Updated March 7, 2013

Attachment A

Test of Three Methodologies for Forecasting MassHealth Caseloads

This document describes:

- Three methodologies for forecasting MassHealth caseloads;
- A pseudo out-of-sample forecast test of the relative effectiveness of the methodologies;
- Interpretation of the findings; and
- Suggestions for proceeding.

Three Modeling Methodologies

The Existing Methodology, Method B

The existing methodology, developed by the Center for Health Policy & Research at the University of Massachusetts Medical School in 2004, assumes that future caseload growth for each population group is best modeled as a continuation of recent growth. This very simple approach was settled on because caseloads did not – and still do not – appear to be statistically related to indicators of trends in either demographics or the economy. In fact, the total caseload in recent years has exhibited essentially constant linear growth, with only minor fluctuations around the trend which appear to be random.

Although the total caseload has exhibited constant growth, individual population group caseloads have not. For most population groups, rates of growth change frequently, often with abrupt discontinuities. Therefore the critical strategy of the existing method involves selecting the most recent period of time for which growth rates were stable, using a method called compounded monthly growth rates (CMGRs) (Center for Health & Policy Analysis, 2004). The method involves a graphical analysis and a subjective choice of choosing the beginning month for the most recent period of stable growth. Although the choice is subjective, and can result in different choices by different analysts, our experience is that the differences between analysts' choices are minor, usually a few months or less.

Once the time period for estimation of the model has been chosen, the current methodology involves estimating a linear regression of the caseload on a time trend. The forecast of the caseload is simply the extrapolation of the estimated linear relationship.

The Autoregressive Model, Method A

An alternative strategy for estimating simple growth models is a time series autoregression. Since caseloads tend to grow over time, the autoregressive model is estimated on the first difference of the natural logarithm of the caseload. The key parameter in this model, the constant, is the trend growth rate. The autoregressive terms, if any, account for the dynamics of short-run deviations from the trend rate of growth. Autoregressive models produce forecasts in which the growth in the initial periods (initial months, in this context) of the forecast is influenced by recent deviations from the trend rate of growth. After several periods, growth returns to the estimated trend rate.

As in the existing methodology, a key strategy in estimating the autoregressive model is to select the period of recent history for which the growth rate is estimated. In this case, we used the QLR test statistic to select the longest period of time for which the growth rate was statistically constant (Stock and Watson, section 14.7). This strategy typically resulted in selecting an historical period for estimation that was much longer than that selected using Method B.

The autoregressive order of the model was chosen using the AIC and BIC criteria, as described in Stock and Watson, section 14.5.

The Linear Model, Method C

A second alternative strategy involves selecting the time period for estimation using the QLR statistic as in Method A, and then estimating a linear regression of caseload on time as in Method B. Like Method B, the forecast is constructed as a linear extrapolation of the estimated regression line, and the constant of the regression line is adjusted so that the extrapolation proceeds from the caseload level at the end of the estimation period.

The Differences between the Methods

The most important difference between the methods, based upon our analysis, was that the alternative methods A and C selected much longer time periods than the existing methodology, Method B. Methods A and C selected an average estimation period of 33 months over which to estimate the model, while method B only selected an average of 11 months.

There are two differences between methods A and C. One is that Method A assumes that the best simple model for caseload growth is a constant growth *rate* – a result of using the logarithm of the caseload rather than the caseload itself, while Method C (and the existing methodology B) assumes that the best simple model for caseload growth is a constant *change* per period in the caseload. We would expect this to be a minor difference.

