Measurement Error Analysis of Forecasts from the Consensus Economic Forecasting Commission

March 26, 2019

Maine Department of Administrative and Financial Services
Office of the State Economist

Sujita Pandey
Economic Analyst

Introduction

The Consensus Economic Forecasting Commission (CEFC), beginning in 1992, provides the Governor, the Legislature, and the Revenue Forecasting Committee with analyses, findings, and recommendations representing state economic assumptions relevant to revenue forecasting. The accuracy of these economic estimates is critical for ensuring an accurate revenue forecast, on which the state budget is based.

In 2009, the first edition of this report evaluated forecast accuracy of three economic variables – wage and salary employment (WSE), personal income (PI), and consumer price index (CPI) – between 1993 and 2008. Four types of average error were calculated: mean error (ME), mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE). Each type of error was calculated twice: once for all years and once with outlier years removed. Each type of error was also calculated for several different categories: year, number of years out, economic conditions (based on employment, income, and CPI), and business cycle (as determined by the National Bureau of Economic Research). The report found while there was room for improvement, forecast errors had been lower in later years. Generally, CPI growth rates were more accurate than employment and income growth rates. Analysis also indicated that there was no particular error bias when broken down into different categories such as economic conditions and business cycle. However, growth rates were more accurate if they were made in the short term compared to four or five years out. Finally, the report also recommended that future analysis should perform a forecast error analysis of the components of personal income and expansion of the database that tracks CEFC forecasts.

This report follows up on the previous analysis and performs forecast error analysis on CEFC forecasts between 2008 and 2017. The report evaluates the CEFC forecast's accuracy using various error measurement metrics. In doing so, it examines trends and if there are any systematic errors in forecasting. While complete accuracy is unattainable, addressing systematic errors helps the Commission correct bias and provide accurate forecasts. Prior analysis of the forecast accuracy focuses on the CPI, WSE, and PI. This report adds another component of personal income as suggested by the first edition, wages and salaries (WS), as a variable of interest.

Data and Methodology

Data for this analysis was collected from historical CEFC reports. This analysis focuses on annual growth rates for four CEFC forecast variables: CPI; WSE; PI; and WS. To measure relative performance of the CEFC forecasts, analysis was broken down into four categories: economic condition; business cycle; years out; and forecast by year. The categories above were based on actual growth rates as opposed to CEFC forecasts. As such, the breakdown attempts to capture the actual status of the economy when the Commission was forecasting. Economic condition relates to a variable's growth rate compared to the previous period. If the growth rate was greater than the previous period, economic condition was designated as 'up'; if the growth rate was lower than the previous period, economic condition was designated as 'down'; and if the growth rate was not only lower but also was negative, the economic condition was designated was 'negative'. Business cycle designation was determined using National Bureau of Economic Research's recession documentation. Years out refers to the number of years between the date the forecast was made and the forecasted year. The CEFC often forecasts growth rates as far as five years out. Additionally, due to lagged release of actual data, the CEFC often forecasts growth rates for the previous year, which is depicted as "-1" year in the tables below. Finally, forecast by year corresponds to a breakdown of the forecasts by calendar year.

This report examines error bias as well as the magnitude of the error using five different metrics for the four CEFC growth rates, namely: Mean Error (ME), Mean Absolute Error (MAE); Mean Absolute Percentage Error (MAPE); Root Mean Squared Error (RMSE); and Mean Absolute Squared Error (MASE).

The measures used in this analysis are described as:

- ME indicates bias – whether errors tended to be positive or negative. This provides information on whether forecasts tended to be higher than actual growth rates (resulting in a positive ME) or lower than actual growth rates (resulting in a negative ME).
- MAE indicates the degree of variance in the errors instead of the bias. A smaller MAE is better than a large one, but MAEs cannot be compared across different bases. The MAE keeps the same units as the original data.
- MAPE also indicates the degree of variance in the errors, but because it uses percentages, the MAPE can be used to compare errors across different bases.
- RMSE error is similar to MAE, in that it also indicates the degree of variance in the errors using the same units as the original data. However, by squaring the errors and then taking a square root, rather than taking their absolute values, RMSE gives disproportionately higher weight to large errors. It can also be used, in conjunction with the MAE, to indicate the degree of variance in the individual errors. When all the errors are of the same magnitude, RMSE will equal MAE. The greater the variance in errors, the greater the difference will be between RMSE and MAE (with RMSE always the same or greater than MAE).
- MASE is scale invariant – meaning it can be used to compare forecast accuracy across different data sets with different scales. This measure also equally penalizes positive and negative as well as small and large errors. A value of greater than one indicates that in-sample one step forecasts using last period's actuals as this period forecast (also known as a naïve forecast) performs better than CEFC forecasts.

