ELSEVIER

International Journal of Forecasting xx (2003) xxx–xxx

international journal of forecasting

www.elsevier.com/locate/ijforecast

The impact of institutional change on forecast accuracy: A case study of budget forecasting in Washington State

Elaine Deschamps*

Washington State Caseload Forecast Council, P.O. Box 40962, Olympia, WA 98504, USA

Abstract

This paper explores the relationship between institutional change and forecast accuracy via an analysis of the entitlement caseload forecasting process in Washington State. This research extends the politics of forecasting literature beyond the current area of government revenue forecasting to include expenditure forecasting and introduces an in-depth longitudinal study to the existing set of cross-sectional studies. Employing a fixed-effects model and ordinary least squares regression analysis, this paper concludes that the establishment of an independent forecasting agency and subsequent formation of technical workgroups improve forecast accuracy. Additionally, this study finds that more frequent forecast revisions and structured domain knowledge improve forecast accuracy.

Keywords: Forecast accuracy; Politics of forecasting; Judgmental forecasting; Government forecasting; Institutional change

1. Introduction

In the state budget process, lawmakers and budget analysts rely heavily on the accuracy of government revenue and expenditure forecasts. Inaccurate forecasts may result in a budget shortfall or perceived wasted opportunity to fund executive or legislative initiatives. The government revenue forecasting process across the states is becoming better understood through recent research, but less attention has been paid to the expenditure side of the forecast process. Fundamental to the expenditure forecast is the ability to accurately predict the demand for entitlement services. An entitlement caseload represents the demand for public services such as Medicaid, where clients who meet the eligibility requirements are entitled to receive the service. The process for estimating expenditures varies from state to state, but most often the caseload estimates provide the foundation for the expenditure forecast.

In 1997 the Washington State Caseload Forecast Council (CFC) was created as an independent agency responsible for the production and oversight of statewide entitlement caseload forecasts. The agency was created with the goals of (1) "promoting the free flow of information and promoting legislative and executive input in the development of assumptions and preparation of forecasts," and (2) "making the caseload forecasts as accurate and as understandable as possible," as stated in the Caseload Forecast Council Strategic Plan in 1997.

* Tel.: +1-360-902-0087; fax: +1-360-902-0084.
E-mail address: Elaine.Deschamps@cfc.wa.gov (E. Deschamps).

0169-2070/$ - see front matter © 2003 Published by Elsevier B.V. on behalf of International Institute of Forecaster.
doi:10.1016/j.ijforecast.2003.11.009

INTFOR-03375; No of Pages 11

This paper addresses the call for more research on how forecast accuracy can be influenced by the specific organizational context in which the forecast is being produced and ultimately used. This research tests hypotheses from the politics of forecasting literature on caseload forecasts produced before and after the creation of the CFC. The goal is to assess the impact of this new institutional arrangement on the accuracy of entitlement caseload forecasts in Washington State. This study employs a quasi-experimental design, using repeated observations over time. A fixed-effects model and a control group are used to reduce threats to internal validity.

This paper begins with an overview of the CFC and how it changed the forecasting process in Washington State. Next is a discussion of the hypotheses and the related literature. Finally, we present the findings and a conclusion.

1.1. Background on the caseload forecast council

The CFC, established in 1997, is charged with forecasting statewide entitlement caseloads. The CFC was modeled after the Economic and Revenue Forecast Council, which was established in Washington State in 1984 to insulate the revenue forecasts from political influences. Similarly, the CFC was created in order to insulate entitlement forecasts from the politics of the budget process, as well as promote a better understanding of the assumptions and methods underlying these forecasts.

The Washington State budget process runs on a 2-year cycle, with a biennial budget passed in the spring of every odd-numbered year. Agencies submit supplemental budgets that are passed by the legislature during the even-numbered years to account for unanticipated expenditures and/or policy changes. Each year, the Governor submits a budget recommendation in December, and the legislature passes a final budget in the spring.

The primary drivers of the state budget are the revenue forecasts that determine the supply-side, and the caseload and expenditure forecasts that estimate the demand-side of the budget equation. Before the CFC was created, the caseload and expenditure forecasts were produced by the executive branch, either in the Governor's budget office (Office of Financial Management) or in the Department of Social and Health Services (DSHS), which administers the human services programs statewide. Entitlement services are presumed to be nonnegotiable in political budget negotiations because by law the state must serve these groups, whether it is through the public school system, Medicaid coverage, or foster homes.

However, because the executive branch produced the caseload estimates, legislative budget analysts were at a disadvantage in not being involved in the forecast process. The Medicaid program area was an exception to this rule because it allowed legislative involvement via a workgroup process before the CFC was established. For assessing the impact of institutional change on forecast accuracy, the Medicaid forecasts serve as a partial control group because there was a technical workgroup process in place both before and after the CFC was created.