The second difference is that Method A, because of its autoregressive structure, will produce forecasts whose rates of growth in the first several periods of the forecast may differ from the trend rate estimated by the model, while Method C (and the existing methodology B) will produce forecasts with constant changes throughout the forecast period. We would also expect these differences to be minor over the course of a two-year forecast, especially given the low-order of autoregressivity exhibited in the models. Twenty-eight of 58 population groups in our analysis did not exhibit any autoregressivity, and 21 of the remaining groups exhibited only first-order autoregressivity.[1]

Comparing the Methods Using Pseudo Out-of-Sample Forecasts

The relative effectiveness of the methodologies was compared using pseudo out-of-sample forecasts (Stock and Watson, pp. 561-563). The idea is to estimate the models over a period of history that ends in the past, say, at time T, and then to use the estimated models to forecast the "future" past, say from T+1 to T+24. Since the actual caseloads are known over this pseudo forecast period, the errors in the forecasts can be calculated and the competing methods can be compared. The extent to which a forecast is in error is summarized in a single number, the root mean squared forecast error (RMSFE). This is calculated by summing the squared deviations of each period's forecast caseload from the actual caseload, taking the average of this sum, and then taking its square root. The RMSFE can be interpreted as the "average" or "typical" error in each period that can be expected in a forecast. In comparing two alternative models – or methodologies – the one with the lower RMSFE is preferred.

In implementing this strategy, for "history" we used the period July 2006 – the earliest period for which caseload month data were available – to June 2009, so that T=June 2009. For each population group, we then estimated three models, one using Method A, one using the existing methodology B, and a third using Method C; and then used each to forecast caseload over the period July 2009 to June 2011. The RMSFE for each forecast was then computed. The results are displayed in Table 1.

According to these out-of-sample forecasts, no methodology clearly dominates. Method A just barely out-performed the existing Method B, with a lower RMSFE than Method B in 30 of the 58 population groups. Method C's advantage over Method B was somewhat larger, with a lower RMSFE in 35 of the 58 population groups. In comparing Methods A and C against each other, each had a lower RMSFE than the other in 29 of the 58 population groups.

Comparing the Forecast of the Total Caseload versus the Sum of the Forecasts of the Population Groups

Once forecasts of the individual population groups are obtained, these can be summed to obtain a single forecast of the total MassHealth caseload. However, the forecast of the total caseload could also be obtained by estimating a model on the total caseload, and then by using the total caseload model to forecast total caseload. We have no a priori reason to prefer one method over the other in making a forecast of the total caseload, so we also performed pseudo out-of-sample forecasts for the total

[1] Three of MassHealth's 61 population groups have very few or no members assigned to them. For this reason, only 58 of the population groups were included in this analysis.

caseload by constructing a total caseload model using each methodology. We then constructed the RMSFE's for each of these three total models and also calculated the RMSFE's for the sum of the 58 population group forecasts. The RMSFE's are displayed in Table 2.

Comparing the pseudo out-of-sample forecasts of the total caseload model versus the sum of population groups, the total caseload model performed better for both Methods A and C, but not for Method B. Among the three total caseload models, Method A had the lowest RMSFE, followed by Method C.

Interpretation of these Findings

The pseudo out-of-sample forecasts suggest a slight preference for either of the "new" Methods A or C over B, but the preference is slight. The primary finding is that all three methods performed badly in forecasting many of the population groups. In many cases, this is because the forecast period exhibited discontinuous changes. In some cases, this involved shifting of cases from one population group to another within the same budget group cluster. Using the QLR statistic to choose the beginning period of model estimation is helpful, but there were a few population groups in which the indicated estimation period appeared to contain a significant change in trend. In these cases expert judgment should be used in addition to the QLR statistic to choose the estimation period. There were also a few population groups in which the BIC and AIC criteria indicated a high autoregressive order. In two of the three cases, this appeared to result from a very short estimation period, and in those cases the BIC and AIC criteria should not be trusted. A zero-order autoregressive model should be estimated instead.

The experience of this exercise suggests that the best procedure for estimating simple growth models would be to combine Methods A or C with expert judgment, and therefore to use the statistical diagnostics as guides rather than rigid rules.

It is very likely that the total caseload model out-performed the sum of individual population groups in Method A and Method C because of large forecast errors in a few population groups for which expert judgment in the selection of the estimation period may have substantially improved the forecast. This means that a more definitive pseudo out-of-sample test would involve comparing the total caseload model against individual population group models estimated with a combination of expert judgment and Method A or Method C.

