial market variables: the paper-bill spread, the federal funds rate, M2. Our finding is consistent with earlier results: the paper-bill spread is superior to both the federal funds rate and M2 explaining real economic activities. However, the recursive regressions help us to detect the prominent and influential observation of 1974:12. We notice that the sharp increase of importance of the paper-bill spread occurs at the 1974:12 observation. Our Granger-causality statistics are especially sensitive to this observation,

when the United States was in the midst of the first OPEC oil-price shock. The individual observation of 1974:12 is shown to heavily influence the in-sample measures of fit and is account for uniformly superior performance of the paper-bill spread in our studies as well as in the previous researches. We re-test the models with the time series excluding this influential observation and we find that the explanatory power of the paper-bill spread is dramatically reduced.

A good in-sample fit cannot guarantee a good the out-of-sample forecasting performance, and even worse in-sample measures of fit can potentially be misleading indicators of a forecasting model's performance in the presence of some influential observations in the sample period. Therefore we proceed to evaluate the forecasting model using the out-of-sample forecasting metrics. RMSE provides a good measurement tool for our purpose. The comparison of the RMSEs of models including the paper-bill spread, the federal funds rate and M2 suggest that, in contrast with the results of the in-sample measures of fit, the paper-bill spread does not appear to have a relative advantage in forecast ability. Moreover, the forecast ability deteriorates to a large extent immediately after 1974:12, when the paper-bill spread has relatively better out-of-sample forecasting performance. In general, the in-sample measures of fit do not provide reliable indicators of out-of-sample fit for these models.

Another important finding in this paper lies in the results that the financial variables contain no incremental information useful in predicting the unemployment rate. In contrast, the pure autoregressive model of the unemployment rate outperformed the models, which include the financial market variables. Given the potential data imperfection on which we test our models, we would conclude that either monetary

policy innovations have no significant real effects on the unemployment rate, or there are still potential problems existing in our current measurement of the money policy.

The different vintages of data matter for evaluating an economic model. We test this claim by running models on two data sets: the real-time data and the February 2002 vintage data. The February 2002 data vintage is revised and updated by incorporating new sources, and the real-time data is the information set available at a particular date to the economists for performing forecasts in real time. We find that the different vintages yield different results on the in-sample measures of fit. Particularly, it biases the granger-causality relationship towards zero in the February 2002 data vintage. The sensitivity of the data to revisions drives us to suspect the robustness of the earlier claims about the Granger-causality between the monetary policy and real output. Thoma and Gray did not address this problem in their paper (1998). It is difficult to conclude which data set we should use as the standard. The economists and researchers use different data vintages as the "actual" values and perform the forecasts. This, no wonder, gives rise to a rich array of explanations on Granger-causality between monetary policy and real output in the pervious literature.

Figure 1

Tests of the Hypothesis that M2, SP, FF Do not Granger-Cause Unemployment Rate
(a). Real-time Data

	SP		M 2
	F F		F(.05)

(b). The Vintage Data of February 2002

	SP		M 2
	F F		F(.05)

Vertical axis: F statistics and .05 critical value.
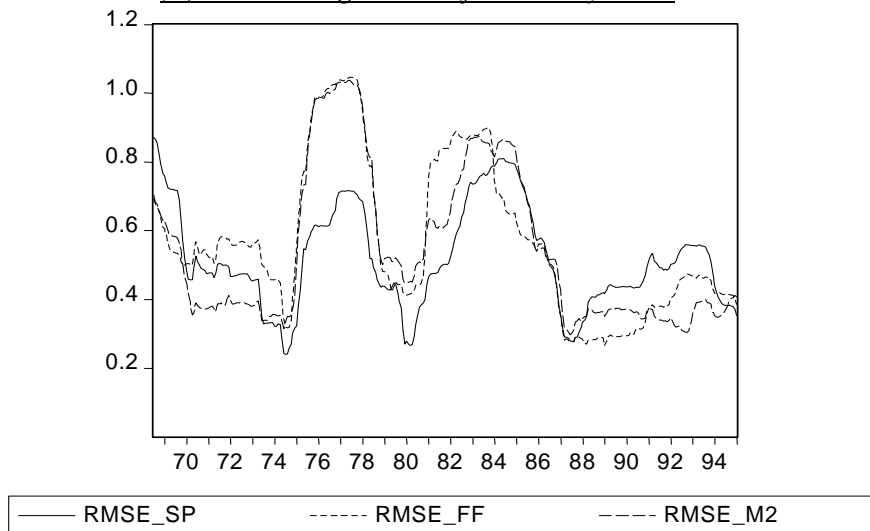Horizontal axis: End-date of the sample
Estimated Equation:

$$ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + \mu_t$$

where *ur* is unemployment rate and *p* is growth gate of the CPI less shelter; *f* is either *M2, SP* or *FF*; *M2* is the growth rate of the monetary aggregate *M2; SP* is the difference between the six-month commercial paper and treasury bill rates; *FF* is the federal funds rate.

Figure 2
Tests of the Hypothesis that M2, SP, FF Do not Granger-Cause Unemployment
Rate:
(a). Real-time Data

(b). The Vintage Data of February 2002
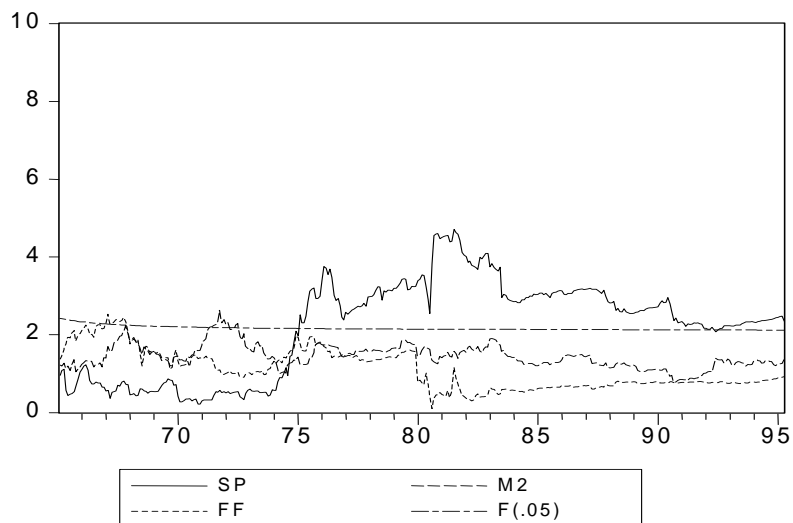
Vertical axis: F statistics and .05 critical value.
Horizontal axis: End-date of the sample
Estimated Equation:

$$ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i M2_{t-i} + \sum_{i=1}^{6} \eta_i SP_{t-i} +$$

$$\sum_{i=1}^{6} \varphi_i FF_{t-i} + \mu_t$$

where *ur* is unemployment rate and *p* is growth gate of the CPI less shelter; *M2* is the growth rate of the monetary aggregate *M2; SP* is the difference between the six-month commercial paper and treasury bill rates; *FF* is the federal funds rate.

Figure 3
Tests of the Hypothesis that M2, SP, FF Do not Granger-Cause Unemployment
Rate: 1974:12 Removed
(a). Real-time Data

(b). The Vintage Data of February 2002

Vertical axis: F statistics and .05 critical value.
Horizontal axis: End-date of the sample
Estimated Equation:

$$ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + \mu_t$$
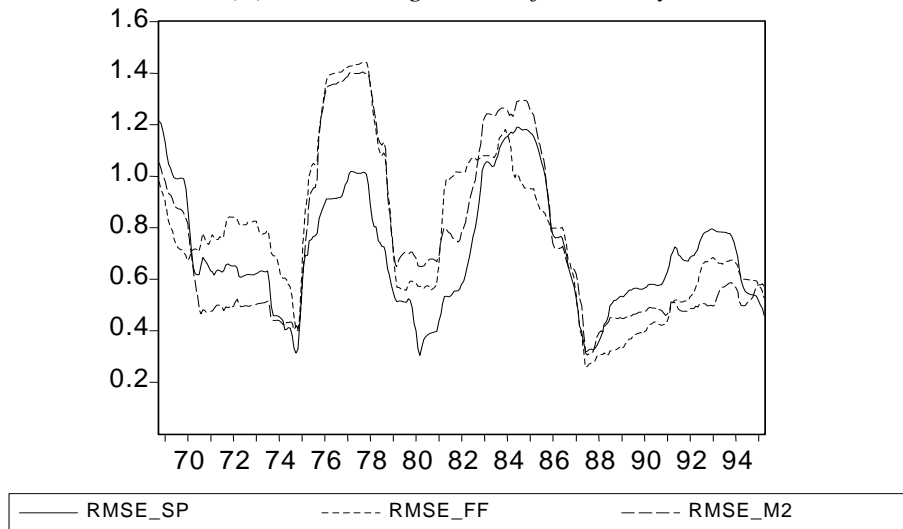
where *ur* is unemployment rate and *p* is growth gate of the CPI less shelter; *f* is either *M2, SP* or *FF*; *M2* is the growth rate of the monetary aggregate *M2; SP* is the difference between the six-month commercial paper and treasury bill rates; *FF* is the federal funds rate.