The creation of the independent, apolitical CFC and the subsequent formation of technical workgroups have significantly altered the budget forecasting process in Washington State. Though the forecasting methods have essentially remained the same (primarily univariate time series), the environment in which they are produced and reviewed has changed substantially. The forecast models and assumptions are open for discussion and debate by a wider group of participants, and the forecasts are now subject to revision up to three times a year. The Governor and the Legislature are bound by the official forecasts, and the Council meetings are open to the media and general public.

The main vehicle for creating a more open forecast process and better understanding of the forecasts is the technical workgroup process. Each program area has its own workgroup consisting of CFC staff, relevant program experts, and budget and forecasting analysts from the legislative and executive branches. Led by CFC staff, these workgroups discuss and review the data, forecast methods, assumptions, and anticipated effects of policy changes. The technical workgroups serve primarily as advisory groups, and while consensus is attempted, CFC staff has ultimate decision-making authority on all aspects of the forecast.

This paper focuses on the entitlement caseload forecasts for social services, which constitute the largest portion of the state budget. The CFC also produces the public school system forecasts (K-12), so together the forecasts produced by the CFC provide the foundation for over 60% of the state operating budget.

2. Hypotheses and related literature

The politics of forecasting is a relatively new area in the forecasting literature. A number of authors have called for more research on forecasting in organizations and real-world settings as opposed to controlled, experimental settings (Jones, Bretschneider, & Gorr, 1997; Makridakis et al., 1982; Schultz, 1992). While much progress has been made in improving forecasting methods, more needs to be understood regarding how forecast accuracy can be influenced by the specific organizational context in which the forecast is being produced and ultimately used. Bretschneider and Gorr (1989) argue that we should not only be looking at improving our forecasting methods, but also considering the organizational environment, culture, and the forecast process. Variations in these factors may have as great an impact on forecast accuracy as variations in forecasting methods.

One area of current research that considers organizational and political influences on forecast accuracy is state and federal revenue forecasting. A number of studies show that political and organizational variables such as the degree of executive and legislative involvement in the forecast process, partisan composition of government, and political pressures on forecasters have an impact on forecast accuracy (Bretschneider & Gorr, 1987; Bretschneider, Gorr, Grizzle, & Klay, 1989; Klay & Grizzle, 1992). However, there is a void for research on expenditure forecasting.

This research contributes to the literature by analyzing the expenditure forecasting process, or the demand-side rather than the supply-side of the budget equation. Most of the studies on the political and organizational influences on forecasting have been cross-sectional across states. In contrast, this research is longitudinal and allows for an assessment of accuracy both before and after an institutional change.

2.1. Independent, apolitical agency

Bretschneider and Gorr (1987) and Bretschneider et al. (1989) argue that organizational design of the forecasting process directly influences the accuracy of the forecasts. Both studies revealed that forecast accuracy increased if a state had independent forecasts produced by both the legislature and the executive, and if a state used a formal consensus procedure to combine these separate forecasts. This is because political positions and forecast assumptions are exposed for debate. A survey by Klay and Grizzle (1992) revealed that some state revenue forecasters feel political pressure to produce a forecast consistent with their political leaders' policy agendas.

The CFC was established, in part, as a means to reduce the perceived bias in the caseload forecast process. This study posits that the creation of the CFC should improve forecast accuracy by making the forecast product the responsibility of an independent, apolitical agency. Opening the forecast process to more participants places the forecast assumptions up for scrutiny and debate, while the oversight by an independent agency minimizes political manipulation of the forecasts.

2.2. Establishment of technical workgroups

The establishment and development of technical workgroups for each program area has been the primary vehicle in achieving the CFC's goal of a more open forecast process. The workgroup meeting provides a forum for the discussion of administrative and policy changes that may impact the caseload and an opportunity to understand, debate, and review the forecast assumptions and methods.

The forecasting literature suggests that group-based forecasting can improve forecast accuracy and legitimacy. White's (1986) study of forecasting in the private sector emphasizes the importance of greater participation in the forecast process.

Steen (1992) emphasizes the utility of team-based forecasting, since no single person has all the necessary information to prepare forecasts. Kahn and Menzer's (1994) analysis of forecasting in the private sector found mixed results on the benefits of team-based forecasting for improving accuracy. However, the team approach led to greater satisfaction with the forecast process. Jenkins (1982) explains that an ideal approach to forecasting in organizations includes a "forecast formulation committee," consisting of both policy makers and forecasters, which enables currently relevant policy assumptions to be passed on to those who produce the forecast models.

