FORECASTING WITH STRUCTURAL BREAKS: THE DISTRICT OF COLUMBIA BEFORE AND AFTER THE FINANCIAL CRISIS*

Fitzroy Lee, Office of Revenue Analysis, Government of the District of Columbia

INTRODUCTION

A KEY ASSUMPTION OF STATISTICAL FORECASTING models is parameter constancy—the assumption that the parameters of a forecasting model remain constant over the sample period and the forecast horizon. But the relationships embodied by the parameters of a model oftentimes change for various reasons. Economies, institutions, and policies can and do change over time, and when they do, relationships between economic variables sometimes change as well. Stock and Watson (1996) tested 608 univariate forecasting equations and 5,700 bivariate relationships for structural breaks in 76 monthly macroeconomic time series. They concluded that "a substantial fraction of forecasting relations are unstable (p. 23)." Some sources of structural breaks are quite subtle, occurring gradually over time (e.g., changes in the demographics of the personal income tax base); others are more abrupt (e.g., a financial crisis). However they occur, some studies find that structural breaks can be detrimental to forecast performance. Clements and Hendry (1999), using Monte Carlo simulations, demonstrate that structural breaks in deterministic parameters (intercepts, trends, and the like) are a cause of forecast failure.[1]

Two decades of turbulence in the economy and finances of the District of Columbia raise the real possibility of structural breaks in the parameters of the statistical models used to forecast the District's taxes. In the early to mid 1990s the District's economy stagnated, the population declined, and the government reached near insolvency. From 1990 to 1996 employment in the District declined from a seasonally adjusted 681,400 to 619,400.[2] Commercial property value fell by 31 percent between 1992 and 1997;[3] and from 1990 to 1996, the District, which had been losing population since the 1960s, lost 32,900 or 5.4 percent of its residents.[4] In all but one year from 1992 to 1996, the District government ran annual budget deficits leading to an accumulated deficit of $518 million in 1996.[5] In 1995 the president and Congress created the District of Columbia Financial Responsibility and Management Assistance Authority (control board) to address the District's fiscal problems. From 1996 to 2001 the control board oversaw the District's finances, instituting many changes in the District's financial policies and administration.

Starting in 1997 the District's economy and finances underwent a remarkable recovery. Between 1997 and 2006 employment in the District grew from 617,100 to 686,300.[6] The value of commercial real property more than doubled over the same period. The District currently enjoys one of the strongest commercial real estate markets in the country. After decades of falling, the latest population figures from the U.S. Census Bureau shows that the population of the District grew from 571,799 in 2000 to 585,459 in 2006.[7] And every year from 1997 to 2006 the District ran annual budget surpluses, ending fiscal year 2006 with an accumulated general fund balance of $1.4 billion.

Given the dramatic changes in the District's economy over the last two decades, it is only natural for a District revenue forecaster to ask: Are there structural breaks in my revenue forecasting model? Are the model parameters stable? The paper investigates these questions for the District's withholding tax forecasting model in two parts. The first part of the paper tests various specifications of a withholding tax forecasting model for structural breaks/parameter stability. This first part also investigates whether some model specifications are more robust to structural breaks or whether some models have more stable parameters than others. The second part compares the forecast performance of the models. Specifically, it uses the root mean squared error (RMSE) and the mean absolute percent error (MAPE) measures of forecast accuracy to assess whether model specifications that are more robust to structural breaks have greater forecasting accuracy.

*The views expressed here are those of the author and not that of the Government of the District of Columbia or the Office of the Chief Financial Officer.

The next section describes the data series used for the withholding tax forecasting model. The third section reviews the literature on structural break and parameter stability testing. The fourth section describes alternative specifications of the withholding tax model. The fifth section presents the results of the tests for structural breaks/parameter stability and the sixth section presents the results of the comparisons of forecast performance. The final section summarizes the findings and outlines some lessons learned.

TIME-SERIES DATA FOR WITHHOLDING TAX AND WAGES AND SALARIES

Figure 1 is a graph of quarterly withholding tax collections from FY1983:Q1 to FY2007:Q3. The series can roughly be divided into three phases: a period of rapid growth from FY1983:Q1 to FY1990:Q2, followed by a period of stagnant growth from FY1990:Q3 to FY1997:Q1, and finally, a resumption of strong growth with a somewhat greater degree of seasonality (bigger bonuses?) from FY1997:Q3 to FY2007:Q3.

Figure 2 shows the equivalent graph of a seasonally adjusted time series of D.C. residents' wages and salaries. The first thing to note about this series is that, like the quarterly withholding collections series, the series can be divided into three phases coinciding roughly with the phases for the withholding collections: a period of strong growth, followed by a slowdown, then a return to strong growth. Another thing to note is that the growth pickup for wages starts a little later than the growth pickup for withholding tax collections.

Figure 1: **Quarterly Withholding Collections (FY 1983Q1-FY2007Q3)**

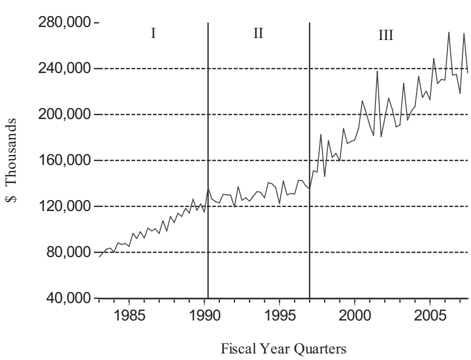
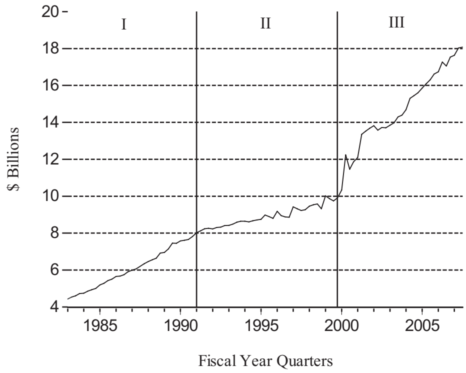


Figure 2: **Quarterly DC Wages and Salaries (FY 1983Q1-FY2007Q3)**



This suggests that the growth pickup in withholding collections might not have been entirely due to economic factors. As there were no major tax policy changes at the time, improvements in the tax administration might have played a role.

