**Calhoun: The NPS Institutional Archive**

**DSpace Repository**

DRMI (Defense Resources Management Institute) Faculty and Researchers' Publications

2005

# Are Errors in Official U.S. Budget Receipts Forecasts Just Noise?

McNab, Robert M.; Rider, Mark; Wall, Kent D.

Monterey, California: Naval Postgraduate School.

https://hdl.handle.net/10945/32537

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

*Downloaded from NPS Archive: Calhoun*

http://www.nps.edu/library

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

Defense Resources Management Institute

# WORKING PAPER SERIES

## 2005/01

**Are Errors in Official U.S. Budget Receipts Forecasts Just Noise?**

Robert M. McNab, Mark Rider, Kent D. Wall

Existing evidence suggests that U.S. Government budget receipts forecasts are unbiased and efficient. Our study is an attempt to examine the veracity of these findings. The time series framework employed in this study is distinguished from previous work in three ways. We find that the U.S. Government's annual, one-year ahead, budget receipts forecasts for fiscal years 1963 through 2003 are biased and inefficient. In addition, we find that these forecasts exhibit serial correlation in their errors and thus do not efficiently exploit all available information. Finally, we find evidence that is consistent with strategic bias that may reflect the political goals of the Administration in power.

Defense Resources Management Institute
School of International Graduate Studies
Naval Postgraduate School
Building 234
699 Dyer Road
Monterey, CA 93943-5138
831-656-2306
www.nps.navy.mil/drmi

*The views herein expressed reflect the views of the author and are not the policy of DRMI, the Naval Postgraduate School, or the U.S. Government.*

# Defense Resources Management Institute
# Naval Postgraduate School

The Defense Resources Management Institute (DRMI) is an educational institution sponsored and supervised by the Secretary of Defense and located at the Naval Postgraduate School in Monterey, California. Since 1965 the Institute has conducted professional education programs in analytical decision making and resources management for military officers of all services and for senior civilian officials of the United States and 149 other countries.

The mission of the Defense Resources Management Institute is to educate civilian and military government officials in the principles and concepts of defense resources management.

The unique course offerings create an interdisciplinary, interactive learning environment that facilitates the exchange of ideas among U.S. and international participants in Monterey, California, and locations worldwide. These programs are presented on a regularly scheduled basis at DRMI in Monterey and by specific arrangement at other locations in the United States and in other countries.

The success of DRMI reflects the breadth and depth of the in-house technical expertise that DRMI can draw upon. The DRMI Faculty, a communication of scholars and practitioners, are experts in defense resource management and has published in major academic and technical journals. The faculty has extensive experience in providing technical assistance and educational programs worldwide over the past 40 years. Our educational strategy is one of collaboration, tailored to the specific environment of the participant.

The Defense Resources Management Institute specializes in four broad educational areas:

- Economic analysis in the public sector
- Operations research of public sector decision problems
- Public budgeting and fiscal management
- Defense organizations and management methods

For more information about our educational and technical assistance activities, please visit our website at **http://www.nps.navy.mil/drmi** or email us at drmiadmin@nps.edu.

# Are Errors in Official U.S. Budget Receipts Forecasts Just Noise?

Robert M. McNab, Mark Rider, Kent D. Wall

August 5, 2005

**Abstract**

Existing evidence suggests that U.S. Government budget receipts forecasts are unbiased and efficient. Our study is an attempt to examine the veracity of these findings. The time series framework employed in this study is distinguished from previous work in three ways. First, we build a model that explicitly admits serial correlation in the residuals by allowing for autoregressive, moving-average, serial correlation. Second, we employ the nonparametric Monte-Carlo bootstrap to free ourselves from reliance on asymptotic distribution theory which is suspect given the short data series available for this study. Third, we control for errors in the macroeconomic and financial assumptions used to produce the U.S. Government's budget forecasts. We find that the U.S. Government's annual, one-year ahead, budget receipts forecasts for fiscal years 1963 through 2003 are biased and inefficient. In addition, we find that these forecasts exhibit serial correlation in their errors and thus do not efficiently exploit all available information. Finally, we find evidence that is consistent with strategic bias that may reflect the political goals of the Administration in power.

## 1 Introduction[1]

Government revenue forecasts are intended to provide information about the fiscal resources available to government for some future time period. However, revenue forecasts, like all predictions of uncertain events, are subject to errors. In fact, the Office of Management and Budget's (OMB) revenue forecasts were overly optimistic throughout much of the 1980s, overly pessimistic during much of the 1990s, and once again overly optimistic during the first part of the 2000s.[2] Since estimates of government revenues are driven by forecasts of uncertain events, this pattern of forecast errors could be the result of accidental errors. In contrast, some observers contend that the OMB's revenue forecasts were deliberately biased during this period to further the policy goals of the administration responsible for preparing the federal government's annual budget. More specifically, the optimistic forecasts of the 1980s and 2000s coincided with the Reagan and Bush II administrations and, according to this view, were useful in garnering support for Republican sponsored increases in defense spending and tax cuts. Likewise, the period of pessimistic forecasts coincided

[1] We would like to thank Kathleen Bailey, Aparna Krishnamoorthy, and Peter Oburu for outstanding research assistance. The views expressed in this paper are solely those of the authors and do not reflect the views of any institutions with which the authors are affiliated.

[2] At the federal level, the Office of Tax Analysis in the Department of Treasury prepares revenue forecasts that the OMB uses in the development of the president's budget. The Bureau of the Budget in the Department of Treasury was reorganized into the OMB in 1970. For ease of reference, we refer throughout this paper to the OMB's revenue forecasts and OMB's revenue forecasting record.

with the Clinton administration and, again, according to this view, furthered the deficit reduction agenda of the Democratic party.

Whether accidental or deliberate, forecasting errors compromise the accuracy, reliability, and trustworthiness of revenue forecasts. Deliberate forecasting errors would be particularly harmful to the budget process. If, for example, the public is unable to distinguish accidental forecasting errors from deliberate ones, elected officials may be able to postpone politically risky decisions to cut expenditures or raise taxes to address fiscal imbalances and avoid political responsibility for any resulting growth in budget deficits by attributing 'unexpected' budget shortfalls to accidental forecasting errors.[3] In short, accurate revenue forecasts play an important role in the budget process by contributing to transparency in government operations, political accountability, and fiscal discipline.

Given the importance of accurate revenue forecasts to the budget process, a number of papers examine the quality of OMB's short-turn revenue forecasting record. Auerbach (1999) finds that the OMB's revenue forecasting errors during the period from FY 1986 through FY 1999 have such large standard errors that it is difficult to conclude that the forecasts were biased during this period. Similarly, Plesko (1988), Blackley and DeBoer (1993), and Campbell and Ghysels (1995) find little evidence of bias in the OMB's short-run revenue forecasts.[4] If these findings are correct, then there is little reason to believe that the OMB's revenue forecasting errors are the result of deliberate bias. Paraphrasing Geweke and Feige (1979), an econometric procedure should be powerful enough to reject the hypothesis of rationality, and it should provide some indication of why the hypothesis is not true. The econometric procedures that we propose and employ are more efficient than most of the traditional techniques that have been used in the study of forecast quality.[5] In light of the importance of accurate revenue forecasts in promoting fiscal discipline and growing concerns about U.S. federal budget deficits, it seems like an opportune time to re-evaluate the OMB's revenue forecasting record, especially now that more data are available. In fact, using nonparametric tests recommended by Campbell and Ghysels and parametric tests for the period FY 1964 through FY 2003, we reject the null hypothesis of mean zero forecast errors at conventional significance levels. This finding allows us to pursue far more interesting questions regarding the source of the bias. Exploring such hypotheses, however, requires a parametric approach that is robust to short time

[3] Ehrbeck and Waldman (2001) provide a formal model that attempts to explain why professional forecasters with incentives to be accurate would produce systematically biased forecasts. In their model, forecasters must balance their aims of minimizing forecast errors and looking good before the outcome is observed by, for example, minimizing forecast revisions. To the best of our knowledge, there has been no attempt to model the 'government budget forecasting game.' However, Blackley and DeBoer (1993) offer some thought provoking speculations in this regard. This would be a fruitful topic for future research.

[4] Feenberg, Gentry, Gilroy, and Rosen (1989) and Gentry (1989) also use a regression based test to assess the quality of state revenue forecasts. They report evidence of downward bias perhaps due to the statutory requirement for the states in question to submit balanced budgets. However, their inferences also are based on asymptotic standard errors that may not be valid given the short times series available for their analysis.

[5] Although intriguing, these studies are not entirely convincing for a variety of technical reasons. Auerbach (1999) uses a regression based test of unbiasedness, but only 14 observations were available for analysis, which means that the asymptotic standards errors on which his inferences are based may not be valid. Plesko (1988) uses a t-test to evaluate the hypothesis of mean-zero forecasting errors or unbiasedness. For a t-test to be valid, however, the series must consist of mutually stochastically independent random variables. This assumption is not satisfied in these data because the actual and forecasted revenue series are serially correlated. Blackley and DeBoer (1993) use a regression based test of unbiasedness, but their results are also suspect because of unaccounted for residual serial correlation and a short time series casting doubt on inferences based on asymptotic standard errors. In contrast, Campbell and Ghysels (1995) use nonparametric tests that are robust to short time series and serial correlation.

series and accounts for residual serial correlation.

In this paper, we look for evidence of politically motivated bias in the OMB's short-run revenue forecasts of the sort described above. Specifically, using a regression based test of weak rationality [Hansen (1980), Brown and Maital (1981), and Zarnowitz (1995), hereafter HBMZ], we evaluate the OMB's one-year ahead, proposed law, budget receipts forecasts for the period beginning in FY1964 and ending in FY 2003. We employ the nonparametric Monte-Carlo bootstrap to avoid appealing to asymptotic theory and model the serial correlation of the error structure to avoid issues of parameter bias and inefficiency. Even after controlling for accidental forecasting errors due to unanticipated economic and financial shocks, we still find evidence of systematically biased revenue forecasts. Furthermore, the direction of the bias during successive administrations over the past twenty-five years conforms to the policy goals of those administrations. These findings suggest that recent administrations may have manipulated official revenue forecasts to further their policy objectives.

