Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 12 (2012) 271 – 275

Procedia Computer Science

Complex Adaptive Systems, Publication 2
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2012- Washington D.C.

Housing Starts Forecast of Retail Sales through the 2007-2009 Recession

Anthony Joseph[a], Maurice Larrain[b], a*

[ab]*Pace University, One Pace Plaza, New York 10038, USA*

Abstract

The expansion following the 2001 recession was in part stimulated by a boom in housing market investments. Many economists were concerned that a severe drop in residential investments would cause a recession throughout the economy because of residential investment's relationship to the gross domestic product and financial markets, and that a decline in housing prices would negatively impact consumer spending. A severe decline in the housing market and the illiquidity in the financial market were both evident by the third quarter of 2007. Some models based on historical data were showing the certainty of a recession in 2007. The best predictors of recessions are respectively interest rate spread, unemployment claims, and building permits. Building permits/new private housing units are a leading economic indicator whereas retail sales are a component of manufacturing and trade sales, which is a coincident economic indicator because it is highly correlated to GDP. This study attempts to use housing starts and past values of retail sales to forecast out-of-sample retail sales values through the period of January 1998 to August 2010, which is inclusive of the recent 2007-2009 recession. If housing Starts are found to be predictive of retail sales, then they are a predictor of a coincident indicator of recessions. This forecast is done in the framework of neural network modeling referenced to robust regression analysis. It was expected that the neural network models would produce better forecasts as ascertained by correlation, root mean square error, and Theil inequality coefficient performance metrics, but the overall result was mixed.

Keywords: Residential investments; Neural network forecasting; Robust regression analysis; Housing starts; Retail sales

1. Introduction

The objective of this study is to forecast retail sales before and through the period covering the 2007-2009 recession using housing starts and past values of retail sales. The forecasting was done with neural network models referenced against robust multilinear regression analysis models. Although it was anticipated that the performance metrics correlation, root mean square, and Theil inequality coefficient would show that neural network models produced overall better forecasts, the results were generally mixed with the neural network forecasts providing better correlation and robust regression providing better errors in rmse and Theil relative to the desired retail sales.

A recession is defined as "a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP [gross domestic product], real income, employment, industrial production, and wholesale-retail sales" [1]. It extends from the peak to the trough of a business cycle, which is a relative alternation in real GDP. The most recent recession, officially started in December 2007 and ended June 2009 [1], is sometimes referred to as the great recession [2, 3, 4, & 5]. It is the longest recession since the great depression – 18 months compared to 43 months [1 & 3]; it shared many of the characteristics of the great depression [4]. In fact, the first year of the 2007-2009 recession was in some ways more severe than the first year of the great depression [4]. In addition, both the great recession and the great depression had global reach and were afflicted by relatively very low interest rates and excesses in the housing and financial markets among other factors [2, 3, 4,

* Corresponding author. Tel.: 212 346 1492.
E-mail address: ajoseph2@pace.edu

 doi:10.1016/j.procs.2012.09.068

& 6]. It is the only recession, except for the one in 1990-1991, to experience increases in unemployment in the financial sector since the 1930s [7]. Moreover, it is the second of two recessions of the 21st century and the 13th since the great depression. The other recession of the 21st century is the relatively mild 8-month 2001 recession [3 & 8]; unlike recessions of the previous 30 years, it was caused by curtailed business investments in "information technology" [8]. Because recessions are a natural consequence of business cycles, accurate forecasting of them is imperative if their negative impacts on the economy in such areas as employment, consumption, investment, and output are to be ameliorated.

