THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, DEPARTMENT OF ECONOMICS

A First Step to Develop a Forecasting Model for the U.S. Government Budget Deficit:

A Regression-Based Method with Quarterly Economic Indicators

Nazire Ozkan

September 2011

Abstract

This study is an initial attempt to develop a forecasting model to predict the quarterly U.S. Government Budget Deficit. Forecast combinations of Augmented Distributed Lag (ADL) Regression Models, each of which is constructed with single quarterly economic indicator, are employed to forecast government expenditures and receipts separately to obtain the deficit forecasts. Findings show that forecast combination of ADL regression models for the deficits provides forecast gains over the traditional autoregressive regression model.

A First Step to Develop a Forecasting Model for the U.S. Government Budget Deficit:

A Regression-Based Method with Quarterly Economic Indicators

Nazire Ozkan

1 Introduction

The United States has been experiencing profound budgetary challenges; the uncertainty of government budget deficit forecasts has become an important public issue. As stated by the Congressional Budget Office (CBO), a nonpartisan government agency whose main goal is to provide accurate forecasts of the budget deficit, the projected deficit of $1.3 trillion for 2011 will be the third largest shortfall in the past 65 years[1]. To deal with current budgetary and economic challenges, the U.S. Congress enacted the Budget Control Act of 2011 on August 2, 2011; and mainly because of this act, the total deficit for the following 10-year period is projected to be $1.45 trillion lower than what was previously projected[2].

Why is the accuracy of government budget forecasts so important for fiscal planning? How do budget forecasts fail? It is crucial for this study to search for the answers to these questions in order to construct a regression model to be able to forecast government budget deficits with high predictive ability, since this paper is an attempt to develop such a model[3].

Government Budget forecasts are vital for fiscal plans and they reflect information about overall macroeconomic well-being. U.S. government agencies such as the Office of Management and Budget (OMB) and the Congressional Budget Office (CBO) provide fiscal forecasts to inform policy makers and researchers. The Congressional Budget Office (CBO) forecasts federal income so as to report the federal budget and well-being of the overall economy to the Congress as a baseline to measure the effects of proposed changes in spending and tax laws. To do so under current law, CBO makes projections of federal receipts and outlays. Since, federal receipts are determined largely by taxes collected on individual and business income, projections of them are the key components of the CBO`s projections of the federal budget; hence, their budget

---

[1] Congressional Budget Office; The Budget and Economic Outlook: An Update (August 2011)
[2] Office of Management and Budget; Fiscal Year 2012 Mid-Session Review (September 2011)
[3] This paper is inspired by Andreou E., Ghysels E., and Kourtellos A. (2010a), “Should Macroeconomic Forecasters Use Daily Financial Data and How?” (November 19, 2010), which introduces regression-based methods for predicting the quarterly U.S. real Gross Domestic Product that use daily financial data.

projections rely on data compiled on individual and business income that are encompassed within the national income and product accounts (NIPAs)[4].

Budget forecast errors may result in badly enacted fiscal policies; for instance, an over-prediction of receipts may lead fiscal policy makers to increase tax cuts for the next fiscal year, and in turn revenues would decrease; which was indeed experienced in 2001. According to Congressional Budget Office (CBO), one of the greatest sources of error in budget projections is in forecasting taxable income which have wages and salaries as its most important and the book profits of corporations as the second important component for revenue projections. CBO reports that during the post World War II period the nontaxable part of labor income rose as a share of total labor compensation since employers and employees preferred untaxed noncash benefits for taxable wages and salaries. But that trend reversed between 1994 and 1999 because of changes in the way health care is provided and the rise in the stock market causing reductions in contributions to benefit pension plans; however, that trend has reversed again after 2001, that is, nontaxable income has risen sharply. Furthermore, in 2002-2003 period there was a change in tax rules for the depreciation of capital goods resulting in a decrease in taxable corporate book profits. Hence, 2001 taxable book profit forecasts were too high. Because of the recent variations in the nontaxable part of the labor income, CBO made forecast erros. For instance, in 2003, due to data revisions, CBO made a 1.2 percentage points forecast error in the income share of wages and profits for the 2003-2004 period[5]. CBO states that there were large errors in forecasts for 1979-1983 period when the economy experienced its hardest recessionary period of the postwar era, and relatively small errors for mid-expansion years of 1985-1987; also, the recession of 2001 and the slow recovery in 2002 caused overpredictions in 2000 and 2001; more recently, there are cyclical or structural changes to fiscal policy that have occurred as a result of or in response to the recession of 2008.

All those forecast errors together with economic fluctuations, recessions, result in cyclical and structural changes in government spending and tax laws. These changes affecting the predictive ability of the models are observed in the present study when evaluating the forecast performance of the proposed models.

Miller (1991) states that CBO has two different divisions responsible for forecasting expenditure and receipts separately; then a third division revises these forecasts and sends the revised forecasts back to their divisions for further investigations until reaching internally consistent forecasts. Then, the deficit forecast is largely determined as a residual. With this information and the evidence reported in Table 1 (which

[4] Congressional Budget Office, Background Paper (Aug 2006); How CBO forecasts Income, available at http://www.cbo.gov/ftpdocs/75xx/doc7507/08-25-Income.pdf

[5] CBO's Economic Forecasting Record: An Evaluation of the Economic Forecasts CBO made from January 1976 through January 2004. (November 2006)

indicates that the deficit forecasts obtained by estimating either random walk or autoregressive regression model with the deficit as the dependent variable have a very low forecast accuracy measured by root mean squared forecast errors; whereas the deficit forecasts obtained after forecasting receipts and expenditures separately then subtracting expenditure forecasts from receipt forecasts have higher forecast performance), throughout this study government budget deficit forecasts are obtained by forecasting expenditures and receipts separately for all regression models. Moreover, this approach allows us to make individual forecast investigations on the two components of the deficit which in turn helps evaluate deficit forecasts as a whole.

Blanchard and Perotti (2002) and Fatás and Mihov (2001) assume that, at a quarterly frequency, government spending does not contemporaneously respond to the realization of other macroeconomic variables, i.e. government spending is determined before other economic variables. Therefore, with this reasonable assumption, all models employed in this study do not include the current value of the macroeconomic predictors, instead their lags are included.

It is reasonable to expect that predictors affecting Expenditure forecasts can be different than those affecting forecasts for Receipts; in addition, the same predictor may influence one of them more than the other or affect in a different way. To investigate the individual impacts of some macroeconomic variables on both Expenditures and Receipts, a bunch of macroeconomic indicators are chosen and each of them is used as a predictor for both variables separately. In order to do so, Augmented Distributed Lag (ADL) regression model, which includes lags of the dependent variable, here Expenditures or Receipts, and the lags of the single predictor as the regressors, is employed. After constructing the longest possible samples with available data, individual forecasts with each single indicator are obtained by using rolling forecast scheme, having multiple forecasts of Expenditures and Receipts. As indicated by Timmermann (2006), it is not reasonable to think that the same individual model dominates all others at all points in time, since forecasting models are considered as local approximations. Stock and Watson (2001, 2004) find that combined forecasts generally outperform forecast performances of the best individual model by employing numerous types of models and variables. Therefore, to obtain more accurate[6] forecasts by using evidence from all individual models forecast combinations method is considered rather than using the best single model.

[6] Timmermann (2006) suggests forecast combinations across different models to have more robust forecasts against misspecification biases and measurement errors in the data underlying the individual forecasts.

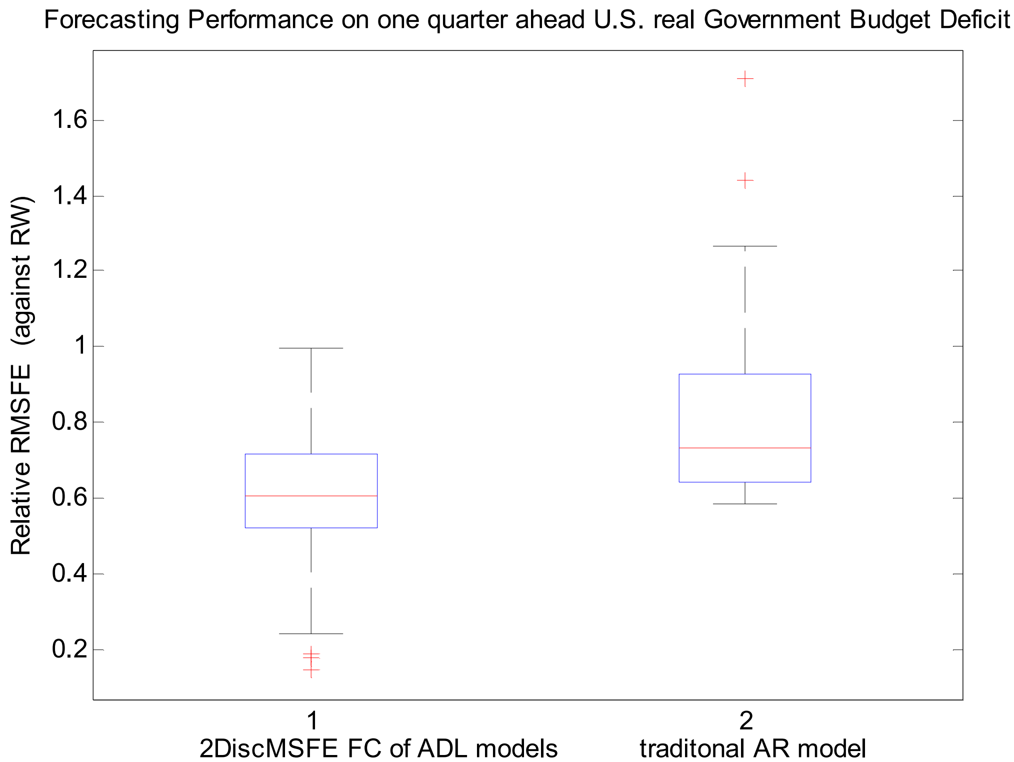
Ramey (2008) argues that defense spending is an important source of the fluctuations in total government expenditures relative to trend and shows that shocks to defense spending accounts for almost all of the unforeseen changes in total government spending by using variance decomposition analysis. Ramey and Shapiro (1998) using a narrative approach build a military dummy variable to identify shocks to government spending. The military dummy variable of Ramey and Shapiro (1998) takes a value of 1 in the following quarters: 1950:Q3, 1965:Q1, 1980:Q1, and 2011:Q3 corresponding with the start of the Korean War, the Vietnam war, the Carter-Reagan buildup, and September 11th, respectively. This dummy is included in all ADL model regressions for U.S. real Government Expenditures, and in the competing models of random walk and autoregressive regressions. Also, Ramey (2008) constructs a military spending variable based on narrative evidence, which is much richer than the dummy variable of Ramey-Shapiro (1998) war dates. For the military dates identified, Ramey (2008)`s military spending variable takes on the present discounted value of the change in anticipated government spending. Moreover, a recent study by Owyang and Zubairy (2010) identifies government spending shocks in vector autoregressions (VARs) as innovations to total government spending; one of these shocks is military spending shocks as defined by Ramey (2008). In light of these findings, this military spending variable defined by Ramey (2008) is also used as one of the predictors for U.S. real Government Expenditures.

Figure 1 provides a concise preview of the forecasting gains of one-step ahead quarterly U.S. Government Budget Deficit by displaying two boxplots; one for the proposed forecasting model in this study, namely forecast combinations of augmented distributed lag models and the other for the competitor model, namely the traditional autoregressive (AR) model. These boxplots present predictive abilities of the two competing models relative to a benchmark model, here taken to be the random walk (RW); forecast performances are measured in terms of Root Mean Square Errors (RMSFEs). Each point in boxplots is attached to each out-of-sample rolling window.

Since smaller RMSFEs reflect better forecast performance, Figure 1 indicates that forecast combination of ADL models outperforms the traditional AR model. It is also evident that combination of ADL models provides better deficit forecasts compared to the bencmark RW model, since relative RMSFEs are smaller than unity throughout the evaluation period.

To sum up, forecast combination of ADL models each of which is constructed by using a single macroeconomic indicator provides forecast gains for U.S. Government Budget Deficit over both the benchmark of random walk and the traditional AR model.

Figure 1: Boxplots Displaying Forecasting Performance on U.S real Government Expenditures for h=1 .



This paper proceeds as follows. In section 2 Augmented Distributed Lag Regression Model and the two competing models are described. Section 3 introduces the data. Forecast combinations method is presented in section 4 while section 5 describes the forecasting tests employed. Empirical results are presented in section 6 with the evidence in favor of the forecast combinations ADL regression models against AR and RW regression models. Section 7 concludes.

2 Augmented Distributed Lag (ADL) Regression Model

Let the quarterly variable of interest to be forecasted for h-step ahead horizon be Y_{t+h}, for instance U.S. real Government Receipts, and a quarterly aggregate of a predictor series, such as quarterly aggregate of monthly U.S. Total Non-Farm Employment be X_t. Then, Augmented Distributed Lag, $\mathrm{ADL}(p_Y, q_x)$, regression model can simply be defined as

$$Y_{t+h} = c + \sum_{j=0}^{p_Y-1} \propto_{j+1} Y_{t-j} + \sum_{k=0}^{q_X-1} \beta_{k+1} X_{t-k} + u_{t+h}$$

which involves p_Y lags of Y_t and q_X lags of X_t, and requires $p_Y + q_X + 1$ regression parameters to be estimated.