The remainder of this paper is organized as follows. In the next section, we describe a regression based test of forecast performance and the data used in this analysis. We discuss our empirical findings in the third section. We conclude with a summary of our findings.

## 2 Model Structure and Data

Having established the importance of accurate, reliable, and trusted revenue forecasts, we need a more precise though perhaps less prosaic formulation of forecast quality. One formulation that is often used, especially in the theoretical literature, is full rationality. Full rationality implies that all available information has been used in an optimal manner in producing the forecast. A related concept is completeness which requires that all available information is actually used in producing the forecast, though perhaps not in an optimal manner. Therefore, completeness is a necessary condition for full rationality. It is often impossible to evaluate a forecast for completeness because we usually do not have a full understanding of all relevant information available at the time the forecast was constructed. HBMZ describe a simple regression based test of forecast quality that does not require any knowledge of the available information set.

Following HBMZ, suppose the one-year ahead forecast of federal revenues, x(t), for fiscal year t constructed at time t-1 is not complete. In other words, suppose x(t) is based on the information set $S_{t-1}$, where $S_{t-1}$ is a proper subset of all relevant information ($I_{t-1}$) available at time t-1, or $S_{t-1} \subseteq I_{t-1}$. A forecast is weakly rational when predictions make optimal use of $S_{t-1}$ or $x(t) = E[y(t)|S_{t-1}]$.[6] Now, let the dependence of x(t) on the information set used to construct the predictions at time t-1 be represented as $x(t) = P_t(S_{t-1})$.[7] The condition of weak rationality implies $x(t) = E[y(t)|x(t)]$. If x(t) possesses this property, then x(t) is said to be an unbiased estimate of y(t). Unbiasedness is thus a necessary condition for weak rationality. This leads to a test of weak rationality based on the following regression equation:

$$y(t) = b_{10} + b_{20}x(t) + e(t). \tag{1}$$

[6] Optimal, in the sense being used here, means that the forecast minimizes the mean squared error associated with the forecast. See, for example, Hamilton (1994) for a proof that the conditional expectation $E[y(t)|S_{t-1}]$ minimizes the quadratic loss function as claimed.

[7] We assume the relation is time invariant, so the same mechanism is used each period to form predictions from the subset of $I_t$.

Under the null hypothesis of weak rationality, it is easily verified that E[e(t)|x(t)] = 0. This is the key requirement for the ordinary least squares estimator of $b_{10} = 0$ and $b_{20} = 1.0$ to be consistent. Thus, (1) provides a test of weak rationality whereby we reject the hypothesis of unbiasedness and by implication weak rationality if regression analysis of (1) leads to rejection of the joint hypothesis $b_{10} = 0$ and $b_{20} = 1.0$. However, OLS estimation of (1) may give inconsistent estimates of the covariance matrix because the residuals e(t) may be serially correlated. The reason being that in the present context y(t) is not known when OMB prepares its one-year ahead forecast of revenues; hence the prior period forecast errors e(t-1) = y(t-1) - x(t-1) may not be known when x(t) is being constructed.[8] Since e(t-1) may not be part of the available information set at time t-1, we cannot rule out the possibility that E[e(t)|e(t-1)] $\neq$ 0 or that cov[e(t), e(t-1)] $\neq$ 0. Thus, the estimation procedure must allow for the possibility that the one-year ahead forecast errors are generated by a lower order moving average process. Since the regressors in each case are not strictly exogenous, generalized least squares (GLS) is likely to yield inconsistent coefficient estimates. The reason [see Hansen (1980) and Brown and Maital (1981)] is that in effect GLS transforms the model to eliminate the serial correlation in the residuals. But the transformed residuals for some particular period are linear combinations of the original residuals with their lagged values. These in turn are likely to be correlated with the transformed data for the same period because these include current values of the variables in the information set. Since neither OLS or GLS is likely to be valid, we propose using maximum likelihood to estimate the parameters of (1) taking into account the correlation structure of the error term.

Assume that the stochastic processes {y(t)} and {x(t)} are jointly stationary and ergodic, then the residuals {e(t)}, as defined above, are covariance stationary. Based on the preceding discussion, we can write

$$cov[e(t), e(t-s)] = \left\{ \begin{matrix} \sigma^2\theta_s, s = 0 \text{ or } 1, \\ 0, s > 1; \end{matrix} \right\}. \tag{2}$$

A covariance matrix consistent with (2) results when the residuals {e(t)} are generated by a first order moving average process. Therefore, in the present context weak rationality and completeness are fully consistent with

$$e(t) = \varepsilon(t) + \rho_1\varepsilon(t-1) \tag{3}$$

where $\varepsilon$(t) is a white-noise error process, perhaps Gaussian. It follows that our empirical approach is to reduce the unexplained error $\varepsilon$(t) to white noise and estimate the model using full information maximum likelihood. The resulting estimates of the coefficients are consistent and efficient, and our test statistics should be uniformly most powerful. To do so, we include variables that account for accidental forecasting errors due to unexpected economic and financial shocks, innovations in the budget process, and politically motivated forecasting errors.

[8] For example, the *FY 2006 Budget of the United States Government* was prepared in the first quarter of FY 2005, which runs from October 1 through December 31, 2004. Therefore, the forecasting errors for FY 2005 were not available when the one-year ahead forecast (FY 2006) was being prepared. The point being that the current period forecasting errors are not in the available information set of the one-year ahead forecast. This may result in an MA(1) error structure, and the estimation procedure should allow for this.

## 2.1 Source Data and Constructed Series

The OMB constructs two revenue forecasts, current law and proposed law revenues. We use the one-year ahead, proposed law, revenue forecast series to make our results comparable to those of previous studies. Further, the Congress generally accepts the administration's aggregate, proposed law, revenue targets, though they often modify the details of the administration's tax proposals. Therefore, we believe that the proposed law revenue forecast provides a superior prediction of expected revenue than the current law forecast. As previously discussed, the sample period begins in FY 1964 and ends in FY 2003.[9] While an uninterrupted series of actual and forecasted federal government revenues extends as far back as FY 1909, information regarding the economic assumptions used to construct the revenue forecast are unavailable before FY 1964.[10] The actual federal revenue series comes from the historical tables of the FY 2004 *Budget of the United States Government*, while the one-year ahead revenue predictions come from the respective FY Budgets.

We also construct an unexpected GDP series and an unexpected net long-term capital gains realizations series because, as Parcell (1999) reports, unexpected growth in GDP and unusually high capital gains realizations account in large part for the 'unexpected' federal government revenue surges in the mid-1990s. We define unexpected GDP in year t as *UGDP(t) = AGDP(t) - PGDP(t),* where AGDP(t) is actual GDP in CY t and PGDP(t) is predicted value for year t constructed in year t-1. Prior to FY 1993, we employ actual and unexpected GNP in the construction of this series because that is the variable reported in the FY Budgets. Similarly, we define unexpected net long-term capital gains realizations in year *t* as *UCG(t) = ACG(t) - PCG(t),* where A*CG(t)* is actual net long-term capital gains realizations in year t and PCG(t) is the predicted value in year t constructed in year t-1. Actual data for GNP, GDP, and the implicit GDP deflator are for the calendar years corresponding to the fiscal year predictions. These data are available from the Bureau of Economic Analysis (BEA) and net long-term capital gains realizations are available from the Statistics of Income Division of the Internal Revenue Service. It is important to note that the BEA frequently revises GDP (GNP) figures several years after the fact. This practice raises the interesting question of whether to use preliminary or revised GDP (GNP) figures. While the administration's forecast of economic conditions, particularly GDP (GNP) and its components, is by necessity based on preliminary figures for the most recent years, actual revenues in a given year depend on the true state of the economy in that year. Consequently, we use the revised figures of GDP (GNP) because these are presumably the best available measures of the true state of the economy in any given year. The fact that the BEA revises GDP (GNP), even years after the fact, means that errors in the GDP (GNP) series may not be known to the revenue forecaster for several years into the future. This may be important for reasons discussed in greater detail below. Finally, data on net long-term capital gains realizations are only available through CY 2003. All series except the implicit GNP/GDP deflator and the capital gains series are expressed in billions of nominal dollars. The base year of the implicit GNP/GDP deflator is CY 2000, and the capital gains series is expressed in millions of nominal dollars.

Since the President's FY Budget does not report the PCG(t) series used in the construction

[9] The start of the federal government's fiscal year was moved 1 July to 1 October with the FY 1976 Budget, which was released in Feburary 1975.

[10] As an interesting aside, the Treasury Department declined to provide a revenue forecast in the FY 1908 Budget proclaiming the futility of the exercise due to the uncertainty surronding federal government revenues. To the best of our knowledge, this is the only time since the federal government began issuing annual budgets in FY 1805 that the budget has not provided a one-year ahead revenue forecast.

of the OMB's federal government revenue forecast, we use a methodology described in Miller and Ozanne (2000) to construct such a series. Based on discussions with current and former members of the Treasury Department's revenue forecasting staff, we believe that this methodology yields consistent estimates of the predictions used to construct the OMB's revenue forecasts. Before describing the methodology, however, it is necessary to explain an adjustment that we made to the actual capital gains realizations series. By way of background, there was a spike in capital gains realizations in CY 1986 in response to the announced increase in the marginal tax rates on net long-term capital gains realizations associated with the Tax Reform Act of 1986 and taking affect in CY 1987. Since our methodology for constructing forecasted realizations is based on a ten-year moving average of actual realizations, as explained in greater detail below, failing to adjust the actual series for this one off event would distort our forecasted realizations series for the ten-year period subsequent to CY 1986. We adjust the actual realizations series by first regressing realized capital gains on a constant and a 0-1 dummy variable, equal to 1 in CY 1986 and zero otherwise. The estimated coefficient of the 1986 dummy variable accounts for the anticipatory behavior of investors. We then multiply this estimated coefficient by realized gains in CY 1986 to estimate the absolute magnitude of the spike. Then we subtract our estimate of the spike from actual CY 1986 realizations to obtain an adjusted or smoothed realizations series. We use this adjusted series to construct the UCG(t) series used in our analysis. Following Miller and Ozanne, we construct a forecasted series of capital gains realizations by first computing the ten-year historical moving average ratio of actual net long-term capital gains realizations and actual GDP for every year in our sample, or

$$r(t) = \frac{1}{10} \sum_{j=t-1}^{j=t-10} \frac{CG(j)}{GDP(j)}. \tag{4}$$

Then, we compute PCG(t) as the product of r(t) and PGDP(t), where the predicted GDP series is obtained from the corresponding FY Budgets. As before, we employ GNP prior to FY 1993.