Besides using different categories to analyze forecast accuracy, two different sample sizes were used to measure forecast accuracy for each category. The first sample includes all observations. However, the second sample drops outliers, which are years with MAPE greater than 100%, to perform forecast accuracy analysis (Table 5 in Appendix). Each variable's outlier years were calculated. Using two different samples allows us to proxy a robustness analysis of the CEFC forecasts. CPI's outlier years were 2009 and 2015 (i.e. MAPE was greater than 100 percent), PI outlier years were 2009 and 2013 and WS outlier years were 2009 and 2010. The analysis without outliers excludes WSE because outlier years for WSE were 2008, 2010, 2011, 2012, 2013 and 2014. Doing an analysis without outliers would only leave us with 2009, 2015, 2016 and 2017. Since the number of outlier years is greater than non-outlier years, we exclude WSE from our analysis without outliers.

Finally, an Access database was built during the process of this analysis to allow for expedited data access and reliable data storage for future analysis. This database houses all variables reported in the CEFC report between 2007 and 2018. In addition to that, data for three variables (CPI, PI, and WSE) going back to 1992 is stored in this database. The goal is to update this database on a timely basis so that future analysis can be conducted by simply extracting variables of interest from this database.

Results

In general, mean errors, albeit small, were negative, meaning that the forecasts were generally underestimating the actual growth rates. Mean errors year by year do not reflect an evident pattern for any of the variables except one. Forecasts for WSE for calendar years 2016 and later seem to be overestimated (mean error is positive). Breakdowns of MAPE and MASE represented in figures 1 and 2 indicate forecasts in recent years have been more accurate than previous years. Most of the volatility around forecast errors can be attributed to the aftermath of the recession. To allow for meaningful figures, MAPE of higher than 100 percent was fixed at 100 percent and MASE of higher than 3 was fixed at 3 (highest MAPE was 3730 percent and highest MASE was 20). Compared across all the variables, forecasts for WSE had the highest errors and outliers. For the full sample, a comparison of MAPE across all variables indicates that forecasts were more accurate when the economic condition was improving (i.e. 'up'). Besides forecasts for WSE, CEFC forecasts during contractions were generally more accurate than expansions. Finally, excluding PI, CEFC forecasts for the immediate future were more reliable than forecasts several years out for all other variables.

