GDP NOWCASTING USING HIGH FREQUENCY ASSET PRICE, COMMODITY PRICE AND BANKING DATA

A Master's Thesis

by
BİNNUR BALKAN

Department of
Economics
İhsan Doğramacı Bilkent University
Ankara
June 2011

GDP NOWCASTING USING HIGH FREQUENCY ASSET PRICE, COMMODITY PRICE AND BANKING DATA

Graduate School of Economics and Social Sciences
of
İhsan Doğramacı Bilkent University

by

BİNNUR BALKAN

In Partial Fulfillment of the Requirements For the Degree
of
MASTER OF ARTS

in

THE DEPARTMENT OF
ECONOMICS
İHSAN DOĞRAMACI BILKENT UNIVERSITY
ANKARA

June 2011

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

Assoc. Prof. Dr. Refet S. Gürkaynak

Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

Assist. Prof. Dr. Taner Yiğit

Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

Assoc. Prof. Dr. Aslıhan Altay Salih

Examining Committee Member

Approval of Graduate School of Economics and Social Sciences

Prof. Dr. Erdal Erel

Director

ABSTRACT

GDP NOWCASTING USING HIGH FREQUENCY ASSET PRICE, COMMODITY PRICE AND BANKING DATA

BALKAN, Binnur

M.A., Department of Economics

Supervisor: Assoc. Prof. Dr. Refet Gürkaynak

June 2011

Knowing the current state of the economy is important especially when we consider that GDP information comes with a lag of quarter. From this perspective, employing high frequency variables in GDP nowcasting may contribute to our knowledge of economic conditions, since they are timelier compared to GDP. This paper deals with nowcasting US GDP using an expectation maximization algorithm in a Kalman filter estimation, which includes asset prices, commodity prices and banking data as explanatory variables together with real variables and price indices. As a result of the estimations, asset prices and other high frequency variables are found useful in nowcasting US GDP contrary to previous studies. Model predictions beat the traditional methods with the medium size model, which includes fifteen variables, yielding the best nowcast results. Finally, this paper also proposes a new route for achieving better nowcast results by changing system specifications of the state variables.

Keywords: Nowcasting, Kalman Filter, EM Algorithm, Asset Prices

ÖZET

VARLIK FİYATLARI, EMTİA FİYATLARI VE BANKACILIK VERİLERİ KULLANILARAK GSYH'NIN ŞİMDİKİ ZAMAN TAHMİNİ

BALKAN, Binnur

Yüksek Lisans, Ekonomi Bölümü

Tez Yöneticisi: Doç. Dr. Refet Gürkaynak

Haziran 2011

GSYH'nin bir çeyreklik gecikme ile öğrenildiği göz önünde bulundurulduğunda, ekonominin bulunduğu durum ile ilgili gerçek zamanda doğru bilgi sahibi olmanın önemi bir kez daha ortaya çıkmaktadır. GSYH'nin şimdiki zaman tahminine yüksek frekanslı varlık fiyatlarını, emtia fiyatlarını ve bankacılık verilerini koymak ekonominin güncel durumu hakkındaki bilgimize katkı sağlayabilir. Bu çalışmada Amerika Birleşik Devletleri GSYH'sinin tahmininde bahsi geçen yüksek frekanslı veriler, düşük frekanslı reel veriler ve fiyat endeksleri ile birlikte bir Kalman filtresi içerisinde beklenti maksimizasyonu algoritması ile kullanılmıştır. Önceki çalışmaların tersine, yüksek frekanslı verilerin GSYH'nin eş zamanlı tahmininde hatayı azalttığı bulunmuştur. Farklı modeller arasında, on beş farklı seri içeren orta ölçekli model diğer modellerden ve geleneksel tahmin metodlarından daha iyi sonuçlar vermiştir. Bunlara ek olarak, bu çalışma farklı model kurguları kullanarak daha iyi tahminlerine ulaşmanın mümkün olduğunu göstermiştir.

Anahtar Kelimeler: Simdiki Zaman Tahmini, Kalman Filtresi, Beklenti Maksimizasyonu Algoritması, Varlık Fiyatları

ACKNOWLEDGEMENTS

I owe my deepest gratitude to Refet S. Gürkaynak for his invaluable guidance, supervision and endless support for all stages of my study. I would like to thank Taner Yiğit as the coadvisor and one of my thesis examining committee members, who provided worthy guidance from beginning to end. This thesis would not be possible without their detailed comments on my work, and their valuable contributions and ideas. I also would like to thank Aslıhan Altay-Salih as an examining committee member for her helpful comments and suggestions. I am indebted to Marta Banbura for providing MATLAB codes and Lucrezia Reichlin for invaluable guidance. I would like to thank TÜBİTAK for its financial support during my study. I would like to thank my friends for their help especially to Kadir Apikoğlu, Burak Alparslan Eroğlu and MuratCan Alkan. Last but not the least, I am grateful to my family for their unconditional support. However, I bear the sole responsibilities for all the mistakes made in the thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

For policymakers and corporate decisiontakers, it is important to make correct inferences on the current state of the economy; especially when we consider GDP information comes with lag of a quarter. Knowing this fact, estimating current quarter's GDP under the light of latest available high frequency data becomes an important task for policymaking and decisiontaking concerns. Depending on this idea, there is a quite new but fastly growing literature on GDP nowcast together with the estimation of business condition or current state of the economy indices, which are constructed to summarize overall current economic activity.

Among the recent studies, works on the US economy are builded on extracting a business condition or current state of the economy indices using high frequency data in a Kalman filter estimation rather than nowcasting GDP or any other named variable. On the contrary of US studies, papers on the EU area deal with nowcasting GDP directly under a similar construction of the Kalman filter but the data preference is over low frequency namely monthly and quarterly variables. In this study, by combining EU and US approaches, I am trying to nowcast US GDP using high frequency data under an expectation maximization algorithm in a Kalman filter estimation. Hence, this study follows EU methodology in the methodology aspect but US studies' high frequency data structure drives the estimation results.