An important aspect of the technical workgroup process is its emphasis on consensus building. Klay and Zingale (1980) found a positive relationship between a consensus-oriented forecast process and perceived improvements in revenue forecast accuracy. A consensus-oriented process is one in which the legislative and executive branches work cooperatively to develop the forecast and agree to using it. In their nationwide survey, respondents from states using this process were more likely to perceive that forecast accuracy had improved. However, a later study by Klay and Grizzle (1986) did not find a significant relationship between consensus-based forecasting and forecast accuracy.

Voorhees (2000) stresses the importance of broad participation in the forecast process for two reasons. First, the broader the consensus and diversity of people involved in the forecast process, the less likely that political bias will affect the forecast. Second, the diversity of the participants and increased competition between perspectives can help to reduce "assumption drag" (Ascher, 1978), which is the tendency to cling to outmoded assumptions. His study concluded that the degree of consensus in the forecast formulation significantly reduced forecast error.

As the primary component of institutional change, the effect of implementing new technical workgroups on forecast accuracy will be assessed by comparing the results for the Medicaid forecasts to the forecasts from the other human services program areas (public assistance, children's services, and long term care). The Medicaid forecasts were produced via a technical workgroup process both before and after the creation of the CFC, so they serve as a partial control group when assessing the impact of institutional change on forecast accuracy.

Institutional Change Hypothesis. Forecast accuracy improves after an independent forecast agency is implemented and Technical Workgroups are established.

2.3. More frequent revisions to forecasts

Prior to the creation of the CFC, the human services forecasts were produced by DSHS once a year in November. The CFC's enabling legislation requires at least three Council meetings per year, so that each forecast may be revised as often as every few months. Thus, the final budget passed by the legislature in the spring may incorporate revised caseload forecasts based on updated data and program knowledge. Though it seems intuitive that revising a forecast more often would improve accuracy, such a practice may increase the risk of adjusting a forecast to random error.

Shkurti and Winefordner's (1989) study of revenue forecasting in Ohio concludes that a mechanism to assure systematic monitoring and revision of previous forecasts assists in assuring more accurate forecasts. They find that monitoring and evaluation are important for both accuracy and acceptance of a forecast, and they also call for additional research aimed at determining the proper frequency of forecast revisions.

Voorhees (2000) found that as forecast frequency increases, forecast accuracy decreases. He concludes that while more frequent forecasts may provide an early warning of a change in trend, it may be difficult to decipher whether the change represents a real trend or just random error.

Mocan and Azad's (1995) survey of general fund revenue forecasts for 20 states found that forecasts revised on a monthly or bimonthly basis were less accurate than forecasts revised either on an annual, biannual, or quarterly basis. Adjusting a forecast as frequently as monthly or bimonthly may lead to misinterpretation of one or two new data points and thus increase forecast error. However, a quarterly or biannual update is based on three to six new data points so is more likely to be based on a real change in the data rather than random variation.

The creation of the CFC forced forecast revisions to be considered roughly every 4 months instead of annually. This study posits that forecast revisions, based on at least 3 months of new data, will improve accuracy. The literature on this subject is sparse and mixed, but this hypothesis reflects the commonly held belief among both technical staff and policy makers involved in the budget process.

Forecast Revision Hypothesis. More frequent revisions improve forecast accuracy.

2.4. Domain knowledge

An important goal of the technical workgroup process is to incorporate more domain knowledge and expertise into the forecasts. There is an extensive literature on whether or not judgment improves forecast accuracy over standard statistical models. There has been considerable evidence to support the integration of judgmental and statistical techniques to improve forecast accuracy (Lawrence, Edmundson, & O'Connor, 1986; Lobo & Nair, 1990; Mathews & Diamantopoulos, 1989; McNees, 1990; Sanders, 1992; Wolfe & Flores, 1990) but also some evidence against it (Carbone, Anderson, Corriveau, & Corson, 1983; Lawrence & O'Connor, 1995, Lim & O'Connor, 1995; Remus, O'Connor, & Griggs, 1995).

Armstrong and Collopy's (1998) review of recent empirical studies on the integration of judgment and statistical methods concluded that judgmental revisions to forecasts work best when forecasters have strong domain knowledge and the revisions are based on structured judgment. Lacking these conditions, judgmental revisions can negatively affect accuracy. Goodwin and Wright (1993) conclude that an important area for future research is field-based rather than laboratory studies on the relationship between judgmental and statistical forecasts.

This study allows for the testing of how the addition of domain knowledge to a statistical forecast affects accuracy in an organizational setting. The technical workgroup journals maintained by CFC staff document discussions of programmatic issues and policy changes that can be used to assess the extent to which domain knowledge was applied to any given forecast. Access to these journals allows for a unique, in-depth assessment of the use of domain knowledge in additional to statistical forecasts.

Some forecasts contain step adjustments separate from the baseline trend that specifically quantify the expected impact of a given policy change. Thus the extent of domain knowledge incorporated into a given forecast ranges from none, to anecdotal, programmatic information, to quantification of the policy impacts in addition to the baseline trend forecast.