The housing starts (or building permits) economic variable is a leading economic indicator. As such, it is capable of foreshadowing activities in the national economy. One example would be a recessionary period in the economy. However, since recessions are different because of the changing nature of the economy over time [8], housing starts' ability to forecast recessionary periods would vary accordingly. A collapse in residential housing market investments preceded most recessions dating back to the great depression [2, 3, & 9]. The peak residential investment that started the slump in the housing market occurred at least two years before the subsequent peak in GDP and the December 2007 start of the great recession [2 & 9]; the decline in investment was over 50% by the end of this recession, June 2009 [2]. Henly and Wolman [2] further stated that "the severe decline in output looks like a delayed response to residential investment shocks that had been accumulating for years." They also stated that any macroeconomic modeling of the great recession should include residential investments. At the microeconomic level, Mian and Sufi [10] provided evidence supporting that the main instigator of the great recession was the high defaulting rate of "highly-leveraged" households. In addition to the housing market, other very good economic variables for individually forecasting recessions are unemployment claims and interest rate spread [9]. These as well as stock prices are also leading economic indicators. In 2006, Leamer [9] had models involving building permits, interest rate spread, and unemployment claims used in various combinations that were predicting a recession in early 2007. However, Labonte reported in [11] that "the 50 private sector forecasters" Blue Chip surveyed one month before the recession officially started did not anticipate the unfolding recession. Besides Estrella and Mishkin [12] found that interest rate spread and stock prices were singly good predictors of recessions with interest rate spread being the most consistent predictor over variable number of horizons.

Retail sales variable as one of the two components of real manufacturing and trade sales is a coincident economic indicator [13]; as such, its movements are contemporaneous with the economy. In fact, the National Bureau of Economic Research (NBER) uses sales as well as the other three coincident economic indicators (industrial production, employment, and personal income) "to establish its business cycle chronology" [8]. Furthermore, the retail sales variable is a proxy of the economy particularly since personal consumption accounts for more than 67% of GDP. It, therefore, provides a good barometer of the state of the economy, suggesting that any prediction of retail sales is a prediction of the state of the economy inclusive of recessions and expansions. Moreover, for some retail sales items, the demand at a particular current price is found to be dependent on the previous prices [14]. Being able to accurately predict the demand of retail sales items is important for retailers and manufacturers to avoid too much or too little inventory stock to support sales; this prediction could be carried out in a neural network setting because neural networks have showed promise in forecasting retail sales [15]. In fact, Shaaf [16] demonstrated that neural network models using interest rate spread as the predictor variable were better forecasters of future recessions than econometric models.

The variables used for this study are 553 raw data of monthly housing starts and retail sales samples. These datasets were preprocessed using a 3-month moving average prior to taking the 24-month first difference of the housing starts variable and 24-month percentage change of the retail sales. Thereafter, they were normalized to within ±1 before the changes in retail sales variable were forecasted using its past values as well as changes in the housing starts variable in the multistep prediction mode using seven models in variable and overlapping forecasting regimes. The overall forecast was the aggregated predictions of the seven models. The contexts for producing the forecasts were Matlab and NeuroSolutions software.

2. Methods and Materials

The 553 samples of housing starts and retail sales raw data used in this study covered the period of August 1964 to August 2010. After taking a 3-month moving average to remove volatility, 24-month backward first difference and the 24-month percentage change of housing starts and retail sales respectively, normalization to within ±1, and allowing for multistep ahead predictions with horizons of three and seven months for past retail sales and correspondingly 10 and 14 months for housing starts, the datasets were respectively reduced to 517 and 513 samples ending in August 2010. About 29-30 percent (152 samples) of the data was reserved for out-of-sample forecasting using seven focused gamma neural network models referenced against an equal number of robust multilinear regression models. The focused gamma neural network model is a member of the class of time lagged recurrent networks with adaptive memory that track the temporal nature of the time series data [17]. The number of samples forecasted by each model ranged from 97 samples for the first model to four samples for the second and fifth models. Moreover, some models output had up to 50% overlap with an adjacent model's output, in which case the median outputs were used. The aggregated output of the seven models produced the overall forecast of retail sales covering the period of January 1998 through August 2010, which is inclusive of the 2007-2009 recession that officially covers the period of December 2007 to June 2009 [1].