As one of the well known studies in the area, Aruoba, Diebold and Scotti (2009) create a current state of the economy index using high frequency data in a Kalman filter. In constructing this index, the authors use quarterly GDP, monthly unemployment, weekly initial jobless claims and the daily yield curve term premium data in order to summarize economic activity in one unobserved factor which is called as the current state of the economy. They use a Kalman filter within a daily frequency framework and propose that using Kalman filter is the efficient way under the arbitrary pattern of missing data which is observed due to simultaneous use of high and low frequency variables in a daily representation such that quarterly variables are observed once in ninety days and consist of eighty nine missing observations in each quarter. As a result of the estimations for current state of economy, ADS finds out that such a model is quite good in catching NBER defined recessions but including daily asset price variable does not alter the estimation outputs significantly. In a later paper on US economy, Aruoba and Diebold (2010) nowcast real activity and inflation at monthly frequency by again using a Kalman filter. Similar to the previous studies on the US economy, the authors estimate one latent factor for real activity and one for inflation separately. In this work, they prefer to do this estimation at monthly frequency in order to ease computational complexity. This computational burden is coming mostly from the low frequency flow variables and missing observation patterns such that when GDP is put into a daily framework, one need to estimate 90 parameters to explain the relationship between lags of the unobserved state variable and one period GDP. This requirement is coming from the GDP's flow structure which means a GDP observation is not a snapshot value for the day it is observed but the summation of ninety days values in that quarter. Similarly, increasing number of variables used in the estimation results in convergence problems for maximum likelihood as Banbura et al. (2010b) suggests. Therefore, working with high number of variables in a daily framework using Aruoba and Diebold's estimation methodology is computationaly hard especially when they are low frequency flow variables.

On the other hand, Banbura et al. (2010b) work on the Euro Area data with an Expectation Maximization Algorithm in the Kalman Filter estimation of latent factors. They prefer to use EM algorithm instead of direct use of MLE because of the computational easiness of this algorithm in a large scale data. Since the EM algorithm first fills-in the missing values for any variable in each step of the estimation, the Kalman filter works as if there is no missing data under the EM algorithm. Therefore, there will be no need for time varying system matrices which results in a significant reduction in the number of parameters to be estimated. In 2010 study, Banbura et al. nowcast EU area GDP and work with different sizes of data sets in terms of the number of observed variables are utilized in the Kalman filter. In this paper, the authors use real, price, money and asset price variables together with some surveys on EU area which are all at either monthly or quarterly frequency. As a result of their estimations, they find that the forecast accuracy of largest data set is a bit worse than the others but all different size estimations outperform traditional methods especially in nowcasting EU area GDP.

As it is stated earlier, in this study, I am proposing to use Banbura et al. (2010b) methodology on the U.S. data with slight modifications. Although the model is estimated under a monthly framework, in order to observe the nowcast performance of higher frequency variables on a daily basis, I take the rolling monthly average of daily, weekly and biweekly variables as they become available. At this point, considering the high frequency variables, motivation of using asset prices in nowcasting is their theoretical ability to summarize all available information in the economy and forward looking structure of asset prices as Stock and Watson suggest and test in "Forecasting Output and Inflation: The Role of Asset Prices" (2000). Moreover, monetary data can provide an important insight to current state of the economy especially considering their policy content and finally banking data summarize the information for the conditions of the commercial banks together with economic condition of the households.

The three main contributions of proposed study are

- nowcasting quarterly GDP of US by an EM algorithm in a Kalman filter

- identifying the significance of high frequency and forward looking variables in nowcasting US GDP,

- and determining the forecast performance of named variables rather than observing the contribution of blocks of variables.

The rest of the paper organized as follows: In the following section, I am discussing data briefly and providing modeling structure. Then, the methodology and estimation of US GDP with the proposed methodology will be discussed in the third section. Nowcast results together with the forecast uncertainty and robustness checks will be provided in the fourth section. Finally, conclusion and the future work closes the paper.

CHAPTER 2

DATA AND MODELING FRAMEWORK

2.1 Data Description

In this study, I collect information consisting of 93 variables where the complete list, flow - stock structure and necessary transformation are available in the Appendix A. For this data set, the observation frequencies are changing from daily to quarterly and it is possible to summarize all information under six groups considering their information content. At this point, it is important to keep in mind that grouping structure will not be carried through the estimation process but as a summary, the variables are distributed into groups as

1. Real Variables: 47 series in different frequencies from weekly to quarterly such as initial jobless claims, industrial production and GDP. Most of the variables under this group are the disaggregated industrial production and GDP data.

2. Asset Prices: 20 variables in several frequencies varying from daily to monthly such as federal funds rate, 4-week T-Bill in the secondary market, or Moody's corporate bonds' index. Different maturity bonds together with yield curve information are included into asset prices.

3. Commodity Prices: Only three variables where two of them are weekly averages and one of them is monthly average which are all formulation gas prices, diesel sales prices and spot oil prices respectively.

4. Banking: Six series, the lowest frequency is quarterly whereas the highest

frequency is weekly such as total checkable deposits and total net loan charge-offs. Although more disaggregated data available for US banking system for different states or financial institution types, only the most aggregated ones are included in this study to reduce sector specific noise content.

5. Price Indices: Eight series which are either monthly or quarterly, all items CPI, house price index are the examples of this block. Other variables are also forms of price indices for different good and service groups and University of Michigan Inflation Expectation survey is the only survey information among all groups.

6. Money: For nine variables, frequency varies from biweekly to quarterly such as velocity of M1, currency component of M1 or monetary base.

Next section follows with the monthly state space representation of model for arbitrary number of variables and observation periods. Although writing monthly or any different frequency state space will not alter the system representation, it will require to adjust different frequency variables into system accordingly as it will be explained in the following chapters.

2.2 State Space Representation

Let's assume $\mathrm{y}_t = [y_{1t},\ y_{2t,}\ ...,\ y_{nt}]'$, $t = 1, ..., T$ denotes the observation matrix where n is the number of observed variables and T denotes the observation period in months and years. Then, all of the above mentioned variables can be written into y with slight modification as it will be explained in the next chapter.

Then, knowing the observed variables, the Kalman Filter extracts the common component among these variables and summarizes this common information into one or multiple state variable(s). These unobserved state variable(s) are designed to follow a prespecified AR process. Finally, relating the observed variable into unobserved state vector f requires the following state space representation where the first system is written for observation and the second one is written for the state equations under the monthly state-space representation.