Two other models employed as forecasting competitors are given as follows.
Random Walk (RW) Model, employed as the benchmark forecasting model:

$$Y_{t+h} = c + Y_t + u_{t+h}$$

Autoregressive, AR(p_Y), Regression Model, employed as the traditional forecasting model:

$$Y_{t+h} = c + \sum_{j=0}^{p_Y-1} \propto_{j+1} Y_{t-j} + u_{t+h}$$

3 Data

Two sample periods of quarterly U.S. real Government Expenditures and quarterly U.S. real Government Receipts are studied in this study. A longer sample period is from the first quarter of 1950 to the second quarter of 2011 while a shorter sample starts at the first quarter of 1960 and ends at the second quarter of 2011. The shorter sample is chosen so as to include some new predictors that are not available in the longer one. Samples are chosen as longer as possible depending on the availability of data in order to investigate historical evaluation of the variables of interest by having long enough observations to use as in-sample and out-of-sample periods. The data series used in this study are listed in Table 4 with their sources. All data are seasonally adjusted; real quantities are obtained by deflating nominal quantities by GDP deflator; and they are transformed to induce stationarity and to insure that the transformed variables correspond to quarterly series observed at the end of the quarter. These transformations are described in Table 5 and corresponding transformations for each variable are reported in column 4 of Table 4.

4 Forecast Combinations

As pointed out by Timmermann (2006), forecast combinations have been viewed as a simple and effective way to obtain more accurate forecasts by using evidence from all models considered rather than using the best single model. Stock and Watson (2001, 2004) find that combined forecasts generally outperform forecast performances of the best individual model by employing numerous types of models and variables. Forecast combinations have been used in many areas such as output growth (Stock and Watson (2004), Andreou, Ghysels, and Kourtellos (2010a)), inflation (Stock and Watson (2008b)), exchange rates (Wright (2008)). Timmermann (2006) lists key reasons for using forecast combinations; forecast combinations provide hedging against model uncertainty, since forecasters may have different information sets, different predictors and modeling structures; forecast combinations can work well under structural breaks while individual forecasts may be very differently affected by them; combining forecasts across different models results in robust forecasts against misspecification biases and measurement errors in the data set underlying the individual forecasts.

Given N individual forecasting models, forecast combinations are time-varying weighted averages of the individual forecasts,

$$\hat{f}_{N,t+h|t} = \sum_{j=1}^{N} \widehat{\omega}_{j,t} \hat{y}_{j,t+h|t}$$

where the weights $\widehat{\omega}_{j,t}$ on the j^{th} forecast in period t depends on the historical performance of the individual forecasts.

In this study, Squared Discounted Mean Square Forecast Errors (henceforth denoted by 2DiscMSFE) and Discounted Mean Square Forecast Errors (henceforth DMSFE) forecast combination methods are utilized (See Stock and Watson (2004) and (2008b)). Each individual predictor is given a weight according to its historical performance, the weight is inversely proportional to the predictor`s 2DiscMSFE or DMSFE (the one providing highest forecast gains is utilized). The discount factor attaches greater weight to the recent predictive ability of the individual predictor. The weights are given as follows.

$$\widehat{\omega}_{j,t} = \frac{\left(\lambda_{j,t}^{-1}\right)^{\kappa}}{\sum_{i=1}^{N}\left(\lambda_{i,t}^{-1}\right)^{\kappa}} \quad , \quad \lambda_{j,t} = \sum_{m=T_0}^{t-h} \delta^{t-h-m} \left(y_{m+h}^{h} - \hat{y}_{j,m+h|m}^{h}\right)^2$$

where $\delta = 0.9$ [7] and κ=1 for DMSFE and 2 for 2DiscMSFE.

[7] Other discount factors of δ=1 and 0.95 are also calculated, but 0.9 is found to be the best giving the highest forecast gains.

In this study, forecast combinations method is employed in three steps:

(1) Forecasts[8] are computed for U.S. real Government Expenditures, U.S. real Government Receipts with single predictors, i.e. forecasts of Expenditures and Receipts are obtained by estimating ADL regression models with single predictors so that for Expenditures and Receipts we have fifteen[9] set of forecasts for each.

(2) For both variables, Expenditures and Receipts, best predictors are picked according to their out-of-sample performance measured by their root mean squared forecast errors (RMSFEs) (See section 6 for details). Figures 2 and 3 display boxplots representing the forecast performance of each predictor, relative to the performance of random walk (RW) model, on Expenditures and Receipts, respectively. Predictors with smaller RMSFEs vis-à-vis RMSFEs for the traditional autoregressive (AR) model are chosen to be able to obtain combined forecasts that are better than the competitor AR model[10] (See Figures 2 and 3, predictors with smaller median and/or lower first- and fourth-quartiles and/or shorter upper whiskers are picked).

(3) Forecasts are combined with the selected predictors for both variables and for both samples. For Expenditures, best combinations, which deliver the highest forecast gains, are obtained by using the weights[11] with $\delta = 0.9$ and $\kappa=1$; whereas for Receipts, forecast combinations that are best are obtained with $\delta = 0.9$ and $\kappa=2$. Hence, for Receipts, forecast combination of ADL models is denoted by 2DiscMSFE and for Expenditures, it is represented by DMSFE; so the forecasting model for the Deficits is denoted by 2DiscMSFE/DMSFE forecast combinations of ADL regression models.

[8] Forecasts are obtained by employing rolling windows forecast method explained in section 6.
[9] These fifteen individual predictors are given in Table 6.
[10] Predictors chosen are displayed in Table 6.
[11] Those forecast combination weights for both Expenditures and Receipts are displayed as snapshots for particular time periods in order to see which predictors have greater weights during that time period. See Figures 5 and 6.

5 Forecasting Tests

Rolling forecasting scheme is employed in this paper. Let T be the total sample size, h be the forecast horizon; and R denotes the size of the estimation window and P= T- R - h + 1 is the out-of-sample size. Consider a sequence of h-step ahead, time-t rolling window out-of sample forecast $\hat{f}_t(\widehat{\beta_t})$ which corresponds to in sample fitted values $\hat{y}_j(\hat{\beta}_t)$ with j=h+1,⋯,T for t=R+1, R+2, ...,T–h.

5.1 Tests of Predictive Accuracy

Let the out-of sample errors be $e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$, and the quadratic loss function be

$$L(y_{t+h}, \hat{y}_{t+h|t}) = {e_{t+h|t}}^2.$$

For the nested comparisons of this study, i.e. random walk, RW, against 2Disc Forecast Combination of ADL regression models, Clark and West (2007), CW test is employed. It follows a standard normal distribution. Suppose model B is a larger model that nests model A. Then define

$$d_{t+h} = e^2_{A,t+h|t} - \left[e^2_{A,t+h|t} - \left(e_{A,t+h|t} - e_{B,t+h|t}\right)^2\right]$$

The CW test statistic defined below is simply the t-statistic used for testing that the expected value of d^{adj}_{t+h} is zero.

$$CW_h = \bar{d}_T \Big/ \sqrt{\widehat{V(\bar{d}_T)}} \quad , \quad \bar{d}_T = \frac{1}{\mathrm{P}} \sum_{\mathrm{t=R}}^{\mathrm{T-h}} d_{t+h}$$

The asymptotic variance V is estimated by the Newey-West HAC estimator, since for h-step ahead forecasting, h > 1, the forecasts errors are assumed to follow a moving average process of at most h-1 order.

CW test concerns forecast losses that are evaluated at the population parameters, it does not directly reflect the effect of estimation uncertainty on relative forecast performance whereas in Giacomini and White (2006), GW test, the losses depend on estimated in-sample parameters, and the expectation is taken to be conditional/unconditional on some information set G_t. Testing the null hypothesis of

$$H_0 \colon E\left[L_{t+h}\left(y_{t+h}, f_{t+h_{|t}}(\hat{\beta}_t)\right) - L_{t+h}\left(y_{t+h}, g_{t+h_{|t}}(\hat{\theta}_t)\right) | G_t\right] = 0$$

GW test statistic is a Wald-type statistic of the following form

$$GW_{R,P}^{\eta} = P\left(P^{-1}\sum_{t=R}^{T-h}\eta_t\,\Delta L_{t+h}\right)'\hat{\Omega}_P^{-1}\left(P^{-1}\sum_{t=R}^{T-h}\eta_t\,\Delta L_{t+h}\right)$$

$$= P\bar{Z}_{R,P}{}'\hat{\Omega}_P^{-1}\bar{Z}_{R,P}$$

Where ΔL_{t+h} is the difference of loss functions at t+h, η_t[12] is referred to as the vector of test functions and $\hat{\Omega}_P$ is a consistent estimator of the asymptotic variance of $\bar{Z}_{R,P}$[13]. Under the null of equal conditional forecast performances, GW test statistic follows a $\chi^2_{\dim(\eta_t)}$ distribution. In this paper GW test is employed to test equal conditional predictive ability of 2DiscMSFE forecast combination of ADL regression models and autoregressive, AR, regression models.

5.2 Fluctuation Tests

A fluctuation test proposed by Giacomini and Rossi (2010) is also considered to investigate possible fluctuations in relative predictive abilities of forecasting models during forecast evaluation period. This test focuses on the entire time path of the models` relative performance and implemented by plotting the standardized sample path of the relative measure of local performance (difference in MSFEs), together with critical values indicating that one of the models outperformed its competitor at some point in time, if crossed.

In this paper, competing models` local relative performance is measured by the out-of-sample MSFE differences computed over rolling windows and to test the null of zero MSFE differentials between the two competing models at each point in time in the forecast evaluation period, critical values are taken from Table 1 of Giacomini and Rossi (2010).

For the rolling scheme employed in this study, there is a sequence of P out-of-sample forecast loss differences $\left\{\Delta L_t(\hat{\beta}_{t-h,R}, \hat{\theta}_{t-h,R})\right\}_{t=R+h}^{T}$ depending on the realizations of the variable and on the in-sample parameter estimates for each model re-estimated at each t= R+h,⋯, T over a window of size R. Then, the local relative loss for the two models is defined over centered rolling windows of size m as

[12] Following Giacomini and White (2006), $\eta_t = (1\ \ \Delta L_t)'$ is used as the test function in this paper, having potential explanatory power for the future difference in forecast performances.

[13] In this study, for h-step ahead forecasts, h>1, $\hat{\Omega}_P$ is a Newey-West HAC estimator.

$$\frac{1}{m}\sum_{j=t-m/2}^{t+m/2-1}\Delta L_j(\hat{\beta}_{j-h,R},\hat{\theta}_{j-h,R}) \ , \qquad t=R+h+\frac{m}{2},\dots,T-\frac{m}{2}+1$$

And fluctuation test statistic is defined as

$$F_{t,m}=\hat{\sigma}^{-1}m^{-1/2}\sum_{j=t-m/2}^{t+\frac{m}{2}+1}\Delta L_j\big(\hat{\beta}_{j-h,R},\hat{\theta}_{j-h,R}\big), \quad t=R+h+\frac{m}{2},\dots,T-\frac{m}{2}+1$$

Where $\hat{\sigma}^2$ is Newey-West HAC estimator of σ^2. The null hypothesis $H_0: E\big[\Delta L_t(\hat{\beta}_{t-h,R},\hat{\theta}_{t-h,R})\big]=0$ for all t=R+h,⋯T. This test statistic is equivalent to Giacomini and White`s (2006) unconditional GW test statistic computed over rolling out-of-sample windows of size m; hence, this test is employed as GW Fluctuation Test in order to test equal predictive ability of 2DiscMSFE forecast combination of ADL regression models and autoregressive, AR, regression models at each point in time in the forecast evaluation period. Moreover, for the nested case, to test equal forecast performance of ADL regression models and the random walk, RW, models, at each point in time, CW Fluctuation Test in which $\Delta L_j\big(\hat{\beta}_{j-h,R},\hat{\theta}_{j-h,R}\big)$ is replaced with its adjusted version as in Clark and West`s (2006) CW test, is utilized.

$$\Delta L_j^{adj}=L_j\left(y_j,f_j(\hat{\beta}_{j-h,R})\right) - \left[L_j\left(y_j,f_j(\hat{\beta}_{j-h,R})\right)-\left(L_j\left(y_j,f_j(\hat{\beta}_{j-h,R})\right)^{1/2}-L_j\left(y_j,g_j(\hat{\theta}_{j-h,R})\right)^{1/2}\right)^2\right]$$

Where the model with parameter β is the larger model (ADL model) that nests the model with parameter θ (RW model).