STRUCTURAL BREAK/PARAMETER STABILITY TESTING

The traditional test for structural break is the test for a single structural break in a linear regression model outlined in Chow (1960). The Chow test splits a sample into two subsamples, estimates the parameters for each subsample, and, using an F-test, tests for equality of the parameters from the two subsamples. It can be shown that, when the breakdate is known, the test statistic is chi-square distributed. That the Chow test requires prior knowledge of the breakdate is a major drawback to the Chow test because the date of the structural break is not always known to the forecaster. If the forecaster picks a candidate breakdate based on features of the data, the Chow test is likely to falsely find breakdates where none exists because the candidate breakdate is now correlated with the data.

The limitations of the Chow test prompted the development of a set of structural break tests that do not require prior knowledge of the breakdate. Early on, Quandt (1960) proposed taking the largest Chow statistic over all possible breakpoints. But with an unknown breakdate, the test statistic is no longer chi-squared distributed and the Quandt test was of limited usefulness until the early 1990s. At that time Andrews (1993) developed critical values for what came to be known as the Quandt Likelihood Ratio (QLR) or the Quandt-Andrews test statistic; Hansen (1997) subsequently computed p-values. Building on the QLR test, Andrews and Ploberger (1994) derived structural break tests of optimal power based on taking simple and exponential weighted averages of the Chow test statistic over all the breakpoints. Unlike the QLR, however, the Andrews and Ploberger test statistic is not informative about the location of the breakdate. Critical values are given in Andrews and Ploberger (1994) and p-values are computed using the techniques developed by Hansen (1997). This family of tests has now largely replaced the Chow test for structural break testing.

Another set of test statistics focuses on parameter stability—whether parameters vary over time—rather than parameter breaks. They include the cumulative sum (CUSUM) and cumulative sum squared (CUSUMSQ) test statistics which are based on cumulative forecast errors. Brown, Durbin, and Evans (1975) originally computed the CUSUM test statistic from recursive residuals.[8] But Ploberger and Kramer (1992) have since shown that the CUSUM test statistic can be computed from ordinary least squares (OLS) residuals. Very often a CUSUM analysis uses a time-series plot of the CUSUM with 95 percent confidence bands. Movement of the CUSUM plot outside the confidence bounds is taken as evidence of parameter instability. Two drawbacks of the CUSUM test are: (1) it is basically a test of the stability of the intercept, and (2) it has relatively low power. A more powerful parameter stability test is the Lagrange multiplier (LM) test of Nyblom (1989). In addition to testing for constancy of all the parameters, the Nyblom LM test is a locally most powerful test against the alternative that the parameters follow a random walk.

The analysis below uses the Quandt-Andrews test for structural breaks and the CUSUM analysis for parameter stability. The Quandt-Andrews test requires the specification of trimming parameters λ_0 and λ_1. The trimming parameters specifies the fraction of the sample near the beginning and the end of a series that are not used in the structural break tests because breaks are hard to identify near the beginning and end of a series. Any knowledge about the location of a break may be used to specify λ_0 and λ_1. If there is no knowledge of the breakdate Andrews (1993) recommends $\lambda_0 = 0.15$ and $\lambda_1 = 0.85$—the values used in this analysis. Critical values for the Quandt-Andrews statistic depend on the trimming parameters as well as the number of variables in the model.

FORECASTING MODELS

This section presents the specifications for the 10 withholding tax forecasting models (4 univariate models, 6 multivariate models) considered in the analysis below. First, I present the specifications for the univariate models starting with a linear trend model:

$$\textbf{(1)}\quad y_t = \alpha + \beta TIME_t + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2),$$

where $TIME_t$ = 1,2,3, ... T. The next model is an autogression of order p ($AR(p)$) in levels:

$$\textbf{(2)}\quad y_t = \alpha + \beta(L)y_{t-1} + \varepsilon_t,$$

where $\beta(L)$ is a pth order lag polynomial. The lag length, p, is chosen using the Bayes-Schwarz information criterion (BIC). Using the BIC, p was chosen to be 4. The $AR(p)$ model is also estimated in differences with 3 lags of the dependent variable:

$$\textbf{(3)}\quad \Delta y_t = \alpha + \beta(L)\Delta y_{t-1} + \varepsilon_t.$$

The last univariate model considered is a trend model with 4 lags of the dependent variables:

$$\textbf{(4)}\quad y_t = \alpha + \beta_1 TIME_t + \beta_2(L)y_{t-1} + \varepsilon_t.$$

The first multivariate (or causal) model is a simple regression model:

$$\textbf{(5)}\quad y_t = \alpha + \beta x_t + \varepsilon.$$

The model of equation (5) fails to address two important specification issues that usually arise in time-series models: nonstationarity and residual serial correlation. Nonstationarity is an issue when the data series being modeled is trending upward or downward over time. It is likely that nonstationarity is an issue for the data series used in this analysis as Figures 1 and 2 above show an upward trend in both the dependent and the explanatory variables. The upward trend in the variables can sometimes lead to the phenomenon of spurious regression,[9] where the coefficient on an explanatory variable is significant even when the variables are unrelated. One method of dealing with trending data series is to allow for a deterministic linear time trend:

(6) $y_t = \alpha + \beta_1\ TIME_t + \beta_2 x_t + \varepsilon_t.$

A second method is to estimate the model in differences:

(7) $\Delta y_t = \alpha + \beta \Delta x_t + \varepsilon.$

The potential for serially correlated residuals is ever present in time-series models as observations from different time periods are typically related to each other. One method of accounting for serial correlation is to directly model the error term as an AR process and estimate the model using a method like the Cochrane-Orcutt procedure:

(8) $$y_t = \alpha + \beta x_t + \varepsilon_t$$
$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \ldots \rho_p \varepsilon_{t-p} + u_t$$
$$u_t \overset{iid}{\sim} N(0, \sigma^2).$$

Another method of accounting for serial correlation is to add lags of the dependent variable to the model:

(9) $y_t = \alpha + \beta(L) y_{t-1} + \beta_1\ x_t + \varepsilon_t.$

Finally, I consider a model with lagged dependent and lagged explanatory variables:

(10) $y_t = \alpha + \beta_1(L) y_{t-1} + \beta_2(L)\, x_{t-1} + \varepsilon.$

This model has the advantage of both addressing nonstationarity as well as serial correlation in the residuals. It is a single equation from a standard vector autoregression (VAR) system.