Observation Equation

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ .. \\ .. \\ .. \\ .. \\ Y_{nt} \end{bmatrix} = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \\ \\ .. \\ \Lambda_{n-1} \\ \Lambda_n \end{bmatrix} \begin{bmatrix} f_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \\ \\ .. \\ .. \\ \varepsilon_{nt} \end{bmatrix}$$

State Equation

$$\begin{bmatrix} f_t \\ f_{t-1} \\ .. \\ .. \\ .. \\ \\ f_{t-p} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & .. & .. & A_p \\ I & 0 & 0 & .. & .. & 0 \\ 0 & I & 0 & .. & .. & 0 \\ & & .. & & & \\ & & .. & & & \\ & & .. & & & \\ 0 & 0 & 0 & .. & I & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ .. \\ .. \\ .. \\ \\ f_{t-p-1} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ 0 \\ 0 \\ .. \\ .. \\ 0 \end{bmatrix}$$

In the state space, Λ's represent the factor loadings for each observed variable separately, where A_q's determine the AR process of the state variable. Finally, all error terms namely ε and u follow random walk process $\varepsilon_{it} \sim i.i.d.N(0, \sigma_\varepsilon^2)$ and $u_t \sim i.i.d.N(0, \sigma_u^2)$. In such a model, there will be no identification problem; since the number of factors are significantly smaller than the number of observations in y_t as Banbura et al. (2010) suggests.

The following section continues with the necessary data transformations conducted on high and low frequency variables. Since a monthly factor model will be employed in the estimation section, it is necessary to transform all information into monthly frequency observations via either aggregation or disaggregation.

2.3 Modeling Quarterly and Daily Variables

Quarterly variables can be thought as partially observed monthly variables in a monthly state space model. In every quarter, first two months constitute missing observations and the last month is the observation period as a convention (Banbura, Modugno, 25). Then, missing observations appearing in the monthly state space representation due to frequency difference can be filled via an approximation on the quarterly growth rate. Assuming y_{1t} denotes the unobserved month to month growth rate of a quarterly variable and $\bar{\mathrm{y}}_{1t}$ represents the quarterly observed growth rate, the approximation can be written as

$$\bar{\mathrm{y}}_{1,t} = y_{1,t} + 2y_{1t-1} + 3y_{1t-2} + 2y_{1,t-3} + y_{1,t-4}$$

a la Banbura and Modugno (2010b). Then, this newly constructed month to month growth rate follows above factor model together with other monthly variables. This means, we can relate the observed quarterly growth rate to state variable as follows

$$\bar{\mathrm{y}}_{1,t} = [\ \Lambda_1 \quad 2\Lambda_1 \quad 3\Lambda_1 \quad 2\Lambda_1 \quad \Lambda_1] \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \end{bmatrix} + \varepsilon_{1t}$$

In addition to the low frequency variables' transformation, high frequency variables are turned into monthly observations via monthly rolling averaging. As an observation becomes available for any daily, weekly or biweekly variable, it is included in the monthly rolling average. For example, for a daily variable, simple average of last 30 available observations are put into estimation as the monthly variable. Therefore, at the end of the month, monthly average of current period's high frequency variable replaces the monthly observation depending on the probable release delays. In the estimation part, only the result for end of month estimations will be provided; therefore, all monthly averages reflect the information only for the current months.

CHAPTER 3

METHODOLOGY AND ESTIMATIONS

3.1 Signal Extraction

Knowing the above state space representation, it is possible to proceed with estimations in Kalman Filter for signal extraction or more clearly for estimating the state variable f. In extracting common component and estimating the factor loadings from the observed variables, the main estimation technique of this study is an application of a specific case of maximum likelihood estimation: expectation maximization (EM) algorithm. The EM algorithm works in two step: in the first step, system parameters are estimated via maximization of the likelihood function; in the second step, missing observations are filled in depending on previous parameter estimations at each observation date as Banbura and Modugno identifies (2010).

As a deficiency of this methodology, estimation process requires fully observed sample at least for the first observation period in order to initialize maximization algorithm and it continues to work with missing data patterns afterwards; since the algorithm fills in these missing observations in the second step. Therefore, in order to fill first row of the observation matrix, random numbers are drawn from a (0,1) normally distributed series to replace the missing observations and since all the variables are normalized previously, no need to specify prior observations separately.

3.2 Estimation

In the reference estimation set, t represents the period from 1960, January to 2010, December. On the other hand, number of observed variables which are chosen from 93 mentioned series will change among different size estimations from six to

23.

Through the estimation steps, all variables are normalized to zero mean unit variance and quarterly variables are written as monthly approximation and other higher frequency variables are written as monthly rolling averages accounting for missing observations as stated in previous chapters. Therefore, there will be no difference between stock and flow variables as it was in mixed frequency Kalman Filter estimation because all of the variables are observed as if they are one-time snapshot observation in the monthly frequency. Then, Kalman filter's estimation of state vector relates every observation to single period unobserved state vector via its loading and fills in the predictions for GDP data using latest available high frequency information.

Since all methodological framework is constructed for arbitrary n and T, it is possible to use different number of observed variables in the estimation of GDP for different time frames. To exploit this property, I conduct three different size estimations in this study with the complete model list is available in the appendix. In the small model specification, only six real variables are used in GDP nowcast. In medium model, seven variables from remaining five categories are added to previous six real variables. Finally, in large model, 22 variables including previous ones are used to nowcast US GDP for the period 1999 Quarter 1 to 2011 Quarter 1.

Following Aruoba and Diebold (2010) variables used in all three estimation sets are dated back to 1960. Moreover, following Banbura and Modugno (2010), four AR(2) unobserved factors are estimated prior to GDP nowcast via Kalman filter and EM algorithm as proposed in the third section. On the other hand, both of this conditions are relaxed in the robustness checks in order to observe the possible distortions created by these specifications.

CHAPTER 4

RESULTS

4.1 Model Predictions and Nowcast Uncertainty

Appendix B summarizes the estimation results of three different models under the proposed framework. In Figure 1, three models' nowcasts of US GDP together with average nowcast value are drawn for the period 1999 quarter 4 to 2011 Quarter 1. Although the lines are so close that it is hard to distinguish one from another, there is strong difference between model performances especially for different subsamples. For example, small model generally overshoots GDP especially for after 2008 crisis period but for the same time frame, large models continuously undershoots as Figure 2 also underlines.

If we consider the latest announced GDP number for the first quarter of 2011, small model continues to hold its leading position with an error little smaller than 50 billion dollars whereas large model overshoots same period GDP by 333 billion dollars. Although one observation is not informative in examining the nowcast performance of different model specifications, it is interesting since it is the only unknown value for the time of estimations.

As a convention, taking average of three models' nowcasts for GDP does not alter results significantly on the contrary of Banbura and Modugno predictions for EU GDP (2010b). Averaging three model nowcasts yields an improvement especially after 2008 period but the medium model beats even the average nowcast prior to 2007 and so in the full sample.