Figure 1

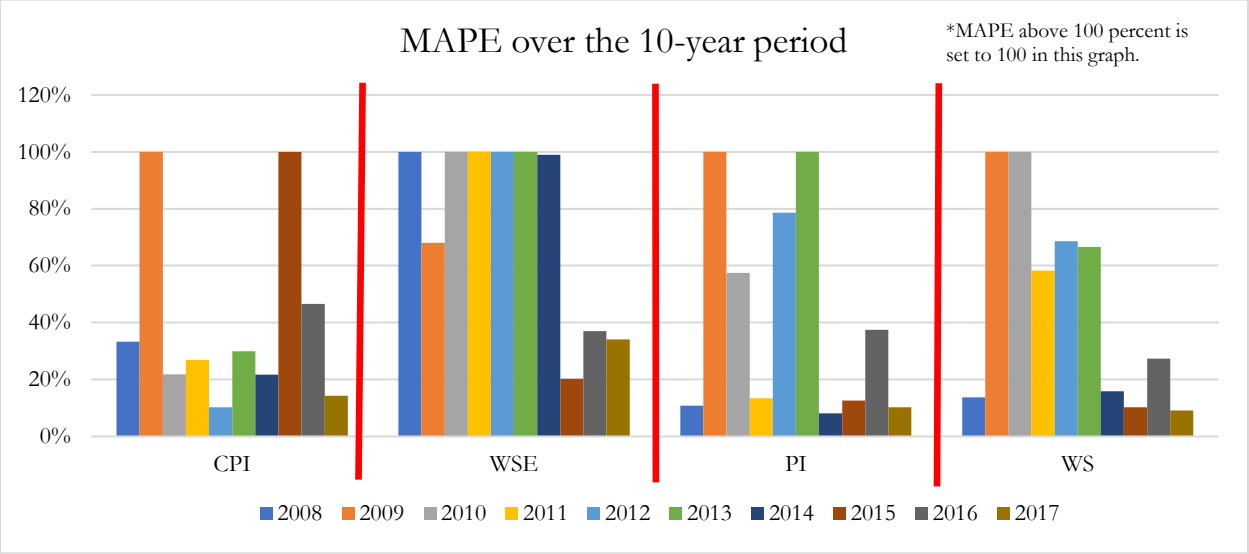
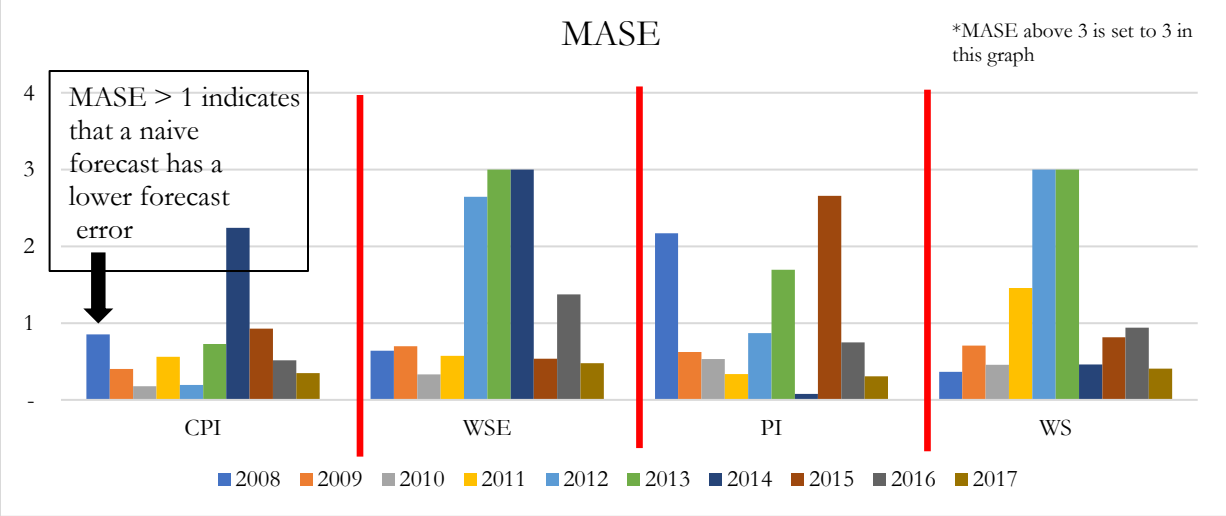


Figure 2



MAE

When considering the full sample**,** MAE indicates that the CEFC forecasts during 'negative' economic condition were less reliable for all the variables. MAE for WSE, PI, and WS indicate that measurement errors during contractions are lower than expansions. When forecasting growth rates for CPI and WSE between -1 to 5 years out, forecasts were more accurate during the short term, -1 to 0 years, rather than forecasts for years that were further out. For PI and WS forecasts 2 and 5 years out were less accurate than other years.

Dropping outlier years sometimes results in no change in an error measurement metric as reflected in the tables below. This is because some of the economic conditions, business cycles, or years out forecasts did not consist of any outliers. For example, MAE for CPI does not change between the full sample and the sample without outliers for an 'up' economic condition because this economic condition did not have any outliers as defined in the methodology section. The sample without outliers results in dropping 'negative' economic condition for all three variables. CEFC forecasts for PI are more accurate during 'up' economic condition and CEFC forecasts for CPI and WS are more accurate during a 'down' economic condition than an 'up' economic condition. In terms of the business cycle, CPI forecasts during expansions are more accurate than during contractions. Conversely, PI and WS forecasts during contractions are more accurate than expansions. In terms of years out, CPI, PI and WS exhibit similar forecast error patterns which is forecasts -1 to 0 years are more accurate than years that are further out than 1 year. Forecasts tend to have larger errors 2-4 years out and forecast 5 years out has a lower forecast error than 2-4 years out.