The environments for performing the experiments were NeuroDimension's NeuroSolutions software for the neural network models and Mathworks' Matlab software for the robust regression models. These two types of models were subjected to similar construction to the extent possible. For example, while the transfer function used in the neural networks models is tanh, the weighting function used in the robust regression models is the related logistic function. The neural network models each had a three layer construction with one input layer, one hidden layer, and an output layer; they were trained under supervised learning control with the Levenberg-Marquardt backpropagation through time algorithm using batch mode for weight update; and the total number of weights used in the models ranged from 11 to 35 with most models using 19 weights. For the regression models, the coefficients with the most dominant weights were associated with the past retail sales values. They were 0.8999, 0.7530, 0.7529, 0.7584, 0.9181, 0.9136, and 0.9349 compared to corresponding weight values of 0.0309, 0.0679, 0.0680, 0.0644, 0.0359, 0.0396, and 0.0459 for the coefficient associated with housing starts. The constant coefficient values ranged from -0.0002 to 0.0021.

The efficacy of the neural network and the regression forecasts were determined using correlation (r), root mean square error (rmse), and Theil inequality coefficient (Theil). Like r, Theil is unitless but it ranges between 0 and 1 with zero indicating a perfect forecast [18]. It is most desirable to have values of r that approach +1 and values of rmse and Theil that approach zero for a forecast.

3. Results

Both numerically and visually, the aggregated outputs produced by the seven neural network and regression models were generally successful in forecasting the desired retail sales inclusive of the period covering the 2007-2009 recession. The forecast of the recession and the subsequent recovery period were done using models 4, 5, and 6. The neural networks models produced generally better forecasts in terms of correlation measures while the regression models produced generally better forecasts in terms of error statistics (rmse and Theil), see Table 1. These results were consistent with the overall forecast of 152 months dating from January 1998 to August 2010. The correlation associated with the neural network overall forecast is r = 0.9565, which is 0.0304 higher than that related to the regression forecast. The rmse values for the forecasts produced by the two types of models are essentially the same, but favor regression – the difference is only 0.0019 between them. The Theil inequality coefficient deemed the regression forecast more accurate by 2:1. It is 0.1221 for the regression forecast and is 0.2619 for the neural network forecast. The decline in economy as shown by the retail sales variable started in February 2006 (sample 98) and ended May 2009 (sample 137), see Fig. 1. The periods between February 2006 and June 2008 (sample 126) showed abrupt declines with some lasting about four months followed by small gradual increases of about one to three months. Throughout most of this period, the neural network forecast appeared to more accurately simulate the desired retail sales variable – May 2006 (sample 101) to June 2008. It was able to capture some of the sample points in retail sales at the moment they occurred, albeit in

Table 1. Results of forecasts

Performance Metrics		Models							
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Overall
r	NN	**0.8986**	08028	**0.9237**	0.8911	**0.9885**	**0.7949**	0.6616	**0.9565**
	RR	0.8915	**0.9360**	0.8560	**0.8938**	0.9573	0.7027	**0.6944**	0.9261
rmse	NN	0.0245	0.0054	**0.0221**	**0.0249**	0.0839	0.0753	0.0155	0.0297
	RR	**0.0155**	**0.0030**	0.0247	0.0340	**0.0698**	**0.0479**	**0.0132**	**0.0278**
Theil	NN	0.6233	0.2375	**0.5972**	**0.2001**	0.3234	0.2187	0.0197	0.2619
	RR	**0.1007**	**0.2086**	0.8560	0.3229	**0.2647**	**0.0911**	**0.0084**	**0.1221**