Comparing three different models errors with basic AR process results in a significant difference between model performances which is measured by squared forecast error (SFE) as Table 3 presents. Referring to same table, ordering the small, medium and large models with respect to squared forecast errors yields the conclusion that adding information into estimation process is found to be useful in this study. There is an improvement in SFE from small model to medium due to inclusion of several asset price, commodity price and banking variables. However, this improvement is dropped down when the information is added continuously i.e. large model's predictions are even worse than small's. Therefore, nowcast uncertainty measure offers the medium model as the main tool for GDP nowcast for the US economy.

On the other hand, when we divide nowcast accuracy into subsamples, first and the most significant observation is coming from the difference between pre-2007 and after 2007 periods. Although, the medium model specification yields the lower SFE for the whole sample, after 2007, small model surpasses other two and create the most reliable nowcasts. Besides, large model undershoots the GDP for all quarters, beginning from the last quarter of 2007 to present except the first quarter of 2011.

From all these observations, it is possible to understand the current crisis results in a gap between asset price and real variable informations. Since there has been a massive policy usage in this period, the US government is able to moderate their GDP loses compare to the significant break down financial system and loses in asset prices. Moreover, increasing noise content in the asset price information may create distortions in the Kalman filter estimation since it becomes harder to extract common component from noisier information. Therefore, including asset prices into GDP nowcast create undershooted predictions for US GDP as expected and including more policy variables may increase the nowcast accuracy of larger models.

4.2 Gains from Asset Prices and Other Non-Real Variables

Considering above uncertainty information, estimations of this work suggest that there is a strong predictive power of asset prices and other non-real variables in US

GDP nowcast for the chosen period. Adding such variables increases the model performance significantly from the squared forecast error perspective. This seems to be a contradicting result with Auroba, Diebold and Scotti (2009) estimations; since, they conclude that adding real variables into estimation increases the model performance in the nowcast of the current state of the economy but adding an asset price does not alter the results significantly.

However, there can be several different reasons behind this result such as methodology differences, estimation period or different variable choices, no need to mention difference between nowcasted variable. As it is stated in the introduction, Auroba et al. (2010) nowcast current state of the economy index and they compare their findings with the NBER defined recession dates. Then, by adding variables, they try to identify possible improvements in the models' ability in catching NBER defined recessions. When they conclude there is no significant nowcast improvement in adding asset price into system, they basically conclude that the graphs are so close that it is hard to distinguish one from another and this graphical interpretation are similar to this works conclusions. However, comparing squared forecast error creates a different understanding and shows that adding asset prices builds significant improvement. As a result, predictions in this paper do not conflict with Auroba et. all (2010) but the lack of comparable variable for current state of economy variable drives such a difference.

4.3 Robustness

The initial GDP estimation in this work and in this literature most of the time depending on several exogenous decisions on the number of state variables to be estimated, AR process these state variables follow and the observation period for the chosen variables. Since most of literature focuses on the doability of GDP nowcast through a Kalman Filter, no attempt to compare different specifications is observed till now.

Beginning with different number of state and different AR process, I also conduct estimations for different sample periods. Two of these estimations can be found in appendix in Table 4 and Table 5 for three model that I estimated previously. For simplicity only squared forecast errors of all specifications are included in the tables. In the first estimation, the number of state variables is decreased to one and the state variable is assumed to follow an AR(3) process rather than four state variables and AR(2) process. From the squared forecast errors, such a specification is found more compatible with US economy regardless of the exact date or the period we refer. On the other hand, this is naturally not the only specification one can impose under this framework. Both AR structure and the number of state can be changed independently or together to have more reliable nowcasts and this result of having better nowcasts with one AR(3) variable is only suggestive rather than being decisive.

In the second estimation, number of factors and AR process that state variable follows are leaved as in the original estimation; however, the information content used in the estimations begin with the fourth month of 1994. There is no theoretical or historical reason behind this choice, it is chosen as the date where whole data set used in all three specification becomes complete, i.e. no missing observations are included in the sample after this date. Again, Table 5 refers, there is an improvement in the nowcast power of the all models but this time the source of improvement is more complicated. The first reason may lie in the possibility that system is not quite compatible with missing observation problem as it is supposed to be. Therefore, filling in missing observations and reestimating the system in EM algorithm and Kalman smoother creates larger distortions than expected. The second possibility is related with possible structural breaks and their effects on the Kalman filter estimation. Since Kalman is relying on smoothing available information, incorporating before break and after break values into one component may decrease the nowcast ability as expected.

As a result, both different AR and state vector specification and smaller sample selection beats the original estimations in GDP nowcasts. As it is stated, these results related with possible different specifications are suggestive rather than decisive so that there may be room for improvement in the nowcast performance of those models without changing any model structure.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this study, US GDP nowcast is constructed for the period 1999 Quarter 1 to 2011 Quarter 1 under EM algorithm in a Kalman smoother. Three different size data set are employed in this construction, where small data set basically includes real and price variables, medium data set adds several asset price, commodity price and banking variables to the small one and finally large model brings the money and more disaggregated information into estimation.

As a result of the estimations, GDP nowcast under Kalman smoother is found more reliable compare to basic AR process and medium model predictions for US GDP beats the large and small models'. However, there is a power reversal among models after 2008 crisis and small model yields lower square forecast error compare to other two. Therefore, we conclude that asset prices and other high frequency non-real variables contribute to the GDP nowcast significantly but this significance disappears after 2008 no matter what the model specification on data set or AR process.

This newly growing literature is also very promising for future works. First, changing modeling decisions on sample data, AR process of state variables or number of states may yield better nowcast results for GDP data as some examples have already provided. Then, constructing a system in daily frequency and allowing missing observations rather than approximating low frequency variables is the first step because such a construction will create the possibility of learning our best guess of quarter GDP with the latest available information without averaging the information content out.

Then, including different countries' information in the US' GDP and inflation nowcasting can be an interesting exercise mainly EU and China data where available. Especially when the information comes with a fewer lags compare to US in one of the other countries, controlling its contribution to US GDP nowcast may create substantial benefit.

Finally, another improvement of this study can be found in disaggregating low frequency data into higher frequency. Backdating methodology which is employed for GDP nowcasting in the scope of this paper can be employed to disaggeragate low frequency data.

BIBLIOGRAPHY

Aruoba, S. B., Diebold, F. X., Scotti, C. (2009), "Real Time Measurement of Business Conditions", Journal of Business & Economic Statistics, 27, 417- 427.

Aruoba, S. B., Diebold, F. X. (2010), "Real-Time Macroeconomic Monitoring: Real Activity, Inflation and Interactions", NBER Working Papers 15657.

Banbura, M., Giannone, D., Reichlin, L.(2010a), "Nowcasting", CEPR Discussion Papers 7883.