Table 1

Full Sample					Sample without Outliers			
Category	**CPI**	**WSE**	**PI**	**WS**	**Category**	**CPI**	**PI**	**WS**
Economic Condition								
Down	0.0070	0.0038	0.0106	0.0073	Down	0.0033	0.0106	0.0073
Negative	0.0169	0.0132	0.0332	0.0438	Negative	n/a	n/a	n/a
Up	0.0050	0.0052	0.0061	0.0103	Up	0.0050	0.0061	0.0093
Cycle								
Contraction	0.0128	0.0037	0.0043	0.0044	Contraction	0.0128	0.0043	0.0044
Expansion	0.0062	0.0065	0.0134	0.0127	Expansion	0.0041	0.0083	0.0092
Years Out								
-1	0.0002	0.0017	0.0091	0.0069	-1	0.0002	0.0075	0.0062
0	0.0036	0.0029	0.0094	0.0077	0	0.0034	0.0077	0.0061
1	0.0084	0.0074	0.0123	0.0135	1	0.0054	0.0069	0.0074
2	0.0082	0.0104	0.0167	0.0168	2	0.0050	0.0096	0.0084
3	0.0068	0.0082	0.0140	0.0142	3	0.0055	0.0091	0.0122
4	0.0071	0.0054	0.0149	0.0136	4	0.0048	0.0082	0.0136
5	0.0102	0.0008	0.0170	0.0064	5	0.0049	0.0036	0.0064

MAPE

Since MAPE is scale invariant we can compare across variables and different categories. In terms of the full sample, the highest MAPE was reported during 'negative' economic condition. The lowest MAPE was reported for PI during 'up' economic condition. CEFC forecasts for PI and WS during contractions in the business cycle had the lowest MAPE. In terms of forecasts -1 years out, CPI had the lowest forecast error and PI had the highest forecast year. Forecast errors generally get larger the further the forecasted year is from the CEFC forecast date. However, there are unique instances, such as MAPE for WSE and WS 5 years out, where MAPE is lower than forecasts 2 to 3 years out.

After dropping the outlier years, MAPE declines considerably. Compared to the full sample, MAPE is lower for the samples without outliers. For CPI and WS, CEFC forecasts during 'down' economic condition were more accurate than 'up' economic condition. However, for PI, CEFC forecasts during 'up' economic condition had lower forecast errors. CEFC forecasts for CPI during expansions in the business cycle were more accurate whereas forecasts for PI and WS were more accurate during contractions in the business cycle. CEFC forecast 2 to 4 years out had higher forecast errors than -1 to 1 and 5 years out.

Table 2

Full Sample					Sample without Outliers			
Category	**CPI**	**WSE**	**PI**	**WS**	**Category**	**CPI**	**PI**	**WS**
Economic Condition								
Down	416.63%	61.73%	42.15%	23.46%	Down	20.56%	42.15%	23.46%
Negative	474.56%	-192.11%	2837.28%	145.84%	Negative	n/a	n/a	n/a
Up	23.72%	89.24%	19.24%	51.28%	Up	23.72%	19.24%	38.36%
Cycle								
Contraction	33.27%	451.50%	10.84%	13.72%	Contraction	33.27%	10.84%	13.72%
Expansion	190.74%	52.01%	562.11%	33.68%	Expansion	22.36%	30.05%	36.93%
Years Out								
-1	3.33%	17.65%	337.45%	14.69%	-1	1.51%	25.24%	23.47%
0	46.55%	5.65%	268.27%	21.01%	0	14.93%	24.57%	22.06%
1	222.53%	42.57%	478.62%	16.47%	1	22.92%	23.35%	28.89%
2	226.98%	64.22%	633.60%	20.75%	2	28.32%	36.48%	34.42%
3	193.84%	89.64%	574.84%	73.64%	3	30.65%	35.90%	50.82%
4	277.79%	92.63%	782.15%	56.30%	4	24.81%	31.41%	56.30%
5	605.05%	11.10%	1461.02%	24.58%	5	30.85%	8.46%	24.58%

RMSE

Like the other measures above, RMSE is high for all variables during 'negative' economic condition indicating high forecast errors. For WSE, PI, and WS CEFC forecasts during expansions had higher forecast errors than during contractions. Forecast error for CPI remained the same during either business cycle. Forecasts -1 years out had the lowest RMSE. Errors for other years out do not exhibit a consistent pattern.

For the sample without outliers forecast errors for CPI and WS were lower during a 'down' economic condition. For PI, forecast error were lower during an 'up' economic condition. While CEFC forecasts for CPI were more accurate during expansions in the business cycle, CEFC forecasts for PI and WS were more accurate during contractions in the business cycle. Similar to the full sample, forecasts -1 years out had the lowest RMSE. Forecasts seems to be less accurate 1-4 years out and then errors are lower again 5 years out.