6 Empirical Results

Using a rolling forecasting method, pseudo out-of-sample forecasts are obtained in order to evaluate the predictive ability of the models for various forecasting horizons h = 1, 2, 3 and 4. The total sample size is T+h, fixed rolling window size is R=80 quarters and for each window, 8 quarters out are used for forecasting and calculating RMSFE for the corresponding window; hence, the period used to evaluate the forecasts is P=T+h-R-8. The estimation periods for the 1950 and 1960 samples are 1950:Q1 to 1970:Q1 and 1960:Q2 to 1980:Q2, while the forecasting periods are 1970:Q1+h to 2008:Q4-h and 1980:Q2+h to 2008:Q4-h, respectively.

The forecast accuracy of each model is assessed using the root mean squared forecast error, RMSFE, obtained as follows:

$$RMSFE_t = \sqrt{\frac{1}{t - T_0 + 1} \sum_{\tau = T_0}^{t} \left(y_{\tau+h}^{h} - \hat{y}_{\tau+h|\tau}^{h}\right)^2}$$

where t = $T_1, \cdots, T_2$. T_o is the point at which the first individual pseudo out-of-sample forecast is computed. For the sample of 1950, T_0=1970:Q1 while for the sample of 1960, T_0=1980:Q2. T_1=1970:Q1 + h and T_1=1980:Q2+h, respectively. And, T_2=2008:Q4- h for all sample periods.

The boxplots used in the Introduction present the relative RMSFEs of the traditional AR model and the 2DiscMSFE/DMSFE forecast combinations of ADL models; a single RMSFE is attached to each out-of-sample rolling window. In line with common practice, results are reported in the form of RMSFEs relative to the RMSFE of a benchmark model, here taken to be the random walk (RW) model. Also, boxplots of relative RMSFEs utilized in order to assess the predictive ability of individual ADL models display relative RMSFEs of the traditional autoregressive (AR) model and of the ADL model for the potential predictor.

Findings of this paper are presented in Tables 2 through 6 and Figures 4 through 9. These tables report the results for the benchmark model of random walk (RW), the traditional autoregressive (AR) model and the 2DiscMSFE/DMSFE forecast combinations of ADL models for the two samples using alternative predictors as discussed in section 4. These results are based on aggregated variables marked by the data availability in two sample periods and evidence is presented for four forecasting horizons, h = 1, 2, 3, and 4 quarters ahead.

6.1 Main Findings

Table 2 having three panels presents RMSFEs of the models for U.S. real Government Expenditures, U.S. real Government Receipts and U.S real Government Budget Deficit[14] for which forecasts are obtained by subtracting forecasts of expenditures from forecasts of receipts. Each panel of the table reports RMSFEs of the benchmark model of random

[14] Here, expenditures and receipts are transformed into quarterly growth rates to induce stationarity. The differential between them is referred to as government budget deficit although it is actually the difference between the two growth rates not the deficit at levels. Hence, negative values of the differential indicate higher growth of expenditures relative to receipts implying increases in deficit at levels.

walk (RW), the traditional autoregressive (AR) regression model and the 2DiscMSFE/DMSFE forecast combinations of ADL models for the corresponding variable of interest. The results are grouped into two samples which correspond to different combinations of ADL models due to data availability; some of the indicators used in ADL modeling are not available for the longer sample as discussed in section 3, hence they are added to the combinations when they become available. Moreover, since indicators used in forecast combinations are picked by comparing their predicting ability as discussed in section 4, the number of predictors in the forecast combinations for the Expenditures model also changes among forecasting horizons, predictors chosen for each sample and horizon are given in Table 6. The evidence (see Table 2) shows that 2DiscMSFE/DMSFE forecast combinations of ADL models, for all three variables, provide stronger forecast gains against their corresponding benchmark of Random Walk (RW)s since their relative RMSFEs are substantially below one; moreover, ADL regression models improve forecasts compared to traditional AR regression models. These results hold for all forecast horizons, and for both samples. It is worth to note that forecast gains for the ADL model of Receipts is less substantial compared to those for the Expenditures model; the forecast performance of ADL model is measured by its forecast gains over the RW model, but in the case of Receipts, the random walk model itself has large forecast errors.

Figures 4 to 9 provide time plots of RMSFEs[15] relative to the RW and forecast combination weights over the evaluation period. These relative RMSFEs for the Deficit show the forecasting gains of traditional AR model and 2DiscMSFE/DMSFE forecast combination of ADL models throughout the evaluation period for h=1 and h=4. Forecast combination weights for expenditures and receipts are given as snapshots for particular time periods, in which forecast performance of the Deficit ADL model changes substantially, to see the possible relation between changes in weights and changes in predictive ability.

Figure 4 shows that on average (global) the relative predictive ability of ADL model on Deficits is much better than that of the AR model, especially for the periods of early 1970s and 2000s; these failures of AR model can be attributed to the important economic and policy changes; for instance, in 2001 nontaxable income has risen sharply reducing revenues. Moreover, Penner (2008) indicates that the largest revenue forecasting error was made by CBO in January of 2001 for fiscal year 2002; hence, revenues were overestimated by 3.2 percent of GDP, which was the reason for federal

[15] These relative RMSFEs are obtained using the rolling method; for the forecast evaluation period, windows of size 15 quarters are used to compute RMSFEs since for each data time I only have forecast errors. Thus, for each rolling out-of-sample window, I take the mean of 15 forecast errors to obtain RMSFEs.

income tax cuts in 2001, and the unpredicted recession of 2001 had not been forecast properly causing further revenue shortfalls. Revenue and deficit forecasts had become much more depressing by 2003, but that did not deter the Congress from passing large tax cuts on dividends and capital gains. The relatively higher performance of forecast combinations of ADL regression models during early 2000s results from the inclusion of personal income and industrial production as predictors for Receipts; personal income is the best predictor for Receipts having the largest weight in this period and industrial production as the second best (See Figure 6.a for the change in combination weights); and since AR model cannot track the changes in nontaxable income, combinations of ADL model performs better. However, the highest weight given to personal income during 2003-2004 period makes ADL forecasts worse than AR foreacasts, because the Congress did not change tax cuts although taxable income changed a lot, so revenues followed an AR model (see Figure 9.b for local changes in predictive ability of ADL model against AR). As reported by CBO there were large errors in their forecasts for 1979-1983 period when the economy experienced its hardest recessionary period of the postwar era[16]. Also, during this period there was substantial changes in President's economic and foreign policies-tax cuts combined with substantial increases in Cold War–era defense spending. Hence, combinations of ADL model for Deficits outperforms the AR model during this period, since both ADL forecasts of Expenditures and Receipts are significantly better than that of the AR model (See Figures 7.b and 8.b for individual forecast gains of ADL model for Expenditures and ADL model for Receipts over their AR counterparts) Industrial production and employment as predictors contribute to this performance of ADL model in this period (See Figure 6.b)

6.2 Forecast Evaluations

This section evaluates the forecasting evidence presented in the previous section by presenting time series statistical inference utilizing a few different tests.

In order to evaluate the hypothesis of equal forecasting accuracy between the traditional forecasting regression models of AR and 2DiscMSFE/DMSFE forecast combinations of ADL regression models, Giacomini and White (2006), GW, test is performed. And to compare the ADL models against the RW Clark and West (2007), CW, test is employed. Table 3.A presents the equal forecasting accuracy test of CW, and it is found that 2DiscMSFE/DMSFE forecast combinations of ADL models for all three

[16] CBO's Economic Forecasting Record: An Evaluation of the Economic Forecasts CBO made from January 1976 through January 2004. (November 2006)

variables and both samples yield significant results against forecasts based on the RW, that is, ADL models for Expenditures, Receipts and Deficits yield significantly lower MSFE than the MSFE of the RW for all forecasting horizons in both samples at 1% size of the test. Table 3.B provides the equal forecasting accuracy test of GW testing for equal forecasting accuracy between forecast combinations of ADL regression models vis-à-vis those obtained from traditional models of AR; it is found that for Expenditures and Deficits ADL regression models yield significant forecast gains over AR models for all forecasting horizons in both samples at 10% size of the test, while for Receipts ADL model' s forecasting accuracy compared to AR model deteriorates in 1960 sample for forecasting horizons of 2-,3- and 4-quarters ahead although it provides better forecasts than AR for 1950 sample for all horizons. This decrease in overall predictive ability of ADL regression model for Receipts in 1960 sample relative to 1950 sample can be explained by the exclusion of forecast gains obtained during the first ten years of the forecast evaluation period of 1950 sample which is not included in the evaluation period of 1960 sample (See Figure 4).

The tests, CW and GW, used in this study for evaluation of forecast performance of ADL regression models against RW and AR regression models, respectively, are based on measures of global performance and hence, they pick the models that have best forecasts on average over the pseudo out-of-sample period, which may result in information losses. Therefore, it is worth to investigate whether there are significant fluctuations in relative predictive ability of ADL regression models against the benchmark of RW and traditional AR models. To do so, the Fluctuation Test proposed by Giacomini and Rossi (2010) that analyzes the evolution of the model' s relative forecast performance over historical samples is employed. This test is implemented for Expenditures, Receipts and Deficits by plotting the standardized sample paths of ADL models' performances measured by their MSFEs relative to MSFEs of their competitors[17], together with one-sided critical values at 5% size of the test (the horizontal lines)[18]. If the test statistic crosses this critical value line, it can be concluded that ADL model outperforms its competitor at some point in time. Figures 7.a), 8.a) and 9.a) depict sequences of differences between the MSFEs of ADL regression models for

[17] These plots display sequences of standardized (rescaled by standard deviation) differences between MSFEs of ADL models and MSFEs of RW or AR, as the Fluctuation test statistic, computed over rolling windows of 30 observations so that rolling window size is a small (not too small) fraction of the out-of-sample size of 150 observations, which is required for Fluctuation test to have good size and power properties relative to full-sample tests of CW and GW as pointed out by Giacomini and Rossi (2010). Positive (negative) values of the test statistic imply that ADL model is better (worse) than its competitor.

[18] From Table 1 of Giacomini and Rossi (2010), one-sided critical value for the ratio of rolling window of size 30 to out-of-sample size of 150 at 5% level is taken to be 2.938.

Expenditures, Receipts and Deficits, respectively, and the MSFEs of their respective RW models as CW Fluctuation Tests together with one sided critical values for forecast horizons of h=1 and h=4. It is strongly evident from full-sample CW test results given in Table 3.A that ADL models beat their respective RW models easily, and these CW Fluctuation Tests also provide consistent results with their full-sample counterparts. Here, these fluctuation tests deserve attention for changes in relative performances of ADL models; for Expenditures, they indicate a decrease in performance of the ADL model relative to RW during the time period of 1985-1995 which can be attributed to Great Moderation Era defined as the reduction in volatility of key U.S. macroeconomic aggregates; whereas for Receipts and Deficits, fluctuations exhibit improvements in forecast performance of ADL models compared to RWs for the same period of time. Figures 7.b, 8.b and 9.b display sequences of differences between the MSFEs of ADL regression models for Expenditures, Receipts and Deficits, respectively, and the MSFEs of their respective AR models as GW Fluctuation Tests for forecast horizons of h=1 and h=4. For Expenditures, there is a significant improvement in predictive ability of ADL model relative to AR during the period of 1995-2004 (Figure 7.b). ADL model for Receipts significantly outperforms the AR model for one-period-ahead forecasts during the two periods of 1980-1984 and 1992-1997 (Figure 8.b left plot). It is noted in previous section that ADL model for Receipts in 1960 sample deteriorates compared to AR model for 2-,3-, and 4-quarters ahead forecasts when full-sample average (global) performance is considered. However, GW Fluctuation Test tells us that although relative performance of ADL model for Receipts decreases, the test statistic stays positive throughout the evaluation period for 1960 sample indicating that ADL model is still locally better than its competitor of AR model (Figure 8.b right plot). For Deficits, GW Fluctuation Test indicates a significant improvement in ADL model' s performance over the AR model during 1979-1984 period and relatively higher positive test statistics after 1995 implying better predictive ability relative to AR vis-à-vis poor relative performance seen during 1985-1995.

7 Conclusion

This study is an initial attempt to develop a forecasting model to predict the U.S. Government Budget Deficit. Evidence shows that forecast combinations of augmented distributed lag models provide forecast gains over the traditional autoregressive models. Therefore, future research is planned so as to develop the forecast model for the U.S. Government Budget Deficit further. Namely, the first step will be the comparison of forecasts from the model proposed in the present study with the forecasts that the Congressional Budget Office (CBO) provides. Then, since the results indicate that there is a room for government securities and interest rates in predicting the quarterly deficit, high frequency data together with some other macroeconomic and financial indicators will be introduced to the forecasting model employing Mixed Data Sampling (MIDAS) regressions. Finally, a state-level analysis is projected.

Tables and Figures

Table 1: RMSFE for RW and AR models

This table presents root mean squared forecast errors (RMSFEs) of random walk (RW) and autoregressive (AR) regression models with U.S. Government Budget Deficit as the dependent variable, and RMSFEs obtained by taking the difference between forecasts of U.S. real Government Receipts and U.S. Government Expenditures for 1-, 2-, 3- and 4-step ahead forecasts for two sample periods: 1950 and 1960. The estimation periods for the 1950 and 1960 samples are 1950:Q1 to 1970:Q1 and 1960:Q2 to 1980:Q2, while the forecasting periods are 1970:Q1+h to 2008:Q4-h and 1980:Q2+h to 2008:Q4-h, respectively.