Banbura, M. and Modugno, M. (2010b), "Maximum Likelihood Estimation of Factor Models on Data Sets with Arbitrary Pattern of Missing Data", NBER Working Papers 1189.

Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2005), "Do financial variables help forecasting real activity and inflation in the euro area?", Journal of Monetary Economics, 50, 1243-1255.

Hamilton, J.D. (1994), "State-Space Models", Handbook of Econometrics, 4, 3041 - 3068.

Stock, J.H., and Watson, M. W. (2000), "Forecasting Output and Inflation: The Role of Asset Prices", Journal of Economic Literature, 41(3), 788-829.

Stock, J. H., and Watson, M. W. (1989), "New Indexes of Coincident and Leading Economic Indicators", NBER Macroeconomics Annual, 4, 351-394

APPENDIX A

Complete Data List in Frequency Order					
Short Name	**Frequency**	**Group**	**Small**	**Medium**	**Large**
Federal Funds Rate	Daily	Asset Price			
Bank Prime Loan	Daily	Asset Price		X	X
Discount Window Primary Credit	Daily	Asset Price			
Discount Window Borrowing	Daily	Asset Price			
1 Year T-Bill Secondary Market	Daily	Asset Price			
6 Months T-Bill Secondary Market	Daily	Asset Price		X	X
4 Week T-Bill Secondary Market	Daily	Asset Price			
30 year nominal treasury	Daily	Asset Price			
20 year nominal treasury	Daily	Asset Price			
10 Year Nominal Treasury	Daily	Asset Price			
7 Year Nominal Treasury	Daily	Asset Price			
5 Year Nominal Treasury	Daily	Asset Price			
3 Year Nominal Treasury	Daily	Asset Price			

2 Year Nominal Treasury	Daily	Asset Price			
1 Year Nominal Treasury	Daily	Asset Price		X	X
6 Months Nominal Treasury	Daily	Asset Price			
3 Months Nominal Treasury	Daily	Asset Price			
SP500	Daily	Asset Price		X	X
US Regular All Formulations Gas Price	Weekly	Commodity			X
US Diesel Sales Price	Weekly	Commodity			X
Initial Jobless Claims	Weekly	Real	X	X	X
M1 Money Stock	Weekly	Money			
Currency Componenet of M1 billions of dollar	Weekly	Money			
M2 Money Stock billions of dollar	Weekly	Money			
M1 Money Multiplier	Biweekly	Money			
St. Louis Adjusted Monetary Base	Biweekly	Money			
Corporate bonds/ Moody's seasoned Aaa	Monthly	Asset Price			
Corporate bonds/ Moody's seasoned Baa	Monthly	Asset Price		X	X
Capacity Utilization	Monthly	Real	X	X	X
Manufacturing (SIC)	Monthly	Real			
Manufacturing (NAICS)	Monthly	Real			
Durable Manufacturing (NAICS	Monthly	Real			
Wood Product (NAICS)	Monthly	Real			
Non-Metalic Mineral Product	Monthly	Real			
Primary Metal	Monthly	Real			
Seasonally Adjusted Industrial Production Index (2007=100)	Monthly	Real	X	X	X
SA Real Retail and Food Services Sales (million dollars)	Monthly	Real			
Spot Oil Price: West Texas Intermediate (dollar per barrel)	Monthly	Commodity		X	X
Manufacturer's Shipments, Inventories, and Orders	Monthly	Real			
University of Michigan Inflation Expectation	Monthly	Price		X	X

Housing Starts: Total: New Privately Owned Housing Units Started	Monthly	Real			X
New Private Housing Units Authorized by Building Permits (thousands	Monthly	Real			
Consumer Price Index for All Urban Consumers: All Items (82-84=100	Monthly	Price	X	X	X
Consumer Price Index for All Urban Consumers: Energy (82-84 =100)	Monthly	Price			
Consumer Price Index for All Urban Consumers: Housing (82-82 =100	Monthly	Price			
Consumer Price Index for All Urban Consumers: Food 82-84=100	Monthly	Price			
Job Openings: Total Nonfarm in thousands	Monthly	Real			
Total Nonfarm Private Payroll Employment in thousands	Monthly	Real	X	X	X
Civilian Unemployment Rate	Monthly	Real			
Average (Mean) Duration of Unemployment	Monthly	Real			
Total Nonfarm Payrolls: All Employees in thousands	Monthly	Real			
Average Hourly Earnings: Total Private Industries	Monthly	Real			
Average Weekly Hours of Production and Nonsupervisory Employees	Monthly	Real			
Total Checkable Deposits	Monthly	Banking			X
Currency Component of M1 billions of dollars	Monthly	Money		X	X
M2 Money Stock billions of dollars	Monthly	Money			X
Savings Deposits - Total billions of dollars	Monthly	Banking		X	X
Producer Price Index: All Commodities 82=100	Monthly	Price			X
Producer Price Index: Finished Energy Goods	Monthly	Price			
Seasonally Adjusted GDP billion dollars	Quarterly	Real	X	X	X
Personal consumption expenditures	Quarterly	Real			
Goods	Quarterly	Real			
Durable goods	Quarterly	Real			
Nondurable goods	Quarterly	Real			
Services	Quarterly	Real			
Gross private domestic investment	Quarterly	Real			
Fixed investment	Quarterly	Real			