RESULTS OF STRUCTURAL BREAK/STABILITY TESTS

The results of the structural break and stability tests are presented in Table 1 and the CUSUM plots of Figure 3. Table 1 presents the results for the Quandt-Andrews structural break tests where the null hypothesis is no structural break. For each model, Table 1 shows both the Andrews critical value as well as the chi-square critical value for the QLR test statistic. As discussed in the third section, the Andrews critical values are the appropriate critical values for the QLR because the breakdate is unknown. The chi-square critical values are included for comparison purposes because it is the applicable critical value for the Chow test where the breakdate is known. As to be expected, the Andrews critical values are larger than the chi-square critical values, reflecting the greater power of the Quandt-Andrews test over the Chow test. Still, there is only one case where the two measures give conflicting results—for the regression model with a trend (#6) the null of no structural break is rejected using the Chi-square critical value but not with the Andrews critical value. Table 1 also reports

Table 1
Results of Quandt-Andrews Breakpoint Tests

Model	*Chi-square Critical Value (5% significance)*	*Andrews Critical Value (5% significance)*	*QLR Statistic*	*Date for max of QLR*	*Durbin-Watson Statistic*
Univariate					
1. $y_t = \alpha + \beta TIME_t + \varepsilon_t$	5.991	11.79	16.51523**	1992Q3	1.48771
2. $y_t = \alpha + \beta(L) y_{t-1} + \varepsilon_t$	11.070	18.35	4.18348	1997Q4	1.90689
3. $\Delta y_t = \alpha + \beta(L)\Delta y_{t-1} + \varepsilon_t$	9.488	16.45	1.73334	1998Q3	1.89922
4. $y_t = \alpha + \beta_1 TIME_t + \beta_2(L) y_{t-1} + \varepsilon_t$	12.592	20.26	4.41409	1997Q4	1.88497
Multivariate					
5. $y_t = \alpha + \beta x_t + \varepsilon$	5.991	11.79	27.75882**	1997Q4	1.65469
6. $y_t = \alpha + \beta_1 TIME_t + \beta_2 x_t + \varepsilon_t$	7.815	14.15	11.24346	1997Q2	1.86115
7. $\Delta y_t = \alpha + \beta \Delta x_t + \varepsilon$	5.991	11.79	0.75467	1990Q3	3.22702
8. $y_t = \alpha + \beta x_t + \varepsilon_t$ $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \ldots \rho_p \varepsilon_{t-p} + u_t$	12.592	20.26	6.33937	1997Q4	1.94162
9. $y_t = \alpha + \beta(L) y_{t-1} + \beta_1 x_t + \varepsilon_t$	12.592	20.26	4.03625	1997Q4	1.93502
10. $y_t = \alpha + \beta_1(L) y_{t-1} + \beta_2(L) x_{t-1} + \varepsilon$	14.067	21.84	4.64076	1997Q4	2.02435