Figure 4
Tests of the Hypothesis that M2, SP, FF Do not Granger-Cause Unemployment
Rate: 1974:12 Removed
(a). Real-time Data

(b). The Vintage Data of February 2002

Vertical axis: F statistics and .05 critical value.
Horizontal axis: End-date of the sample
Estimated Equation:

$$ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i M2_{t-i} + \sum_{i=1}^{6} \eta_i SP_{t-i} + $$

$$\sum_{i=1}^{6} \varphi_i FF_{t-i} + \mu_t$$

where *ur* is unemployment rate and *p* is growth gate of the CPI less shelter; *M2* is the growth rate of the monetary aggregate *M2; SP* is the difference between the six-month commercial paper and treasury bill rates; *FF* is the federal funds rate.
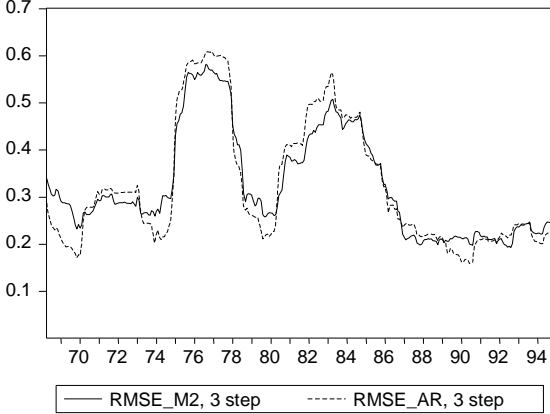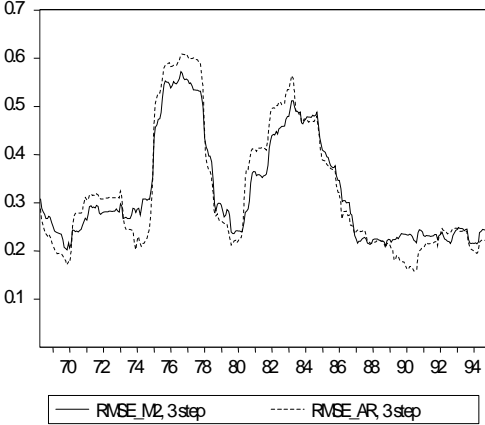
Figure 5
Root-Mean-Square-Errors of Models including M2, SP, or FF:
3-month Forecast Horizon
(a). Real-time Data
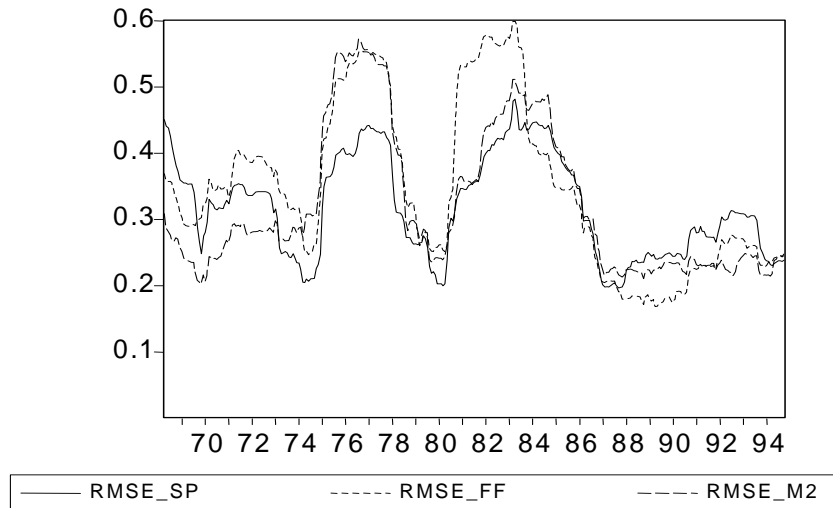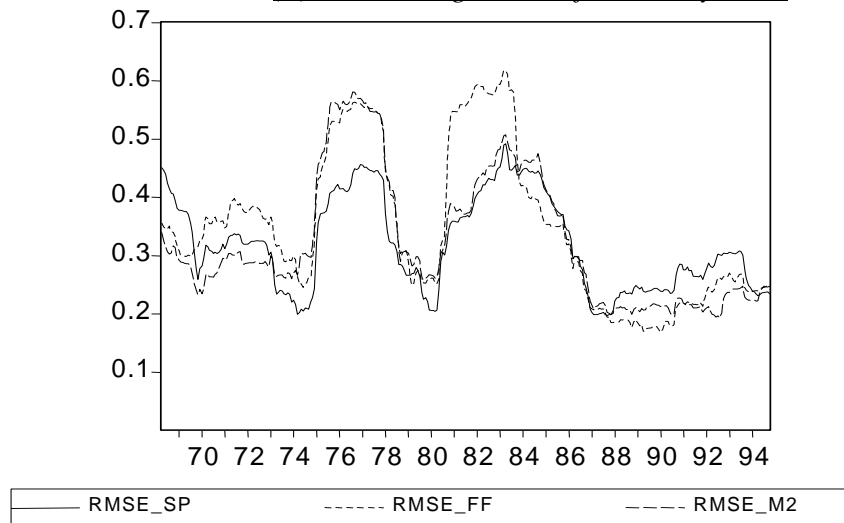
| RMSE_SP | RMSE_FF | RMSE_M2 |

