



Decomposition of time-series by level and change



Thomas H. Tessier ^{a,*}, J. Scott Armstrong ^{b,c,**}

^a CFO, eCycle, Inc., 1800 North 12th, Reading, PA 19604, USA

^b The Wharton School, University of Pennsylvania, 700 Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, USA

^c Ehrenberg-Bass Institute, University of South Australia

ARTICLE INFO

Article history:

Received 1 August 2014

Received in revised form 1 March 2015

Accepted 1 March 2015

Available online 14 April 2015

Keywords:

Accuracy

A priori analysis

Conditional regression

Nowcasting

ABSTRACT

This article examines whether decomposing time series data into two parts – level and change – produces forecasts that are more accurate than those from forecasting the aggregate directly. Prior research found that, in general, decomposition reduced forecasting errors by 35%. An earlier study on decomposition into level and change found a forecast error reduction of 23%. The current study found that nowcasts consisting of a simple average of estimates from preliminary surveys and econometric models of the U.S. lodging market, improved the accuracy of final estimates of levels. Forecasts of change from an econometric model and the improved nowcasts reduced forecast errors by 29% when compared to direct forecasts of the aggregate. Forecasts of change from an extrapolation model and the improved nowcasts reduced forecast errors by 45%. On average then, the error reduction for this study was 37%.

© 2015 Published by Elsevier Inc.

1. Introduction

Decomposition for forecasting involves breaking a problem into pieces, forecasting each piece, and then reassembling the forecast pieces. Decomposition allows a forecaster to use different methods and data for each component. Decomposition can be multiplicative, such as forecasting sales by forecasting market size and market share, then multiplying the two components. It can also be additive such as to decompose a sales forecast by region, forecast each region, and then add the regional forecasts.

A meta-analysis by Armstrong, Green and Graefe (2015—in this issue) found 16 studies prior to this current study that examined reduction in forecast error due to decomposition. In all studies, decomposition led to improved accuracy. In the eight studies that assessed the amount of improvement, the average error reduction was 35%. This article examines the additive decomposition of time-series data by level and change. Little comparative research has been done on this type of decomposition.

2. Prior research on decomposition by level and change

Forecasters have long been aware that errors in estimating current levels are common. Morganstern (1963) describes the problems that

economists face in assessing current levels. The errors in estimating current levels are often substantial. Runkle (1998) analyzes deviations between current and revised estimates of quarterly GDP growth from 1961 to 1996. There were upward revisions of as much as 7.5% and downward revisions of as much as 6.2%. Obviously, the errors in estimating levels affect the forecasts. For example, Zarnowitz (1967) reports that about 20% of the error in predicting the next year's GNP in the U.S. arose from errors in estimating the current U.S. GNP figure. Cole (1969) estimated that 40% of the errors for one-year ahead U.S. GNP forecasts are due to errors in estimating the starting levels.

Given the concern over the introduction of error due to poor estimation of the current levels, interest in how to improve the estimates – referred to as “nowcasting” – is strong. A Google Scholar search for “nowcasting” in March 2015 found almost 1700 hits.

One approach for dealing with errors caused by poor estimation of the current status is to make adjustments. Mechanical adjustments, such as adding one-half of the most recent forecast error to the estimate of current level, is a simple, low cost, and objective approach. Another approach is to use judgmental adjustments as they can include recent information that is not already incorporated into the data, such as recent stock-outs for a product. While McNees (1990) found that judgmental and mechanical adjustments each tend to improve accuracy of economic forecasts, the improvements were modest. Moreover, judgmental adjustments are risky as they increase the likelihood that biases would be introduced.

A search for studies that assessed the performance of decomposition forecasting using nowcasting-plus-trend found only one such study, Armstrong (1970). That study analyzed annual sales of photographic equipment averaged over 1955–1960 data for 17 countries by using a cross-sectional regression model. The econometric estimates were

* Corresponding author.

** Correspondence to: J. Scott Armstrong, The Wharton School, University of Pennsylvania, 700 Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, USA.