**QLR test statistic exceeds the Andrews critical value.

Figure 3: **CUSUM Plots**

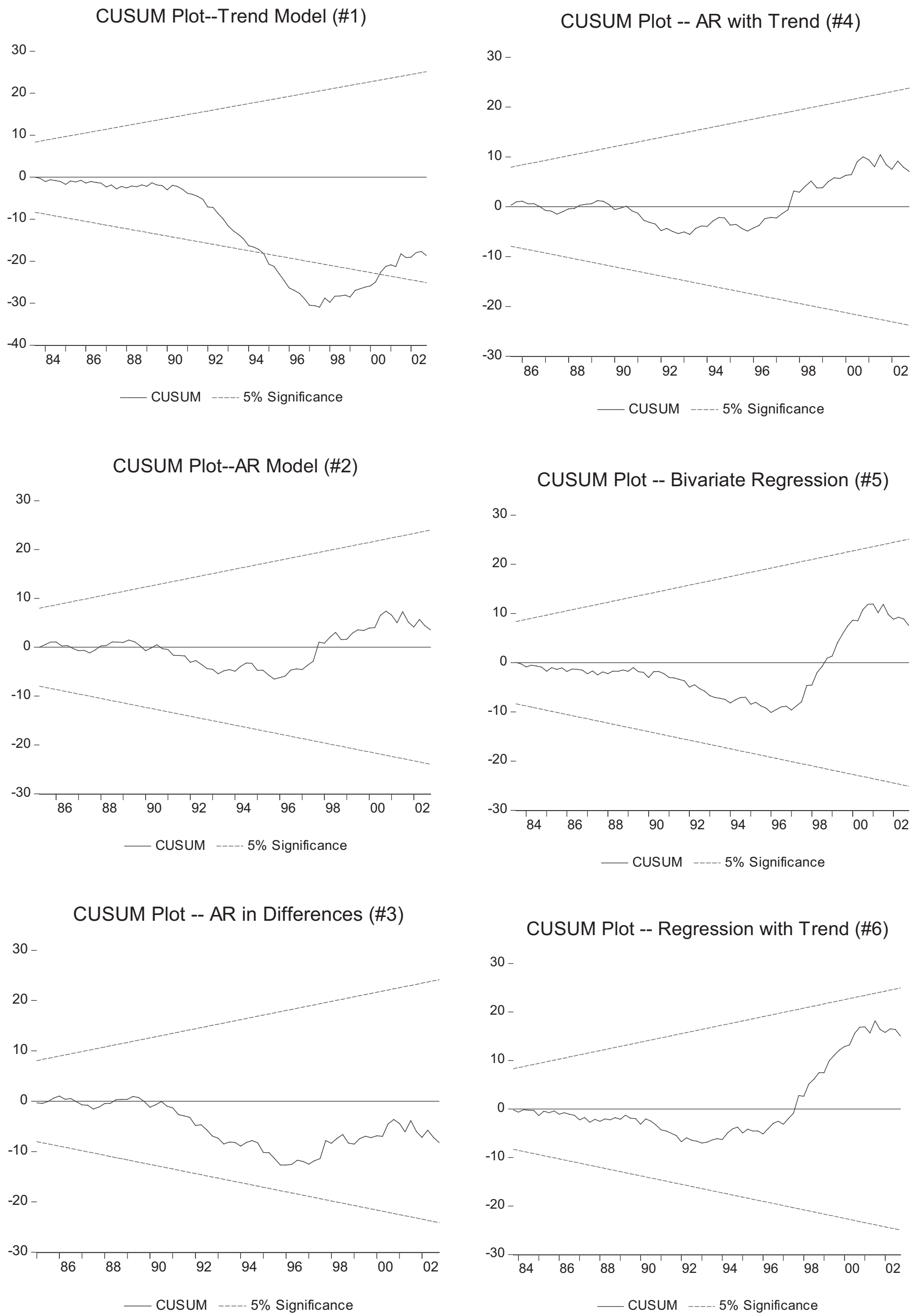
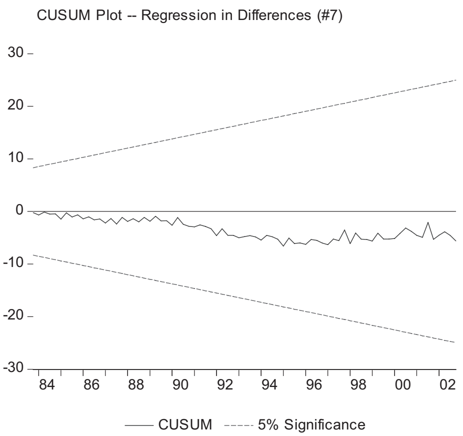
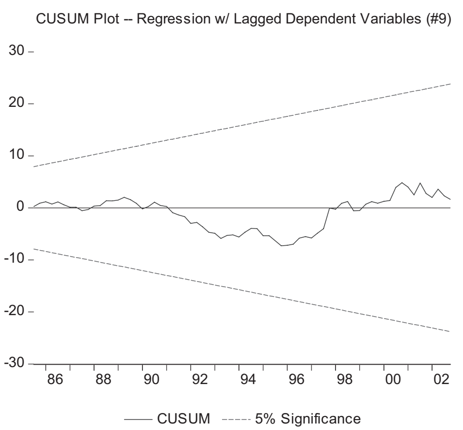
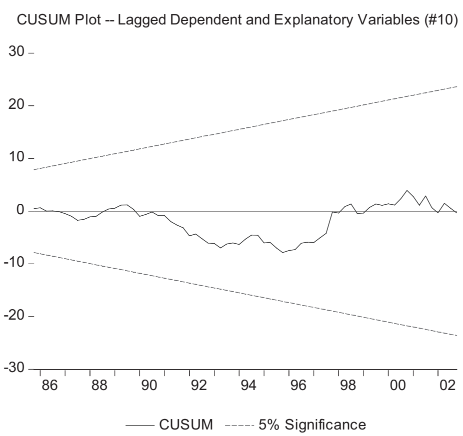


Figure 3: **CUSUM Plots (continued)**







the date of the maximum QLR statistic (it is the likely breakdate when the statistic is significant) and the Durbin-Watson test statistic.

The results in Table 1 show that, using the Quandt-Andrews test, the null hypothesis of no structural break was rejected for only two of the models that were analyzed—the trend model and the simple regression model. The results of the CUSUM plots of Figure 3 are more or less consistent with the results of the Quandt-Andrews tests. Although the plot crosses the 5 percent significance line only for the trend model, the plot approaches the 5 percent significance line both for the simple regression and the regression with trend models, indicating a high degree of instability, if not a break, in the parameters of those models. Interestingly, the two models for which the Quandt-Andrews test finds structural breaks appear to be misspecified, judging from their Durbin-Watson statistics. In light of the identified misspecification there are two ways to interpret the results. First, it could be the case that the identified structural breaks are the artifacts of model misspecification—that is, because of model misspecification the structural breaks tests are finding parameter breaks/instability where none exists (type II error). Alternatively, it could be that a structural break exists but a well-specified model does a better job of making the model more robust to structural breaks. There is some evidence to support the latter explanation. As pointed out previously, the time trend of withholding collections show three distinct phases, in effect “breaks” in the withholding collections time trend. And Figure 4 shows a scatter plot of withholding collections against D.C .wages and salaries. The gap in the plot indicates that there is indeed a “break” in the relationship.

A final note on the results of the structural break and parameter stability tests. The Durbin-Watson statistic for the regression model in differences indicates residual correlation, one symptom of a misspecified model. The discussion of the results in the previous paragraph suggests that structural break and parameter stability tests are more likely to find structural breaks in misspecified models. Yet, neither the Quandt-Andrews test nor the CUSUM plot finds evidence of structural breaks or parameter instability in the model. This seems to lend some support to Clements and Hendry (1999) that recommend differencing as one way of making forecasting models more robust to structural breaks. They argue that differencing, in the particular case of step shifts (change in the intercept of the model), changes a break into a blip, which is not detectable as a break by the structural break tests.