(b). The Vintage Data of February 2002

| RMSE_SP | RMSE_FF | RMSE_M2 |

Vertical axis: RMSE of 36 forecasts generated form the equation.
Horizontal axis: Date of last forecast error used in calculating RMSEs.
Estimated equation:

$$ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + \mu_t$$

where *ur* is unemployment rate and *p* is growth gate of the CPI less shelter; *f* is either *M2, SP* or *FF*; *M2* is the growth rate of the monetary aggregate *M2; SP* is the difference between the six-month commercial paper and treasury bill rates; *FF* is the federal funds rate.

Figure 6
Root-Mean-Square-Errors of Models including M2, SP, or FF:
6-month Forecast Horizon
(a). Real-time Data
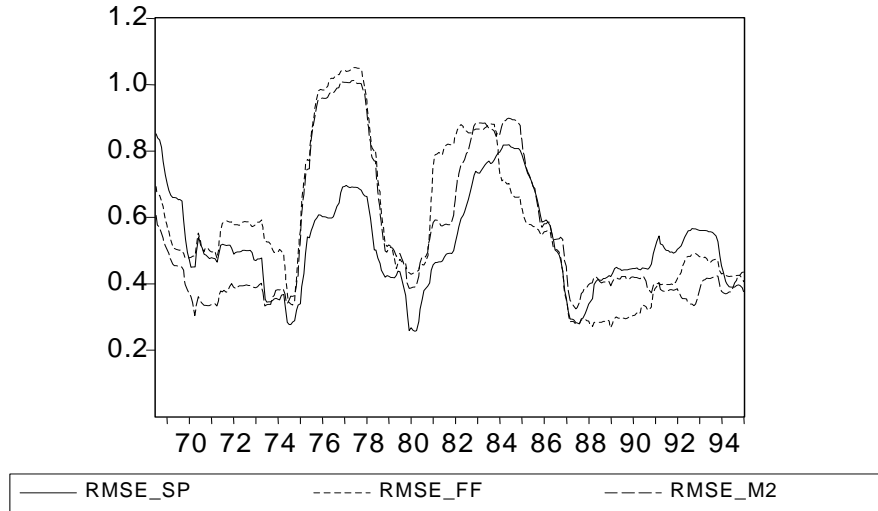
| RMSE_SP | RMSE_FF | RMSE_M2 |