E-mail addresses: ttessier@ecycle.com (T.H. Tessier), armstrong@wharton.upenn.edu (J.S. Armstrong).

Decomposition Testing: Lodging Market

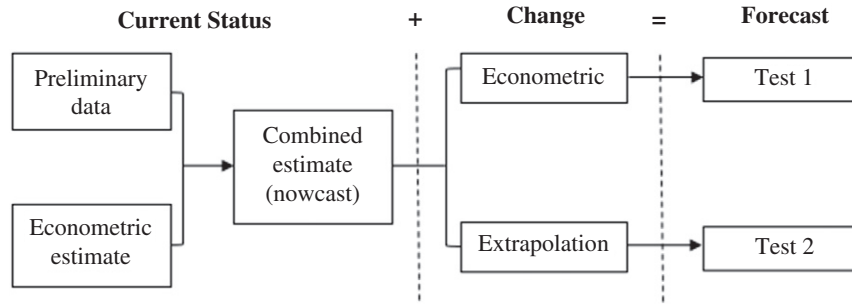


Fig. 1. Decomposition testing: lodging market.

combined with trade and production data from surveys of producers to provide estimates of the current levels. Backcasts (forecasting backwards in time) were then made for average annual sales in 1953–55. One approach started with the survey data and added the change over time where the change was forecast by an econometric model. Another approach used an average of the estimates from the survey data and the econometric estimates of the current level. The a priori weights – two-thirds on survey and one-third on the econometric estimate – reduced the backcast error for 14 of the 17 countries. On average, across the countries, the mean absolute percentage error (MAPE) was reduced from 30% to 23%, an error reduction of about 25%. No matter the weights, the combination was always more accurate than forecasts based on survey data alone. For a further assessment on the effects of decomposing by level and change, this current study reanalyzes data from an MBA thesis by the first author (Tessier, 1974). These data relate to the U.S. lodging market.

3. Testing decomposition using U.S. lodging market data

The data include room, food, and beverage sales in constant dollars for the U.S. lodging market – e.g. hotels and motels – for 1958 through 1970. The U.S. Department of Commerce conducted annual surveys of lodging sales and published them in the *U.S. Industrial Outlook*. The Department of Commerce used various sources to make its estimates. During 1958, 1963, and 1967, the business census was the primary source. During non-census years, sales were estimated from sample surveys, tax returns, and information from private sources. Estimates of lodging sales were made at the end of each year.

Sales estimates were revised in following years as additional data became available. For example, at the end of 1968, the Department of Commerce estimated that 1968 lodging sales were approximately \$7.3 billion. In 1969, the 1968 estimate was revised to \$7.6 billion. In 1970, it was revised to \$7.1 billion. The most current (1971) and presumably “final” estimate of 1968 lodging sales is \$6.5 billion. In this example, the preliminary estimate of \$7.3 billion made in 1968 was 11.0% higher than the final estimate made in 1971. The MAPE of the Department of Commerce's preliminary estimates between 1964 and 1970 was 11.0%.

This study decomposes the Department of Commerce's sales estimates into level and change, and focuses on improvement in nowcasting. The effect that decomposition and nowcasting has on forecast accuracy is then examined. The approach is summarized in the Figure.

3.1. Developing an econometric model for nowcasting

A number of alternative approaches could be used to estimate the current level. This study uses an econometric model. Because few data on the lodging market were available for estimating the coefficients,

development of the model relied primarily on estimates from prior econometric studies.

The first step in the a priori analysis was to identify the causal variables relevant to lodging sales. In broad terms, lodging demand is determined by market size, ability to buy, and needs. Given the available data, the following five variables were chosen for the model: U.S. population (market size), corporate profits and lodging rates (ability to buy), and aircraft speed and intercity passenger miles (measures of needs). The model is specified in constant dollars.