Domain Knowledge Hypothesis. Domain knowledge improves forecast accuracy over statistical models alone.

3. Research design

We test these hypotheses using multiple regression models estimated via ordinary least squares (OLS). The data consist of actuals and forecasts produced both before and after the creation of the CFC, from November 1996 to February 2000. The unit of analysis is the individual caseload forecast, and the data set contains 180 observations, based on 20 separate caseloads and 9 forecasts for each caseload. We employ a fixed-effects model by using dummy variables to control for the nuisance variation among the caseloads.

The first model addresses whether forecast accuracy improved after the CFC was established, via a pre-CFC vs. post-CFC analysis. During the pre-CFC time frame, the forecasts were produced only once a year and never subject to revision. Each caseload has a November 1996 and November 1997 forecast for this period. During the post-CFC time frame, the forecasts were subject to revision three times a year, so each caseload has a February, June, and November forecast for 1998 and 1999, and a February 2000 forecast. The second part of the analysis examines only the post-CFC time period and tests the revision and domain knowledge hypotheses.

The data used to assess forecast accuracy will be based on a comparison of actual and forecasted values. The actuals are monthly caseload data from administrative databases generated by DSHS. For the Medicaid caseloads, the actuals are run through a data completion process called lag adjustment, which is necessary because recent historical data are often under-reported. The lag adjustment process is essentially a forecast of the completed data, based on an analysis of the data as they become more complete from month to month.

3.1. Dependent variable: Forecast error

The measurement for accuracy is mean absolute percentage error (MAPE) that indicates the absolute error regardless of direction. This measurement is commonly used in the forecasting literature to measure overall accuracy of the forecast (Bretschneider & Gorr, 1987; Frank & Wang, 1994; Kamlet, Mowery, & Su, 1987; Welch, Bretschneider, & Rohrbaugh, 1998). Armstrong and Collopy (1992) evaluated measures for making comparisons of errors across 191 annual and quarterly time series and recommended the use of MAPE for comparing forecast accuracy in all cases except when few series are available.

The MAPE for each forecast is calculated for both a 6- and 12-month forecast horizon. MAPE6 is calculated by taking the mean of the absolute percentage errors, defined as 100 × (forecast-actual)/actual, for the first 6 months after the forecast was produced, and MAPE12 is the mean of the absolute percentage errors for the first 12 months after the forecast was produced.

3.2. Independent variables and related hypotheses

The Institutional Change Hypothesis are tested using a variable indicating the number of years since the CFC was created. The number ranges from 0 for the "pre-CFC" period to 2 since the time frame of this analysis only covers up to 2 years "post-CFC". This variable, called "time," is expected to be inversely related to forecast error and thus have a negative coefficient in the regression equation.

The Forecast Revision Hypothesis is tested using a dummy variable to indicate whether or not the forecast was revised from the last forecast (= 1) or if the previous forecast was retained (= 0). This variable is also expected to be inversely related to forecast error, and applies only to the post-CFC period because prior to the creation of the CFC the forecasts were produced only once a year.

The Domain Knowledge Hypothesis is measured via two dummy variables. The first one measures anecdotal program information: 1 if programmatic information was used in making the decision between statistical forecast models, and 0 if the decision was made without the use of program knowledge. A second dummy variable will indicate whether a step adjustment or quantifiable estimate of a specific policy change was incorporated into the model.

The step adjustment is distinguished from anecdotal program knowledge because it represents a concerted attempt to measure or quantify the expected impact of a policy change. Thus, it is a much more structured form of domain knowledge than anecdotal information.

While we believe that structured domain knowledge should improved forecast accuracy, the existence of a step adjustment adds another element of complexity to the forecast. Many times, the policy change that the step adjustment is attempting to capture represents a significant divergence from prior policies and historical patterns in the data. Since there typically is no historical precedence from which to base these estimates, the step adjustment introduces another element of potential forecast error to the model. Due to the increased complexity of the process being predicted, it is possible that forecasts with step adjustments may be less accurate than other forecasts.

3.3. Control variables

In order to control for the impact of the technical workgroup, the primary driver of institutional change, separate models will be run for the Medicaid forecasts vs. the other forecasts, in addition to models run on the full sample. The expectation is that the impact of institutional change on forecast accuracy will be stronger for the non-Medicaid programs than for the Medicaid programs because the latter already had a workgroup process in place prior to the creation of the CFC.

The model also includes a variable to capture the volatility in the actual data used to produce the Medicaid forecasts. The Medicaid data change from month to month due to a substantial lag time in the billing process and retroactive eligibility for Medicaid. These data are run through a process that essentially forecasts what the final values will be for the data when they become complete. However, even this process is subject to forecast error, so this additional error inherent in the actuals is controlled for via a data volatility variable that captures the amount by which the historical data have changed from the time the forecast was produced to 12 months later when forecast accuracy is measured. The absolute value of this data volatility measurement will be used to make it consistent with the dependent variable MAPE.