(b). The Vintage Data of February 2002

| RMSE_SP | RMSE_FF | RMSE_M2 |

Vertical axis: RMSE of 36 forecasts generated form the equation.
Horizontal axis: Date of last forecast error used in calculating RMSEs.
Estimated equation:

$$ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + \mu_t$$

where *ur* is unemployment rate and *p* is growth gate of the CPI less shelter; *f* is either *M2, SP* or *FF*; *M2* is the growth rate of the monetary aggregate *M2; SP* is the difference between the six-month commercial paper and treasury bill rates; *FF* is the federal funds rate.
.

Figure 7
Root-Mean-Square-Errors of Models including M2, SP, or FF:
9-month Forecast Horizon
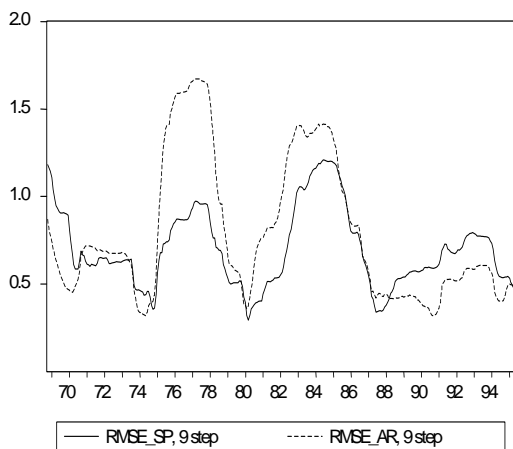*(a). Real-time Data*
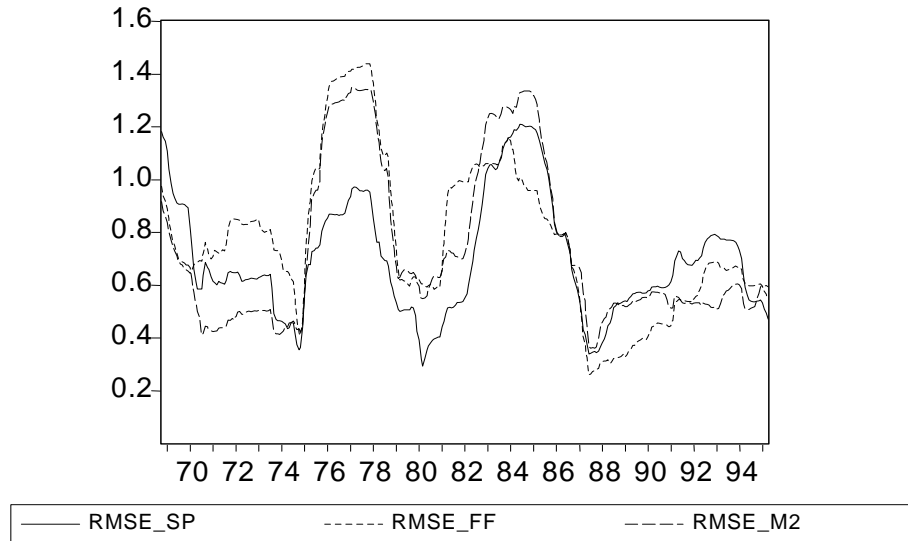
—— RMSE_SP	------ RMSE_FF	– – – RMSE_M2

(b). The Vintage Data of February 2002

—— RMSE_SP	------ RMSE_FF	– – – RMSE_M2

Vertical axis: RMSE of 36 forecasts generated form the equation.
Horizontal axis: Date of last forecast error used in calculating RMSEs.
Estimated equation:

$$ur_t = \alpha + \sum_{i=1}^{6} \lambda_i ur_{t-i} + \sum_{i=1}^{6} \gamma_i p_{t-i} + \sum_{i=1}^{6} \beta_i f_{t-i} + \mu_t$$

where *ur* is unemployment rate and *p* is growth gate of the CPI less shelter; *f* is either *M2, SP* or *FF*; *M2* is the growth rate of the monetary aggregate *M2; SP* is the difference between the six-month commercial paper and treasury bill rates; *FF* is the federal funds rate.

