7-2020

Forecasting Singapore GDP using the SPF data

Tian XIE

Jun Yu
Singapore Management University, yujun@smu.edu.sg

SMU ECONOMICS & STATISTICS

Forecasting Singapore GDP using the SPF Data

Tian Xie, Jun Yu

July 2020

Paper No. 17-2020

Forecasting Singapore GDP using the SPF Data

Tian Xie and Jun Yu[1]

In this article, we use econometric methods, machine learning methods, and a hybrid method to forecast the GDP growth rate in Singapore based on the Survey of Professional Forecasters (SPF). We compare the performance of these methods with the sample median used by the Monetary Authority of Singapore (MAS). It is shown that the relationship between the actual GDP growth rates and the forecasts from individual professionals is highly nonlinear and non-additive, making it hard for all linear methods and the sample median to perform well. It is found that the hybrid method performs the best, reducing the mean squared forecast error (MSFE) by about 50% relative to that of the sample median.

1. Introduction

A very large body of applied works in economics have tried to foresee key macroeconomic indicators, including GDP growth rates, unemployment rates, and inflation rates. A straightforward reason to justify these extensive studies is that these macroeconomic variables are vital to many decision-makers in the economy. In this paper, we focus our attention to predicting the GDP growth rate in Singapore using the Survey of Professional Forecasters (SPF).

SPF is a leading macroeconomic forecast consensus in Singapore. It has been run by the Monetary Authority of Singapore (MAS) since the last quarter of 1999 and is made available to the public at https://www.mas.gov.sg/monetary-policy/MAS-Survey-of-Professional-Forecasters.[2] The survey is conducted quarterly following the release of economic data for the previous quarter by the Ministry of Trade and Industry of Singapore. It contains forecasts for 15 key economic indicators; see the MAS's SPF. The first of the indicators is the GDP growth rate (year-on-year growth in percentage terms and constant prices). It should be noted that the SPF results do not represent MAS's own views or forecasts.

Every quarter MAS reports the sample median and the empirical density of the forecasts which are available in the survey. In this article, we denote the sample median as the

[1] Tian Xie is Associate Professor in College of Business, Shanghai University of Finance and Economics, Shanghai, China. Jun Yu is Lee Kong Chian Professor of Economics and Finance in School of Economics and Lee Kong Chian School of Business, Singapore Management University. The views in this article are solely those of the authors and should not be attributed to MAS or SHUFE or SMU.

[2] There are some similar surveys internationally with different starting dates. Two well-known examples are the SPF produced by the Federal Reserve Bank of Philadelphia since the late 1960s and the SPF collected by the European Central Bank for the eurozone since the late 1990s.

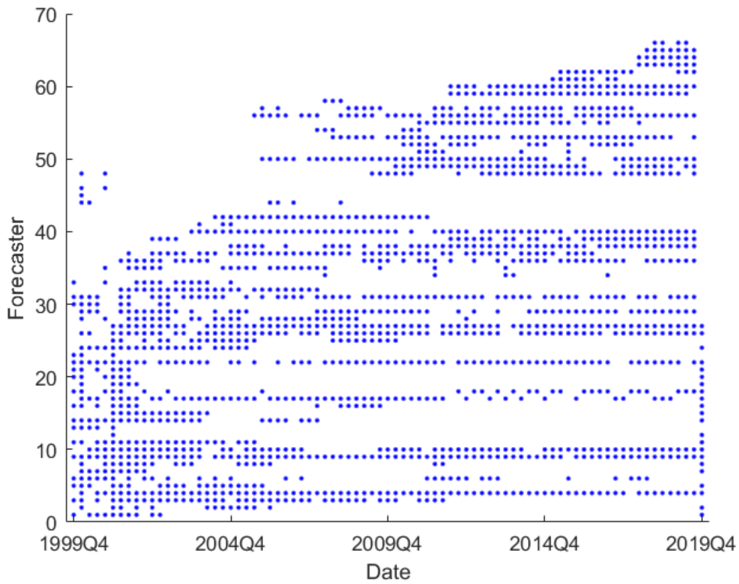
benchmark forecast. In the literature, Genre et al. (2013) employ the sample mean as the benchmark. We find the difference between the sample median and the sample mean is negligible in the SPF.

We first describe the data in Section 2. In Section 3 we introduce alternative methods for obtaining the forecasts and discuss criteria to evaluate those forecasts. Section 4 provides an empirical analysis to contrast the performance of alternative methods and the benchmark method. Section 5 concludes.

2. Data

In the article, we consider utilizing the individual forecasts from the SPF, denoted as $\{x_{1t}, \ldots, x_{pt}\}$, to predict the real GDP growth rate, denoted as y_{T+1}. Here i represents the ith forecasters and t represents the period t and $t = 1, \ldots, T$. From the last quarter of 1999 to the last quarter of 2019, the SPF collects one-month-ahead quarterly predictions of the real GDP growth rate from 66 different forecasters.[3] At period T, the sample median of $\{x_{1T}, \ldots, x_{pT}\}$, acting as the final forecast of y_{T+1}, is the "middle" number of these numbers when they are listed ascendingly.