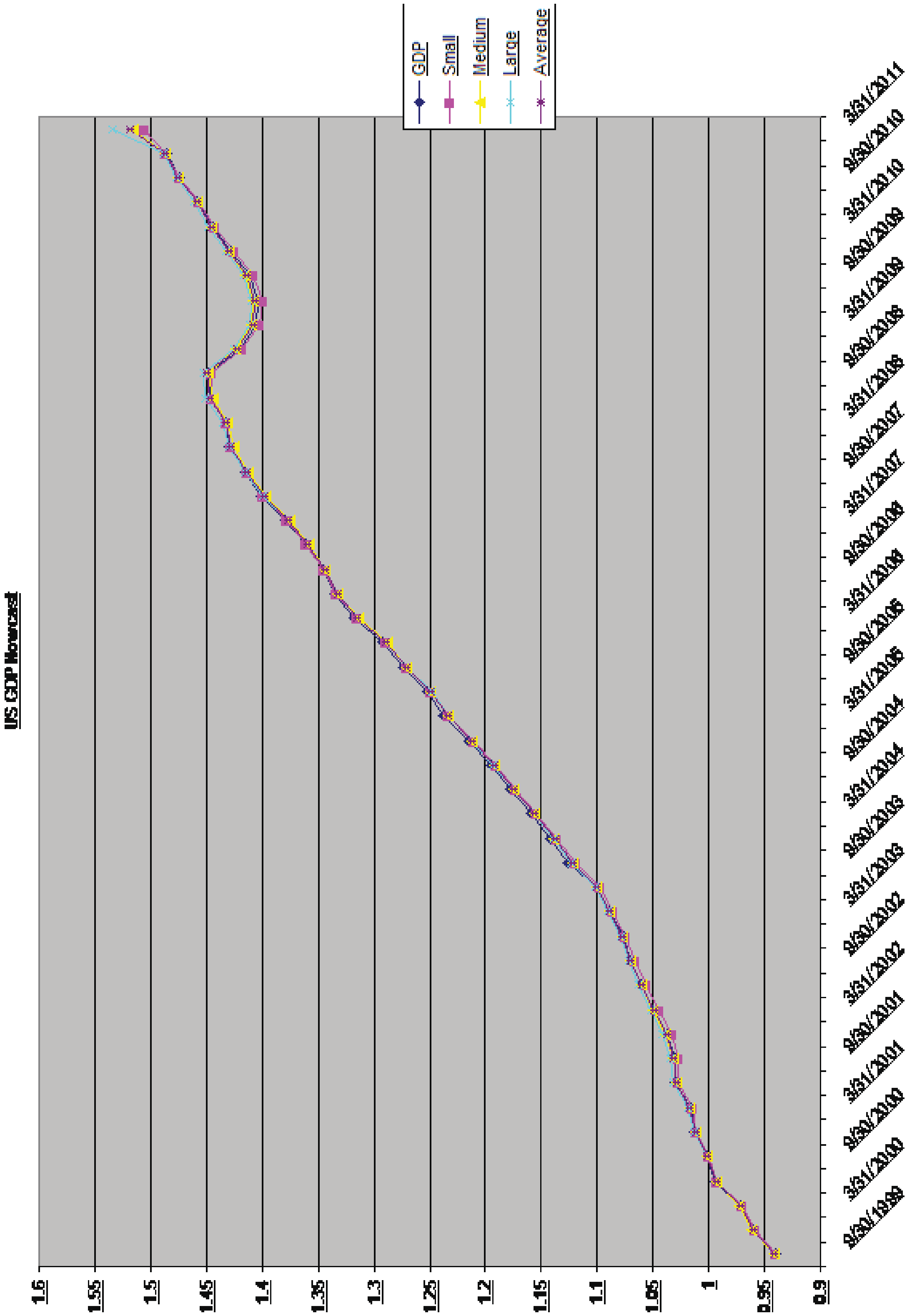
Nonresidential	Quarterly	Real			
Structures	Quarterly	Real			
Equipment and software	Quarterly	Real			
Residential	Quarterly	Real			
Change in private inventories	Quarterly	Real			
Net exports of goods and services	Quarterly	Real			
Exports	Quarterly	Real			
Goods	Quarterly	Real			
Services	Quarterly	Real			
Imports	Quarterly	Real			
Goods	Quarterly	Real			
Services	Quarterly	Real			
Government consumption expenditures and gross investment	Quarterly	Real			
Federal	Quarterly	Real			
National defense	Quarterly	Real			
Nondefense	Quarterly	Real			
State and local	Quarterly	Real			
Federal Government Debt: Total Public Debt (million dollars)	Quarterly	Real			
House Price Index (Not a Press Release) (index: 1980 Q1=100)	Quarterly	Price			
Household Debt Service Payments as Percent of Disposable Persona	Quarterly	Banking			
Nonperforming Total Loans (ratio to total loans)	Quarterly	Banking			
Net Interest Margin for all U.S. Banks	Quarterly	Banking			X
Velocity of M2 Money Stock	Quarterly	Money			
Velocity of M1 Money Stock	Quarterly	Money			
Total Net Loan Charge-offs	Quarterly	Banking		X	X