The direction of each relationship in the model is based on standard economic theory: The coefficients for corporate profits and intercity passenger miles are positive, while the coefficients for lodging rates and aircraft speed are negative. The functional form of the model is multiplicative (log–log), which assumes constant elasticities, following standard econometric practice. The effect of market size is fixed a priori at 1.0 by transforming the dependent variable values into per capita figures.

The ranges of plausible values for each of the four remaining elasticities are subjective estimates based on previous studies on similar products and services (Houthakker & Taylor, 1970). While the subjective estimates are highly uncertain, prior research shows that the accuracy of econometric models is not sensitive to magnitudes of the relationships as estimated by regression analysis. That research began at least as far back as 1971; Graefe (2015–in this issue) provides a review.

The a priori analysis yielded a range of subjective estimates of the elasticity of each variable shown here:

	A priori range
B = corporate profits per capita in constant dollars	1.0 to 2.0
M = miles of intercity passenger travel per capita	0.6 to 1.0
A = lodging rates in constant dollars	–0.5 to –0.9
S = aircraft speed in miles per hour	–0.4 to –0.7.

Data from 1958 to 1964 (see Table 1) were used to update the model. Note that only seven years of data were available. Revising the model with only seven years of data was possible only by using a priori information. The regression analysis provided an estimate of the constant and additional information on the coefficients.

A method called “conditional regression analysis” was used to update the coefficients. Wold and Jureen (1953) describe this approach. It was also called a “poor man's Bayesian regression analysis” when used in Armstrong and Grohman (1972).

The procedure was as follows: First, historical data for each independent variable (corporate profits, intercity passenger miles, rates and speed) were regressed against the dependent variable (lodging sales per capita). Second, results were examined for verification of a priori

Table 1
Data on the U.S. lodging market (final estimates).

Year	Lodging sales ^a	Corporate profits ^b	Intercity passenger miles ^c	Lodging rates ^d	Aircraft speed ^e	Consumer price index ^b	U.S. population ^b
t	Y	B	M	A	S		
1958	3644	22.3	702	6.95	219	0.866	175
1959	3996	28.5	763	7.40	223	0.873	178
1960	4248	26.7	782	7.76	235	0.883	181
1961	4372	27.2	788	7.92	253	0.896	184
1962	4616	31.2	815	8.27	274	0.906	187
1963	4667	33.1	849	8.59	287	0.917	189
1964	5031	38.4	892	9.58	297	0.929	192

* subjective estimate.

^a U.S. Department of Commerce, *U.S. Industrial Outlook*. Data in millions of current dollars.

^b Council of Economic Advisors, *Economic Report of the President*. Data on profits are after-tax profits in billions of current dollars. Data on population given in millions. Consumer price index based on 1967 = 1.00.

^c Automobile Manufacturer's Association, *Automobile Facts and Figures*, Detroit, Michigan. Includes auto, air, bus, and train miles in billions.

^d Computed from Harris, Kerr, Forster & Co., *Trends in the Hotel Business* and U.S. Census Bureau data; rates represent daily gross income per occupied room in current dollars.

^e U.S. Department of Transportation, *FAA Statistical Handbook of Aviation*. Data in miles per hour.

estimates of sign and magnitude of the elasticities. Third, succeeding regressions were run with various parameters fixed, one at a time, based on their original a priori values and from information obtained from previous regression runs. The updated model is as follows:

$$Y_t = 557 * B^{0.9} * M_t^{0.9} * A_t^{-0.6} * S_t^{-0.3}$$

(see Table 1 for a description of the variables).

3.2. Testing the value of combining multiple methods for nowcasting

The econometric estimates for the nowcasts were made for the years 1965 through 1970 by inserting values of the causal variables into the econometric model for the appropriate years. These data, shown in Table 2, are the preliminary estimates that would have been available at the time of the forecasts.

Table 3 shows that the survey estimates differed from the final estimates of current status by 10.5%, while combinations of the survey estimates and the econometric estimates were off by only 4.4%. The difference represents an error reduction of 58% for the combination.