RMSFE		Sample of 1950 horizon				Sample of 1960 horizon			
		1	2	3	4	1	2	3	4
RW	Receipts minus Expenditures	3.23	3.22	3.25	3.34	2.64	2.68	2.81	2.82
	Deficit	79.08	81.24	82.37	82.72	85.09	87.77	89.18	89.61
AR	Receipts minus Expenditures	2.43	2.47	2.49	2.44	2.11	2.16	2.22	2.27
	Deficit	65.40	71.59	72.40	75.94	71.27	79.12	80.18	84.53

Table 2: RMSFE for RW, AR and 2DiscMSFE/DMSFE forecast combinations of ADL Models

This table presents RMSFEs of the benchmark model of random walk (RW), and RMSFEs of the traditional autoregressive (AR) models and the 2DiscMSFE/DMSFE forecast combination of ADL models for U.S. real Government Receipts and DMSFE forecast combination of ADL models for U.S. real Government Expenditures the difference between them relative to the RMSFE of Random Walk (RW) for 1-, 2-, 3- and 4-step ahead forecasts for two sample periods: 1950 and 1960. The estimation periods for the 1950 and 1960 samples are 1950:Q1 to 1970:Q1 and 1960:Q2 to 1980:Q2, while the forecasting periods are 1970:Q1+h to 2008:Q4-h and 1980:Q2+h to 2008:Q4-h, respectively. Entries below 1 imply improvements compared to the benchmark.

	Sample of 1950 horizon				Sample of 1960 horizon			
Panel A: ***Expenditures***								
Forecast Horizon	1	2	3	4	1	2	3	4
RW	1.332	1.342	1.348	1.352	1.282	1.316	1.330	1.339
AR	0.712	0.723	0.727	0.728	0.692	0.704	0.706	0.705
ADL	0.693	0.680	0.680	0.698	0.620	0.624	0.617	0.644
Panel B: ***Receipts***								
Forecast Horizon	1	2	3	4	1	2	3	4
RW	2.987	3.012	3.024	3.026	2.604	2.632	2.648	2.647
AR	0.806	0.856	0.855	0.900	0.848	0.934	0.934	1.006
ADL	0.570	0.568	0.566	0.568	0.588	0.609	0.602	0.593
Panel C: ***Receipts - Expenditures***								
Forecast Horizon	1	2	3	4	1	2	3	4
RW	3.343	3.372	3.388	3.393	2.814	2.860	2.884	2.891
AR	0.767	0.793	0.794	0.815	0.789	0.841	0.842	0.878
ADL	0.584	0.588	0.589	0.590	0.588	0.633	0.625	0.614

Table 3: Time Series Test for predictive ability

A. This table presents Clark-West (CW) statistics for testing whether the difference in the MSFEs of 2DiscMSFE/DMSFE forecast combinations of ADL models and the Random Walk (RW) is zero for U.S. real Government Expenditures, U.S. real Government Receipts and the difference between them for both samples.

RW vs. ADL	Sample of 1950 horizon				Sample of 1960 horizon			
Panel A: ***Expenditures***	1	2	3	4	1	2	3	4
CW	7.430	4.477	4.244	4.044	5.988	3.420	3.224	3.089
p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: ***Receipts***	1	2	3	4	1	2	3	4
CW	6.393	3.420	3.245	3.159	6.871	3.794	3.586	3.450
p-val	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001
Panel C: ***Receipts-Expenditures***	1	2	3	4	1	2	3	4
CW	5.147	2.775	2.596	2.647	7.215	4.080	3.936	4.035
p-val	0.000	0.003	0.005	0.004	0.000	0.003	0.005	0.004

B. This table presents Giacomini-White (GW) statistics for testing for equal forecasting accuracy between 2DiscMSFE/DMSFE Forecast Combinations of ADL models against the traditional AR models for U.S. real Government Expenditures, U.S. real Government Receipts and the difference between them for all three samples.

AR vs. ADL	Sample of 1950 horizon				Sample of 1960 horizon			
Panel A: Expenditures	1	2	3	4	1	2	3	4
GW	3.417	4.597	3.429	4.799	2.824	3.827	3.502	2.997
p-val	0.003	0.000	0.003	0.000	0.019	0.001	0.002	0.011
Panel B: Receipts	1	2	3	4	1	2	3	4
GW	4.703	3.618	3.368	2.834	2.837	1.830	2.090	2.085
p-val	0.000	0.001	0.003	0.018	0.018	0.188	0.113	0.114
Panel C: Receipts - Expenditures	1	2	3	4	1	2	3	4
GW	6.188	4.942	4.216	3.510	3.445	2.352	2.289	2.209
p-val	0.000	0.000	0.000	0.002	0.003	0.063	0.073	0.087

Table 4: Data Series

This table lists the data series used in this study, all are seasonally adjusted. Government securities of (with IDs listed in this table) GS10, FEDFUNDS are not available for 1950 sample, so these two securities and the spreads constructed with them are included only in 1960 sample. Third column lists codes for transformation methods applied on the corresponding data series to aggregate them in quarterly frequency and/or to ensure stationarity. Transformation methods are described in Table 5.

ID	Title	Frequency	Transformation	Source Link
GEXPND	Government Current Expenditures	Quarterly	0&2	http://research.stlouisfed.org/fred2
GRECPT	Government Current Receipts	Quarterly	0&2	http://research.stlouisfed.org/fred2
INDPRO	Industrial Production Index	Monthly	2	http://research.stlouisfed.org/fred2
GDPC96	Real Gross Domestic Product, 3 Decimal	Quarterly	2	http://research.stlouisfed.org/fred2
PIECTR	Real personal income excluding current transfer receipts	Monthly	2	http://research.stlouisfed.org/fred2
NPPTTL	Total Nonfarm Private Payroll Employment	Monthly	2	http://data.bls.gov/cgi-bin/surveymost?bls
PPIFGS	Producer Price Index: Finished Goods	Monthly	3	http://data.bls.gov/cgi-bin/surveymost?bls
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	Monthly	3	http://data.bls.gov/cgi-bin/surveymost?bls
GDPDEF	Gross Domestic Product: Implicit Price Deflator	Quarterly	na	http://research.stlouisfed.org/fred2
AAA	Moody's Seasoned Aaa Corporate Bond Yield	Monthly	1	http://research.stlouisfed.org/fred2
BAA	Moody's Seasoned Baa Corporate Bond Yield	Monthly	1	http://research.stlouisfed.org/fred2
FEDFUNDS	Effective Federal Funds Rate	Monthly	1	http://research.stlouisfed.org/fred2
GS10	10-Year Treasury Constant Maturity Rate	Monthly	1	http://research.stlouisfed.org/fred2
TB3MS	3-Month Treasury Bill: Secondary Market Rate	Monthly	1	http://research.stlouisfed.org/fred2
MICH	Michigan Survey of Consumers: Index of consumer sentiment	Quarterly	1	http://www.sca.isr.umich.edu/main.php
RAMVAR	Ramey defense news variable from Ramey (2008)	Quarterly	1	http://weber.ucsd.edu/~vramey/research.html#data
RAMSHPRO	Ramey-Shapiro military date from Ramey and Shapiro (1998)	Quarterly	na	http://weber.ucsd.edu/~vramey/research.html#data

na: no transformation is applied

Table 5: Transformations of the Data Series

This table describes transformation methods applied on the data series to aggregate them in quarterly frequency and/or to ensure stationarity.

Code	Transformation	Description
0	$\mathrm{x_{it}}/GDP\ deflator$	Nominal to Real
1	$\mathrm{x_{it}} = (1 - L^3)(1 + L + L^2)X_{it}$ for monthly data $\mathrm{x_{it}} = (1 - L^3)X_{it}$ for quarterly data	Quarterly difference
2	for monthly data $\mathrm{x_{it}} = (1 - L^3)(1 + L + L^2)\log(X_{it}) * 100$ for quarterly data $\mathrm{x_{it}} = (1 - L^3) * 100$	Quarterly growth rate
3	for monthly data $\mathrm{x_{it}} = (1 - L^3)(1 + L + L^2)(1 - L^{12}) \log(X_{it}) * 100$ for quarterly data $\mathrm{x_{it}} = (1 - L^3)(1 - L^{12}) \log(X_{it}) * 100$	Quarterly difference of yearly growth rate

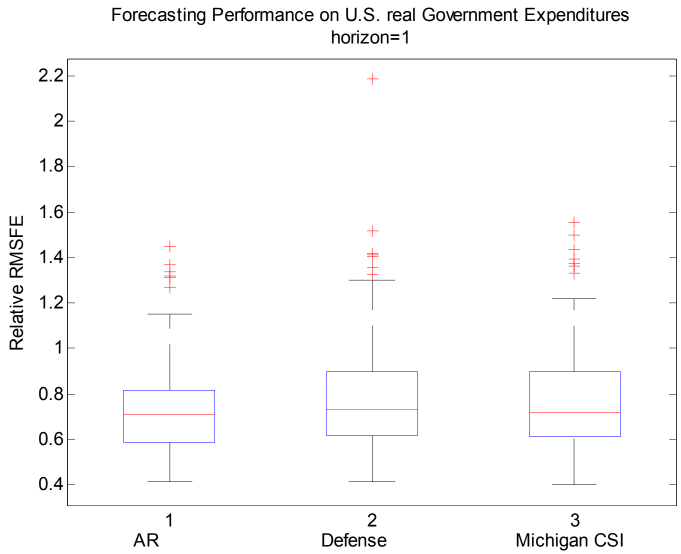
Table 6: Selected Predictors for Forecast Combinations

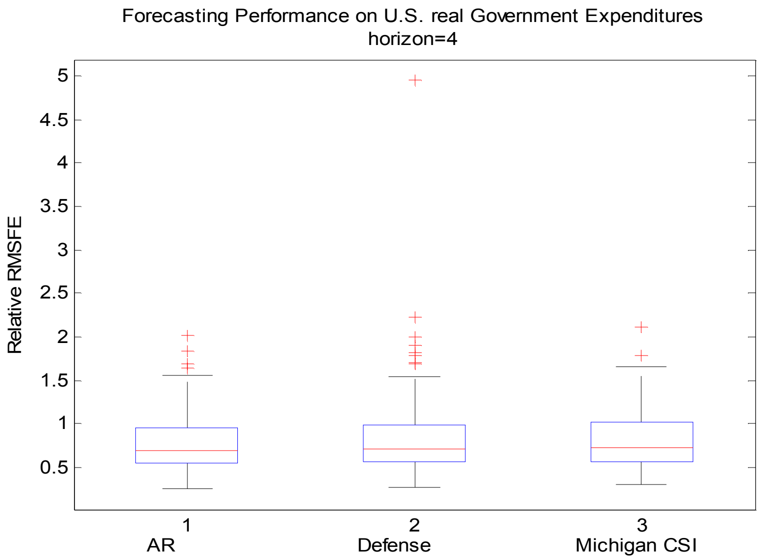
This table reports the predictors used in this study with individual ADL regression models; the ones with + sign are those picked according to their relative predictive performances in order to use in forecast combinations and they are included in combinations of forecasts for their corresponding variable and sample. – sign is given to those with relatively low forecast performances and/or to those that are not available in 1950 sample. Second column displays abbreviations of predictors that are used to represent them in figures. Since, with abbreviations given in the second column of this table, ff and 10y are not available in 1950 sample, predictors ff, 10y, 10y-3m and 10y-ff are not included in 1950 sample.