Note: NN is neural network and RR is robust regression

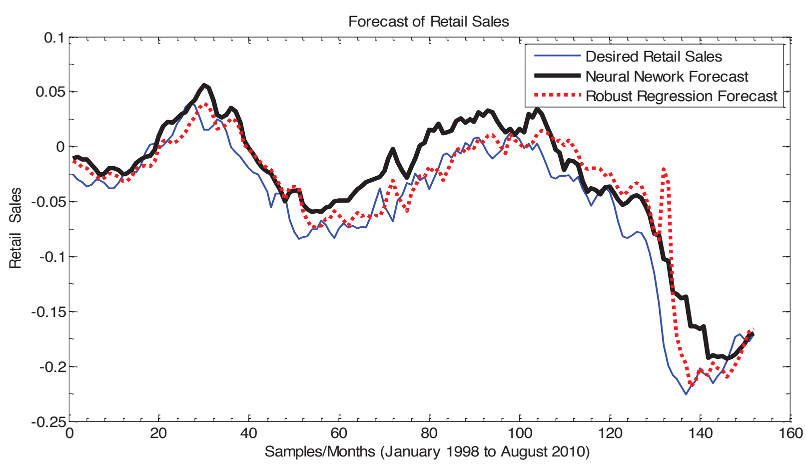

Fig. 1. Neural network and regression models' forecasts of retail sales over a period that includes the 2007-2009 recession

some instances with much higher values. For example in August 2006 (sample 104), the desired retail sales value was 0.0025 and the corresponding neural network forecasted value is 0.0345. A similar correspondence occurred in June 2008 (sample 126), wherein the retail sales value was -0.0774 and the neural network forecasted value is -0.0445. During this 28 month period, the regression forecast accurately replicated retail sales with a delay of six to seven months. In fact, for the small two month rise that occurred in retail sales between April 2008 (sample 124) and June 2008 (sample 126) the corresponding regression forecast occurred between November 2008 (sample 131) and January 2009 (sample 133), peaking in December 2008 (sample 132) at a value of -0.0204 relative to the desired retail sales value of -0.0774. The forecasts matching the period of February 2006 to June 2008 were produced by model 2, model 3, and model 4 with model 2 favoring regression and model 3 favoring neural networks, see Table 1. Model 1 forecasted 97 samples (January 1998 to January 2007). The neural network and regression models produced similar forecasts, except the regression forecast experienced relatively better error statistics. The forecasted responses were about three months late with the neural network forecast having comparatively higher values in some periods; it was also smoother in the regions corresponding to December 2001 (sample 48) and September 2003 (sample 69). The Theil for model 1 regression forecast is six times smaller than that for neural network forecast – 0.1007 compared to 0.6233.

The forecasts of the recession and the subsequent recovery period are revealed in the retail sales data as being December 2007 to May 2009 and June 2009 to August 2010 (end of data). The end of the recession revealed in the desired retail sales data is one month earlier than the June 2009 date officially announced by NBER [1]. During the recession/recovery period, the neural network forecast is more complex and volatile than the regression forecast, yet it seemed to be more aligned with the desired retail sales over the recession period and part of the recovery. Except for an apparent unaccounted and misplaced peak at sample 140 (August 2009), which coincided with an actual peak of retail sales during the recovery period, the regression forecast indicated economic activities occurring one to seven months later than they actually occurred, see Fig. 1. For example, the regression forecast showed that the recession ended in June 2009, the official end of the recession, but one month later than indicated in the desired retail sales data. Otherwise, the regression forecast is a reasonably good replica of the retail sales data. By comparison, the neural network forecast picked up the start of the recession in December 2007 (sample 120) with a value of -0.0362 relative to the actual value of -0.041. The decline in the forecast over the next three months was not as pronounced as in the desired retail sales -- -0.0534 relative to -0.0816. The next month decline in retail sales was not captured in the neural network forecast that showed an increase in the expected retail sales. The subsequent increase in retail sales over April 2008 (sample 124) and June 2008 (sample 126) followed by the beginning of the abrupt decline in retail sales, which signaled towards the end of the recession, was captured in the neural network forecast in time but not in magnitude, see Fig 1. The neural network forecast between April 2008 and August 2008 (sample 128) represented a better replica of retail sales in time, trend, and magnitude. Like the regression forecast, the neural network forecast recognized the end of the recession one month late, June 2009, relative to retail sales. Moreover, the neural network forecast while indicating an abrupt negative change of 0.0285 corresponding to the decline of 0.0015 of retail sales from September 2009 (sample 141) to October 2009 (sample 142) during the recovery period, it did not pick up the increasing retail sales between June 2009 and August 2009 (sample 140). It remained essentially flat for the next four months ending in February 2010 (sample 146) – in most of which time retail sales were increasing -- before it recognized the upward trend of the last six months (March 2010 to August 2010) of the recovery evidenced in the retail sales data.