COMPARISON OF FORECAST PERFORMANCE

Table 2 presents a comparison of the forecast performance of the models using both the root mean square error (RMSE) and the mean absolute percent error (MAPE) measures of forecast performance. The table also shows forecast performance rankings of the models by each of the measures. Figure 5 shows graphs comparing actual withholding collections with forecast, along

Figure 4: **Scatterplot of Withholding Collections v DC Wages**

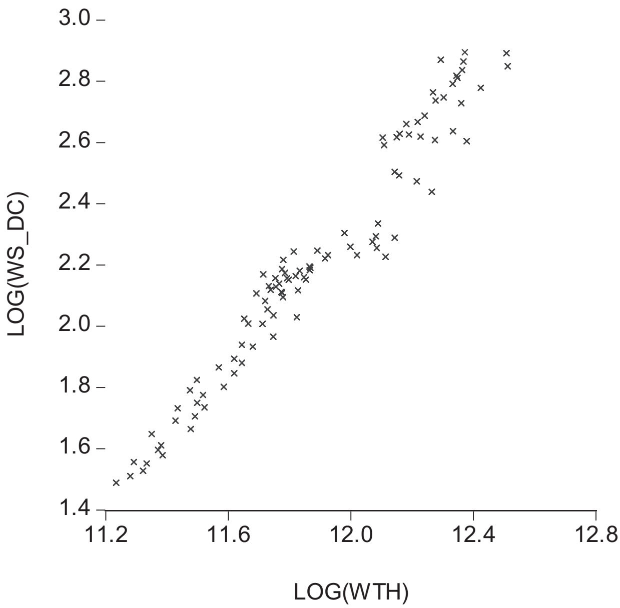
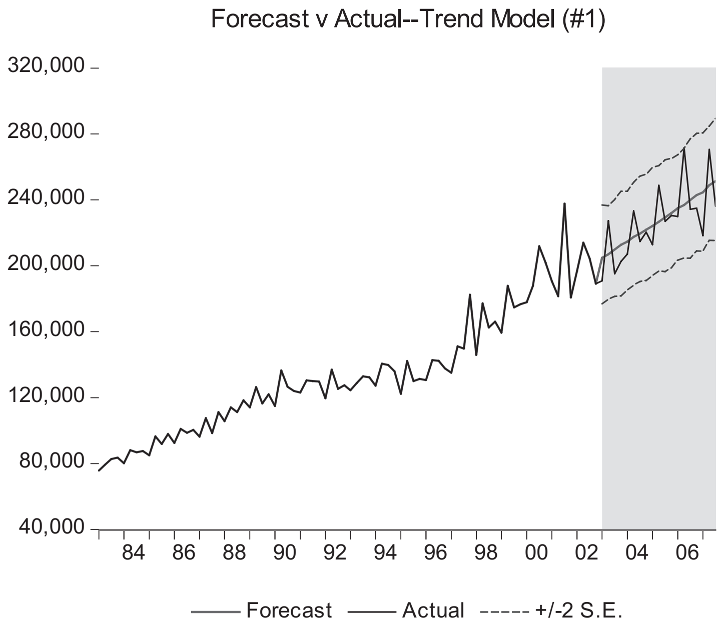


Table 2
Comparison of Forecast Performance

Model	*RMSE ($ thousands)*	*MAPE*	*Rank (RMSE)*	*Rank (MAPE)*
Univariate				
1. $y_t = \alpha + \beta TIME_t + \varepsilon_t$	15,581	5.4%	1	4
2. $y_t = \alpha + \beta(L)y_{t-1} + \varepsilon_t$	17,256	5.2%	5	3
3. $\Delta y_t = \alpha + \beta(L)\Delta y_{t-1} + \varepsilon_t$	16,013	6.0%	3	5
4. $y_t = \alpha + \beta_1 TIME_t + \beta_2(L)y_{t-1} + \varepsilon_t$	15,968	5.0%	2	2
Multivariate				
5. $y_t = \alpha + \beta x_t + \varepsilon$	19,311	7.9%	7	9
6. $y_t = \alpha + \beta_1 TIME_t + \beta_2 x_t + \varepsilon_t$	17,048	7.0%	6	8
7. $\Delta y_t = \alpha + \beta \Delta x_t + \varepsilon$	21,715	6.6%	9	7
8. $y_t = \alpha + \beta x_t + \varepsilon_t$				
$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \ldots \rho_p \varepsilon_{t-p} + u_t$	16,560	5.0%	4	1
9. $y_t = \alpha + \beta(L)y_{t-1} + \beta_1 x_t + \varepsilon_t$	32,228	10.7%	10	10
10. $y_t = \alpha + \beta_1(L)y_{t-1} + \beta_2(L)x_{t-1} + \varepsilon$	19,875	6.0%	8	6

Figure 5: **Graphical Comparison of Forecast Performance**



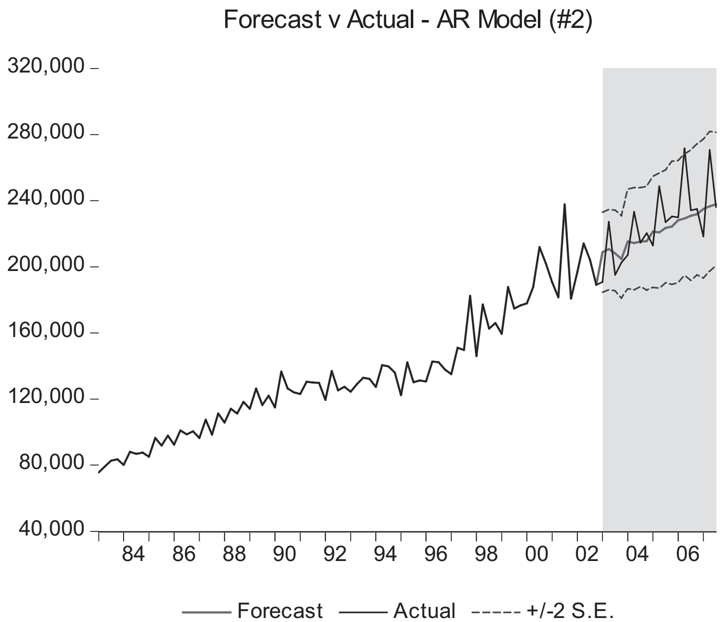
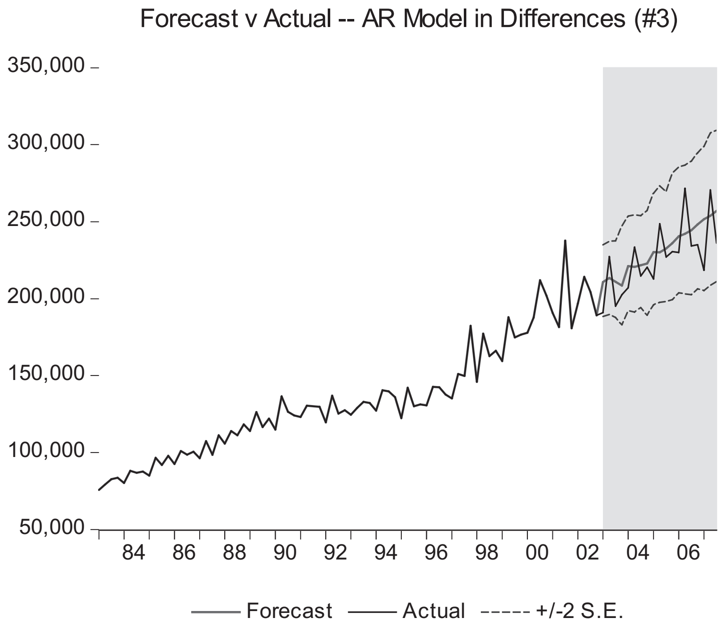