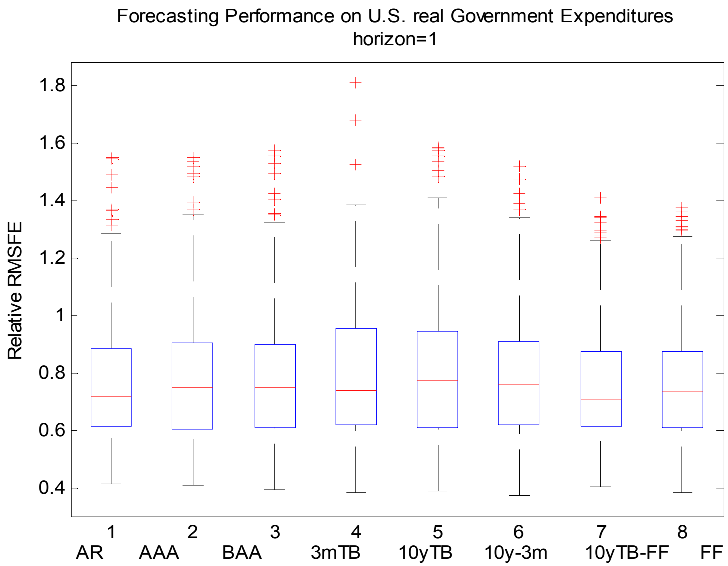
		Expenditures				***Receipts***			
		1950 sample		**1960 sample**		**1950 sample**		**1960 sample**	
Predictor	***Abbreviation***	**forecast horizons**				**forecast horizons**			
		h=1	**h=2,3,4**	**h=1**	**h=2,3,4**	**h=1**	**h=2,3,4**	**h=1**	**h=2,3,4**
real gdp growth	gdp	+	+	+	+	+	+	+	+
industrial production	ind	+	+	+	+	+	+	+	+
total non-farm employment	emp	+	+	+	+	+	+	+	+
personal income minus transfer payments	pinc	-	-	-	-	+	+	+	+
inflation (producer price index)	inf	+	-	+	-	-	-	-	-
inflation (consumer price index)	inf(cpi)	-	-	-	-	-	-	-	-
AAA corporate bond yield-aaa	aaa	-	+	-	+	+	+	+	+
BAA corporate bond yield	baa	-	+	-	+	-	-	-	-
three month treasury bill-3mTB	3m	+	+	+	+	+	+	+	+
ten year treasury bond yields-10yTB	10y	-	-	-	+	-	-	+	+
effective federal funds rate- FF	ff	-	-	+	+	-	-	+	+
10yTB minus 3mTB	10y-3m	-	-	+	+	-	-	+	+
10yTB minus FF	10y-ff	-	-	+	+	-	-	+	+
Defense News Variable (Ramey,2008)	mil or defense	+	+	+	+	-	-	-	-
Michigan Consumer Sentiment Index	mich	-	-	+	+	-	-	+	+

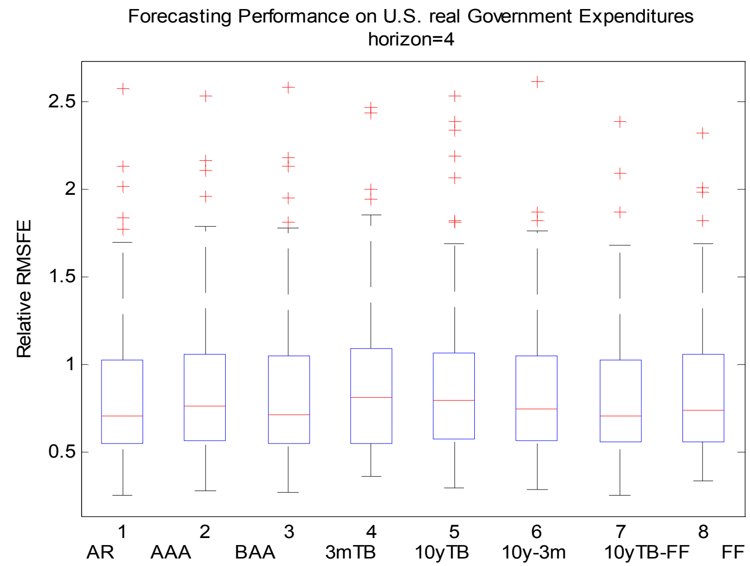
Figure 2: Boxplots Displaying Forecasting Performances of individual indicators on U.S real Government Expenditures for h=1 and h=4.

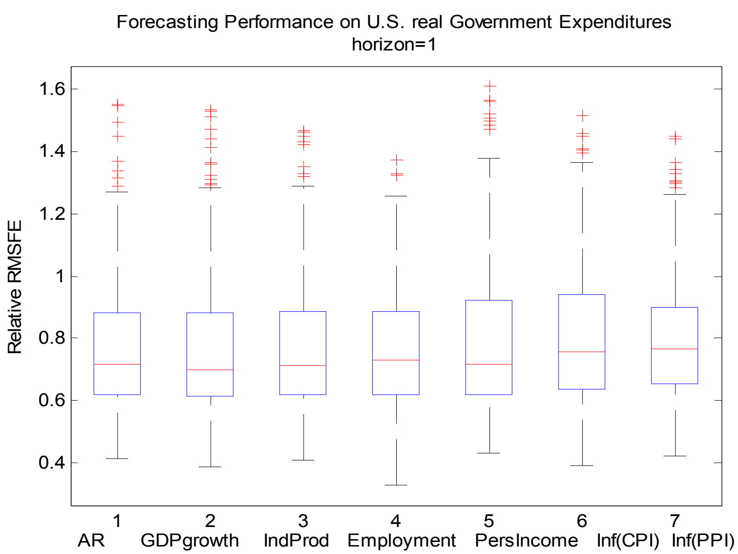
a)






Forecasting Performance on U.S. real Government Expenditures
horizon=1
Relative RMSFE
AR AAA BAA 3mTB 10yTB 10y-3m 10yTB-FF FF


Forecasting Performance on U.S. real Government Expenditures
horizon=4
Relative RMSFE
AR AAA BAA 3mTB 10yTB 10y-3m 10yTB-FF FF


Forecasting Performance on U.S. real Government Expenditures
horizon=1
Relative RMSFE
AR GDPgrowth IndProd Employment PersIncome Inf(CPI) Inf(PPI)

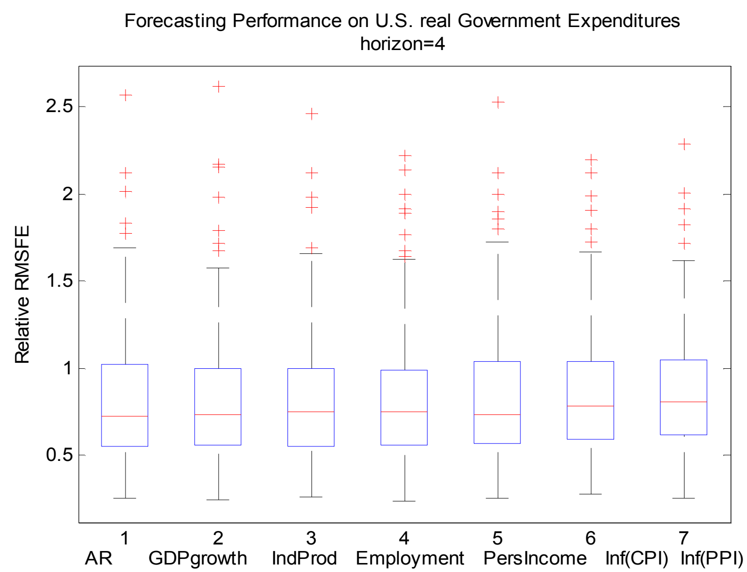
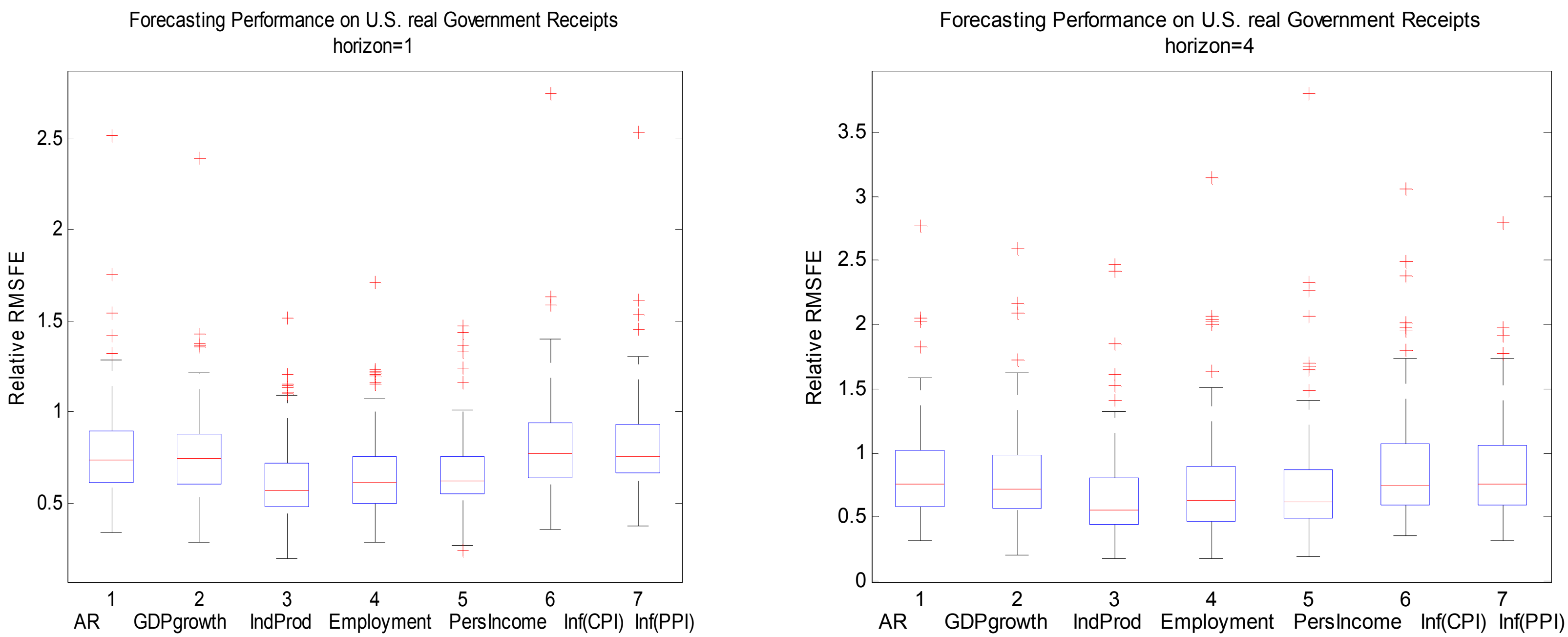
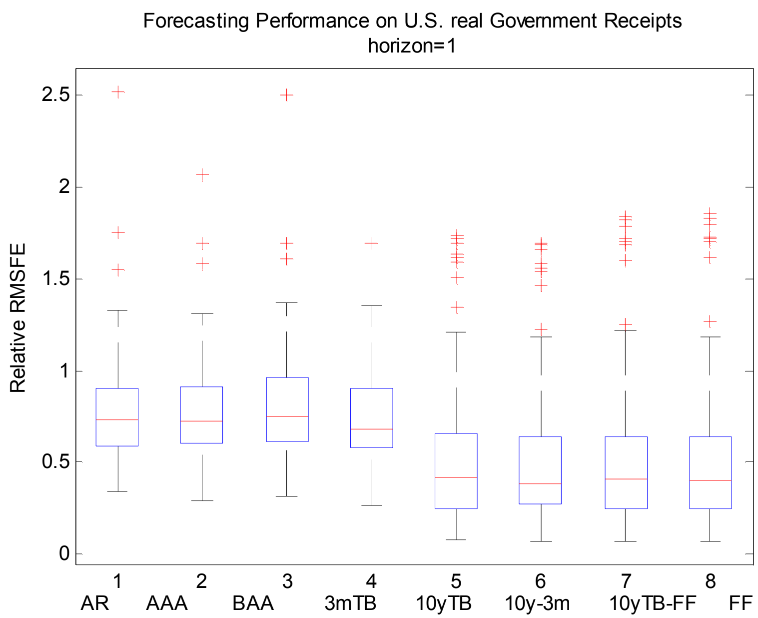
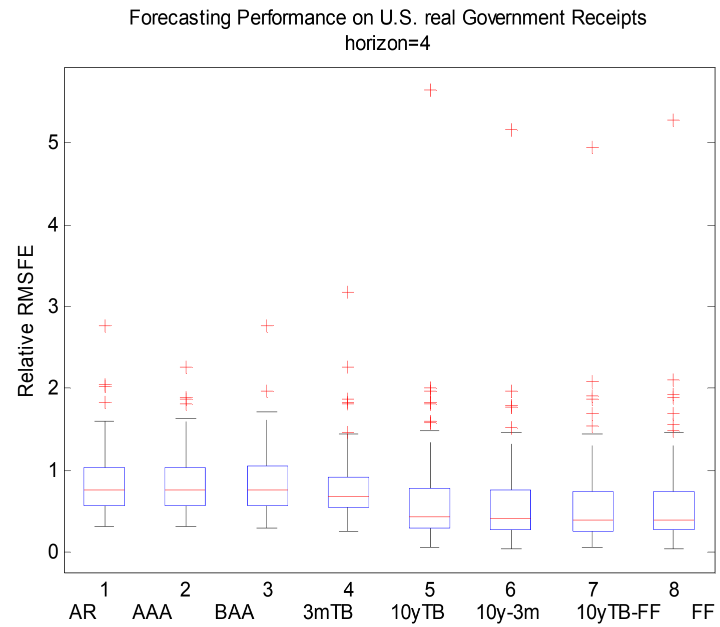

Forecasting Performance on U.S. real Government Expenditures
horizon=4
Relative RMSFE
AR GDPgrowth IndProd Employment PersIncome Inf(CPI) Inf(PPI)

Figure 3: Boxplots Displaying Forecasting Performances of individual indicators on U.S real Government Receipts, for h=1 and h=4