3.3. Forecasting tests

As shown in Fig. 1, two trend-forecasting models were devised to test whether the improved estimates of the current levels improve the forecasts of lodging sales: an econometric model (Test 1) and an extrapolation model (Test 2). Forecasts from the models were derived using econometric estimates for the current level and using equally weighted combinations of the preliminary survey and econometric estimates. The accuracy of those forecasts is compared to that of forecasts from the two models derived using the preliminary survey estimates of the current level only.

Table 2
Data for testing the lodging sales model.

Year	Corporate profits	Intercity passenger miles	Lodging rates	Aircraft speed	Consumer price index	U.S. population
	B	M	A	S		
1965	46.5	917	9.91	314	0.945	194
1966	49.9	968	10.72	320	0.972	197
1967	46.6	1017	11.15	354	1.000	199
1968	47.8	1075	11.46	373	1.042	201
1969	44.8	1134	11.93	390	1.098	203
1970	40.2	1181	12.47	400 ^a	1.163	205

Sources and units: same as in Table 1.

^a Subjective estimate.

The econometric trend-forecasting model, here has a functional form of a multiplicative or log–log model. It consists of the same variables as the nowcasting model, including a scaling constant. This functional form was selected because the exponents can readily be interpreted as the demand elasticities. In addition, the assumption of constant elasticities appears to be a reasonable representation of human behavior. The model used a similar process to that used for estimating current levels. The coefficients in the model were not updated when each successive starting year is used; only the current sales level is changed. Forecasts of the independent variables are derived from linear extrapolations from data that would have been available at the time of the forecast.

$$Y_{t+f} = (1.01)^f * Y_t \left(\frac{B_{t+f}}{B_t}\right)^{0.8} * \left(\frac{M_{t+f}}{M_t}\right)^{0.7} * \left(\frac{A_{t+f}}{A_t}\right)^{-0.6} * \left(\frac{S_{t+f}}{S_t}\right)^{-0.5}$$

(See Table 1 for description of variables and f is the number of years in the future.)

The extrapolation change-forecasting model is the average of forecasts from two sub-models: a constant unit-change model developed from a five-year moving average of the yearly unit changes and a constant percentage change model developed from a five-year moving average of the yearly percentage changes. Data from 1958 up to the year of the first forecast were used to develop these extrapolations. Then, for the subsequent starting years, data from the years 1965 to 1970 were used to update the extrapolation model. Only data that would have been available at the time of the forecast are used for forecasting.

Forecasts were obtained for the years 1965 through 1970. To obtain a larger sample size, we used successive updating. Thus, current status for 1964 was used to forecast each year through 1970. The 1965 data were then included in the data and the process was repeated—and so on until the next-to-last observation was reached. This provided 28

Table 3
Errors of survey, econometric, and combined estimates of current lodging sales: Adjusted MAPEs^a (bolded figures show most accurate forecasts for each year).

Year	Preliminary Survey	Econometric Model	Combined Equal Weights
1965	5.4	10.2	2.7
1966	7.6	2.6	2.4
1967	2.5	5.8	1.6
1968	11.1	0.3	5.9
1969	19.4	0.2	10.3
1970	16.7	11.9	3.4
Average	10.5	5.2	4.4

^a Adjusted MAPE = |F – A| / ((F + A) / 2) where F = forecasted value and A = actual value.

Table 4
Adjusted MAPEs for direct vs. combined estimate of current sales.

Horizon, years (forecasts, N)	Test 1			Test 2		
	Econometric model			Extrapolation model		
	Current status estimated by:			Current status estimated by:		
	Survey	Equal weights	Model	Survey	Equal weights	Model
1 (7)	15.9	7.2	4.4	11.6	6.4	4.3
2 (6)	18.7	11.9	9.9	13.3	8.4	5.4
3 (5)	18.0	14.8	13.8	11.3	8.2	6.0
4 (4)	14.4	14.4	14.4	8.3	5.7	7.1
5 (3)	10.7	12.1	16.8	7.8	1.7	6.4
6 (2)	8.3	7.8	12.5	10.0	2.7	5.5
7 (1)	10.6	1.7	6.5	18.3	3.4	5.8
Weighted average ^a	15.4	11.0	10.7	11.2	6.2	5.6

^a Weighted by number of forecasts.

forecasts: seven for a one-year horizon, six for a two-year horizon, and so on.