A fixed effects model is used to control for caseload specific variation. Dummy variables representing 19 caseloads are used in the model to control for the variation specific to each caseload, for which there are repeated observations.

4. Results

The first set of models test for the impact of institutional change on forecast accuracy. Table 1 presents the results of the regression analysis on forecast error (MAPE) for the entire sample of 180

Table 1
Regression results for "pre- vs. post-CFC" models

Independent variable	Full sample		Medicaid programs		Non-Medicaid programs	
	MAPE6	MAPE12	MAPE6	MAPE12	MAPE6	MAPE12
Intercept						
Beta coefficient	3.71***	6.37***	3.78***	5.19***	3.82***	6.77***
Standard error	0.96	1.16	1.35	1.55	0.63	0.87
Time—years since CFC was established (0 for pre-CFC)	− 0.70**	− 0.96**	0.11	0.33	− 0.84***	− 1.48***
	0.27	0.32	0.45	0.51	0.24	0.34
Data volatility	N/A	N/A	1.26***	1.46***	N/A	N/A
	0.27	0.31				
N	180	180	90	90	90	90
Adjusted R-squared	0.39	0.40	0.46	0.46	0.35	0.46
Pr > F	0.0001	0.0001	0.0001	0.0001	0.0002	0.0001
Durbin–Watson	1.41	1.57	1.40	1.63	2.08	1.94

*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

observations and for models run on Medicaid programs and non-Medicaid programs separately.

4.1. Pre-CFC vs. post-CFC hypothesis

The results support the Institutional Change Hypothesis that forecast accuracy improved after the CFC was created. In the full sample model, forecast error was reduced by almost 1 percentage point over a 12-month forecast horizon for each year since the CFC was established. The results are significant for both the full sample and the non-Medicaid programs for both the 6-month and 12-month forecast horizons. Results were not significant for the Medicaid forecasts.

The significant findings for the non-Medicaid programs and nonsignificance of the Medicaid program coefficients support the Institutional Change Hypothesis because the Medicaid forecasts served as a partial control group. Improvements in forecast accuracy were not apparent in the Medicaid forecast group, which already had a technical workgroup process in place prior to the creation of the CFC. However, for the non-Medicaid program areas, forecast error over a 12-month forecast horizon was reduced by 1.5 percentage points per year. Forecast error measured at 6-month rather than 12-month intervals was reduced by 0.8 percentage points for each year since the CFC was created.

The results establish a link between the creation of the CFC and reduced forecast error. While we cannot determine the extent to which the transfer of forecast responsibility to an independent agency improved forecast accuracy, the results indicate that the technical workgroups played an important role in the reduction of forecast error. The non-Medicaid forecasts, subject to a newly established technical workgroup process, improved in accuracy over the Medicaid forecasts, which already had a workgroup process in place before the CFC was established.

While these results support the Institutional Change Hypothesis that forecast accuracy has improved since the CFC was created, there are a multitude of factors that could explain why forecast accuracy improved. The next series of models confine the sample to only those forecasts produced after the CFC was created, allowing for a more detailed analysis of the role of forecast revisions and domain knowledge that are major components of the technical workgroup process. Table 2 summarizes OLS regression results for the full sample and separate models for Medicaid programs and non-Medicaid programs. Note that of both tables, only the results for the full sample and non-Medicaid program regressions in Table 1 have significant positive serial correlation. Thus, the significance of coefficients is likely overstated for those two regressions.

4.2. Post-CFC hypotheses

The results indicate strong support for the Forecast Revision Hypothesis across all samples. The coefficients are significant and suggest that when a forecast

Table 2
Regression results for "post-CFC" models

Independent variable		Full sample		Medicaid programs		Non-Medicaid programs	
		MAPE6	MAPE12	MAPE6	MAPE12	MAPE6	MAPE12
Intercept	beta coefficient	4.15***	5.81***	6.21***	9.63***	3.51***	5.32***
	standard error	1.04	1.17	1.54	1.76	0.67	0.79
Revision made to forecast		− 2.00***	− 1.97***	− 2.06*	− 1.74	− 1.10**	− 1.25*
		0.47	0.53	0.84	0.96	0.40	0.48
Step adjustment (domain knowledge)		− 1.63*	− 2.57**	0.02	− 1.30	− 2.33***	− 2.99***
		0.79	0.89	1.47	1.67	0.62	0.73
Program information (domain knowledge)		− 0.71	− 1.09	− 0.06	− 0.43	− 0.73	− 1.16*
		0.54	0.62	0.95	1.09	0.45	0.53
Data volatility		N/A	N/A	1.03***	1.13**	N/A	N/A
				0.29	0.32		
N		140	140	70	70	70	70
Adjusted *R*-squared		0.51	0.57	0.52	0.54	0.50	0.60
Pr > *F*		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Durbin–Watson statistic		2.24	2.13	2.20	2.00	2.23	2.13

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

is revised, forecast error is reduced by 1.1–2.1 percentage points, depending on the program area and forecast horizon. This hypothesis is the only one supported by the Medicaid data, indicating that when a forecast is revised, error is reduced by as much as 2.1 percentage points. For the non-Medicaid data, more frequent forecast revisions improve forecast accuracy by 1.1–1.3 percentage points, depending on the forecast horizon.