4. Discussion

The 2007-2009 recession was difficult to accurately predict based on historical data. In Joseph et al [19], the neural network and regression forecasts hardly showed evidence of the great recession as indicated by the purchasing managers' index data covering December 2007 and December 2008 and they totally missed the abrupt decline in the economy between July 2008 and December 2008. The current study provides better forecasts of the recession as revealed in the forecast of retail sales data that covered December 2007 to May 2009. The neural network forecast was better at capturing the precise times and trends of certain economic activities during this recessionary period. The regression forecast also captured the recession, but with less precision with respect to exact times of economic occurrences – the forecast provided a delayed response of movements in the economy. Why was the great recession of 2007-2009 difficult to predict from historical data? One reason might be the general lack of historical data for coincident economic indicators that cover periods of severe recessions [1 & 2]. Since the great depression of 1929 -- 1933, the 2007-2009 recession was the most severe and longest one with a duration of 18 months. The other two recessions that came close in duration are the 16-month ones of 1973 -- 1975 and 1981 – 1982. Moreover, the severity and length of the 2007-2009 recession were not only induced by the bust in the housing market resulting from excesses within it, but also the corrections and deepening crisis in the financial market relative to the preceding boom years following the 2001 recession and the slowdown in consumer spending [2, 3, 6, & 10]. Henly and Wolman [2] reported that while a residential investment peak typically leads a business cycle peak in recessionary times, in the 2007-2009 recession the length of time between peaks was uncharacteristically long. This appeared to be the consequence of GDP peaking in the second half of 2008, almost three years after residential investment peaked [2]. The rise in GDP during the first half of the recession was greatly influenced by continued consumer spending and nonresidential fixed investment that peaked concurrently with GDP [2 & 3]. Moreover, at about the time that GDP peaked, the crisis in the financial sector deepened to calamitous proportions, thereby creating a compounding effect. From the second half of 2008, the compounding effect of falling residential and nonresidential investments, declining consumer spending, and the collapsing financial market acted to worsen the severity of the recession. This helps to explain why further into the recession the housing starts variable effect was possibly not sufficiently strong to produce a forecast that would continue to pick up sample points in retail sales when they were occurring. It also helps to explain why some economists doubted the likelihood of an impending recession, even when their models suggested otherwise [9]; perhaps because the 2001 recession was not preceded by a bust in the housing market [2 & 9].

5. Conclusion

Both the neural network and regression out-of-sample forecasts were reasonably good predictors of the desired retail sales variable, which moves coincidentally with the economy. The regression forecast always occurred later than the desired retail sales variable. This seemed to have been caused by the dominant weight of the coefficient associated with past retail sales values. Although the first 99 samples of the neural network forecast was delayed by about three months, suggesting a relatively weaker effect of housing starts on retail sales, from May 2006 (sample 101) the neural network forecast is graphically better in terms of accurate temporal alignment. Post May 2006, the neural network models appeared to incorporate the effect of housing starts on retail sales and the economy. Numerically, however, none of the forecasts is superior. A future study of retail sales as a proxy of economic activity covering a period inclusive of the 2007-2009 recession will be more comprehensive. In addition to housing starts and past retail sales values, it will include stock prices and interest rate spread data as predictors. Furthermore, the models to be used will include support vector regression, generalized regression neural network, or kernelized partial least squares to avoid the neural network potential problems of local minima during the model training phase.