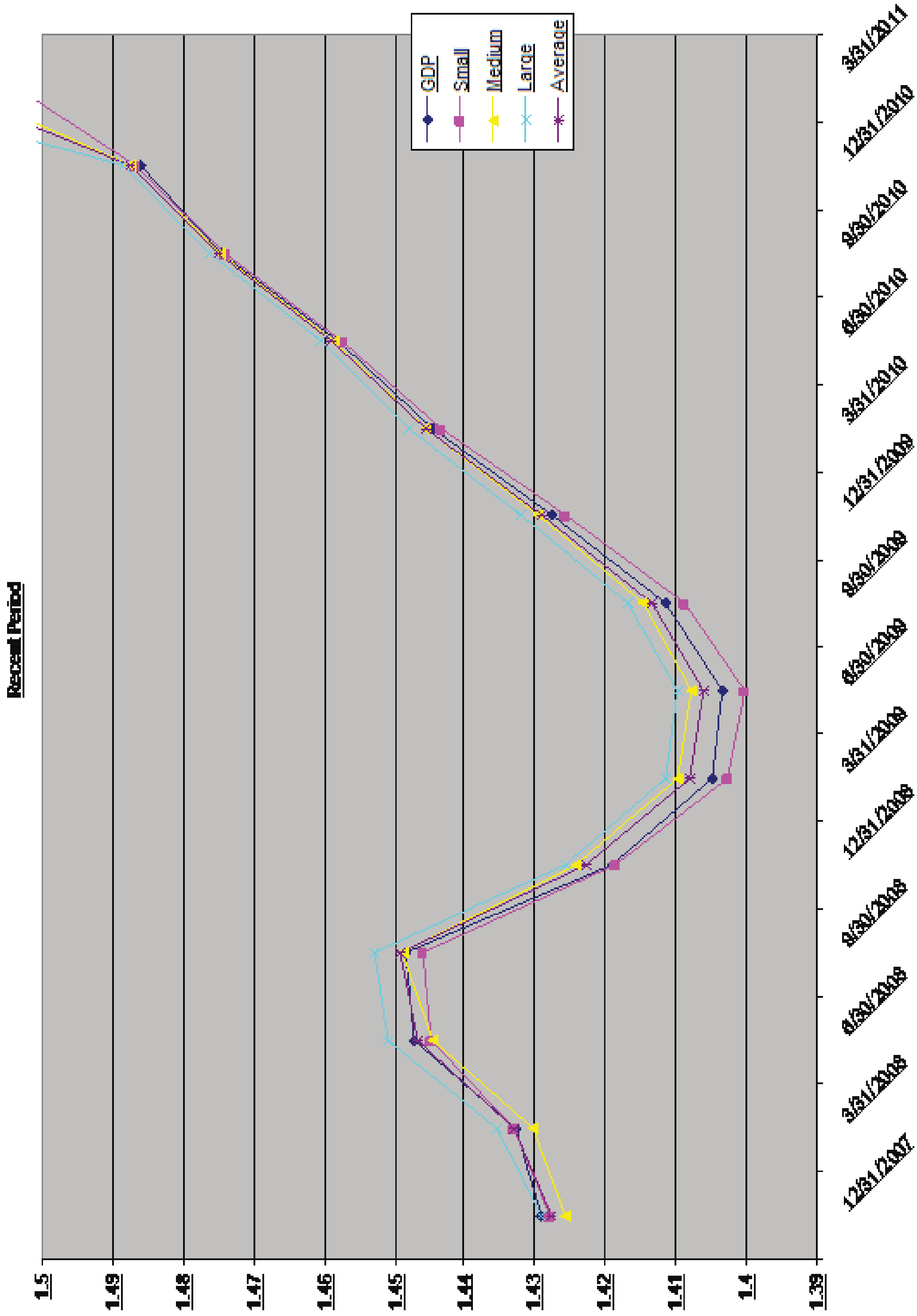
APPENDIX B

Predictions in Billion Dollars

		Large		Medium		Small
9/30/1999		9412		9419		9419
12/31/1999		9597		9609		9609
3/31/2000		9721		9727		9727
6/30/2000		9925		9932		9932
9/30/2000		10012		10012		10012
12/31/2000		10128		10119		10119
3/31/2001		10190		10169		10169
6/30/2001		10318		10295		10295
9/30/2001		10347		10320		10320
12/31/2001		10411		10386		10386
3/31/2002		10523		10502		10502
6/30/2002		10625		10609		10609
9/30/2002		10722		10707		10707
12/31/2002		10795		10777		10777
3/31/2003		10911		10890		10890
6/30/2003		11022		11001		11001
9/30/2003		11238		11222		11222
12/31/2003		11392		11384		11384
3/31/2004		11564		11562		11562
6/30/2004		11750		11746		11746
9/30/2004		11929		11925		11925
12/31/2004		12121		12119		12119
3/31/2005		12341		12342		12342
6/30/2005		12485		12486		12486
9/30/2005		12709		12705		12705
12/31/2005		12896		12886		12886
3/31/2006		13151		13143		13143
6/30/2006		13327		13317		13317
9/30/2006		13444		13431		13431
12/31/2006		13593		13579		13579
3/31/2007		13772		13753		13753
6/30/2007		13992		13967		13967
9/30/2007		14152		14123		14123
12/31/2007		14292		14256		14256
3/31/2008		14354		14304		14304
6/30/2008		14509		14446		14446

9/30/2008		14529		14488		14488
12/31/2008		14256		14242		14242
3/31/2009		14116		14099		14099
6/30/2009		14099		14079		14079
9/30/2009		14169		14148		14148
12/31/2009		14320		14296		14296
3/31/2010		14479		14455		14455
6/30/2010		14605		14588		14588
9/30/2010		14764		14748		14748
12/31/2010		14884		14875		14875
3/31/2011		15344		15154		15154


US GDP Nowcast
GDP
Small
Medium
Large
Average
1.6
1.55
1.5
1.45
1.4
1.35
1.3
1.25
1.2
1.15
1.1
1.05
1
0.95
0.9
9/30/1999
3/31/2000
9/30/2000
3/31/2001
9/30/2001
3/31/2002
9/30/2002
3/31/2003
9/30/2003
3/31/2004
9/30/2004
3/31/2005
9/30/2005
3/31/2006
9/30/2006
3/31/2007
9/30/2007
3/31/2008
9/30/2008
3/31/2009
9/30/2009
3/31/2010
9/30/2010
3/31/2011


Recent Period
GDP
Small
Medium
Large
Average
1.5
1.49
1.48
1.47
1.46
1.45
1.44
1.43
1.42
1.41
1.4
1.39
12/31/2007
3/31/2008
6/30/2008
9/30/2008
12/31/2008
3/31/2009
6/30/2009
9/30/2009
12/31/2009
3/31/2010
6/30/2010
9/30/2010
12/31/2010
3/31/2011

Squared Forecast Erros

		Large		Medium		Small
9/30/1999		49		196		100
12/31/1999		144		1		289
3/31/2000		36		324		36
6/30/2000		49		289		361
9/30/2000		0		25		121
12/31/2000		81		121		225
3/31/2001		441		16		196
6/30/2001		529		36		841
9/30/2001		729		225		841
12/31/2001		625		169		1225
3/31/2002		441		9		2601
6/30/2002		256		49		1521
9/30/2002		225		25		1225
12/31/2002		324		100		529
3/31/2003		441		4		1024
6/30/2003		441		49		1521
9/30/2003		256		1156		4489
12/31/2003		64		1089		3600
3/31/2004		4		1225		3136
6/30/2004		16		1024		2401
9/30/2004		16		625		1681
12/31/2004		4		676		1369
3/31/2005		1		1444		1521
6/30/2005		1		961		361
9/30/2005		16		1369		961
12/31/2005		100		900		729
3/31/2006		64		1600		1089
6/30/2006		100		961		289
9/30/2006		169		484		9
12/31/2006		196		1089		81
3/31/2007		361		1296		81
6/30/2007		625		1681		225
9/30/2007		841		1225		169
12/31/2007		1296		1225		144
3/31/2008		2500		576		9
6/30/2008		3969		676		625
9/30/2008		1681		9		676
12/31/2008		196		2601		25
3/31/2009		289		2401		484
6/30/2009		400		1936		1024
9/30/2009		441		1089		729
12/31/2009		576		361		400
3/31/2010		576		81		196
6/30/2010		289		81		16
9/30/2010		256		9		16
12/31/2010		81		196		49
SFE	**Full**	**439.0217**		**688.7826**		**853.0435**
	Pre-08	262.9706		637.2941		1029.147
	After-08	795.8182		858.1818		385.4545