Figure 5: **Graphical Comparison of Forecast Performance (continued)**



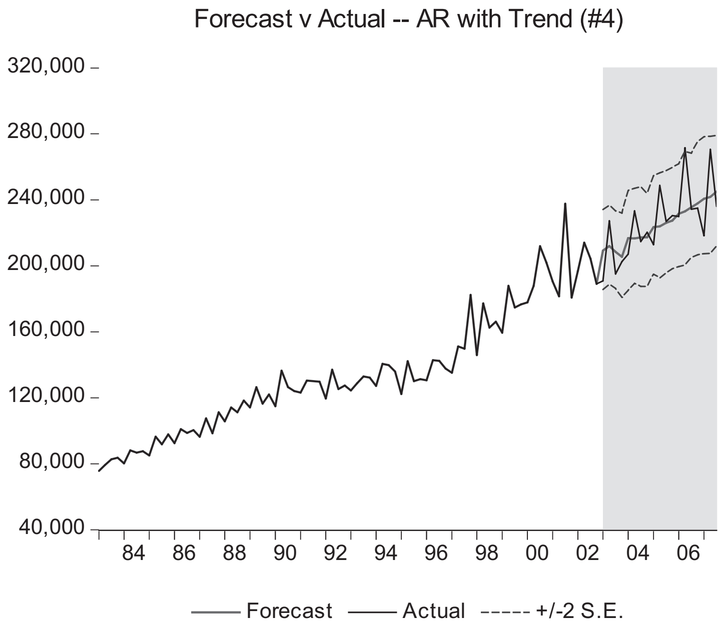
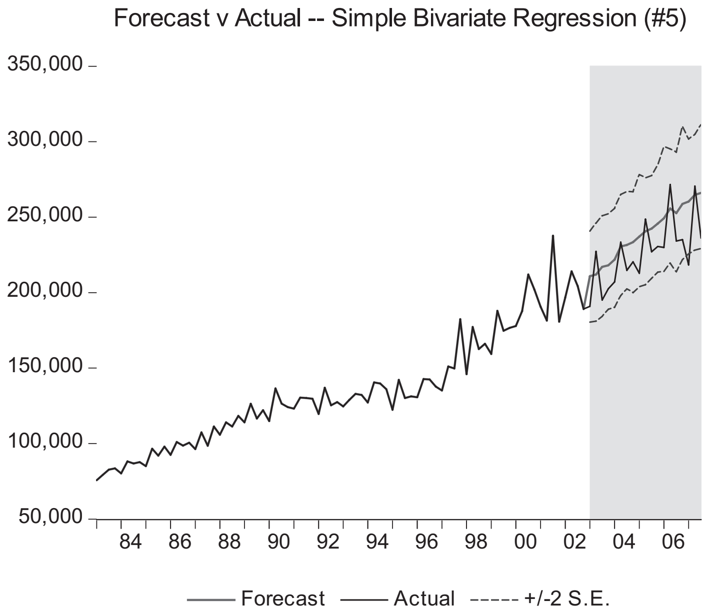


Figure 5: **Graphical Comparison of Forecast Performance (continued)**



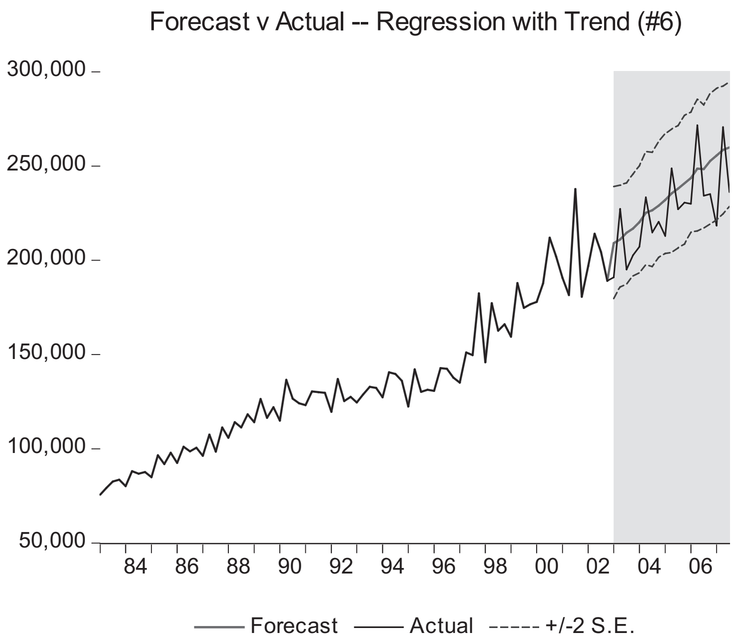
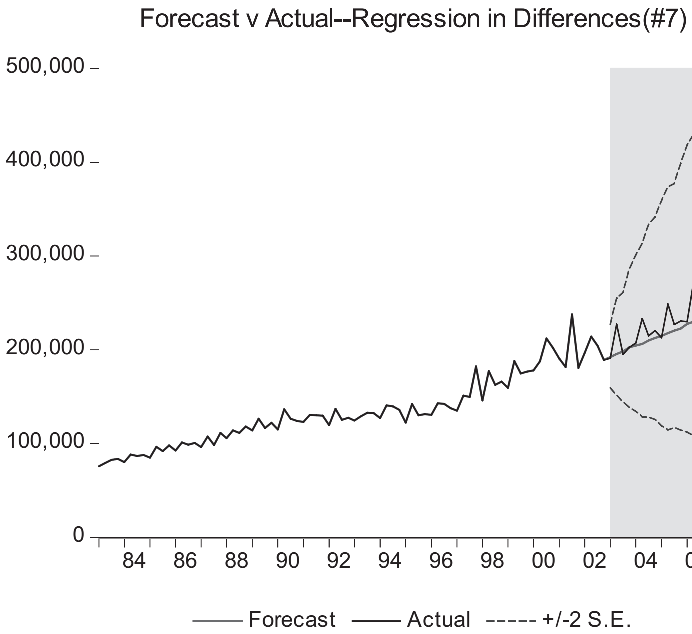