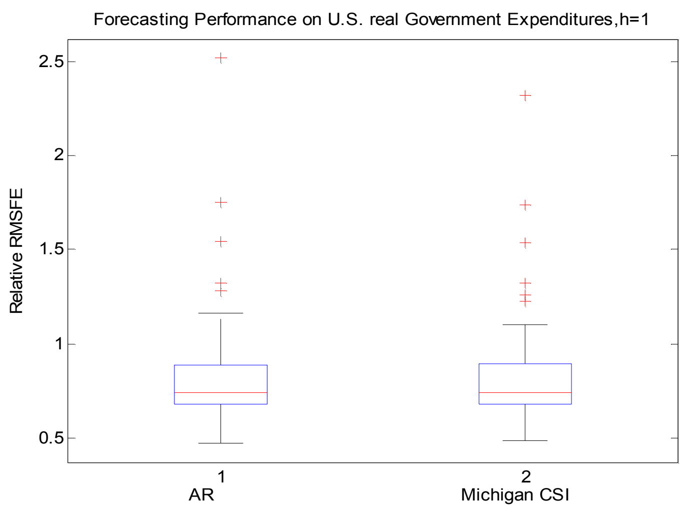




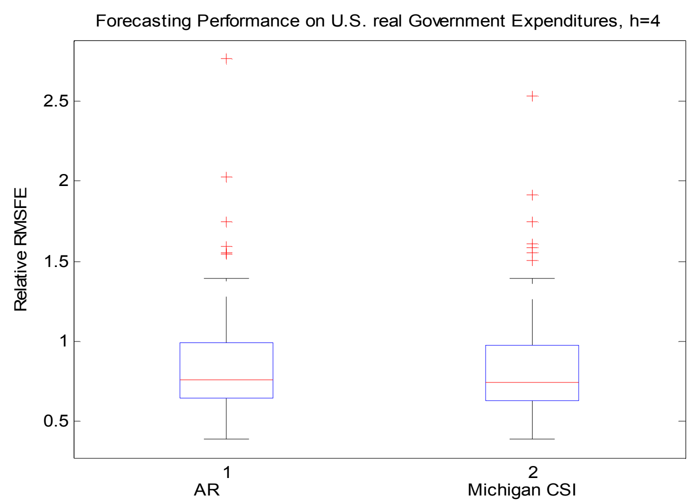
Forecasting Performance on U.S. real Government Receipts
horizon=1
Relative RMSFE
0
0.5
1
1.5
2
2.5
1 2 3 4 5 6 7 8
AR AAA BAA 3mTB 10yTB 10y-3m 10yTB-FF FF



Forecasting Performance on U.S. real Government Receipts
horizon=4
Relative RMSFE
0
1
2
3
4
5
1 2 3 4 5 6 7 8
AR AAA BAA 3mTB 10yTB 10y-3m 10yTB-FF FF

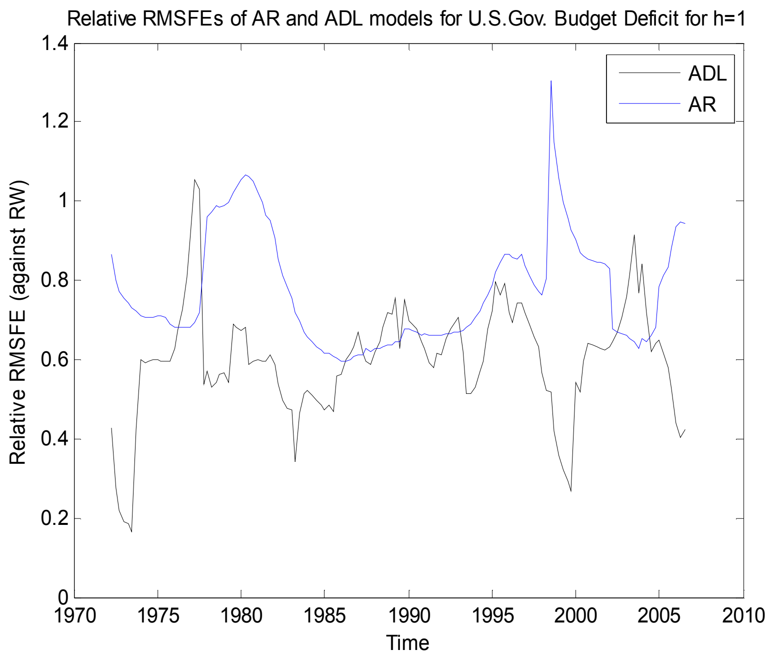


Forecasting Performance on U.S. real Government Expenditures,h=1
Relative RMSFE
0.5
1
1.5
2
2.5
1
AR
2
Michigan CSI



Forecasting Performance on U.S. real Government Expenditures, h=4
Relative RMSFE
0.5
1
1.5
2
2.5
1
AR
2
Michigan CSI

Figure 4: Relative RMSFEs of AR and 2DiscMSFE/DMSFE Forecast Combinations of ADL models against the benchmark RW for U.S. real Government Budget Deficit for h=1 and h=4.



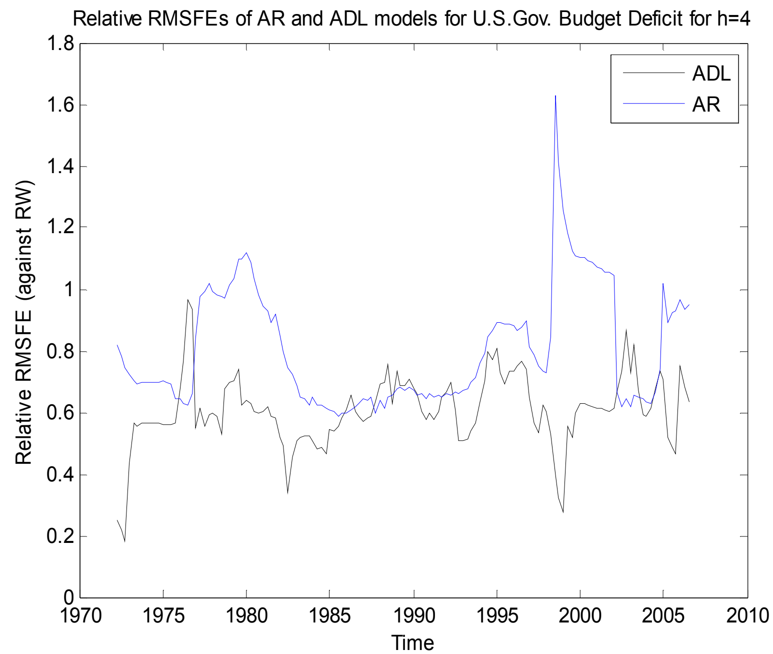
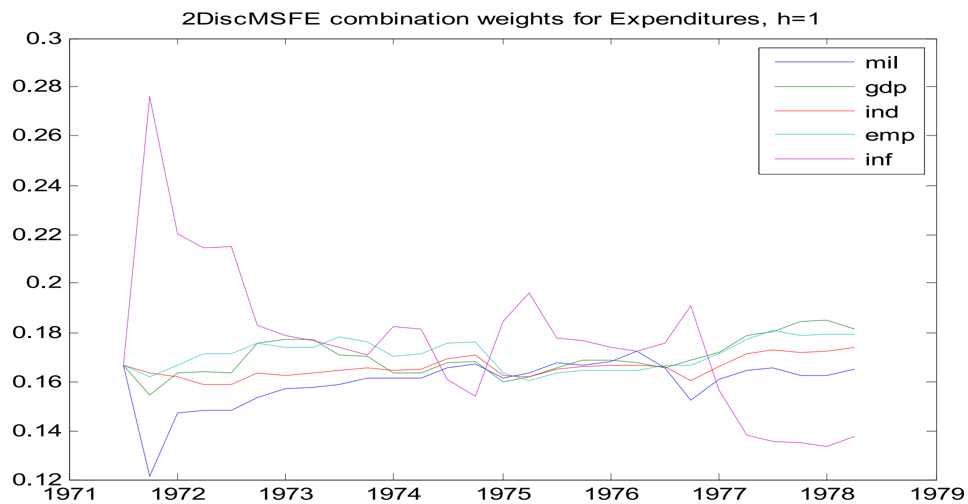


Figure 5: Snapshots of DMSFE combination weights for Expenditures for particular time periods, h=1.



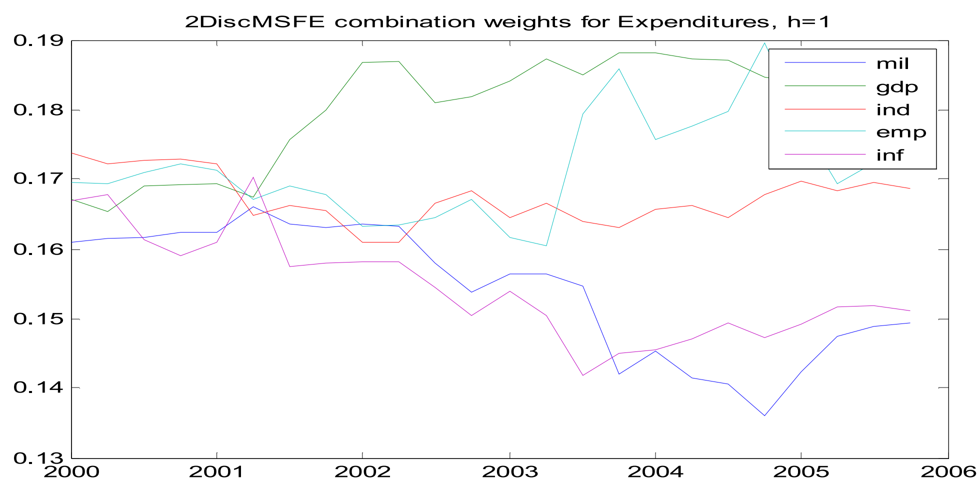
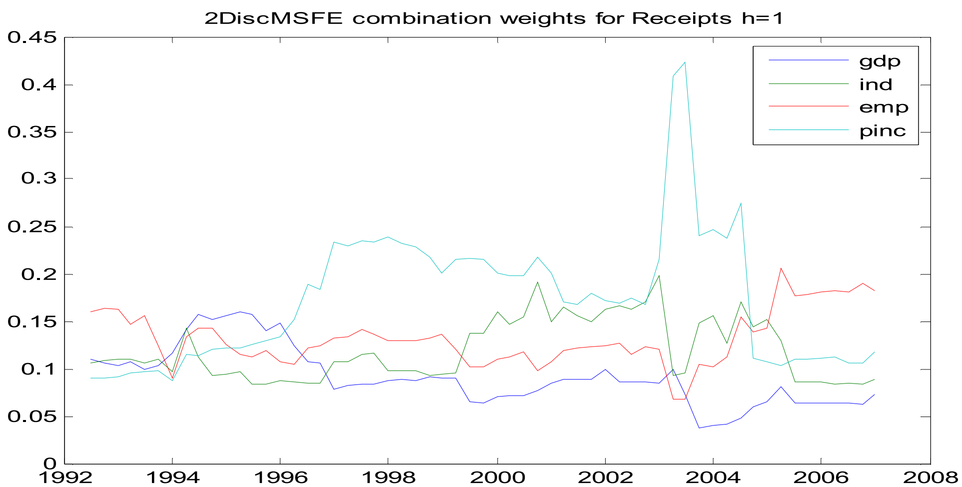
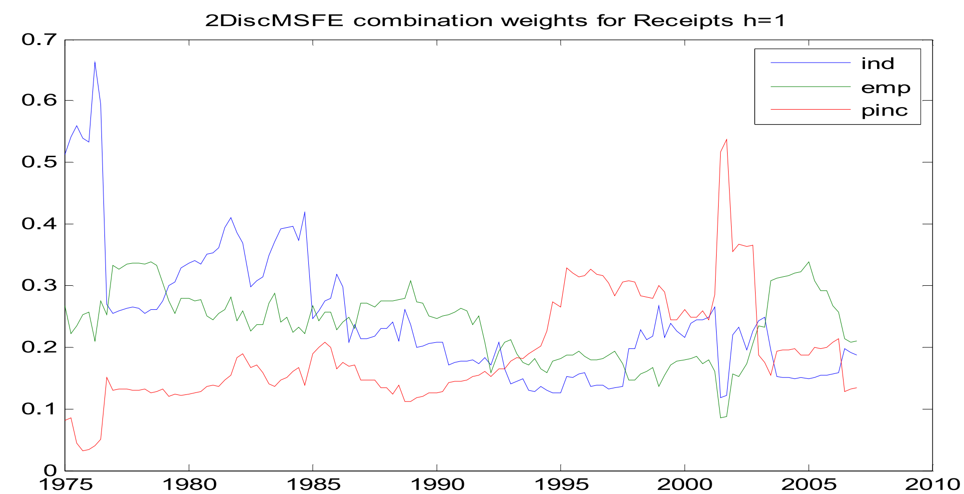


Figure 6: Snapshots of 2DiscMSFE combination weights for Receipts for particular time periods, h=1.