For the reader's convenience, we provide the source data and constructed series in Table 1. The series used in our analysis are displayed in Figure 1. It is worth commenting on some of the more salient features of the data that are evident in Figure 1. As previously remarked, OMB's revenue forecast generally lies above actual revenues in the 1980s, below actual revenues in the 1990s, and above actual revenues in the early part of the 2000s. Further, the actual and forecasted revenue series appear to be nonstationary. The bull markets in the latter part of the 1960s, mid-1980s, and the latter part of the 1990s are evident in the graph of unexpected capital gains. Finally, the unexpected GDP series is greater than zero in nearly every year in our sample, but there are large downward spikes in this series in recession years, most notably FY 1982, FY 1991, and FY 2001.[11] These downward spikes are dramatic illustrations of the failure of official GDP forecasts to account for economic recessions.[12]. As one would expect, the overly optimistic forecasts of GDP in recession years is reflected in Figure 1 as upward spikes in forecasted revenues.

[11] Given the downward bias of projected GDP, one might expect that projected receipts, which are, in part, based on projected GDP, would also exhibit downard bias. As projected receipts do not appear to exhibit such bias, an unanswered question is how the apparent bias in projected GDP is adjusted by forecasters. We leave this question to future research.

[12] The government's economic forecast which drives the revenue forecast is produced by the Council of Economic Advisors, Department of Treasury, and Office of Management and Budget or the troika in the parlance of Washington, D.C. A number of studies have evaluated the quality of the troika's economic forecasting record, including Kamlet, Mowery, and Su (1987), Belongia (1988), Plesko (1988), and McNees (1995).

Before turning to the discussion of our results, we briefly describe the estimation algorithm used in our analysis. We employ a Box-Jenkins (1976) framework for our parametric examination of the intertemporal correlation structure of *e(t)*. We argue that the Box-Jenkins approach is most appropriate for evaluating the hypotheses of interest because the transfer function form can accomodate exogenous explanatory variables as well as autoregressive moving average (ARMA) error structures. For the parametric component of our study, we employ the Gaussian Maximum Likelihood (GML) estimator, seeking the parameter vector yielding a minimum of the negative of the Gaussian log-likelihood function. Minimization of the negative log-likelihood is done iteratively by the Davidon-Fletcher-Powell (DFP) algorithm. Now, we proceed with a discussion of our empirical investigation in the following section.[13]

# 3 Estimation Results

We begin by noting that the actual federal government revenue series, *y(t)* and the forecasted revenue series, *x(t)* appear to be nonstationary (Figure 1). The OMB forecast errors also appear to exhibit serial correlation, heteroskedasticity, and bias over short intervals (Figure 2). While these observations are certainly suggestive, we submit them to rigorous tests below.

## 3.1 Model 0

We proceed by evaluating the weak rationality of the revenue forecast series using the Brown-Maital-Zarnowitz regression based test. We begin by estimating a restricted regression that only incorporates a constant term or

$$y(t) = b_{10} + x(t) + \varepsilon(t). \tag{5}$$

Consistent with the null hypothesis of unbiasedness, we explicitly assume that the slope coefficient is equal to 1.0. If the null hypothesis is true, imposing this restriction increases the efficiency of the test. As reported in Table 2, we cannot reject the null hypothesis that the estimated intercept is equal to zero at conventional significance levels. This finding is consistent with those of previous studies that find no evidence of bias in the OMB's one-year ahead, revenue forecasts. As expected from the preceding discussion, the residual autocorrelation and partial autocorrelation functions suggest the presence of an MA(1) process, suggesting that a more appropriate form of the model is

$$e(t) = y(t) - x(t) = b_{10} + [1 + \theta_1 L]\varepsilon(t). \tag{6}$$

Campbell and Ghysels (1995) suggest using nonparametric statistics to investigate the relationship between $x(t)$ and $y(t)$.They contend that short data series make suspect any approach like Box-Jenkins that is based on asymptotic arguments. Further, they argue that heteroskedasticity, which one may infer from Figures 1 and 2, and a lack of compelling reasons to support distributional assumptions are additional factors favoring a nonparametric approach. Therefore, we use the nonparametric tests to evaluate each model. Table 3 presents the probability values for the sign and signed rank tests for each of the first three models. Regarding the present case, we find

[13] DFP is a quasi-Newton method using a numerical approximation to the inverse Hessian of the negative log-likelihood function, updated at each iteration. It exhibits quadratic convergence near a local minimum. All gradient information is generated numerically. We employ this approach to treat linear and nonlinear parameterizations with equal ease. See, for example, Hamilton (1994) for a more detailed discussion of the DFP algorithm, and Shanno (1970) for further details on the BFGS method.

that these tests corroborate the information conveyed by our residual diagnostics, i.e., the residual autocorrelation and partial autocorrelation functions.[14] These tests indicate nonzero correlation for lag-1 with the signed rank test providing the more forceful message. There is also evidence of a statistically significant lag-6 correlation. We set this finding aside for the time being, however, as there appears to be little reason to include such a long lag correlation in our modeling effort at this stage of the investigation.

It is worth noting that we have almost twice the number of observations as Campbell and Ghysels; we work in levels rather than growth rates; and we make no distributional assumptions. Regarding the latter point, we use GML estimation and take comfort in its desirable properties in many non-Gaussian environments.[15] Finally, we do not rely on asymptotic arguments for our hypothesis testing because we conduct our tests using a nonparametric bootstrap, as discussed in greater detail below.

### 3.2 Model 1

The next model removes the restriction $b_{20} = 1.0$ to see if this accounts for the suggested MA(1) error process. We thus estimate the following simple OLS model:

$$y(t) = b_{10} + b_{20}x(t) + \varepsilon(t). \tag{7}$$

Again, the results given in Table 2 appear to justify the assertion that OMB forecasts are unbiased. The intercept is not statistically significantly different than zero at conventional levels, and the slope coefficient is not statistically significantly different than unity at conventional levels, at least insofar as inferences based on asymptotic standard errors are valid. However, our residual diagnostics and the nonparametric tests reported in Table 3 reveal the presence of significant residual correlation consistent with an MA(1) error process. Thus, this OLS model is not valid and, therefore, is not a reliable framework within which to conduct hypothesis tests. In fact, little has changed relative to our previous effort. Relaxing the restriction on the slope coefficient does nothing to remove the residual correlation: an MA(1) error structure appears necessary. As previously discussed, knowledge of e(t-1) is not available when x(t) is being constructed. Consequently, an MA(1) error structure is not inconsistent with completeness or full rationality of the forecast, much less weak rationality.

### 3.3 Model 2

Model 2 elaborates on Model 1 by adding an MA(1) error to the OLS specification:

$$y(t) = b_{10} + b_{20}x(t) + [1 + \theta_1 L]\varepsilon(t). \tag{8}$$

The model parameter estimates given in Table 2 continue to indicate no statistically significant intercept and a slope coefficient that is not significantly different from unity, both at conventional significance levels. However, the story regarding the residuals has changed. The estimate of $\theta_1$ is more than seven times its asymptotic standard error, and our residual diagnostics indicate that

[14] Due to space considerations, we do not present the autocorrelation and partial autocorrelation functions. These are available upon request from the authors.

[15] See, for example, Hamilton (1994) for a guide to the literature and a discussion of the properties of GML estimation when the density is misspecified.

we have eliminated all serial correlation. Furthermore, the nonparametric tests of Table 3 provide further corroboration of this conclusion. Except at lag-8 the message is the same. As before, we find little to justify attempting to fit such a long lag term, but we keep this finding in mind for subsequent evaluation.

### 3.4 Model 3

Figure 3 provides line graphs of the residuals of the various specifications estimated thusfar. As evidenced in Figure 3, there are several salient features of the residuals from Model 2 that require our attention. There are relatively large *negative* residuals in FY 1982, FY 1990, FY 1991, and FY 2000-2002. All these years correspond to recessions. Furthermore, there are persistent *positive* residuals from FY 1995 through FY 2000. These years correspond to the boom in the equity markets. Both the positive and negative residuals in these years can be attributed to *unexpected* economic and financial shocks; phenomena that are impossible for the OMB to predict. To the extent that these unforeseen events contribute to the model residuals, they should be taken into account in our modeling effort. Therefore we expand the MA(1) model to include $UGDP(t)$ and $UCG(t)$:

$$y(t) = b_{10} + b_{20}x(t) + b_{30}UGDP(t) + b_{40}UCG(t) + [1 + \theta_1 L]\varepsilon(t). \tag{9}$$

The estimates of Model 3, which are reported in Table 2, indicate that both additional explanatory variables are statistically significant at conventional levels, and the residual variance decreases by more than 40 percent. Also, the intercept is now statistically significantly different from zero at conventional levels, but the slope coefficient remains indistinguishable from unity also at conventional levels. The MA(1) coefficient is now only marginally significant at conventional levels. The residual correlation structure indicates that there may be statistically significant serial correlation at lag-3, however, the nonparametric correlation tests indicate that this is not significant at conventional levels. Clearly, the addition of the unexpected effects has altered the situation. This suggests that omitted variables may completely explain the remaining residual serial correlation. It is possible that Model 3 is, in reality, a model still subject to omitted variable problems where the variables omitted represent the political goals of each administration. Thus we turn to an investigation of the potential influence of strategic political considerations on the construction of OMB revenue forecasts.