Figure 5: **Graphical Comparison of Forecast Performance (continued)**



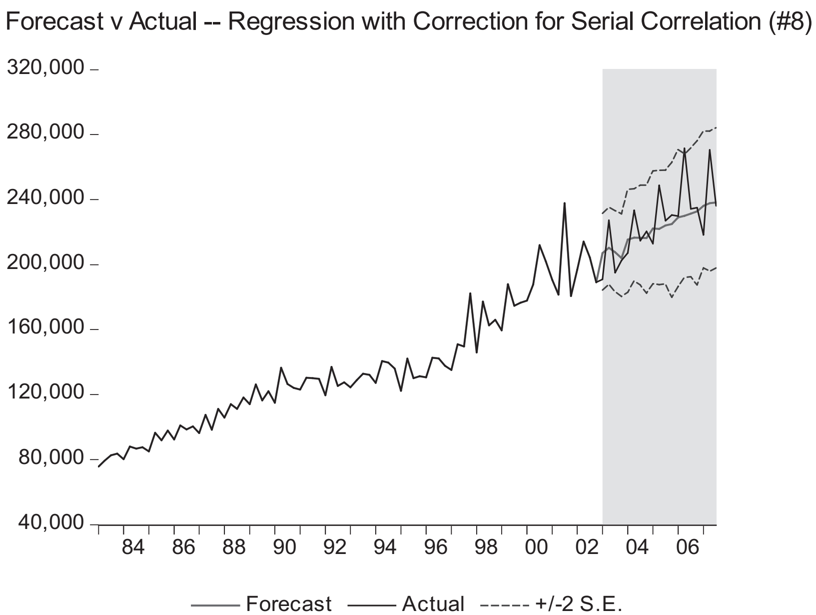
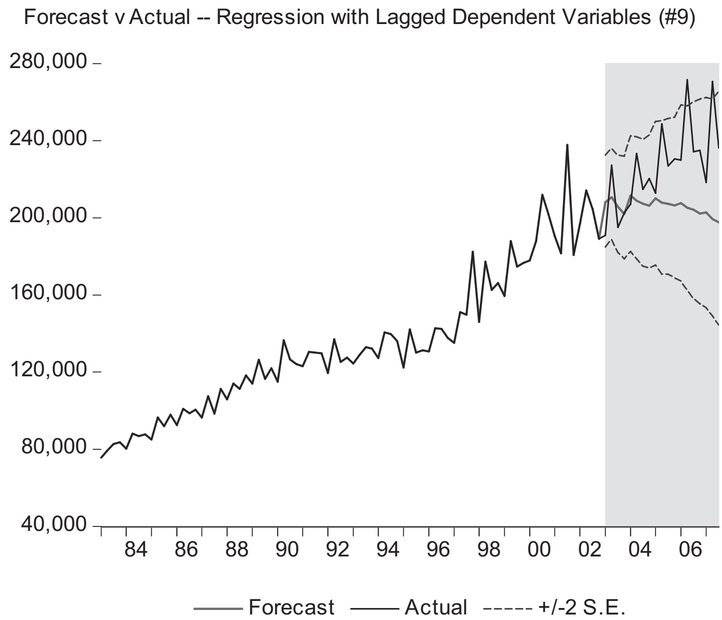
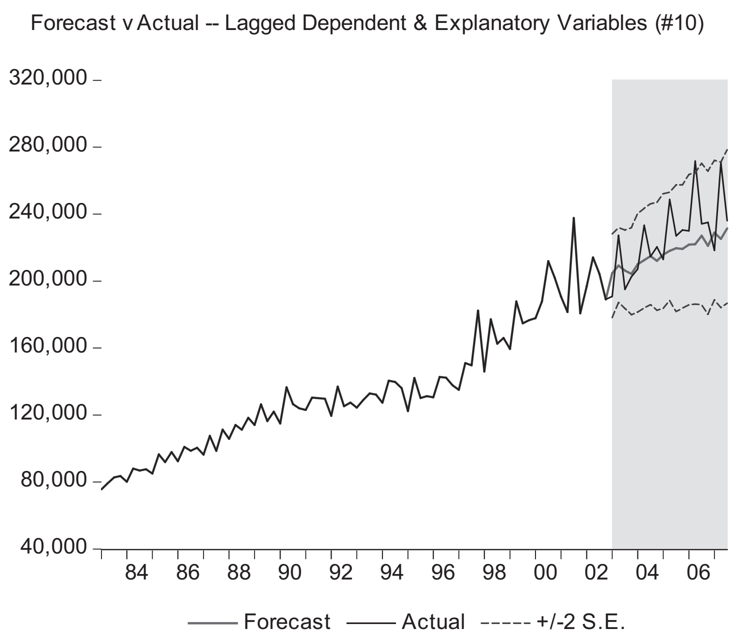


Figure 5: **Graphical Comparison of Forecast Performance (continued)**





with 95 percent confidence bands for each of the models. The forecast performance statistics are computed over the out-of-sample period 2003Q1 to 2007Q3 (the models were estimated with data for the quarters 1983Q1 through 2002Q4). On the basis of the RMSE, the simple trend model outperformed all the other models—univariate and causal models alike—despite the finding of a structural break reported in the previous section. The MAPE measure had slightly different rankings but the trend model still fell in the top 5. Another interesting aspect of the results in Table 2 is that, in general, the univariate models outperformed the causal models with the exception of the bivariate regression model with correction for residual correction. Finally, the graphs of Figure 5 corroborate the RMSE first place ranking of the trend model forecast performance—the point forecast of the trend model was the most centered on the actual withholding collections and its 95 percent confidence band was the narrowest, just encompassing the fluctuations in actual withholding collections.