The results also support the Domain Knowledge Hypothesis, though not consistently across all models. This hypothesis has two components: the step adjustment variable indicates a quantifiable measurement of an expected policy change, and the program information variable that specifies anecdotal information was used in addition to a statistical forecast.

The step adjustment variable is significant for the full sample and for non-Medicaid programs. Over a 12-month forecast horizon, the inclusion of a quantifiable estimate of a policy impact in a forecast reduced forecast error by 2.6 percentage points for the full sample and 3.0 percentage points for the non-Medicaid programs sample.

The program information variable, capturing anecdotal evidence, is significant only for non-Medicaid programs, reducing forecast error by 1.2 percentage points for a 12-month forecast horizon. The fact that the step adjustment, which is the more structured form of domain knowledge, yielded more significant results than the less structured, anecdotal form supports findings from the literature that structured judgmental inputs are most effective.

The lack of significance of both domain knowledge variables for Medicaid programs may be due to the overriding volatility in the actual data used to generate the forecasts, as evidenced in the significance of the data volatility variable. The data volatility variable indicates that a 1-percentage-point increase in the error inherent in the administrative data yields an equivalent increase in forecast error over a 6-month horizon and a slightly larger (1.1 percentage point) increase in forecast error over a 12-month horizon.

Table 3 summarizes the results of both the pre-CFC vs. post-CFC and the post-CFC analyses in terms of the hypotheses addressed in this paper.

Table 3
Summary of results

Hypothesis	Full sample	Medicaid programs	Non-Medicaid programs
Institutional Change	×		×
Forecast Revision	×	×	×
Domain Knowledge	×		×

5. Conclusion

This study addressed a specific gap in the politics of forecasting literature by examining institutional influences on the demand-side rather than the supply-side of government forecasting, and by expanding this body of literature to include a longitudinal study in addition to the existing set of cross-sectional studies.

The results indicate a relationship between institutional change and forecast accuracy. This study posited that forecast accuracy would improve as a result of the creation an independent agency and the establishment of technical workgroups that de-politicized the forecast process and improved communication between program experts, forecasters, and budget writers. Results showed that forecast accuracy improved after the CFC was created by as much as 1.5 percentage points per year.

The fact that accuracy improved for non-Medicaid programs but not for Medicaid programs served to isolate the importance of the technical workgroups in improving forecast accuracy. For Medicaid programs, a technical workgroup had already been established prior to the implementation of the CFC workgroups, and we did not see a reduction in forecast error for this control group. For the non-Medicaid programs, the reduction in forecast error was significant for both the 6-and 12-month forecast horizons.

This study also supported findings in the judgmental forecasting literature, but with a unique twist. The use of technical workgroup journal notes to obtain in-depth information on the degree to which domain knowledge was incorporated into each individual forecast allowed for the unparalleled opportunity to examine the relationship between domain knowledge and forecast accuracy in a real-world, organizational setting.

The results highlight the importance of adding domain knowledge to statistical models, in the form of both anecdotal program knowledge and specific quantitative adjustments to account for expected policy changes. However, the evidence favored step adjustments over anecdotal information to improve forecast accuracy. For non-Medicaid programs, quantifiable domain knowledge was over twice as effective in reducing forecast error as anecdotal input, improving accuracy by as much as 3 percentage points over a 12-month forecast horizon.

The fact that quantifiable estimates of policy changes were more significant than anecdotal input supports findings from the judgmental forecasting literature that domain knowledge is best utilized when it is incorporated into the forecast in a more structured manner. Additionally, the quantifiable step adjustment is useful to policy makers because it allows for the possibility of evaluating the impact of a policy once the actual data are known.

Finally, the strongest finding of this study was that revising forecasts significantly improved forecast accuracy across all samples, by as much as 2.1 percentage points over a 12-month forecast horizon. These results suggest that the frequency of revisions used by the CFC might serve as a middle ground between monthly and annual revisions to forecasts, which is consistent with Mocan and Azad (1995). Given the sparseness of the literature in this area, these results can be used to further test under what circumstances a forecast should be revised. In the CFC process, forecasts are subject to revision three times per year, but are only updated if the technical workgroup agrees that sufficient evidence exists to merit a revision.