### 3.5 Presidential Administration Effects

Figure 3 provides a useful perspective on our modeling efforts thus far. The journey from Model 1 through Model 3 shows progressive improvement in the residuals with each elaboration of the model. At each stage we introduce variables that account for certain behavior in the residuals. The residuals of Model 3 are no less informative. As evident in Figure 3, there is a sustained series of negative residuals from FY 1981 through FY 1986 and a sustained series of positive residuals from FY 1995 through FY 1998. This brings to mind the previous discussion about an administration using deliberate forecasting errors to garner support for their policy goals. More specifically, the sustained negative series of residuals clearly result from $x(t)$ values that are too large or an overly optimistic forecast. Similarly, the sustained series of positive residuals result from $x(t)$ values that are too small or an overly pessimistic forecast. The timing of both sequences of residuals brings to mind the Reagan and Clinton administrations. An effort to justify increases in defense spending and tax reductions could have been well served by overly optimistic revenue predictions, and efforts to

rein-in government spending could have been well served by overly pessimistic revenue projections. Model 3 is complete enough to serve as the basis for testing of these hypotheses.

To test for such influences, we construct a set of 0 - 100 dummy variables for each administration over the sample period. For example, we represent the Reagan presidency by a series with zero values from FY 1963 through FY 1981, then with ones from FY 1982 through FY 1989, and zeros again from FY 1990 through FY 2003. Table 4 presents the estimation of Model 2 with these 'presidential dummy variables' as each is inserted in the model as $z_5(t)$. To save space, we use the following abbreviations: Ke = Kennedy, Jo = Johnson, Ni = Nixon/Ford, Ca = Carter, Re = Reagan, B1 = Bush I, Cl = Clinton, and B2 = Bush II to identify the different models in Table 4.

The $b_{50}$ estimates reported in Table 4 indicate that the Clinton and Bush II administrations have the greatest impact. The Reagan administration appears to have a much weaker effect. While weak, this effect has a sign consistent with a desire to overstate revenues for the purposes of creating the appearance of greater fiscal space for a defense build-up and tax cuts than subsequent facts would show to be the case. The Reagan administration effect may be diluted by the Gramm-Rudman-Hollings Act which went into effect in FY 1986. The UGDP and UCG effects are present in all cases, and their coefficients are relatively stable except in the Clinton and Bush II administrations. The MA(1) error is not as well defined in all but the Bush II administration. It appears that there are presidential administration effects which may affect the serial correlation structure of the residuals. Taken together, these findings suggest that the Reagan, Clinton, and Bush II effects should be included in the model. Furthermore, the Reagan effect requires inclusion of a dummy variable representing the Gramm-Rudman-Hollings (GRH) Act.

### 3.6 Model 4

To pursue the issues raised above, we re-estimate Model 3 and include three presidential effects: Reagan, Clinton, and Bush II. We also introduce a 0-100 dummy variable representing the GRH Act. This dummy variable is zero for all years except FY 1986 where it takes on the value 100. GRH was in effect for only one fiscal year before it was declared unconstitutional by the Supreme Court. It was amended in 1987 and no longer had the original effect or potency.

Model estimation with the expanded set of explanatory variables is summarized in Tables 5, 6, and 7. The weakness of the MA(1) residual structure with presidential administration effects is tested by re-estimating the model assuming an OLS specification. The residual autocorrelations and partial autocorrelations show large values at lag-2 and lag-6 but little correlation at lag-1. These conclusions are further corroborated by the nonparametric test results reported in Table 6. Estimation of the MA(1) specification reinforces these findings: $\theta_1$ is not statistically significant at conventional levels (Table 5) and the nonparametric tests indicate lag-5 or lag-6 residual correlations (Table 6). These findings lead us to re-estimate the model with a restricted MA(6) specification in which $\theta_2$ and $\theta_6$ are nonzero. The model with moving-average lag-2 and lag-6 terms is selected as our baseline model which we henceforth refer to as Model 4. Model 4 is used to conduct hypothesis tests.

The estimated parameters and the p-values of the nonparametric tests of Model 4 are reported in Tables 5 and 6, respectively. Model 4 has a stronger Reagan effect than before, and the GRH dummy appears well defined. The traditional asymptotic statistics suggest Reagan, Clinton, and GRH are statistically significant at conventional levels, while Bush II is only marginally significant at conventional levels. Line graphs of the residual series for all three versions of Model 4 are provided in Figure 3.

This model controls for unexpected economic and financial shocks, strategic political effects, GRH effect, and residual serial correlation. Until now traditional asymptotic standard errors have been used to draw inferences from the sample, but this approach is suspect given that only 40 observations are being used to estimate up to 11 parameters (10 coefficients and $\sigma_\varepsilon^2$). By way of explanation for this approach, recall that our modeling effort up to now has been focused on reducing the residuals to white noise, and all inferences regarding statistical properties of the residuals have been corroborated by nonparametric tests that are robust to short time series and other violations of the asymptotic theory. Clearly, however, we are at the limits of the applicability of asymptotic theory and relying only on this theory is tenuous at best. Thus we now turn to an investigation of statistical significance based on the nonparametric bootstrap. The residuals of Model 4 comprise a sequence for which we cannot reject serial interrelatedness or exchangeability, two key requirements of the bootstrap we use for hypothesis testing.

# 4 Bootstrap Analysis

The bootstrap is a resampling technique developed by Bradely Efron (1979, 1982) that is valuable in assessing the sampling distribution of an estimator when data series are 'short' and reliance on asymptotic theory is suspect. This is important in the present context because our series consists of only 40 observations.

## 4.1 Bootstrap Confidence Intervals

The bootstrap is most valuable because of the distributional information it provides. This allows us to construct confidence levels and conduct hypothesis tests without resorting to asymptotic theory. 5,000 realizations of the estimated parameter vector $\boldsymbol{\lambda}^*(b)$ permit the construction of reliable estimates of the sampling distribution and the relative frequency histograms for the parameter estimates. Both GML estimation and the bootstrap indicate strong correlation between $b_{10}$ and $b_{20}$.Our GML estimation gives a correlation of $-0.833$, and the bootstrap estimate is $-0.825$. High correlation between $b_{10}$ and $b_{20}$ requires a joint approach to hypothesis testing which we take up in due course. First, however, we consider confidence intervals based on the marginal distribution information reported in Table 8. Confidence intervals can be established easily from the marginal distributions using the bootstrap percentile method. At the usual 95 percent confidence level we find $b_{10}$ insignificant, but we must reject $b_{20} = 1.0$ at this confidence level. All other parameters are significant at the 95 percent level. The evidence of estimated parameter bias reported in Table 7, however, requires that we consider Efron's "Bias Corrected" (BC) or "Bias Corrected Accelerated" (BCa) methods. These results are summarized in Table 9, where we give the traditional central 90 percent and 95 percent confidence intervals. Based on the confidence intervals reported in Table 7, we find little evidence to reject $b_{10} = 0$ at conventional significance levels, but we do find evidence against $b_{20} = 1.0$ at conventional significance levels. The political effects are all statistically significant at conventional levels. The estimated effects of Reagan and Bush II dummy variables are negative, while the effect of the Clinton dummy variable is positive. The respective signs of the three administration effects conform to the policy aims of these administrations, as discussed above. All results so far derive from consideration of only marginal distributions and ignore correlation among the parameters. Since both the GML estimation and the bootstrap show the coefficients to be correlated, a joint test is more appropriate – a point to which we turn to now.

### 4.2 Joint Hypothesis Testing

The strong correlation between $b_{10}$ and other parameters, most notably $b_{20}$, requires testing *joint hypotheses.* The bootstrap distribution of $\boldsymbol{\lambda}^*(b) - \widehat{\boldsymbol{\lambda}}$ is an estimate of the sampling distribution of $\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}$; where $\widehat{\boldsymbol{\lambda}}$ is our sample parameter estimate, thus the joint behavior is already available for our use. Figure 3 shows the joint sampling distribution of $b_{10}$ and $b_{20}$ from the bootstrap. Note that only the first 1,500 bootstrap values are plotted because the entire 5,000 point postscript file is too large to load into this document; the picture with all 5,000 points is even more dramatic. The null hypothesis $H_0$: $\{b_{10} = 0, b_{20} = 1.0\}$ is indicated by the crosshairs in this figure and corresponds to a point so far removed from the scatter of points that a probability value of zero appears appropriate. This picture illustrates how important information is concerning the correlation between parameter estimates. To evaluate model 4, however, the testing framework must be more than bivariate and go beyond the capabilities of a two-dimensional scatter diagram.

Testing in this situation requires a multivariate framework that allows us to employ traditional probability value computations *after* we have taken account of the multivariate correlations. This is done by transforming the data using a Cholesky factorization [see, for example, Hamilton (1994)] to diagonalize the parameter variance-covariance matrix (see Appendix 3 for details). The magnitudes of the elements of the resulting vector provide strong evidence for rejecting the joint hypothesis $b_{10} = 0$, $b_{20} = 1.0$, and Reagan = Clinton = Bush II = 0. The joint test and the estimates reported in Table 9 suggest that $b_{10} > 0$ and $b_{20} < 1.0$. This has the following interpretation: forecasters tended to overestimate the underlying trend by some constant amount. For percentage changes that exceed (fall short of) the trend, predictions thus overstated (understated) the actual increase. Hence, predicted changes were more volatile around the trend than actual changes. The estimated effects of the Reagan and Bush II dummy variables are negative which is consistent with the notion that the revenue forecasts were deliberately biased upward to garner support for increases in defense expenditures and tax cuts by suggesting that there is greater fiscal space for such actions than optimal use of all available information might otherwise support. A positive Clinton effect suggests that pessimistic revenue forecasts would further the policy objective of deficit reduction by suggesting that there was less fiscal space than an optimal use of all available information would support. Systematic bias could emerge for technical problems in forecasting, such as erroneously specified models, but it should not vary with the administration in power. That the algebraic signs of the administration effects conform with the policy aims of successive administrations, even after controlling for unexpected economic and financial shocks, provides evidence in support of the hypothesis that revenue forecasts were deliberately biased for strategic reasons during the past twenty-five years.