CONCLUSIONS

The results of the structural break tests suggest that structural break and parameter stability testing cannot be separated from more general model misspecification considerations. The results suggest that the first thing to do on finding a structural break is to probe more broadly for other sources of model misspecifications. This is not to say that the structural break is a manifestation of broader misspecification problems. As discussed previously, other evidence indicates that the structural breaks in the data are in fact real. But it does seem to be the case that better specified models are more robust to structural breaks.

The results of the forecast competition were unexpected but interesting. The expectation was that the better specified models—the ones that showed no structural breaks—would outperform the misspecified models. This turned out not to be the case. Without further analysis one can only speculate as to why the results turned out the way they did. I offer two alternatives: it may be that, other things equal, simpler models—with less parameters to estimate and therefore less ways to make errors—will always perform better than more complex models. Or, it may be that the trend model is the most faithful representation of the withholding collection time series—the graph of the withholding time series shows that it is essentially a trend with seasonal variations—and the identified breaks are not dramatic enough to seriously impact its forecast performance.

So what are the lessons for the model builder/forecaster? First, structural break and parameter stability testing should be an important part of the model evaluation and testing process; even if accounting for the structural breaks yields no improvements in forecast performance, it will still yield valuable insights and understanding of the model. Second, having a well-specified model does not guarantee good forecast performance and therefore it is important to evaluate the forecast performance of alternative models. Third, even though theory tells us that structural breaks can be detrimental to forecast performance, this is not always the case. Finally, it is hard to make general rules; as such, the model builder should emphasize testing and evaluation.

Notes

1 Clements and Hendry (1999) define forecast failure as "significant mis-forecasting relative to the previous record (in-sample, or earlier forecasts)…" (p. 37). They draw a distinction between forecast failure and poor forecasting which is judged relative to some standard, absolute (a policy requirement for accuracy) or relative (a rival model).

2 U.S. Bureau of Labor Statistics (2007).

3 District of Columbia's Comprehensive Annual Financial Report (Fiscal Year 2001).

4 U.S. Census Bureau (2002).

5 District of Columbia's Comprehensive Annual Financial Report (Fiscal Years 1990, 1991, 1992, 1993, 1994, 1995, 1996).

6 U.S. Bureau of Labor Statistics, 2007.

7 U.S. Census Bureau (2007).

8 Recursive residuals are the residuals obtained from a recursive estimation procedure where a model is estimated beginning with a small sample, an observation is added, the model re-estimated, and continuing in this manner until all the observations are used up.

9 See Granger and Newbold (1974) and Phillips (1986).

References

Andrews, Donald W. K. Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61(July 1993): 821-856.

Andrews, Donald W. K., and Werner Ploberger. Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative. *Econometrica* 62 (November 1994): 1383-1414.

Brown, R. L., James Durbin, and J. M. Evans. Techniques for Testing the Constancy of Regression Relationships Over Time With Comments. *Journal of the Royal Statistical Society*, *Series B (Methodological)* 37 (1975): 149-192.

Chow, Gregory C. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* 28 (July1960): 591-605.

Clements, Michael P., and David F. Hendry. *Forecasting Non-stationary Economic Time Series*. Cambridge, MA: MIT Press, 1999.

District of Columbia.

Comprehensive Annual Financial Report of the Government of the District of Columbia (Fiscal Years 1990, 1991, 1992, 1993, 1994, 1995, 1996). Exhibit A-6: General Fund Schedule of Budgetary Basis Revenues and Expenditure. Washington, D.C.

Comprehensive Annual Financial Report of the Government of the District of Columbia (Fiscal Year 2001). Exhibit S-7: Assessed Value, Construction and Bank Deposits: Last Ten Fiscal Years. Washington, D.C.

Granger, Clive W. J., and Paul Newbold. Spurious Regression in Econometrics. *Journal of Econometrics* 2 (July 1974): 111-120.

Hansen, Bruce E. Approximate Asymptotic P Values for Structural-Change Tests. *Journal of Business & Economic Statistics* 15 (January 1997): 60-67.

Nyblom, Jukka. Testing for the Constancy of Parameters Over Time. *Journal of the American Statistical Association* 84 (March 1989): 223-230.

Philips, Peter C. B. Understanding Spurious Regressions in Econometrics. *Journal of Econometrics* 33 (December 1986): 311-340.

Ploberger, Werner, and Walter Kramer. The CUSUM Test With OLS Residuals. *Econometrica* 60 (March 1992): 271-286.

Quandt, Richard E. Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistic Association* 55 (June 1960): 324-330.

Stock, James H. and Mark W. Watson. Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business & Economic Statistics* 14 (January 1996): 11-30.

U.S. Bureau of Labor Statistics. State and Area Employment, Hours, and Earnings for 1990-2006 (SMS 11000000000000001). Washington, D.C. State and Metropolitan Area Employment (Current Employment Statistics). Washington, D.C., 2007. http://data.bls.gov/home.htm#data

U.S. Census Bureau. Population Division.

Table CO-EST2001-12-11-Time Series District of Columbia Intercensal Population Estimates by County: April 1, 1990 to April 1, 2000. Washington, D.C., 2002. http://www.census.gov/popest/archives/2000s/vintage_2001/CO-EST2001-12/CO-EST2001-12-11.html

Table 1: Annual Estimates of the Population for the United States, Regions, States, and Puerto Rico: April 1, 2000 to July 1, 2007 (NST-EST 2007-01). Washington, D.C., 2007. http://www.census.gov/popest/states/NST-ann-est.html