Figure 8
Comparing the RMSEs of the three variable M2 Model
and the Autoregressive Model

Real-time Data   *The Vintage Data of February 2002*

8a.

8b
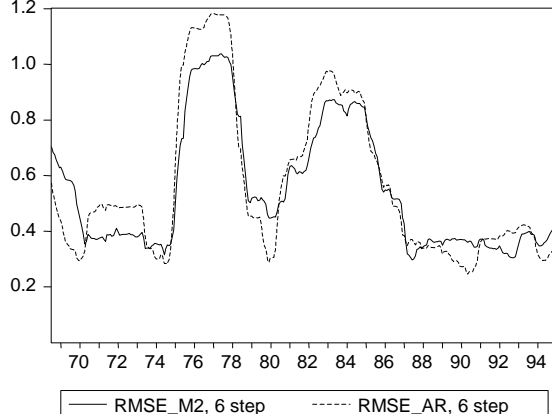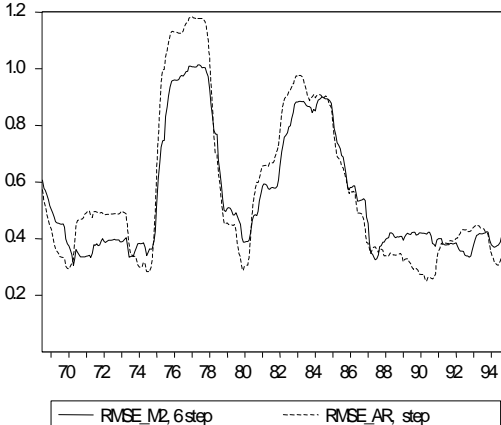
8c
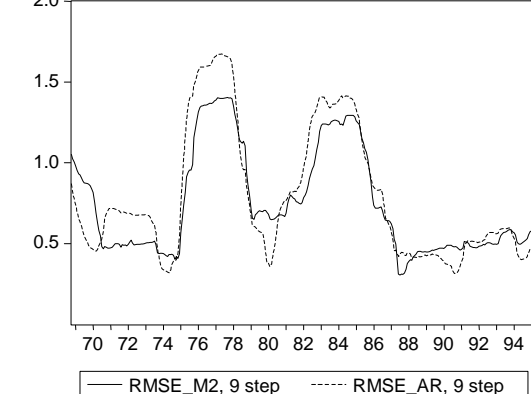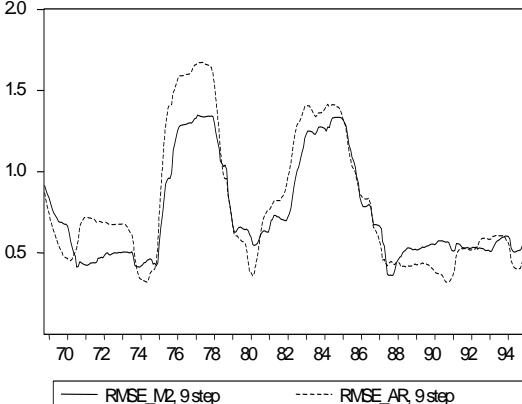
Figure 9
Comparing the RMSEs of the three variable SP Model
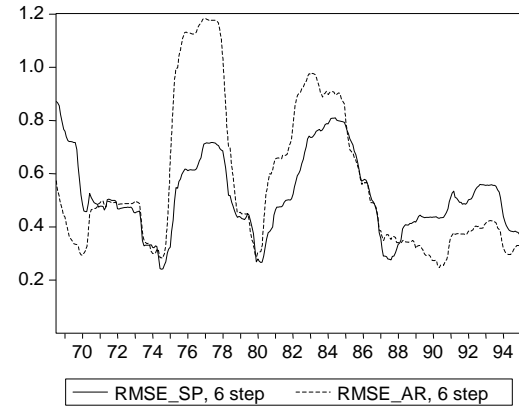and the Autoregressive Model