As previously noted, the GRH Act provided for automatic spending cuts, if the president and Congress failed to reach specified targets according to the OMB's budget forecasts. A positive GRH effect suggests an administration seeking to reduce the chances that the president and Congress would not achieve the targets triggering automatic expenditure cuts. As many have pointed out, *ex ante* deficit rules, like the GRH Act and for that matter some balanced-budget rules, are not a substitute for a political commitment to fiscal discipline. Rather, such rules may simply give rise to budget gimmicks like inflated revenue forecasts to give the appearance of fiscal restraint. Finally, the estimated coefficients of the lag terms are negative and rather small. It is difficult to attach special meaning to these particular lags, except to note that information about forecasting errors 2 and 6 years in the past are available at the time that the forecast is constructed. As previously noted, however, the BEA may revise GDP (GNP) figures several years after the fact, and perhaps

the indicated lag structure is an artifact of such revisions.

# 5 Summary

Since any pattern of forecast errors could simply reflect accidental forecasting errors, it is difficult to provide convincing evidence of deliberate and politically motivated bias in the OMB's short-run revenue forecasts. In contrast to previous studies, we find strong evidence for rejecting the hypothesis of unbiasedness and thus weak rationality of the OMB's short-run revenue forecasts. A nonparametric bootstrap approach to evaluate the joint significance of the parameter estimates soundly rejects the hypothesis of unbiasedness or $b_0 = 0$ and $b_1 = 1.0$ and, consequently, weak rationality. Even when we obtain optimal least-squares estimates of $b_0$ and $b_1$, the basic model, $e(t) = y(t) - b_0 - b_1 x(t)$, is not supported by the data. We continue to find evidence of serial correlation in our model even after we have controlled for unexpected economic and financial shocks, the GRH Act, and administration effects. This serial correlation is evident in both a parametric, Box-Jenkins approach, *and* nonparametric Campbell-Ghysels approach. The form of the serial correlation is important because information about forecast errors at lag 2 and lag 6 is available at the time the forecast is constructed. Thus, the evidence of MA(2) and MA(6) error processes in our baseline model are not consistent with completeness. In other words, there is information in the forecast errors that could be used to improve the forecast. The elaborations on the model that include variables representing unexpected events have an effect on the form of the serial correlation in the residuals but does not serve to eliminate it. Though it may come as little or no surprise to some, we find evidence of administration effects in the OMB's revenue forecasts, even after controlling for unexpected economic and financial shocks. The direction of the administration effects conform to the policy aims of successive administrations over the past twenty-five years. This is consistent with the hypothesis that the OMB's short-run revenue forecasts are deliberately biased to further the policy aims of an administration. Based on this evidence, formal modeling of the dynamics of the budgeting process may be an interesting and fruitful area for future research. To the extent that the current budget process has an inherent bias favoring deficit spending or growth in the size of government, as some contend, then formal modeling and further empirical research may suggest reforms of the budget process that would foster greater fiscal discipline.

# 6 References


1. Auerbach, Alan J., "On the Performance and Use of Government Revenue Forecasts." *National Tax Journal*, December 1999, 52 (4), pp. 767-82.

2. Belongia, Michael T., "Are Economic Forecasts by Government Agencies Biased?" Federal Reserve Bank of St. Louis Review, November-December 1998, 70: 15-23.

3. Blackley, Paul R. and Larry DeBoer, "Bias in OMB's Economic Forecasts and Budget Proposals." *Public Choice*, July 1993, 76, pp. 215-232.

4. Box, G. E. P. and G. Jenkins, *Time Series Analysis,* revised edition, San Francisco: Holden-Day, 1976.

5. Brown, Bryan W. and Shlomo Maital. "What Do Economists Know? An Empirical Study of Experts' Expectations." *Econometrica*, March 1981, 49(2), pp. 491-504.

6. Campbell, Bryan and Eric Ghysels, "Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency." *The Review of Economics and Statistics*, February 1995, 77 (1), pp. 17-31.

7. Efron, Bradely, "Bootstrap Methods: Another Look at the Jackknife", *Annals of Statistics, 1979,* Vol. 7, No. 1, pp 1-26.

8. Efron, Bradely, *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: SIAM, 1982.

9. Ehrbeck, Tilman and Robert Waldmann, "Why are Professional Forecasters Biased? Agency Versus Behavioral Explanations," *The Quarterly Journal of Economics*, February 1996, pp. 21-40.

10. Feenberg, Daniel R., William Gentry, David Gilroy, and Harvey S. Rosen, "Testing the Rationality of State Forecasts." *The Review of Economics and Statistics*, May 1989, 71 (2), pp. 300-08.

11. Gentry, William M., "Do State Revenue Forecasters Utilize Available Information?" *National Tax Journal*, December 1989, 42 (4), pp.429-39.

12. Geweke, Joh and Edgar Freige, "Some Joint Tests of the Efficiency of Markets for Forward Foreign Exchange." *The Revenue of Economics and Statistics*, August 1979, 61, pp. 334-41.

13. Hamiltion, James D., *Time Series Analysis*, Princeton: Princeton University Press, 1994.

14. Hansen, Lars Peter and Robert J. Hodrick, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." *The Journal of Political Economy*, October 1980, 88 (5), pp. 829-53.

15. Kamlet, Mark C., David C. Mowery, Tsai-Tsu Su. "Whom Do You Trust? An Analysis of Executive and Congressional Economic Forecasts." Journal of Policy Analysis and Management, April 1987, 6(3), pp. 365-84.

16. Miller, Preston and Larry Ozanne, "Forecasting Capital Gains Realizations." Technical Paper Series, 2000-5, Congressional Budget Office, August 2000

17. McNees, Stephen, "An Assessment of the "Official" Economic Forecasts," *New England Economic Review*, July/Agust 1995, pp. 13-23.

18. Parcell, Ann D., "Challenges and Uncertainties in Forecasting Federal Individual Income Tax Receipts." *National Tax Journal*, September 1999, 52 (3), pp. 325-38.

19. Plesko, George A., "The Accuracy of Government Forecasts and Budget Projections." *National Tax Journal*, December 1988, 41 (4), pp. 483-01.

20. Shanno, D.F., "Conditioning of Quasi-Newton Methods for Function Minimization," *Mathematics of Computing*, Vol. 24, pp 647-656, 1970.

21. Stoffer, David S. and Kent D. Wall, *Journal of the American Statistical Association*, Vol. 86, No. 416, December 1991, pp 1024 - 33.

22. Theil, Henri, *Principles of Econometrics*,New York: John Wiley & Sons, 1971

23. Wall, Kent D. and David S. Stoffer, "A State Space Approach to Bootstrapping Conditional Forecasts in ARMA Models." *Journal of Time Series Analysis*, vol. 23, No. 6, Nov. 2002, pp.733-751.

24. Zarnowitz, Victor, "Rational expectations and macroeconomic forecasts." *Journal of Business and Economic Statistics* 3, October 1985, 293-311.