For forecasting purposes, use of the total caseload model would still require the estimation of individual population group models in order to distribute the total caseload forecast to the individual population groups. The best procedure is probably to compare the total caseload forecast to the sum of the population groups to see if the forecasts are consistent with each other; and if not, to inspect the individual population group models to identify groups with suspect forecasts.

Attachment A
The most striking observation from this exercise has been the instability of the trends of the individual population groups, even after accounting for shifting from one population group to another within budget group. One would expect caseload trends to respond gradually to changes in demographics, with changes around these trends that are coherent with cyclical changes in the economy. The changes in trend, however, appear to be too abrupt and too short in duration to be the result of demographic trends and business cycles. The changes in trend may be dominated by changes in policy and administrative actions, such as changes in rates of re-determinations and processing. These could account for the frequency and (lack of) duration of trends that we are observing, and also explain why the time series methodology, which happens to be suggesting longer estimation periods, tends to perform marginally better than the existing methodology.

Suggestions for Proceeding

What we have learned suggests two courses of action for improving caseload forecasts:

1. Investigate what information is available about changes in rates of re-determinations and processing, and case openings and closings, and see if these help explain changes in caseloads. If these do, then it is possible that the caseloads adjusted for these administrative actions might correlate better with economic indicators.
2. Use existing data sources, such as the American Community Survey, the Current Population Survey, national surveys of health, and information from the Division of Employment and Training and the Health Connector to estimate the size and characteristics of the eligible population and likely trends in the size of the eligible population. These may be very helpful in forecasting changes in the trend of caseload growth.

Alan Clayton-Matthews , Susan Jureidini, and Mohamed Sesay

March 7, 2013

References

Center for Health Policy & Research, University of Massachusetts Medical School. (2004). Stepwise Methodology for Creating the MassHealth Caseload Member Month Forecast.

Stock, J. H., & Watson, Mark W. (2011). *Introduction to Econometrics* (3rd ed). Boston: Pearson Education, Addison Wesley.