Table 3

Full Sample						Sample without Outliers			
Category	**CPI**	**WSE**	**PI**	**WS**		**Category**	**CPI**	**PI**	**WS**
Economic Condition									
Down	0.0099	0.0047	0.0123	0.0084		Down	0.0045	0.0123	0.0084
Negative	0.0205	0.0186	0.0363	0.0507		Negative	n/a	n/a	n/a
Up	0.0066	0.0068	0.0081	0.0138		Up	0.0066	0.0081	0.0127
Cycle									
Contraction	0.0135	0.0049	0.0044	0.0057		Contraction	0.0135	0.0044	0.0057
Expansion	0.0091	0.0099	0.0189	0.0190		Expansion	0.0055	0.0103	0.0123
Years Out									
-1	0.0002	0.0021	0.0114	0.0084		-1	0.0003	0.0081	0.0078
0	0.0052	0.0036	0.0113	0.0096		0	0.0051	0.0085	0.0078
1	0.0119	0.0116	0.0168	0.0215		1	0.0077	0.0078	0.0094
2	0.0112	0.0150	0.0225	0.0270		2	0.0061	0.0121	0.0125
3	0.0085	0.0095	0.0206	0.0180		3	0.0064	0.0122	0.0160
4	0.0093	0.0071	0.0223	0.0161		4	0.0056	0.0112	0.0161
5	0.0128	0.0008	0.0256	0.0084		5	0.0055	0.0036	0.0084

MASE

MASE is scale invariant as well. Therefore, MASE values can be compared across variables and categories. MASE higher than 1 indicates that a naïve forecast would have performed better than the CEFC forecast. In the full sample, 'negative' economic condition has particularly high MASE value compared to other measures. In terms of the business cycle, besides MASE value for WS during contractions in the business cycle, measurement errors are high as well. Forecasts for near-term years, -1 to 1 years, have MASE of less than one.

The sample without outliers has lower MASE value for economic conditions. Dropping outliers does not improve MASE value during contractions in the business cycle. Forecasts for WS did not perform well compared to naïve forecast even at -1 to 0 years out. However, CEFC forecasts for CPI and PI between -1 to 1 years out had lower forecast errors than a naïve forecast. Overall, despite the sample size, it seems like naïve forecast is a reliable benchmark for CEFC members to begin their forecast with and to adjust accordingly. Doing so might address some of the volatility around economic conditions and business cycles.

Table 4

Full Sample						Sample without Outliers			
Category	**CPI**	**WSE**	**PI**	**WS**		**Category**	**CPI**	**PI**	**WS**
Economic Condition									
Down	0.6538	1.2231	0.9388	0.7417		Down	0.3961	0.9388	0.7417
Negative	2.0122	0.5236	1.1301	0.7057		Negative	n/a	n/a	n/a
Up	0.2697	1.5555	0.2802	1.0547		Up	0.1772	0.2802	1.5604
Cycle									
Contraction	0.8523	0.6423	2.1683	0.3658		Contraction	0.8523	2.1683	0.3658
Expansion	0.4985	0.8825	0.6535	0.9262		Expansion	0.4352	0.4535	1.4464
Years Out									
-1	0.0129	0.2122	0.4921	0.5057		-1	0.0234	0.5803	1.0927
0	0.2631	0.3435	0.4662	0.4734		0	0.3345	0.4447	0.8424
1	0.6238	0.8847	0.6084	0.8330		1	0.5299	0.4011	1.0174
2	0.6143	1.1990	0.7503	1.0062		2	0.5219	0.4925	1.2768
3	0.6688	1.3670	0.7082	1.2889		3	0.5680	0.4661	1.8569
4	0.7250	1.5471	0.8709	2.2967		4	0.5424	0.5064	2.2967
5	1.0284	0.2695	1.3450	1.2764		5	0.6652	0.4444	1.2764

Conclusion

Generally, the CEFC had difficulty forecasting growth rates for all variables during 'negative' economic condition. 'Negative' economic condition also accounted for most of the outlier years. Once the outliers were dropped, MAPE drops drastically. For CPI, CEFC forecasts performed better than naïve forecasts, likely because CPI is a national measure and generally exhibits less volatility. WSE's MASE indicates that naïve forecasts would have been more accurate than CEFC forecasts during 'up' and 'down' economic conditions. For WS, 'up' economic condition and expansions also called for naïve forecasts since the value of MASE was greater than one. Finally, for PI, naïve forecasts performed better than the CEFC during contractions.