# 7 Appendix 1: Tables

Table 1
Source Data and Constructed Series

| Fiscal Year | Receipts | | Economic Assumptions | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Actual (CY) | | | | Forecast (CY) | | | Unexpected | | |
| | Actual | Forecast | GNP | GDP | CG | GDP Deflator | GNP | GDP | CG | U-GNP | U-GDP | U-CG |
| 1964 | 112.6 | 112.2 | 622.2 | 617.7 | 17,431 | 22.1 | 578 | | 13091.7 | 44.2 | | 4,339 |
| 1965 | 116.8 | 119.7 | 668.5 | 663.6 | 21,484 | 22.5 | 623 | | 14312.0 | 45.5 | | 7,172 |
| 1966 | 130.8 | 123.5 | 724.4 | 719.1 | 21,348 | 23.2 | 660 | | 15537.5 | 64.4 | | 5,811 |
| 1967 | 148.8 | 145.5 | 792.9 | 787.8 | 27,535 | 23.9 | 722 | | 17674.6 | 70.9 | | 9,860 |
| 1968 | 153.0 | 168.1 | 838.0 | 832.6 | 35,607 | 24.9 | 787 | | 20477.7 | 51.0 | | 15,129 |
| 1969 | 186.9 | 178.1 | 916.1 | 910.0 | 31,439 | 26.2 | 846 | | 23611.9 | 70.1 | | 7,827 |
| 1970 | 192.8 | 198.7 | 990.7 | 984.6 | 20,848 | 27.5 | 921 | | 26257.7 | 69.7 | | -5,410 |
| 1971 | 187.1 | 202.1 | 1,044.9 | 1,038.5 | 28,341 | 28.9 | 985 | | 27865.7 | 59.9 | | 475 |
| 1972 | 207.3 | 217.5 | 1,134.7 | 1,127.1 | 35,869 | 30.2 | 1,065 | | 29681.6 | 69.7 | | 6,187 |
| 1973 | 230.8 | 220.8 | 1,246.8 | 1,238.3 | 35,757 | 31.9 | 1,145 | | 32598.2 | 101.8 | | 3,159 |
| 1974 | 263.2 | 256.0 | 1,395.3 | 1,382.7 | 30,217 | 34.7 | 1,267 | | 36350.2 | 128.3 | | -6,133 |
| 1975 | 279.1 | 295.0 | 1,515.5 | 1,500.0 | 30,903 | 38.0 | 1,390 | | 39031.2 | 125.5 | | -8,128 |
| 1976 | 298.1 | 297.5 | 1,651.3 | 1,638.3 | 39,492 | 40.2 | 1,498 | | 40431.0 | 153.3 | | -939.02 |
| 1977 | 355.6 | 351.3 | 2,051.2 | 2,030.9 | 45,338 | 42.8 | 1,890 | | 50009.4 | 161.2 | | -4,671 |
| 1978 | 399.6 | 393.0 | 2,316.3 | 2,294.7 | 50,526 | 45.8 | 2,092 | | 53115.9 | 224.3 | | -2,590 |
| 1979 | 463.3 | 439.6 | 2,595.3 | 2,563.3 | 73,443 | 49.6 | 2,335 | | 55292.8 | 260.3 | | 18,150 |
| 1980 | 517.1 | 502.6 | 2,823.7 | 2,789.5 | 74,132 | 54.1 | 2,565 | | 59892.8 | 258.7 | | 14,239 |
| 1981 | 599.3 | 600.0 | 3,161.4 | 3,128.4 | 80,938 | 59.1 | 2,842 | | 68179.6 | 319.4 | | 12,758 |
| 1982 | 617.8 | 711.8 | 3,291.5 | 3,255.0 | 90,153 | 62.7 | 3,312 | | 79686.7 | -20.5 | | 10,466 |
| 1983 | 600.6 | 666.1 | 3,573.8 | 3,536.7 | 122,773 | 65.2 | 3,524 | | 84364.6 | 49.8 | | 38,408 |
| 1984 | 666.5 | 659.7 | 3,969.5 | 3,933.2 | 140,500 | 67.7 | 3,566 | | 88543.8 | 403.5 | | 51,956 |
| 1985 | 734.1 | 745.1 | 4,246.8 | 4,220.3 | 171,985 | 69.7 | 3,974 | | 104873.9 | 272.8 | | 67,111 |
| 1986 | 769.2 | 793.7 | 4,480.6 | 4,462.8 | 143,431 | 71.3 | 4,285 | | 122465.3 | 195.6 | | 20,965 |
| 1987 | 854.4 | 850.4 | 4,757.4 | 4,739.5 | 148,449 | 73.2 | 4,629 | | 135451.3 | 128.4 | | 12,998 |
| 1988 | 909.3 | 916.6 | 5,127.4 | 5,103.8 | 162,592 | 75.7 | 4,816 | | 145257.5 | 311.4 | | 17,334 |
| 1989 | 991.2 | 964.7 | 5,510.6 | 5,484.4 | 154,040 | 78.6 | 5,113 | | 159226.2 | 397.6 | | -5,186 |
| 1990 | 1,032.0 | 1,059.3 | 5,837.9 | 5,803.1 | 123,783 | 81.6 | 5,570 | | 173179.4 | 267.9 | | -49,396 |
| 1991 | 1,055.0 | 1,170.2 | 6,026.3 | 5,995.9 | 111,592 | 84.5 | 6,002 | | 183489.8 | 24.3 | | -71,898 |
| 1992 | 1,091.3 | 1,165.0 | 6,367.4 | 6,337.7 | 126,692 | 86.4 | 6,095 | | 181944.6 | 272.4 | | -55,253 |
| 1993 | 1,154.4 | 1,169.1 | 6,689.3 | 6,657.4 | 152,259 | 88.4 | 6,319 | 6,307 | 183416.7 | 370.3 | 350 | -31,158 |
| 1994 | 1,258.6 | 1,251.3 | 7,098.4 | 7,072.2 | 152,727 | 90.3 | | 6,594 | 183982.2 | | 478 | -31,255 |
| 1995 | 1,351.8 | 1,353.8 | 7,433.4 | 7,397.7 | 180,130 | 92.1 | | 7,118 | 188637.3 | | 280 | -8,507 |
| 1996 | 1,453.1 | 1,415.5 | 7,851.9 | 7,816.9 | 260,696 | 93.9 | | 7,507 | 186559.8 | | 310 | 74,136 |
| 1997 | 1,579.3 | 1,495.2 | 8,337.3 | 8,304.3 | 364,829 | 95.4 | | 8,008 | 202922.7 | | 296 | 161,906 |
| 1998 | 1,721.8 | 1,566.8 | 8,768.3 | 8,747.0 | 455,223 | 96.5 | | 8,313 | 221125.8 | | 434 | 234,097 |
| 1999 | 1,827.5 | 1,742.7 | 9,302.2 | 9,268.4 | 552,608 | 97.9 | | 8,772 | 250879.2 | | 496 | 301,729 |
| 2000 | 2,025.2 | 1,883.0 | 9,855.9 | 9,817.0 | 644,285 | 100.0 | | 9,199 | 292068.3 | | 618 | 352,217 |
| 2001 | 1,991.2 | 2,019.0 | 10,171.6 | 10,128.0 | 349,441 | 102.4 | | 10,156 | 367444.1 | | -28 | -18,003 |
| 2002 | 1,853.2 | 2,191.7 | 10,514.1 | 10,487.0 | 268,615 | 104.1 | | 11,004 | 415621.1 | | -517 | -147,006 |
| 2003 | 1,798.3 | 2,048.1 | 11,059.2 | 11,004.0 | 288,630 | 106.0 | | 11,073 | 424428.1 | | -69 | -135,798 |

Table 2
Parameter Estimation Results
(standard errors are reported in parentheses)

| Variable | Model 0 | Model 1 | Model 2 | Model 3 |
|---|---|---|---|---|
| Intercept | -14.7 | 27.4 | 51.5 | -67.3 |
| | (13.8) | (36.4) | (50.0) | (28.2) |
| OMB Forecast | 1.0 | 0.962 | 0.941 | 0.975 |
| | | (0.030) | (0.041) | (0.020) |
| Unexpected GDP | — | — | — | 0.233 |
| | | | | (0.038) |
| Unexpected CG | — | — | — | 0.0366 |
| | | | | (0.093) |
| $\theta_1$ | — | — | 0.771 | 0.368 |
| | | | (0.095) | (0.212) |
| $\sigma_\varepsilon$ | 87.1 | 84.5 | 64.1 | 36.3 |
| | | (9.2) | (7.4) | (4.1) |
| $\ell(\lambda)$ | 235.4 | 234.2 | 223.2 | 200.4 |

Table 3
Nonparametric Residual Correlation Tests
(p-values for the sign and signed rank tests)

| Lag length | Test | Model 0 | Model 1 | Model 2 | Model 3 |
|---|---|---|---|---|---|
| 1-period | Signed | 0.06 | 0.11 | 1.00 | 1.00 |
| | Signed Rank | 0.03 | 0.04 | 0.88 | 0.79 |
| 2-periods | Signed | 0.87 | 0.87 | 0.63 | 0.87 |
| | Signed Rank | 0.90 | 0.75 | 0.42 | 0.79 |
| 3-periods | Signed | 1.00 | 1.00 | 0.51 | 0.19 |
| | Signed Rank | 0.83 | 0.94 | 0.99 | 0.14 |
| 4-periods | Signed | 0.41 | 0.87 | 0.87 | 0.87 |
| | Signed Rank | 0.41 | 0.77 | 0.63 | 0.71 |
| 5-periods | Signed | 0.18 | 0.74 | 1.00 | 0.74 |
| | Signed Rank | 0.17 | 0.40 | 0.84 | 0.22 |
| 6-periods | Signed | 0.03 | 0.61 | 0.39 | 0.23 |
| | Signed Rank | 0.04 | 0.23 | 0.16 | 0.26 |
| 7-periods | Signed | 1.00 | 0.61 | 0.39 | 0.30 |
| | Signed Rank | 0.48 | 0.23 | 0.16 | 0.93 |
| 8-periods | Signed | 0.86 | 0.86 | 0.05 | 0.38 |
| | Signed Rank | 0.50 | 0.45 | 0.08 | 0.39 |

Table 4
Parameter Estimates with Presidential Administration Effect
(standard errors are reported in parentheses)

| Variable | Ke | Jo | Ni | Ca | Re | B1 | Cl | B2 |
|---|---|---|---|---|---|---|---|---|
| Intercept | -68.2 | -87.8 | -70.5 | -68.3 | -61.1 | -68.1 | -20.0 | -79.7 |
| | (28.7) | (35.8) | (33.2) | (28.6) | (27.1) | (29.8) | (23.8) | (26.0) |
| OMB Forecast | 0.976 | 0.987 | 0.977 | 0.975 | 0.975 | 0.976 | 0.926 | 1.006 |
| | (0.020) | (0.024) | (0.022) | (0.020) | (0.018) | (0.021) | (0.018) | (0.020) |
| Unexpected GDP | 0.234 | 0.246 | 0.234 | 0.244 | 0.229 | 0.235 | 0.209 | 0.190 |
| | (0.040) | (0.042) | (0.041) | (0.041) | (0.040) | (0.040) | (0.033) | (0.036) |
| Unexpected CG | -0.365 | 0.340 | 0.368 | 0.354 | 0.391 | 0.358 | 0.246 | 0.295 |
| | (0.097) | (0.102) | (0.095) | (0.094) | (0.091) | (0.100) | (0.077) | (0.089) |
| Kennedy Dummy | -0.276 | - | - | - | - | - | - | - |
| | (1.489) | - | - | - | - | - | - | - |
| Johnson Dummy | - | 0.235 | - | - | - | - | - | - |
| | - | (0.255) | - | - | - | - | - | - |
| Nixon Dummy | - | - | 0.044 | - | - | - | - | - |
| | - | - | (0.235) | - | - | - | - | - |
| Carter Dummy | - | - | - | -0.172 | - | - | - | - |
| | - | - | - | (0.283) | - | - | - | - |
| Reagan Dummy | - | - | - | - | -0.253 | - | - | - |
| | - | - | - | - | (0.183) | - | - | - |
| H. Bush Dummy | - | - | - | - | - | -0.065 | - | - |
| | - | - | - | - | - | (0.265) | - | - |
| Clinton Dummy | - | - | - | - | - | - | 0.906 | - |
| | - | - | - | - | - | - | (0.212) | - |
| W. Bush Dummy | - | - | - | - | - | - | - | -1.308 |
| | - | - | - | - | - | - | - | (0.394) |
| $\theta_1$ | 0.367 | 0.379 | 0.356 | 0.355 | 0.239 | 0.369 | 0.161 | 0.444 |
| | (0.207) | (0.221) | (0.210) | (0.197) | (0.266) | (0.207) | (0.200) | (0.153) |
| $\sigma_\varepsilon$ | 36.2 | 35.9 | 34.3 | 36.1 | 35.4 | 36.2 | 30.2 | 31.7 |
| | (4.2) | (4.3) | (3.9) | (4.0) | (4.3) | (4.4) | (3.5) | (3.6) |
| $\ell(\lambda)$ | 200.4 | 199.9 | 200.4 | 200.2 | 199.5 | 200.4 | 193.0 | 195.1 |