SFE Comparison for AR(3) Process

	Large Model	Large Model AR(1)	Medium Model	Medium Model AR(1)	Small Model	Small Model AR(1)
9/30/1999	49	0.9614	196	1.2132	100	6.6064
12/31/1999	144	2.389	1	2.7086	289	23.1461
3/31/2000	36	3.101	324	3.6531	36	1.1045
6/30/2000	49	9.4362	289	6.3921	361	38.6686
9/30/2000	0	6.3908	25	1.8493	121	3.4393
12/31/2000	81	0.0002	121	3.4575	225	5.9046
3/31/2001	441	51.1076	16	21.1799	196	0.1906
6/30/2001	529	25.4089	36	5.4787	841	42.8659
9/30/2001	729	50.798	225	22.361	841	14.6676
12/31/2001	625	44.6705	169	5.7618	1225	36.3418
3/31/2002	441	21.4387	9	2.5136	2601	47.7564
6/30/2002	256	18.6131	49	1.8426	1521	23.0543
9/30/2002	225	15.0327	25	0.9191	1225	9.534
12/31/2002	324	22.3671	100	1.6837	529	1.2074
3/31/2003	441	12.5371	4	0.6299	1024	9.3995
6/30/2003	441	8.2797	49	0.2044	1521	20.3927
9/30/2003	256	0.9074	1156	0.7397	4489	100.9753
12/31/2003	64	1.9828	1089	1.1493	3600	46.0882
3/31/2004	4	8.0227	1225	3.8858	3136	31.5192
6/30/2004	16	10.8972	1024	5.6063	2401	21.5476
9/30/2004	16	10.2537	625	5.425	1681	8.4336
12/31/2004	4	19.6352	676	10.0586	1369	9.8459
3/31/2005	1	40.232	1444	19.5441	1521	18.0214
6/30/2005	1	33.6816	961	18.8555	361	2.4562
9/30/2005	16	41.4933	1369	25.3284	961	3.7897
12/31/2005	100	39.9688	900	24.9842	729	0.909
3/31/2006	64	64.025	1600	38.2821	1089	10.1835
6/30/2006	100	57.5653	961	31.0783	289	1.6691
9/30/2006	169	58.1687	484	28.9303	9	0.9943
12/31/2006	196	60.9074	1089	29.772	81	3.2219
3/31/2007	361	51.4664	1296	29.3571	81	3.4731
6/30/2007	625	76.8556	1681	38.9983	225	7.6054
9/30/2007	841	34.9417	1225	27.9665	169	2.2446
12/31/2007	1296	4.8424	1225	17.5519	144	0.4978
3/31/2008	2500	18.9039	576	5.5609	9	2.2001
6/30/2008	3969	7.6995	676	15.3142	625	0.7188
9/30/2008	1681	48.6003	9	1.8815	676	11.5057
12/31/2008	196	376.19	2601	56.7538	25	15.6337
3/31/2009	289	247.0046	2401	74.3689	484	134.2274
6/30/2009	400	124.2067	1936	52.6955	1024	131.6527
9/30/2009	441	74.815	1089	40.332	729	67.5606
12/31/2009	576	43.1259	361	24.3321	400	45.7149
3/31/2010	576	26.6537	81	15.6661	196	39.8371
6/30/2010	289	20.1593	81	13.0262	16	23.5166
9/30/2010	256	14.8386	9	10.7606	16	27.7789
12/31/2010	81	17.1527	196	13.0015	49	3.877
Mean	**439.0217**	**41.90716**	**688.7826**	**16.58816**	**853.0435**	**23.0865**

SFE Comparison for 1994 Vintage

	Large Model	Large Model 1994	Medium Model	Medium Model 1994	Small Model	Small Model 1994
9/30/1999	49	1.0412	196	0.9153	100	0.0573
12/31/1999	144	4.6296	1	5.081	289	2.9963
3/31/2000	36	1.1402	324	1.5545	36	5.2573
6/30/2000	49	9.5924	289	12.823	361	31.2465
9/30/2000	0	0.5729	25	1.1545	121	2.129
12/31/2000	81	6.3151	121	7.6127	225	17.6165
3/31/2001	441	5.5656	16	4.9725	196	0.0052
6/30/2001	529	8.5774	36	11.0824	841	3.5067
9/30/2001	729	3.5785	225	0.7613	841	1.951
12/31/2001	625	8.2989	169	9.3062	1225	4.6078
3/31/2002	441	41.4361	9	46.346	2601	4.3735
6/30/2002	256	8.0334	49	6.682	1521	2.2675
9/30/2002	225	2.0456	25	12.5875	1225	26.556
12/31/2002	324	29.0709	100	0.655	529	16.5831
3/31/2003	441	5.3033	4	38.033	1024	13.988
6/30/2003	441	50.406	49	56.9222	1521	117.9837
9/30/2003	256	150.4709	1156	240.7717	4489	240.9853
12/31/2003	64	119.9627	1089	134.0221	3600	26.5732
3/31/2004	4	105.438	1225	111.1424	3136	7.3533
6/30/2004	16	47.8771	1024	44.7439	2401	10.0604
9/30/2004	16	22.0121	625	6.5118	1681	10.3059
12/31/2004	4	36.1014	676	41.775	1369	26.9857
3/31/2005	1	154.2534	1444	168.8882	1521	149.4721
6/30/2005	1	73.4033	961	62.1187	361	0.5241
9/30/2005	16	134.2989	1369	150.163	961	31.5295
12/31/2005	100	44.6108	900	24.3144	729	7.1288
3/31/2006	64	225.805	1600	175.6948	1089	3.6976
6/30/2006	100	140.0232	961	86.1399	289	7.7388
9/30/2006	169	103.4789	484	38.8883	9	7.7722
12/31/2006	196	270.875	1089	178.3953	81	93.3392
3/31/2007	361	183.5125	1296	186.8468	81	59.637
6/30/2007	625	187.9827	1681	251.2374	225	23.5184
9/30/2007	841	69.0955	1225	177.811	169	0.809
12/31/2007	1296	6.312	1225	122.1562	144	5.1022
3/31/2008	2500	169.4501	576	11.3995	9	267.23
6/30/2008	3969	40.5717	676	1.0486	625	93.7469
9/30/2008	1681	0.0163	9	0.1397	676	6.3222
12/31/2008	196	215.0391	2601	3.9508	25	142.3127
3/31/2009	289	514.7612	2401	42.6666	484	585.64
6/30/2009	400	94.6817	1936	28.0264	1024	145.2616
9/30/2009	441	8.3715	1089	42.9852	729	213.2888
12/31/2009	576	2.3839	361	25.9133	400	77.7892
3/31/2010	576	0.0928	81	12.8408	196	21.2586
6/30/2010	289	8.7203	81	21.7257	16	6.8097
9/30/2010	256	0.0463	9	0.1082	16	83.7512
12/31/2010	81	65.6886	196	41.8547	49	7.458
Mean	**439.0217**	**73.49878**	**688.7826**	**57.62542**	**853.0435**	**56.83754**