Attachment A

Table 1: Pseudo Out-of-Sample Forecast Results, Three Methods

		Pseudo Out-of-Sample Forecast									
		Method A				Method B			Method C		
Pop Group	Title	AR Order	Beginning Date for Stable Mean Difference	Mean Error	RMSFE	Begin Date for Stable Trend	Mean Error	RMSFE	Break in Trend	Mean Error	RMSFE
0	Other	0	2006m8	181.1	214.2	2008m10	-114.3	120.9	2006m8	872.8	1,015.4
1	HMO Disabled Children	0	2006m8	253.0	452.4	2007m12	701.9	906.4	2006m8	795.4	1,006.3
2	HMO Disabled Adults	1	2006m9	213.7	1,890.8	2008m1	511.6	2,130.2	2006m8	1,089.4	2,555.7
3	PCC Disabled Children	0	2006m8	487.3	506.2	2008m4	577.0	604.9	2006m8	552.7	578.0
4	PCC Disabled Adults	1	2006m9	-455.4	736.7	2008m9	-1,116.0	1,296.1	2006m8	22.7	631.6
5	TPL Disabled Children	0	2006m8	-86.9	173.1	2008m8	-70.9	156.6	2006m8	-52.9	140.8
6	TPL Disabled Adults	0	2006m8	-1,152.0	1,289.2	2008m6	-1,020.3	1,120.2	2006m8	-786.3	861.2
7	FFS Disabled Children	0	2008m7	-1,264.0	1,631.8	2008m8	-1,334.8	1,708.4	2008m7	-1,289.8	1,658.7
8	FFS Disabled Adults	2	2007m1	-1,759.4	4,000.3	2008m10	-2,568.4	4,695.1	2006m1	-1,821.7	4,018.0
9	HMO Non-Disabled Children	1	2006m9	-5,472.4	6,750.1	2008m7	-2,440.5	4,210.8	2006m8	768.2	2,653.1
10	HMO Non-Disabled Adults	1	2006m9	-4,371.5	5,696.9	2008m7	-1,861.7	3,547.9	2006m8	983.2	2,227.2
11	PCC Non-Disabled Children	1	2006m9	7,775.0	8,284.1	2008m7	10,067.0	10,751.2	2006m8	9,332.6	9,930.7
12	PCC Non-Disabled Adults	1	2006m10	1,610.6	2,042.6	2008m11	3,382.5	3,896.0	2006m9	2,233.0	2,619.9
13	TPL Non-Disabled Children	0	2006m10	925.7	1,308.6	2008m11	2,731.7	3,323.2	2006m1	963.1	1,352.1
14	TPL Non-Disabled Adults	0	2007m12	1,840.6	2,437.5	2008m8	2,732.7	3,454.9	2007m1	2,434.3	3,120.3
15	FFS Non-Disabled Children	1	2006m9	-1,088.6	2,090.8	2008m7	-827.9	2,565.1	2006m8	-3,512.9	3,796.3
16	FFS Non-Disabled Adults	1	2006m9	3,777.7	4,399.5	2008m7	-4,154.2	4,198.0	2006m8	-5,119.1	5,170.3
17	FFS Newborns	0	2006m8	4,354.4	4,488.5	2008m11	3,949.7	4,089.9	2006m8	4,518.5	4,653.1
18	Community Seniors	0	2006m8	-188.9	290.7	2008m11	2,144.1	2,422.8	2006m8	-329.9	436.1
19	Institutional Seniors	2	2006m10	638.0	699.3	2008m10	321.0	369.9	2006m8	813.4	882.9
20	Buy-in (Aged)	0	2006m8	1,064.6	1,270.2	2008m4	-465.8	542.8	2006m8	1,362.0	1,606.3
21	Buy-in (Disabled)	6	2007m2	484.0	491.4	2008m2	2,233.1	2,551.6	2006m8	431.6	493.3
22	HMO Basic	1	2006m9	-1,208.4	1,272.0	2008m11	-820.0	892.5	2006m8	-993.7	1,053.3
23	PCC Basic	0	2006m8	-633.6	716.1	2008m11	-717.7	784.7	2006m8	-413.9	585.4
24	Unenrolled Basic	0	2006m8	-3,588.8	3,662.8	2008m7	-3,733.6	3,808.1	2006m8	-3,759.1	3,833.9
25	EAEDC	0	2006m8	846.7	967.2	2008m3	921.5	1,052.7	2006m8	807.3	922.4
26	Basic Premium Assistance	0	2006m8	-6.2	6.7	2008m7	-8.8	9.6	2006m8	-5.2	5.6
27	CommonHealth Working Adults	0	2006m11	80.2	125.8	2008m7	296.1	374.2	2006m1	384.9	474.3
28	CommonHealth Non-Working Adults	1	2006m9	-1.8	30.8	2008m8	-130.3	156.3	2006m8	30.5	37.6
29	CommonHealth Children	0	2006m8	220.4	274.3	2008m5	207.2	259.5	2006m8	229.7	284.9