Table 5
Parameter Estimates with Combined Set of Explanatory Variables
(standard errors are reported in parentheses)

| Variable | OLS | MA(1) | MA(6) |
|---|---|---|---|
| Intercept | -34.6 | -41.2 | 15.3 |
| | (19.1) | (24.2) | (8.5) |
| OMB Forecast | 0.953 | 0.960 | 0.923 |
| | (0.020) | (0.025) | (0.015) |
| Unexpected GDP | 0.195 | 0.196 | 0.134 |
| | (0.030) | (0.031) | (0.026) |
| Unexpected CG | 0.263 | 0.247 | 0.293 |
| | (0.061) | (0.073) | (0.062) |
| Reagan Dummy | -0.244 | -0.233 | -0.280 |
| | (0.128) | (0.151) | (0.123) |
| Gramm-Rudman-Hollings Dummy | 0.801 | 0.727 | 0.912 |
| | (0.285) | (0.282) | (0.280) |
| Clinton Dummy | 0.646 | 0.605 | 0.878 |
| | (0.217) | (0.252) | (0.196) |
| W. Bush Dummy | -0.611 | -0.754 | -0.610 |
| | (0.368) | (0.398) | (0.344) |
| $\sigma_\varepsilon$ | 26.7 | 26.3 | 23.9 |
| | (3.1) | (3.0) | (2.7) |
| $\theta_1$ | — | 0.232 | — |
| | | (0.256) | |
| $\theta_2$ | — | — | -0.282 |
| | | | (0.107) |
| $\theta_6$ | — | — | -0.409 |
| | | | (0.102) |
| $\ell(\lambda)$ | 188.1 | 187.6 | 183.6 |

Table 6
Nonparametric Residual Correlation Tests
(p-values for sign and singed rank tests)

| Lag length | Test | OLS | MA(1) | MA(6) |
|---|---|---|---|---|
| 1-period | Signed | 0.75 | 0.75 | 0.20 |
| | Signed Rank | 0.52 | 0.91 | 0.58 |
| 2-periods | Signed | 0.04 | 0.63 | 0.87 |
| | Signed Rank | 0.03 | 0.32 | 0.34 |
| 3-periods | Signed | 0.51 | 0.74 | 0.74 |
| | Signed Rank | 0.50 | 0.53 | 0.73 |
| 4-periods | Signed | 0.41 | 0.41 | 0.87 |
| | Signed Rank | 0.61 | 0.87 | 0.89 |
| 5-periods | Signed | 0.31 | 0.90 | 0.74 |
| | Signed Rank | 0.11 | 0.10 | 0.95 |
| 6-periods | Signed | 0.23 | 0.23 | 0.61 |
| | Signed Rank | 0.03 | 0.01 | 0.61 |
| 7-periods | Signed | 0.73 | 1.00 | 1.00 |
| | Signed Rank | 0.87 | 0.92 | 0.49 |
| 8-periods | Signed | 0.86 | 0.60 | 0.86 |
| | Signed Rank | 0.86 | 0.99 | 0.65 |

Table 7
Bootstrap and ML Estimates of Model 4
(standard errors are reported in parentheses)

| Variable | MLE | Bootstrap ($\overline{\boldsymbol{\lambda}}^*$) | $\widehat{\text{bias}}_B$ |
|---|---|---|---|
| Intercept | 15.3 | 23.8 | 8.5 |
| | (8.5) | (18.7) | |
| OMB Forecast | 0.923 | 0.916 | -0.006 |
| | (0.015) | (0.019) | |
| Unexpected GDP | 0.134 | 0.129 | -0.006 |
| | (0.026) | (0.030) | |
| Unexpected CG | 0.293 | 0.297 | 0.004 |
| | (0.062) | (0.060) | |
| Reagan Dummy | -0.280 | -0.273 | 0.007 |
| | (0.123) | (0.117) | |
| Gramm-Rudman-Hollings Dummy | 0.912 | 0.879 | -0.033 |
| | (0.280) | (0.251) | |
| Clinton Dummy | 0.878 | 0.921 | 0.043 |
| | (0.196) | (0.212) | |
| W. Bush Dummy | -0.610 | -0.582 | 0.028 |
| | (0.344) | (0.335) | |
| $\theta_2$ | -0.282 | -0.397 | -0.115 |
| | (0.107) | (0.176) | |
| $\theta_6$ | -0.409 | -0.458 | -0.049 |
| | (0.102) | (0.137) | |
| $\sigma_\epsilon$ | 23.9 | 20.0 | -3.8 |
| | (2.7) | (2.5) | |

TABLE 8
Bootstrap Quantiles for Model 4

| Percentile | $b_0$ | $b_1$ | $\theta_1$ | $\theta_2$ | $\sigma_\varepsilon^2$ | $c_{03}$ | $c_{04}$ | Re | GRH | Cl | B2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.010 | −18.5 | 0.871 | −0.822 | −0.773 | 14.4 | 0.059 | 0.156 | −0.55 | 0.33 | 0.42 | −1.34 |
| 0.020 | −13.2 | 0.877 | −0.772 | −0.737 | 15.0 | 0.067 | 0.173 | −0.51 | 0.38 | 0.48 | −1.26 |
| 0.025 | −11.3 | 0.879 | −0.754 | −0.726 | 15.3 | 0.069 | 0.178 | −0.50 | 0.41 | 0.50 | −1.23 |
| 0.050 | −6.2 | 0.885 | −0.687 | −0.680 | 16.0 | 0.079 | 0.198 | −0.47 | 0.48 | 0.56 | −1.13 |
| 0.500 | 23.6 | 0.916 | −0.392 | −0.461 | 20.0 | 0.129 | 0.297 | −0.27 | 0.87 | 0.92 | −0.58 |
| 0.950 | 55.5 | 0.947 | −0.113 | −0.234 | 24.1 | 0.176 | 0.397 | −0.08 | 1.30 | 1.26 | −0.03 |
| 0.975 | 63.4 | 0.952 | −0.058 | −0.182 | 24.9 | 0.184 | 0.414 | −0.04 | 1.38 | 1.33 | 0.06 |
| 0.980 | 65.6 | 0.954 | −0.044 | −0.161 | 25.5 | 0.185 | 0.419 | −0.03 | 1.41 | 1.35 | 0.09 |
| 0.990 | 71.5 | 0.960 | 0.007 | −0.125 | 25.7 | 0.195 | 0.435 | 0.01 | 1.47 | 1.41 | 0.18 |

TABLE 9
Bootstrap BCa Confidence Intervals for Model 4

| Variable | 90 percent | 95 percent |
|---|---|---|
| Intercept | [-17.700, 38.800] | [-21.3, 46.400] |
| OMB Forecast | [0.869, 0.933] | [0.863, 0.938] |
| Unexpected GDP | [0.061, 0.163] | [0.049, 0.170] |
| Unexpected CG | [0.181, 0.383] | [0.158, 0.400] |
| Reagan Dummy | [-0.487, -0.101] | [-0.526, -0.066] |
| Gramm-Rudman-Hollings Dummy | [0.444, 1.261] | [0.383, 1.352] |
| Clinton Dummy | [0.500, 1.199] | [0.435, 1.263] |
| W. Bush Dummy | [-1.221, -0.111] | [-1.324, -0.017] |