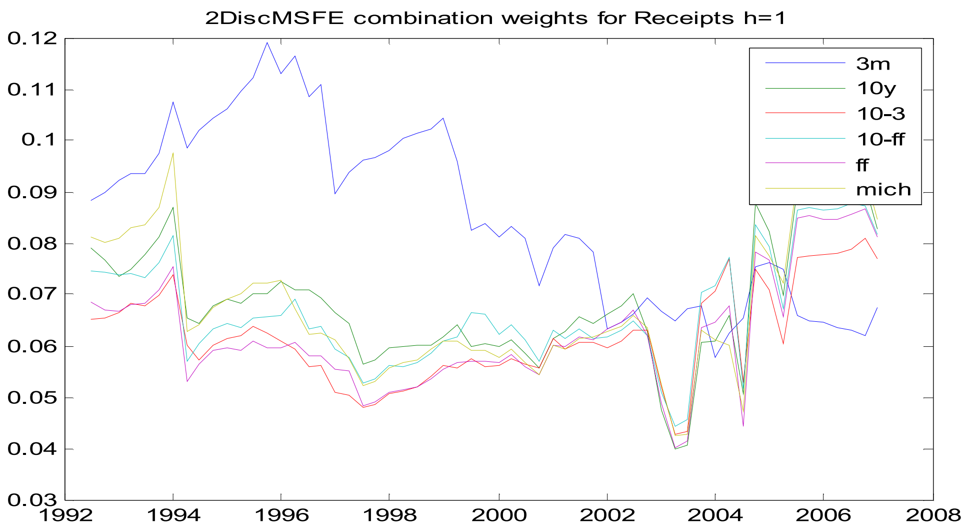
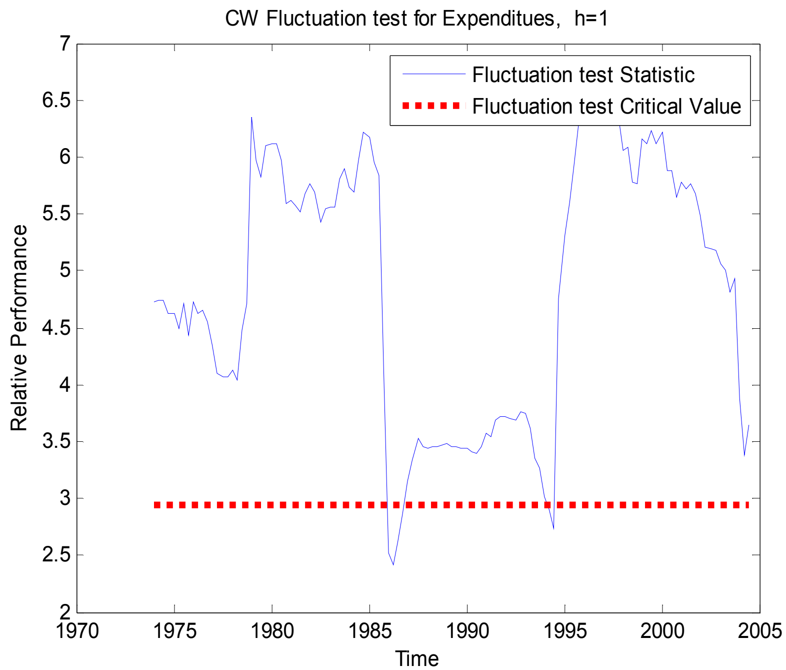
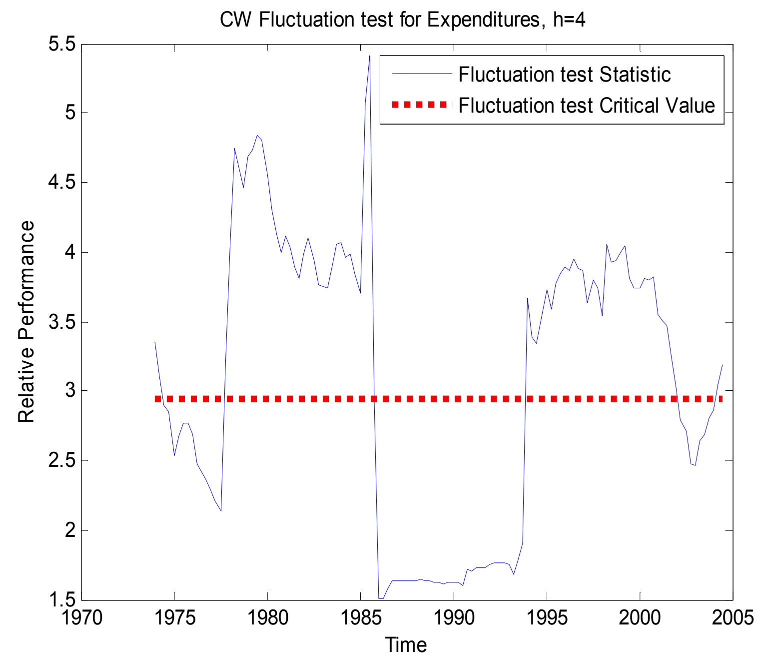

2DiscMSFE combination weights for Receipts h=1
0.12
0.11
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
1992
1994
1996
1998
2000
2002
2004
2006
2008
3m
10y
10-3
10-ff
ff
mich

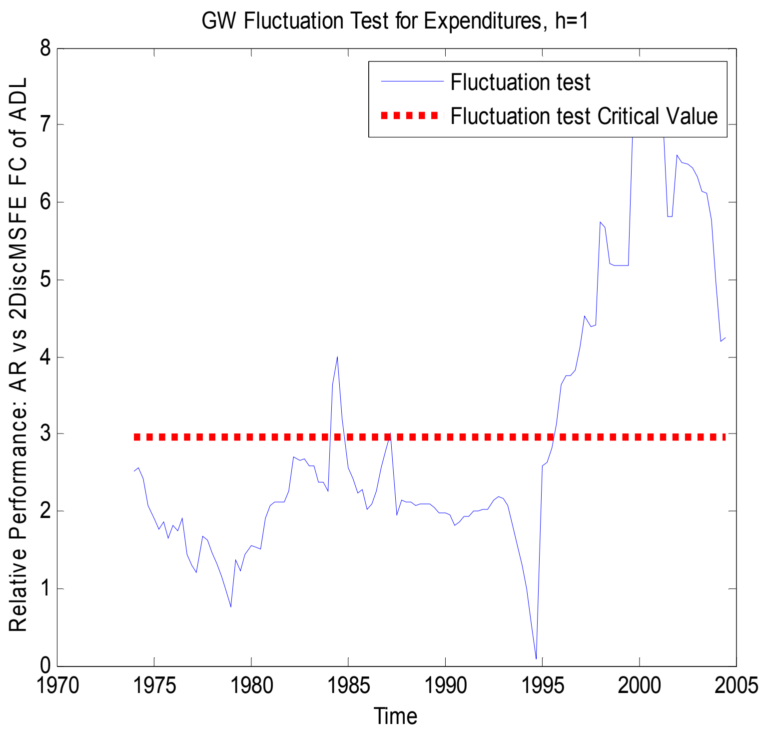
Figure 7:

a) CW Fluctuation Tests for Expenditures, for h=1 and h=4.





b) GW Fluctuation Tests for Expenditures, for h=1 and h=4.



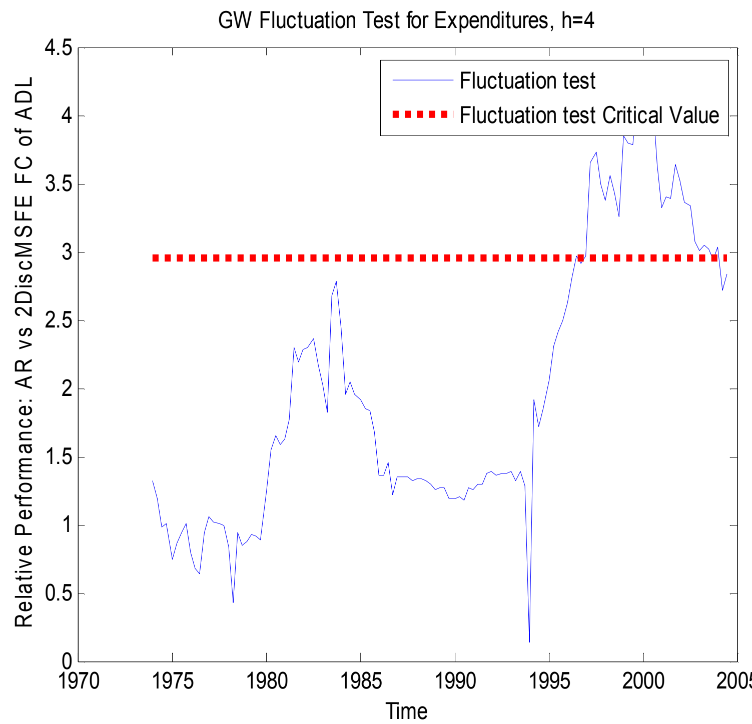
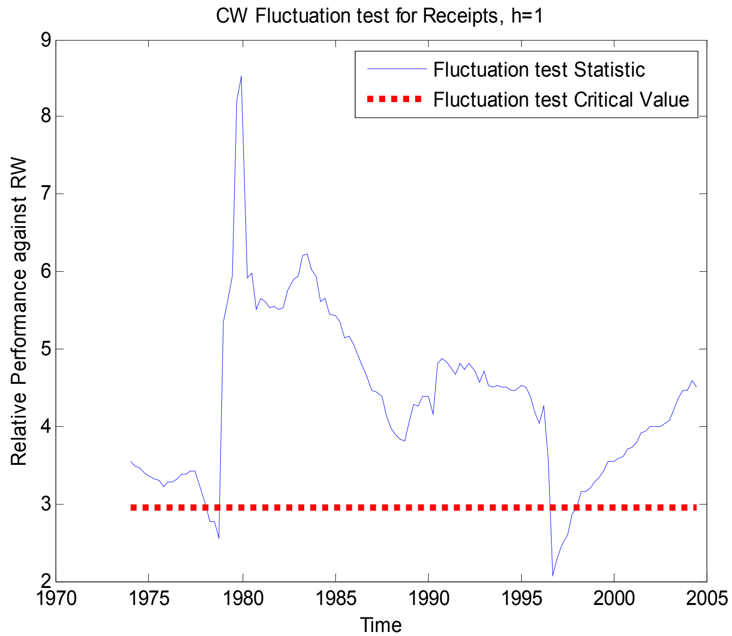
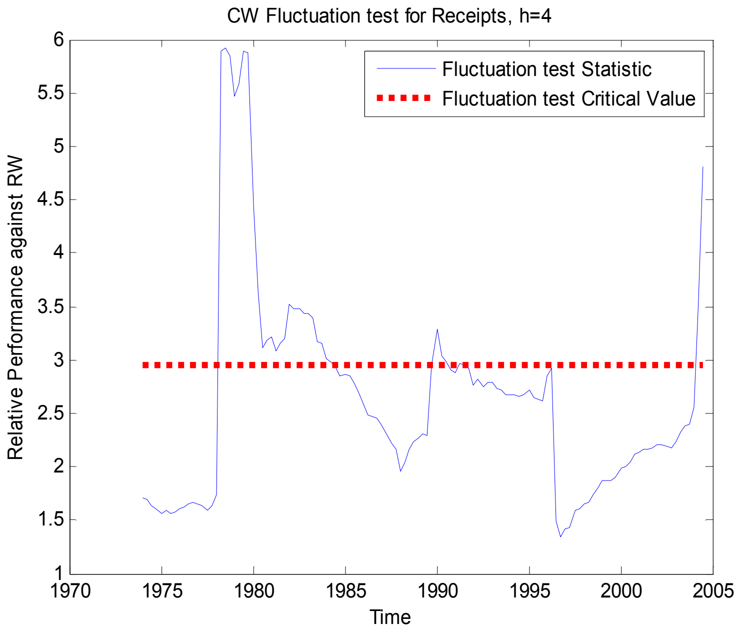


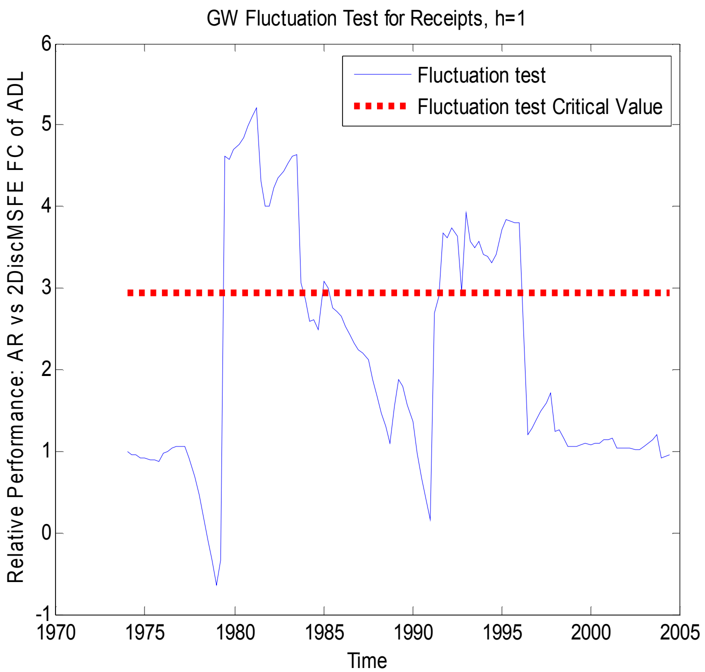
Figure 8:

a) CW Fluctuation Tests for Receipts, for h=1 and h=4.





b) GW Fluctuation Tests for Receipts, for h=1 and h=4.