Future research will involve expanding the current database and revisiting these hypotheses, as well as adding a qualitative component involving an in-depth analysis of the individual workgroups and the dynamics within that may impact both the forecast process and outcome. Subjects to address in the qualitative analysis include the translation of group preferences into outcomes and the consensus formation process, the distribution of resources and expertise among workgroup members, and how these differences between workgroups affect the legitimacy of the forecast process and ultimately the accuracy of the forecasts.

Acknowledgements

Gratitude goes to Stuart Bretschneider for his dedication to analysis of politics and institutions in the forecasting literature, and for encouraging me to submit this paper for publication in the first place, and thanks go to both Stuart and Wilpen Gorr for their very insightful feedback that greatly improved the quality of this paper.

Appendix A. Descriptive statistics for variables used in regression analysis

Variable name	Variable description	N	Mean	Standard deviation	Minimum	Maximum
Dependent variables						
MAPE6	mean absolute percent error: average of first 6 months after forecast was produced	180	3.2%	3.6%	0.03%	20.8%
MAPE12	Mean absolute percent error: average of first 12 months after forecast was produced	180	4.2%	4.4%	0.27%	20.5%
Independent Variables						
REV	indicates whether a given forecast was revised or the previous forecast carried forward	140[a]	0.51	0.50	0	1
STEP	indicates whether or not a given forecast contains a step adjustment, which is a quantifiable estimate of a specific policy or program change	140[b]	0.19	0.39	0	1
PROG	indicates whether or not anecdotal program information/domain knowledge was used to inform the workgroup in deciding on a forecast	140[c]	0.27	0.45	0	1
TIME	the number of years since the CFC was established (0 for pre-CFC)	180	0.78	0.79	0	2
Control variables						
DATAVOL	data volatility measure: the absolute value of the change in the historical actuals from when the forecast was produced to when accuracy was measured	90[d]	1.8%	1.7%	0.01%	7.7%
ADATSA	alcohol and drug addiction treatment support medical	180	0.05	0.22	0	1
AS	adoption support	180	0.05	0.22	0	1
CNAGED	Medicaid categorically needy aged	180	0.05	0.22	0	1
CNBD	Medicaid categorically needy blind/disabled	180	0.05	0.22	0	1
FC	foster care	180	0.05	0.22	0	1
GAU	general assistance unemployable medical	180	0.05	0.22	0	1
GAUX	general assistance grant	180	0.05	0.22	0	1
HCS	home and community long-term care services	180	0.05	0.22	0	1
MNAGED	Medicaid medically needy aged	180	0.05	0.22	0	1
MNBD	Medicaid medically needy blind/disabled	180	0.05	0.22	0	1
MPCADULT	Medicaid personal care for developmentally disabled adults	180	0.05	0.22	0	1
MPKKIDS	Medicaid personal care for developmentally disabled children	180	0.05	0.22	0	1
NH	nursing homes	180	0.05	0.22	0	1
OKIDS	Medicaid categorically needy children < 200% FPL	180	0.05	0.22	0	1
PREGW	Medicaid categorically needy pregnant women	180	0.05	0.22	0	1
SOLT18	state-funded medical for non-Medicaid children	180	0.05	0.22	0	1
SSIA	supplemental security income for aged	180	0.05	0.22	0	1
SSIB	supplemental security income for blind	180	0.05	0.22	0	1
SSID	supplemental security income for disabled	180	0.05	0.22	0	1
TANF	Medicaid TANF/categorically needy family medical	180	0.05	0.22	0	1

[a] This variable is only relevant to the "post-CFC" period because the question of whether or not to revise a forecast more than once a year was not a factor until the CFC was created.

[b] The "pre-CFC" forecasts are excluded from this sample because the documentation is unavailable to determine whether or not a step adjustment was included for each forecast.

[c] This variable is only used in the "post-CFC" period because it is based on CFC technical workgroup meeting notes.

[d] This variable is only relevant to the Medicaid caseloads. Actuals do not change from month to month for the non-Medicaid programs.

References

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, *8*, 69–80.

Armstrong, J. S., & Collopy, F. (1998). Integration of statistical methods and judgment for time series forecasting: Principles from empirical research. In G. Wright, & P. Goodwin (Eds.), *Forecasting with judgment*. Wiley (pp. 269–293).

Ascher, W. (1978). *Forecasting: An appraisal for policy-makers and planners*. Baltimore: Johns Hopkins.

Bretschneider, S., & Gorr, W. (1987). State and local government revenue forecasting. In S. Makridakis, & S. Wheelwright (Eds.), *The handbook of forecasting* (2nd ed.). New York: Wiley (pp. 118–134).

Bretschneider, S., & Gorr, W. (1989). Forecasting as a science. *International Journal of Forecasting*, *5*, 305–306.