References

1. National Bureau of Economic Research (NBER), US Business Cycle Expansions and Contractions. Retrieved 5/30/2012: http://www.nber.org/cycles/cyclesmain.html.
2. S. Henly and A. Wolman, Housing and the Great Recession: A VAR Accounting Exercise. Economic Quarterly 97 (First Quarter 2011) 45-66.
3. M. Labonte, The 2007-2009 Recession: Similarities to and differences from the Past. Congressional Research Service (2010). Retrieved 4/6/2012: http://www.fas.org/sgp/crs/misc/R40198.pdf.
4. H. Shomali and G. Giblin, The Great Depression and the 2007-2009 Recession: The First Two Years Compared. International Research Journal of Finance and Economics 59 (2010) 15-22.
5. S. Verick and I. Islam, The great Recession of 2008-2009: Causes, Consequences, and Policy Responses. IZA Discussion Paper No. 4934 (May 2010). Retrieved 5/2/2012: http://www.iza.org/en/webcontent/publications/papers/searchResults.
6. M. Guillén and S. Suárez, The Global Crisis of 2007-2009: Markets, Politics, and Globalizations. (2009). Retrieved 4/10/2012: http://apps.business.ualberta.ca/mlounsbury/markets%20on%20trial%20abstracts/guillen%20paper.pdf.
7. U.S. Bureau of Labor Statistics, The Recession of 2007-2009. BLS Spotlight on Statistics (February 2012). Retrieved 4/10/2012: http://www.bls.gov/spotlight/2012/recession/pdf/recession_bls_spotlight.pdf.
8. J. Stock and M. Watson, How Did Leading Indicators Forecasts Do During the 2001 Recession? Federal Reserve Bank of Richmond Economic Quarterly (Summer 2003) 71-90.
9. E. Leamer, Is a Recession Ahead? The Models Say Yes, but the Head Says No. Economists' Voice 4 (January 2007) 1-5. Retrieved 4/6/2012: DOI:10.2202/1553-3832.1239.
10. A. Mian and A. Sufi, Household Leverage and the Recession of 2007 to 2009. IMF 10th Jacques Polak Annual Research Conference (November 2009). Retrieved 5/2/2012: http://www.imf.org/external/np/res/seminars/2009/arc/pdf/mian.pdf.
11. M. Labonte, Would a Housing Crash Cause a Recession? Congressional Research Service (2007). Retrieved 4/6/2012: http://assets.opencrs.com/rpts/RL34244_20071107.pdf.
12. A. Estrella and F. Mishkin, Predicting U.S. Recessions: Financial Variables as Leading Indicators. The Review of Economics and Statistics 80 (February 1998) 45-61.
13. M. Larrain, Do Interest Rates Lead Real Sales and Inventories? A Special Analysis Approach. Business Economics (April 2002) 33-43.
14. M. Pesendorfer, Retail Sales: A Study of Pricing Behavior in Supermarkets. Journal of Business 75 (January 2002) 33-66.
15. M. Orra and G. Serpen, An Exploratory Study of Neural Network Forecasting of Retail Sales Trends Using Industry and National Economic Indicators. Retrieved 4/10/2012: http://www.eng.utoledo.edu/~gserpen/Publications/CIEF%202005%20Manuscript%20by%20Orra%20and%20Serpen.pdf.
16. M. Shaaf, Predicting Recession Using the Yield Curve: An Artificial Intelligence and Econometric Comparison. Eastern Economic Journal 26 (Spring 2000) 171-190.
17. J. Principe, N. Euliano, and W. Lefebvre, Neural and Adaptive Systems: Fundamentals Through Simulations. John Wiley & Sons, USA, New York, 2000.
18. R. Leuthold, On the Use of Theil's Inequality Coefficients. American Journal of Agricultural Economics 57 (May 1975) 344-346.
19. A. Joseph, M. Larrain, C. Turner, Forecasting Purchaser's Managers' Index with Compressed Inrerest Rates and Past Values. Elsevier Procedia Computer Science 6, Complex Adaptive Systems 1, (2011) 213-218.