# 8 Appendix 2: Figures

Figure 1: Time Series Data

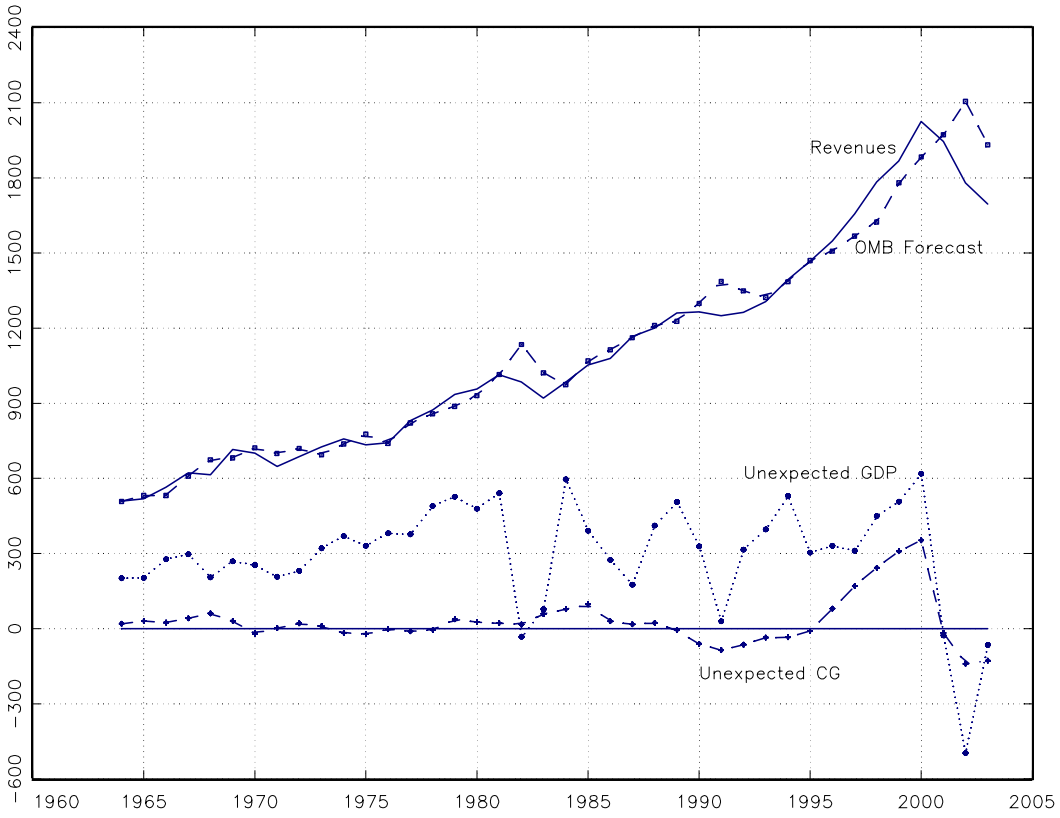


Figure 2: OMB Forecast Error

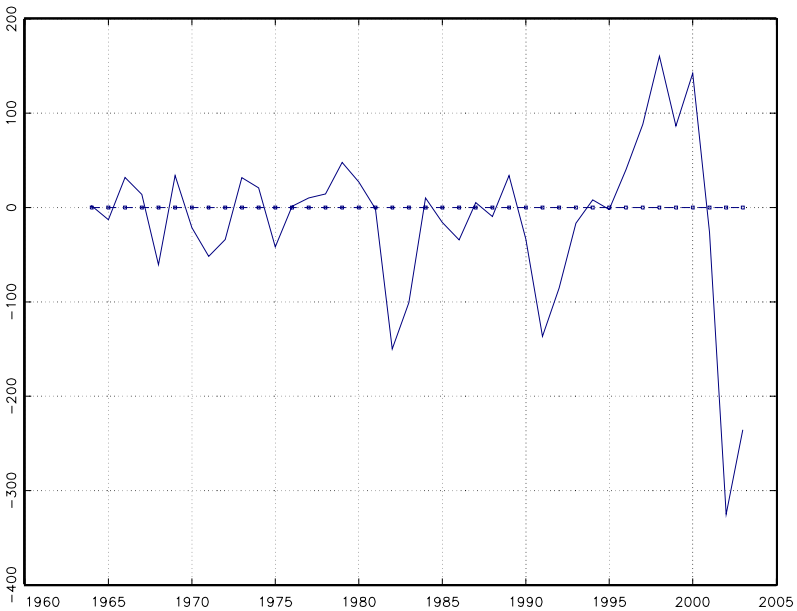


Figure 3: Estimated Model Residuals

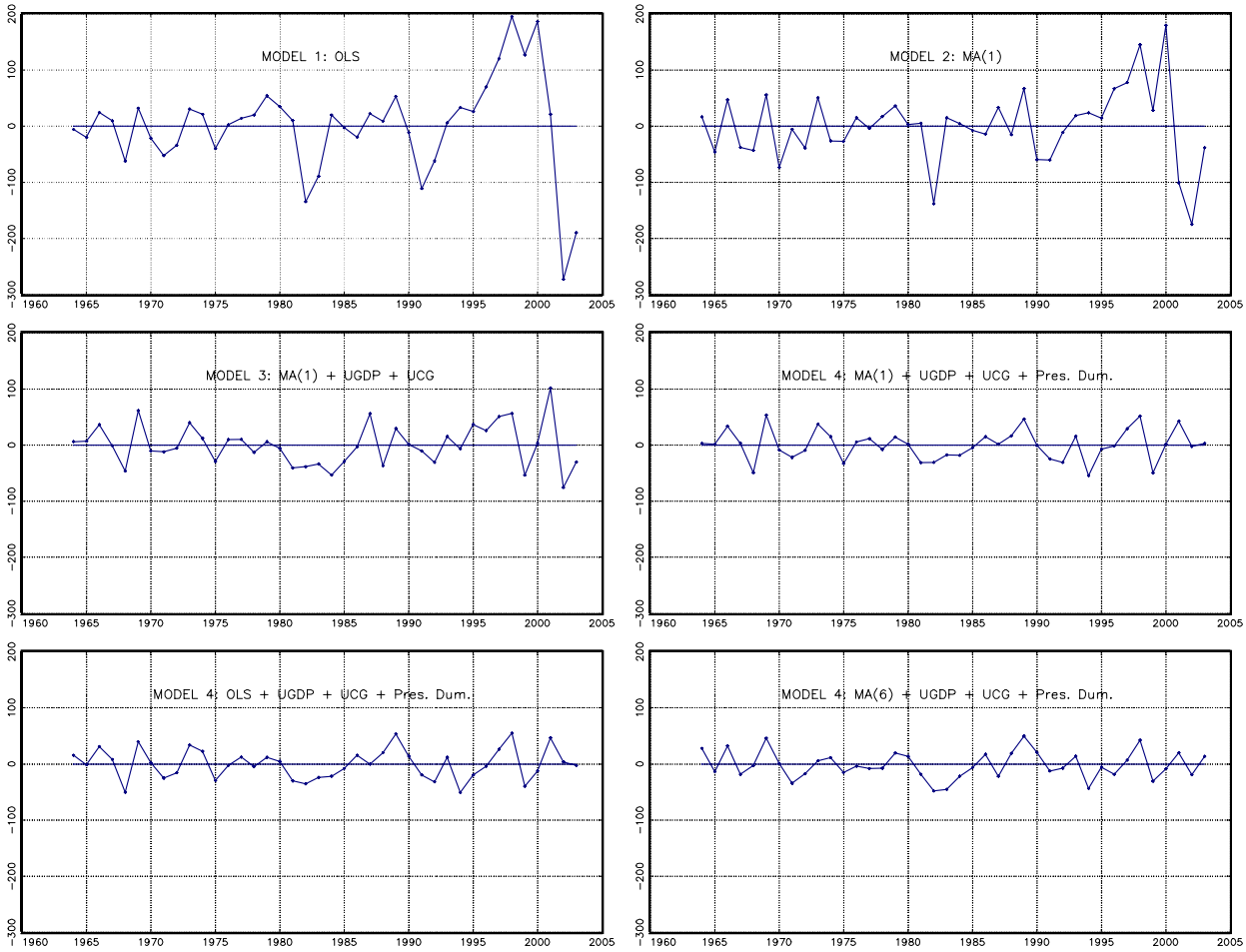
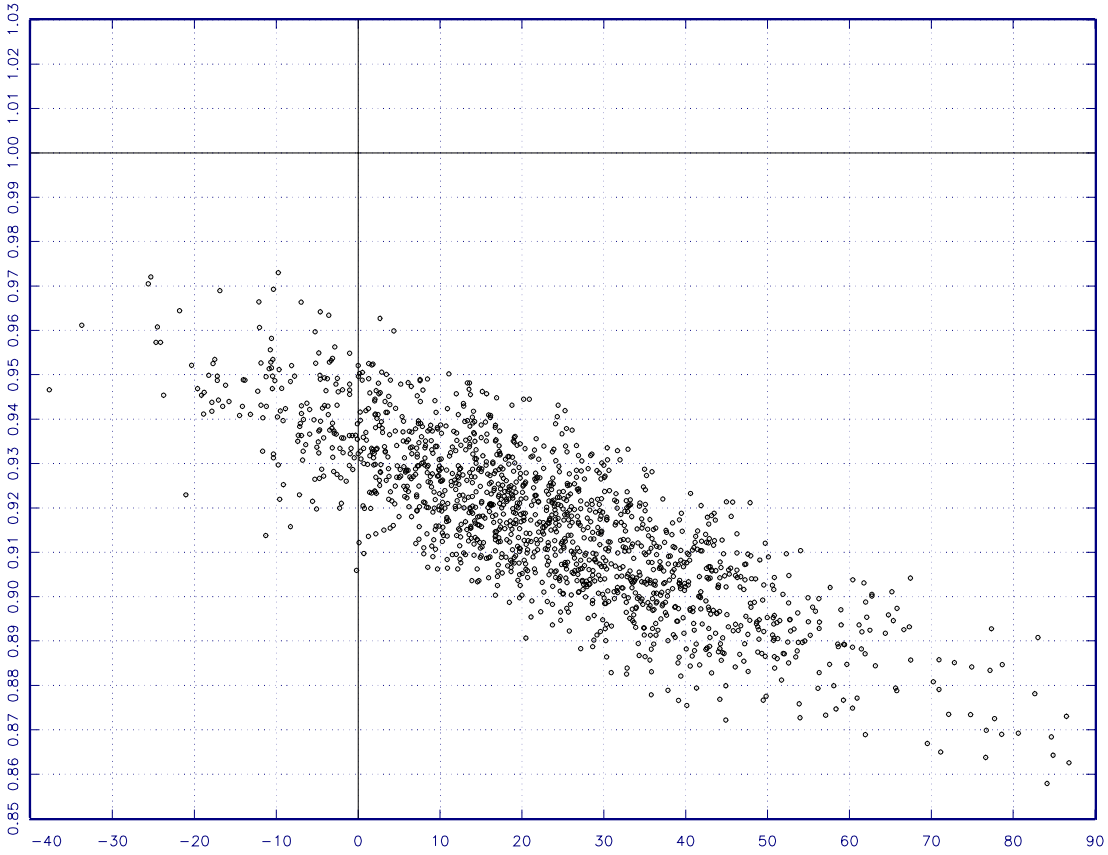


Figure 4: Bootstrap Scatter Plot for $b_1$ (vertical axis) vs. $b_0$ (horizontal axis)



# 9 Appendix 3: Joint Hypothesis Test

Following Theil(1971), we calculate $\mathbf{P}$ such that. $\mathbf{C}^{-1}= \mathbf{P}'\mathbf{P}$, where $\mathbf{C}$ is the parameter variance-covariance matrix. Next, we transform the data to $\mathbf{z}^*(b) = \mathbf{P}[\boldsymbol{\lambda}^*(b) - \overline{\boldsymbol{\lambda}}^*]$, and we transform the null hypothesis to $\mathbf{z}^0 = \mathbf{P}[\boldsymbol{\lambda}^0 - \overline{\boldsymbol{\lambda}}^*]$. This amounts to a multivariate standardization of $\boldsymbol{\lambda}^*(b)$. We now have a spherical distribution with unit variance and zero mean. Hypothesis testing now asks how likely is an observation within a specified radial distance from the origin. The original joint hypothesis test is transformed into a test of how likely is $\left\|\mathbf{z}^0\right\|$ given our observations, $\mathbf{z}^*(b)$. The Kolmolgorov-Smirnov goodness of fit test on each element of $\boldsymbol{\lambda}^*(b)$ indicates only $b_{10}$ is not Gaussian distributed. We thus can conduct significance tests using the standard Gaussian distribution for all elements of $\mathbf{z}$. The transformed null hypothesis vector, corresponding to $b_{10} = 0$, $b_{20} = 1.0$, and Re = GRH = Cl = B2 = 0 is

$$\mathbf{z}^0 = \quad [9.27 \quad 2.99 \quad -14.22 \quad -9.09 \quad 4.52 \quad -3.42 \quad -5.88 \quad 1.74]'.$$