Given these results, breakdown by years compared to economic conditions and business cycles do not portray a systemic bias in CEFC forecasts. As mentioned earlier, the highest forecast errors were during negative growth condition in the economy, which were also characterized as outlier years. In terms of the business cycle, forecast errors were generally smaller during contractions compared to expansions. This analysis finds that, similar to the first edition, forecasts for immediate years are more reliable than forecasts several years out. There were a few cases where forecasts further out had small forecast errors. A second similarity to the first edition is that in terms of MAPE it appears that CEFC forecasts in recent years have lower errors. This is also corroborated by MASE since CEFC forecasts have a lower measurement error than a naïve forecast. However, since naïve forecasts out-performed some of the commission's forecasts, the commission should consider using naïve forecasts as a benchmark for WSE, PI, and WS forecasts and adjust accordingly. In general, CEFC forecasts during volatile economic conditions, such as the 2008 recession and years following it, have higher forecast errors. In addition to that, in 2011, the CEFC commission members and the forecasting methodology changed. The CEFC, in order to improve forecast stability, decided to give greater weight to the existing CEFC forecast as opposed to starting with the new forecast from one of the national forecasting models. Doing so mirrors a naïve forecast, which in turn seems to have decreased forecast errors in the more recent years. Overall, consistent monitoring of CEFC forecasts allows the committee to ensure accurate forecasts and to be cognizant regarding any forecast bias.

Appendix

Tables in this section shows different metrics used to measure forecast error by year. The results from Table 6 were used to drop outlier years for each variable. Outlier years are highlighted in light grey.

Table 5

MAE - Full Sample				
Category	**CPI**	**WSE**	**PI**	**WS**
2008	0.0128	0.0037	0.0043	0.0044
2009	0.0169	0.0231	0.0261	0.0438
2010	0.0036	0.0095	0.0138	0.0191
2011	0.0085	0.0048	0.0054	0.0117
2012	0.0021	0.0075	0.0165	0.0151
2013	0.0044	0.0062	0.0373	0.0153
2014	0.0035	0.0056	0.0036	0.0055
2015	0.0140	0.0018	0.0053	0.0041
2016	0.0059	0.0046	0.0105	0.0085
2017	0.0030	0.0025	0.0043	0.0037

Table 6

MAPE - Full Sample				
Category	**CPI**	**WSE**	**PI**	**WS**
2008	33.27%	451.50%	10.84%	13.72%
2009	474.56%	68.09%	1307.48%	145.84%
2010	21.78%	170.94%	57.47%	159.44%
2011	26.86%	168.60%	13.39%	58.29%
2012	10.29%	131.94%	78.62%	68.65%
2013	29.97%	103.70%	3729.67%	66.58%
2014	21.76%	98.61%	8.08%	15.82%
2015	1175.76%	20.25%	12.65%	10.21%
2016	46.58%	37.01%	37.42%	27.35%
2017	14.29%	34.07%	10.27%	9.18%

Table 7

RMSE - Full Sample				
Category	**CPI**	**WSE**	**PI**	**WS**
2008	0.0135	0.0049	0.0044	0.0057
2009	0.0205	0.0285	0.0322	0.0507
2010	0.0049	0.0099	0.0147	0.0205
2011	0.0096	0.0060	0.0067	0.0133
2012	0.0035	0.0096	0.0176	0.0176
2013	0.0053	0.0087	0.0386	0.0192
2014	0.0046	0.0064	0.0046	0.0079
2015	0.0158	0.0025	0.0059	0.0045
2016	0.0065	0.0049	0.0107	0.0093
2017	0.0034	0.0029	0.0058	0.0048

Table 8

MASE - Full Sample				
Category	**CPI**	**WSE**	**PI**	**WS**
2008	0.8523	0.6423	2.1683	0.3658
2009	0.4024	0.6976	0.6226	0.7057
2010	0.1790	0.3339	0.5305	0.4555
2011	0.5592	0.5754	0.3347	1.4574
2012	0.1959	2.6464	0.8690	7.5514
2013	0.7259	20.6746	1.6953	15.3141
2014	2.2424	15.1243	0.0790	0.4615
2015	0.9282	0.5353	2.6570	0.8170
2016	0.5142	1.3720	0.7485	0.9421
2017	0.3505	0.4799	0.3082	0.4078