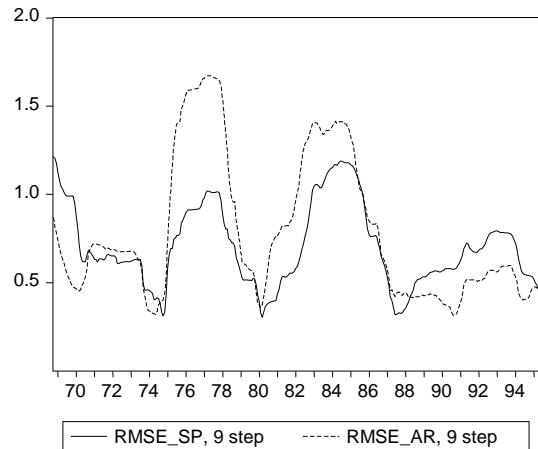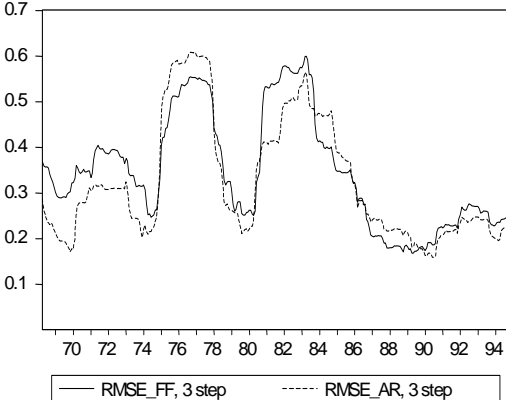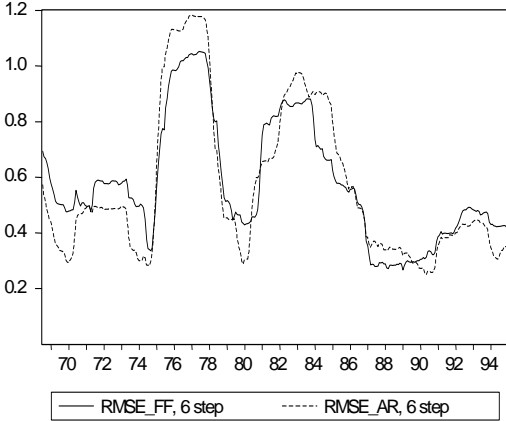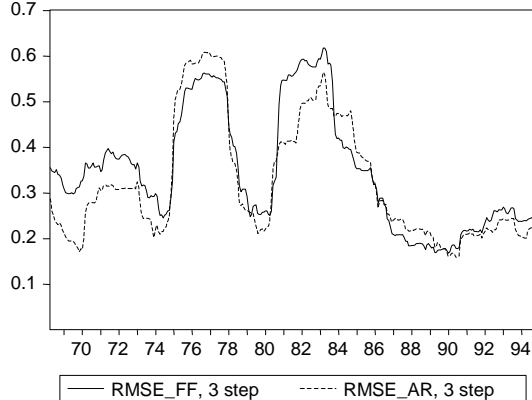*Real-time Data*   *The Vintage Data of February 2002*
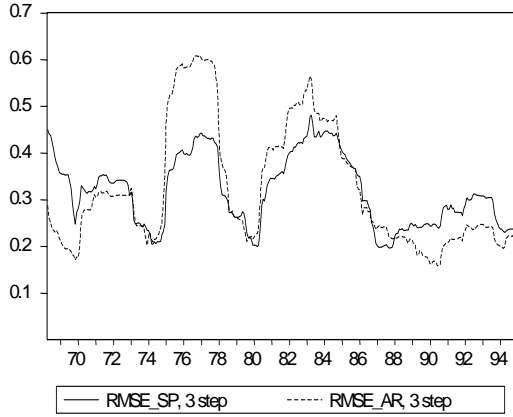
9a
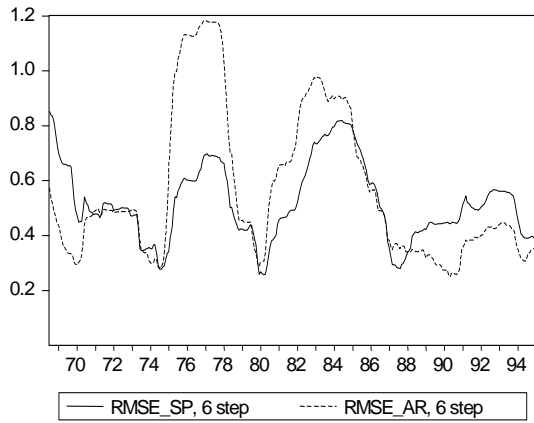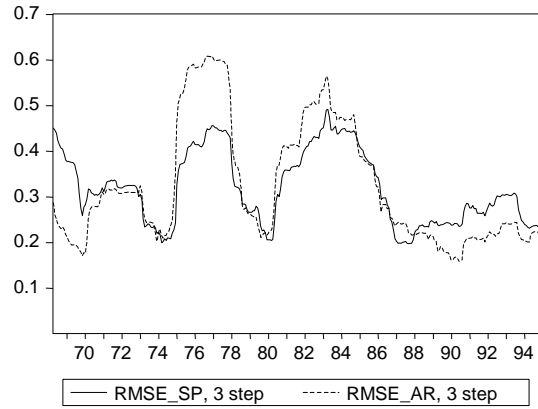
9b