30	HMO Family Assistance Children	1	2006m10	-8,803.9	9,688.4	2008m8	-5,955.3	6,406.6	2006m9	-5,729.8	6,156.2
31	HMO Family Assistance Adults	6	2008m7	-5,613.3	13,776.8	2008m8	409.3	486.1	2008m1	382.1	455.2
32	PCC Family Assistance Children	1	2006m10	-2,042.8	2,194.0	2008m8	-846.2	933.8	2006m9	-1,193.6	1,283.4
33	PCC Family Assistance Adults	1	2006m9	-579.6	659.9	2008m8	68.9	82.8	2006m8	-17.8	91.0
34	Unenr Family Assistance Children	2	2006m10	254.7	439.1	2008m8	-1,425.6	1,628.5	2006m8	33.7	303.1
35	Unenr Family Assistance Adults	1	2006m9	387.1	391.5	2007m7	-294.8	295.6	2006m8	-318.0	319.4
36	FA IP Premium Assistance Children	0	2007m12	-1,285.4	1,469.0	2007m11	-1,160.7	1,320.3	2007m1	-1,059.8	1,205.5
37	FA IP Premium Assistance Adults	6	2008m10	-262.8	297.8	2008m11	-184.7	201.6	2008m4	-54.5	72.2
38	Limited Children	1	2006m11	893.9	1,045.8	2008m11	248.7	378.0	2006m1	1,240.9	1,385.7
39	Limited Adults	1	2006m11	-2,690.0	3,305.6	2008m07	1,648.8	1,894.6	2006m1	-991.4	1,644.5
40	Prenatal/Presumptive Eligibility	1	2006m9	-80.4	92.1	2008m07	-309.2	325.6	2006m8	-172.1	178.3
41	LTC <65	3	2006m11	-57.1	79.7	2008m11	-78.9	95.6	2006m8	-36.5	71.6
42	Kaleigh-Mulligan/A4E Adoption	1	2006m9	-8.4	13.0	2008m10	12.4	14.6	2006m8	-7.2	11.1
46	Standard Disabled Adult - Prem. Assist.	0	2006m8	-1.9	5.7	2008m9	-5.2	8.7	2006m8	1.0	4.3
47	Standard Non-Disabled Children - Prem.	2	2006m11	-1,659.9	1,927.0	2008m9	-1,571.3	1,806.8	2006m9	-1,102.0	1,271.8
48	Standard Non-Disabled Adult - Prem.	0	2006m8	-729.8	855.7	2008m8	-595.0	687.1	2006m8	-428.6	497.8
49	Commonhealth Premium Assistance	1	2006m12	-91.1	119.0	2008m10	75.0	84.0	2006m11	-90.8	111.3
50	PA Essential	1	2006m9	136.1	144.4	2008m10	107.9	114.6	2006m8	176.6	190.2
51	PCC Essential	1	2006m9	-2,590.4	10,325.8	2008m11	-5,230.0	12,142.1	2006m8	-91.6	8,545.0
52	SCO Community	0	2008m5	1,472.9	1,646.0	2008m8	1,814.9	2,053.2	2008m5	1,775.3	2,008.0
53	SCO Institutional	0	2007m5	-939.7	1,095.1	2008m11	-139.3	146.8	2007m5	-339.9	367.2
54	MCO CommonHealth Working Adult	0	2006m8	178.3	319.8	2008m10	121.3	269.8	2006m8	213.3	354.6
55	PCC CommonHealth Working Adult	2	2006m10	-120.4	302.3	2008m10	-324.6	482.4	2006m8	-89.7	265.8
56	MCO CommonHealth Non Working Adult	0	2006m8	8.2	18.3	2008m11	-5.0	12.3	2006m8	7.2	17.6
57	PCC CommonHealth Non Working Adult	0	2006m8	-9.4	19.9	2008m10	2.8	18.3	2006m8	80.8	95.8
58	MCO CommonHealth Children	0	2006m8	-64.1	70.8	2008m11	-64.6	71.4	2006m8	-63.9	70.7
59	PCC CommonHealth Children	0	2006m8	158.0	175.0	2008m10	164.6	182.6	2006m8	170.3	189.2
60	MCO Essential	0	2006m10	6,873.2	11,061.8	2008m11	9,620.8	14,835.1	2006m10	9,533.4	14,744.5

Attachment A

Table 2: RMSFEs of Individual Models and Total Caseload Models

	RMSFE of Individual Models*	RMSFE of Total Caseload Model**
Method A	19,679	4,015
Method B	7,601	13,157
Method C	14,975	8,167

*58 separate forecasts made, errors summed across forecasts to calculate one rmse

**58 groups added together to make one total caseload forecast