Bretschneider, S., Gorr, W., Grizzle, G., & Klay, E. (1989). Political and organizational influences on forecast accuracy: Forecasting state sales tax receipts. *International Journal of Forecasting*, *5*, 307–319.

Carbone, R., Anderson, A., Corriveau, Y., & Corson, P. (1983). Comparing for different time series methods the value of technical expertise, individualized analysis, and judgmental adjustment. *Management Science*, *29*, 559–566.

Frank, H., & Wang, X. (1994). Judgmental vs. time series vs. deterministic models in local revenue forecasting: A Florida case study. *Public Budgeting and Financial Management*, *6*(4), 493–517.

Goodwin, P., & Wright, G. (1993). Improving judgmental time series forecasting: A review of guidance provided by research. *International Journal of Forecasting*, *9*, 147–161.

Jenkins, G. M. (1982). Some practical aspects of forecasting in organizations. *Journal of Forecasting*, *1*, 3–21.

Jones, V., Bretschneider, S., & Gorr, W. (1997). Organizational pressures on forecast evaluation: Managerial, political, and procedural influences. *Journal of Forecasting*, *16*, 241–254.

Kahn, K. B., & Menzer, J. T. (1994). The impact of team-based forecasting. *Journal of Business Forecasting*, *13*(2), 18–21 (Summer).

Kamlet, M. S., Mowery, D. C., & Su, T. (1987). Whom do you trust? An analysis of executive and congressional economic forecasts. *Journal of Policy Analysis and Management*, *6*, 365–384.

Klay, E., & Grizzle, G. (1986). *Revenue forecasting in the states: New dimensions of budgetary forecasting*. Presented at the National Conference of the American Society for Public Administration, Anaheim, CA.

Klay, E., & Grizzle, G. (1992). Forecasting state revenues: Expanding the dimensions of budgetary forecasting research. *Public Budgeting and Financial Management*, *4*, 381–405.

Klay, E., & Zingale, J. (1980). *Revenue estimating as seen from an administrative perspective*. Presented to the 35th Annual Conference on Revenue Estimating. National Association of Tax Administrators.

Lawrence, M. J., Edmundson, R. H., & O'Connor, M. J. (1986). The accuracy of combining judgmental and statistical forecasts. *Management Science*, 32.

Lawrence, M., & O'Connor, M. (1995). The anchor and adjustment heuristic in time-series forecasting. *Journal of Forecasting*, *14*, 445–464.

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*, 149–168.

Lobo, G. J., & Nair, R. D. (1990). Combining judgmental and statistical forecasts: An application to earnings forecasts. *Decision Sciences*, *21*, 446–460.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 111–153.

Mathews, B. P., & Diamantopoulos, A. (1989). Judgmental revision of sales forecasts: A longitudinal extension. *Journal of Forecasting*, *8*, 129–140.

McNees, S. K. (1990). The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting*, *6*, 287–299.

Mocan, H. N., & Azad, S. (1995). Accuracy and rationality of state general fund revenue forecasts: Evidence from panel data. *International Journal of Forecasting*, *11*, 417–427.

Remus, W., O'Connor, M., & Griggs, K. (1995). Will reliable information improve the judgmental forecasting process? *International Journal of Forecasting*, *11*, 285–293.

Sanders, N. R. (1992). Accuracy and judgmental forecasts: A comparison. *OMEGA: The International Journal of Management Science*, *20*, 353–364.

Schultz, R. L. (1992). Fundamental aspects of forecasting in organizations. *International Journal of Forecasting*, *7*, 409–411.

Shkurti, W. J., & Winefordner, D. (1989). The politics of state revenue forecasting in Ohio, 1984–1987: A case study and research implications. *International Journal of Forecasting*, *5*, 361–371.

Steen, J. G. (1992). Team work: Key to successful forecasting. *Journal of Business Forecasting*, 22 (Summer).

Voorhees, W. (2000). *The impact of political, institutional, methodological, and economic factors on forecast error*. PhD dissertation, Indiana University.

Welch, E., Bretschneider, S., & Rohrbaugh, J. (1998). Accuracy of judgmental extrapolation of time series data: Characteristics, causes, and remediation strategies for forecasting. *International Journal of Forecasting*, *14*, 95–110.

White, H. R. (1986). *Sales forecasting: Timesaving and profit making strategies that work*. London: Scott, Foresman and Company.

Wolfe, C., & Flores, B. (1990). Judgmental adjustment of earnings forecasts. *Journal of Forecasting*, *9*, 389–405.

Biography: Elaine DESCHAMPS is Senior Forecaster for the Washington State Caseload Forecast Council. She received her PhD in political science from Indiana University. Her research interests include political and organizational influences on forecast processes and outcomes, government forecasting, and Medicaid forecasting.