Table 4 presents the adjusted MAPE for each forecast horizon from 1 to 7 years. The combined estimates yielded a substantial reduction in the forecasting error. The adjusted MAPE was reduced from 15.4 to 11.0 in the test using the econometric model – a 29% reduction – and in the test using the extrapolation model, adjusted MAPE was reduced from 11.2 to 6.2—a 45% reduction.

4. Summary and conclusions

Our study of the U.S. lodging market finds that a simple average of preliminary survey estimates and econometric estimates reduced the error in estimating final survey values from 10.5% to 4.4%, a reduction of 58%. A simple average of the two estimates of current status provided nearly optimum results. Any nowcast incorporating information from the econometric estimates was superior to one that used only the preliminary survey.

The primary purpose of this study was to test the effect of decomposition by level and change on the accuracy of forecasts. We conducted two tests, both using the improved nowcasts, but each using a different way of predicting change. Decomposition using the econometric model

for forecasting change reduced the adjusted MAPE by 29%, and decomposition using the extrapolation for change reduced the adjusted MAPE by 45%. Averaging across the tests yielded an error reduction of 37%.

The findings are consistent with the Armstrong et al. (2015—in this issue) estimate of a 35% error reduction from decomposition. Nevertheless, little direct comparative research exists on the value of decomposing by level and change, so researchers should be skeptical and conduct studies addressing issues such as the following: Under what conditions is decomposition by current level and change most useful? What other procedures can be used for nowcasting and for forecasting change?

Given the evidence to date, decomposition by level and change is expected to improve forecast accuracy. Practitioners forecasting time series would be well advised to consider this level and change decomposition.

Acknowledgments

Kay A. Armstrong, Robert Fildes, Paul Goodwin, Andreas Graefe, and Kesten Green provided reviews. Laura Blaggrave, Emma Hong, Jennifer Kwok, and Lynn Selhat edited the article.

References

- Armstrong, J. S. (1970). An application of econometric models to international marketing. *Journal of Marketing Research*, 7, 190–198.
- Armstrong, J. S., Green, K., & Graefe, A. (2015). Golden Rule of Forecasting: Be Conservative. *Journal of Business Research*, 68(8), 1717–1731 (in this issue).
- Armstrong, J. S., & Grohman, M. C. (1972). A comparative study of methods for long-range market forecasting. *Management Science*, 19(2), 211–221.
- Cole, R. (1969). Data errors and forecasting accuracy. In J. Mincer (Ed.), *Economic forecasts and expectations: Analyses of forecasting behavior and performance*. New York: National Bureau of Economic Research.
- Graefe, A. (2015). Improving forecasts using equally weighted predictors. *Journal of Business Research*, 68(8), 1792–1799 (in this issue).
- Houthakker, H. S., & Taylor, L. D. (1970). *Consumer demand in the United States, 1929–1970: Analyses and projections*. Cambridge: Harvard University Press.
- McNees, S. K. (1990). The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting*, 6, 287–299.
- Morganstern, O. (1963). *On the accuracy of economic observations*. N. J.: Princeton University Press.
- Runkle, D. E. (1998). Revisionist history: How data revisions distort economic policy research. *Federal Reserve Bank of Minneapolis Quarterly Review*, 22(4), 3–12.
- Tessier, T. H. (1974). Conditional lodging sale forecasts through 1980, Wharton School MBA thesis <http://tinyurl.com/tessier74>
- Wold, H., & Jureen, L. (1953). *Demand analysis: A study in econometrics*. New York: John Wiley and Sons.
- Zarnowitz, V. (1967). *An appraisal of short-term economic forecasts, occasional paper 104*. New York: National Bureau of Economic Research.