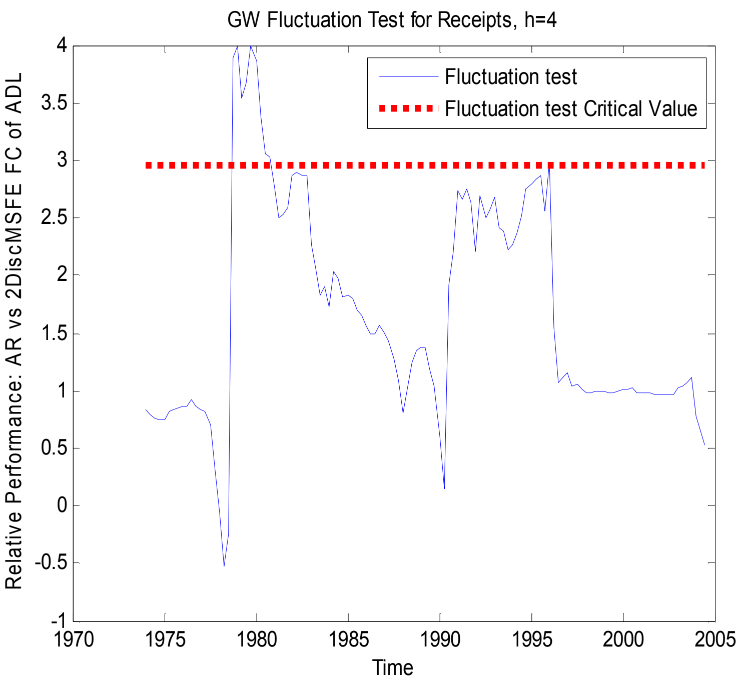
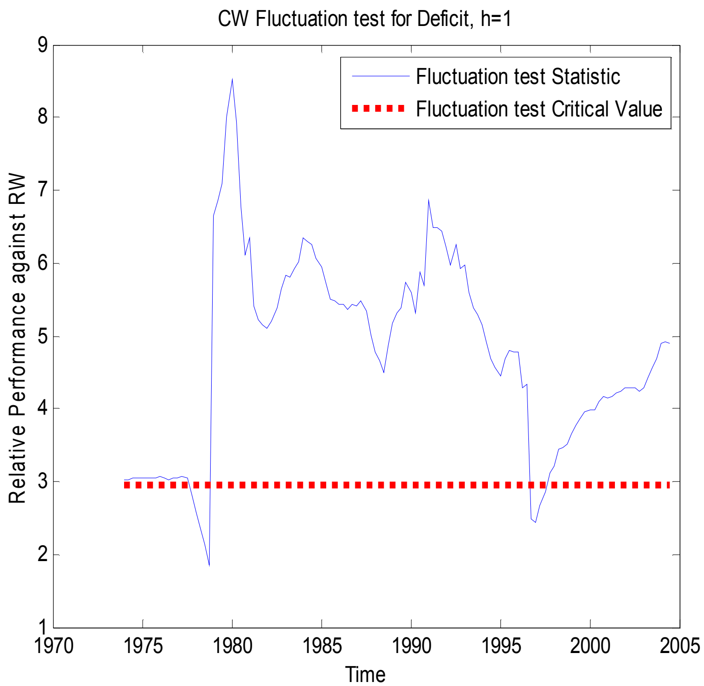
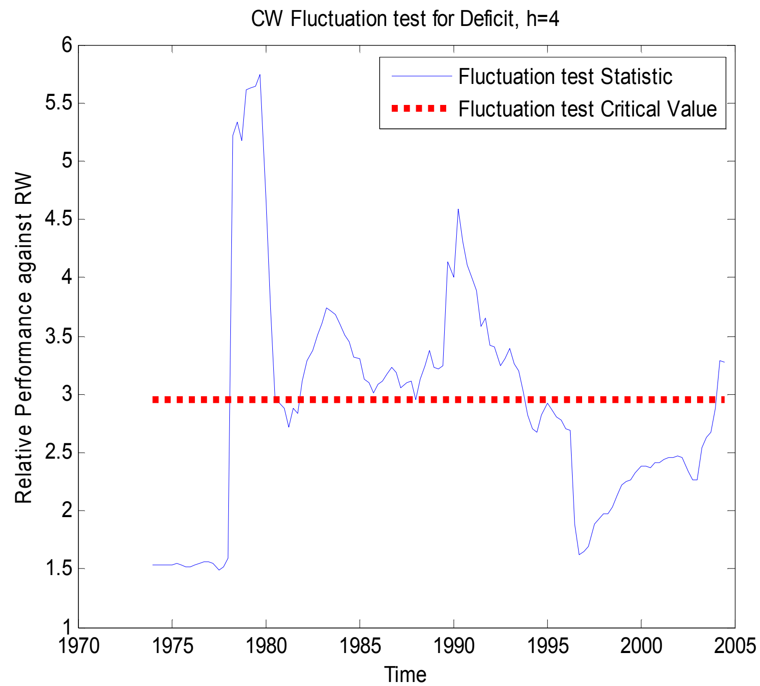


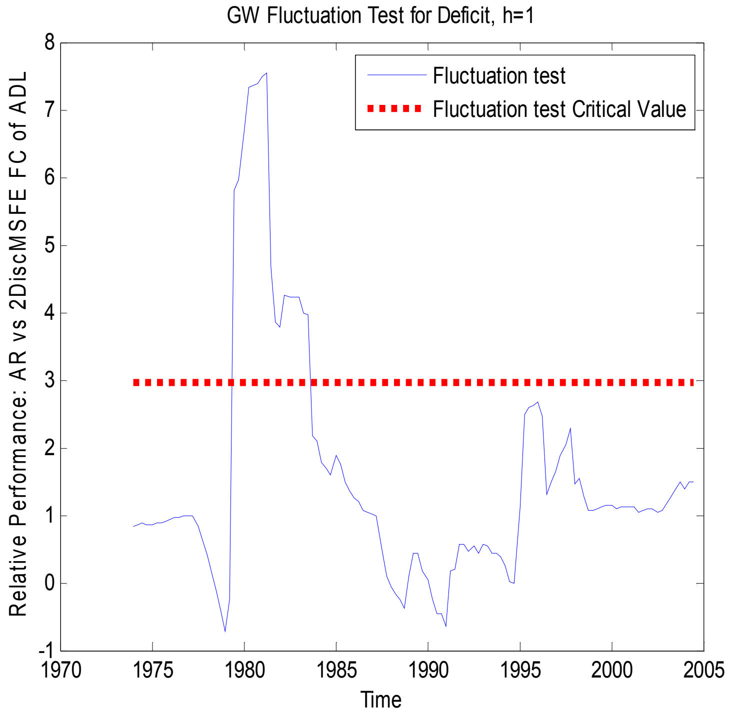
Figure 9:

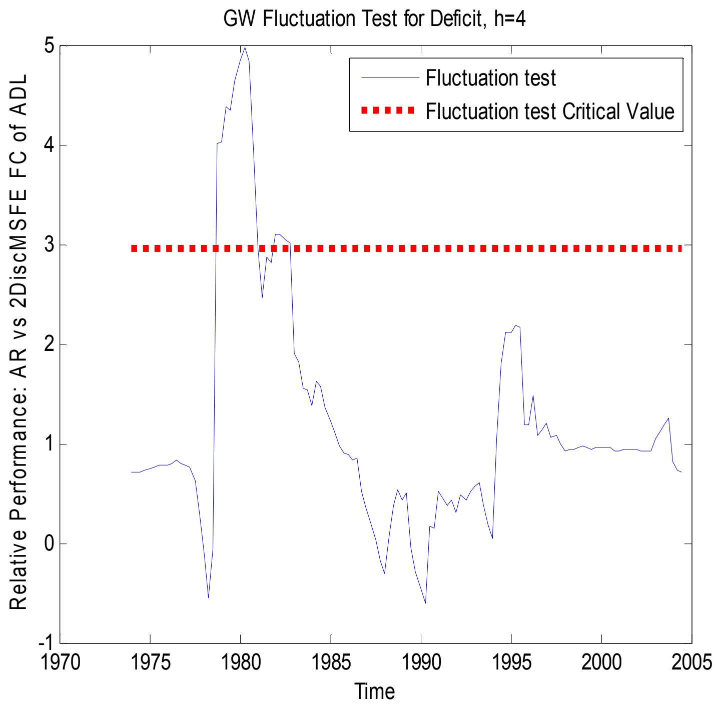
a) CW Fluctuation Tests for Deficits, for h=1 and h=4.





b) GW Fluctuation Tests for Deficits, for h=1 and h=4





References

Aiolfl, Marco, and Allan Timmermann (2006), Persistence in forecasting performance and conditional combination strategies, *Journal of Econometrics* 135, 31 53.

Andreou E., Ghysels E., and Kourtellos A. (2010a), Should Macroeconomic Forecasters Use Daily Financial Data and How? (November 19, 2010). Available at SSRN: http://ssrn.com/abstract=1711899

________ 2010b, Technical appendix: Should macroeconomic forecasters use daily financial data and how?, Available at http://www.unc.edu/~eghysels/working_papers.html.

Blanchard O., and R. Perotti (2002), An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output, *Quarterly Journal of Economics*, 117(4), 1329 68

Clark, T., and K. D. West (2006), Using Out-of-sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis*, Journal of Econometrics* 135, 155-186.

Clark, T. E., and K. D. West (2007), Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291 311.

Congressional Budget Office (2011), The Budget and Economic Outlook: An Update, *Washington, D.C: U.S. Government Printing Office.*

___________(2006), An Evaluation of the Economic Forecast CBO made from January 1976 through January 2004, *Washington, D.C: U.S. Government Printing Office*.

___________(2006), The Budget and Economic Outlook: Fiscal Years 2007 to 2016, *Washington, D.C: U.S. Government Printing Office*.

___________ (2006), How CBO Forecasts Income, *Washington, D.C: U.S. Government Printing Office.*

Diebold, F. X., R. S. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Edwards R.D. (2010), Forecasting Government Revenue and Expenditure in the U.S. Using Data on Age-Specific Utilization, *Nationoal Transfer Accounts Project Working Paper* WP10-01, February.

Engemann, K. M., M. T. Owyang, and S. Zubairy (2008), A Primer on the Empirical Identification of Government Spending Shocks, *Federal Reserve Bank of St. Louis Review*, 90(2), 117–31.

Fatás A., and I. Mihov (2001), The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence, *CEPR Discussion Paper* No. 2760.

Giacomini, R. and H. White (2006), Tests of Conditional Predictive Ability, *Econometrica*, 74, 1545-1578.

Giacomini R., Rossi B. (2010), Forecast Comparisons in Unstable Environments, *Journal of Applied Econometrics* 25(4), April 2010, 595-620.

Hendry, D. F. and M. P. Clements (2004), Pooling of forecasts, *Econometrics Journal* 7, 1 131.

Miller S.M (1991), Forecasting Federal Budget Deficits: How Reliable Are U.S. Congressional Budget Office Projections?, *Applied Economics,* vol. 23, pp. 1789-1799.

Newey W. and K. West (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.

Office of Management and Budget (2011), Fiscal Year 2012 Mid-Session Review, Budget of the U.S. Government

Owyang M. T., Zubairy S. (2009), Who Benefits from Increased Government Spending? A State-Level Analysis Working Paper 2009-006B http://research.stlouisfed.org/wp/2009/2009-006.pdf

Penner R.G. (2008), Federal Revenue Forecasting, Chapter reproduced with permission from *Handbook of Government Budget Forecasting* (New York: Taylor & Francis).

Ramey, V. A. (2008), Identifying Government Spending Shocks: It' s All in the Timing, Working Paper, University of California, San Diego.

Ramey, V. A., and M. D. Shapiro (1998), Costly Capital Reallocation and the Effects of Government Spending, *Carnegie-Rochester Conference Series on Public Policy*, 48, 145–94.

Stock, James H., and Mark W. Watson (2004), Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting* 23, 405 430.

__________ (2003a), Forecasting Output and Inflation: The Role of Asset Prices, *Journal of Economic Literature.*

__________ (2008b), Phillips curve inflation forecasts, *NBER* Working paper.

__________ (2001), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in R.F. Engle and H. White, eds., Festschrift in Honour of Clive Granger , *Cambridge University Press*, Cambridge, 1-44.

Timmermann, Allan (2006), Forecast combinations, in G. Elliott, C. Granger, and A. Timmermann, ed.: *Handbook of Economic Forecasting* vol. 1. pp. 136 196 (North Holland: Amsterdam).

Wright, Jonathan (2008), Bayesian model averaging and exchange rate forecasting, *Journal of Econometrics* 146, 329 341.