9c

Figure 10
Comparing the RMSEs of the three variable FF Model
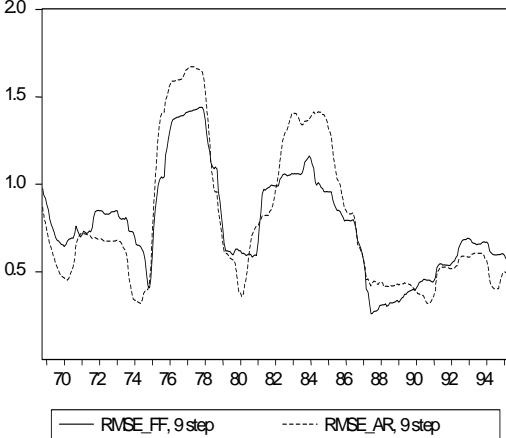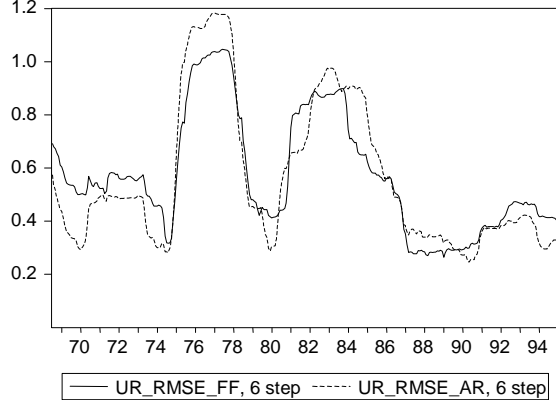and the Autoregressive Model

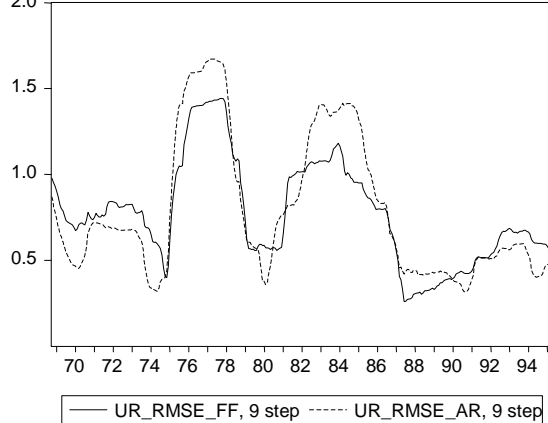*Real-time Data*   *The Vintage Data of February 2002*

10a

10b

10c

References:

Amato, Jeffery D. and Swanson, Norman R (2000), "The Real-time Predictive Content of Money for Output," *Journal of Monetary Economics* 48, 2001, 3-24

Clements, Michael P. (2002). "Why Forecasting Performance Does Not Help Us Choose A Model?" working paper, Department of Economics, University of Warwick.

Croushore, Dean and Stark, Tom (2000), "A Real-Time Data Set for Macroeconomists: Does Data Vintage Matter for Forecasting?" working paper, Research Department, Federal Reserve Bank of Philadelphia, June 2000.

Thoma, Mark A. and Gray, Jo Anna (1998), "Financial Market Variables Do Not Predict Real Activity," *Economic Inquiry*, Vol. XXXVI, October 1998, 522-539.