Figure 1: An illustration of the entries and exits of individual forecasters



However, an initial data cleaning is necessary since a specific forecaster may or may not submit a survey response each time throughout the whole period. Figure 1 describes the entries and exits of individual forecasters over the survey period. A blue dot represents a

[3] Take Q1 as an example. Questionnaires are sent out to forecasters in the middle of February and forecasting results must be submitted before the end of February.

specific forecaster (labeled in the vertical axis) if he or she submitted a survey response and a blank space indicates otherwise.

The data clearly exhibit severe sparsity in the submission of forecasters. To avoid the issues caused by missing observations, we follow Genre et al. (2013) to first remove irregular respondents if he or she misses more than near 50% of the observations. In the end, we narrow down to $p = 15$ qualified forecasters. Then the missing observations for each forecaster i are filled using the approach suggested in Genre et al. (2013).

3. Methods

Let $\mathbf{X}_\mathrm{t} = [1, x_{1t}, \ldots, x_{pt}]'$. If all the p forecasters are employed, and the relationship between y_{t+1} and all the elements in $\boldsymbol{X}_t$ is linear and additive, the following linear model can then be presumed:

$$y_{t+1} = \boldsymbol{\beta}'\mathbf{X}_t + \varepsilon_t, \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of slope parameters and ε_t is the error term. There are p+1 slope parameters in Equation (1). In practice, p can be very large and therefore the estimation error can be large as well. If $p > T - 2$, it is not viable to estimate $\boldsymbol{\beta}$ by least squares.

In practice, we do not know if all the p forecasters are beneficial *ex-ante*. If most of the variables in $\mathbf{X}_t$ are not useful, which means there is sparsity in Equation (1), one needs to deal with the problem of variable selection and parameter estimation simultaneously. Furthermore, there is no reason to believe why the relationship between y_{t+1} and $\mathbf{X}_t$ should be linear and additive. Although it is theoretically possible to specify a general functional form to relate y_{t+1} and $\mathbf{X}_t$ as follows

$$y_{t+1} = f(\mathbf{X}_t) + \varepsilon_t, \tag{2}$$

the nonparametric estimation of $f(\mathbf{X}_t)$ incurs the well-known problem of the curse of dimensionality even when p is of a moderate magnitude.

In this section, we review 4 methods employed to forecast the Singapore GDP based on the SPF survey outcomes. Other than the benchmark method of the sample median, we also use the complete subset regression of Elliott et al. (2013), the elastic net method of Zou and Hastie (2005), the LSSVR method of Suykens and Vandewalle (1999), the Mallows-type model averaging LSSVR method of Qiu et al. (2020). The first method is a conventional econometric method. The second method is a variable selection method. The third method is a machine learning technique. The last method combines an econometric method with a machine learning method. A more extensive survey of both econometric methods and machine learning methods for a forecasting purpose can be found in Xie et al (2020)

3.1 Complete Subset Regression

The complete subset regression (CSR) of Elliott et al. (2003) is a method for mixing forecasts from all possible linear regression models, each of which has a fixed number of predictors from a given set of potential predictor variables. The weight assigned to each model can be the same or different.

To explain the idea, let the number of predictor variables be fixed at 1, although we use 5 predictor variables in our empirical study. In this case, the equally weighted forecast of y_{T+1} is given by

$$\hat{y}_{T+1} = \frac{1}{p}\sum_{i=1}^{p}\left[\hat{\beta}_{0i} + \hat{\beta}_{1i}x_{iT}\right], \tag{3}$$

where $\hat{\boldsymbol{\beta}}_i = \left[\hat{\beta}_{0i}, \hat{\beta}_{1i}\right]'$ is the least squares estimate of $\boldsymbol{\beta}_i = [\beta_{0i}, \beta_{1i}]'$ from the following linear regression model

$$y_{t+1} = \beta_{0i} + \beta_{1i}x_{it} + \varepsilon_t, \qquad t = 1, \ldots, T-1. \tag{4}$$

One of the successful applications of CSR in economics and finance is Rapach et al. (2010) where each of potentially valuable predictors is used to predict stock returns.

3.2 Elastic Net

When the number of predictors p is large and a significant subset of predictors is not informative in predicting y, Model (1) and the least squares method does not perform well out-of-sample. Many penalized regressions have been proposed to select predictors which in turn can improve the predictive performance. One of the successful methods is the elastic net of Zou and Hastie (2005). The idea of the elastic net is to shrink the slope parameter towards zero if the corresponding predictor is not significant.

The elastic net imposes a constraint on the sum of squared coefficients excluding intercept, that is,

$$\hat{\beta}_* = \underset{\beta_*}{\operatorname{argmin}}\left\{\sum_{t=1}^{T-1}\left[y_{t+1} - \beta_0 - \sum_{i=1}^{p}\beta_i x_{it}\right]^2 + \lambda\left[\alpha\sum_{i=1}^{p}|\beta_i| + (1-\alpha)\sum_{i=1}^{p}\beta_i^2\right]\right\},$$

where the second term in the bracket is the penalty that contains two components (one is the L_1-penalty and the other is L_2-penalty), λ is a tuning parameter that determines the severity of the penalty, and α is a mixing parameter that determines the trade-off between two penalty terms. The penalty term is used to shrink the slope parameters to accommodate possible sparsity in potential predictors.

3.3 Least Squares Support Vector Regression

Instead of locating a consistent estimator of $f(\mathbf{X}_t)$ in Equation (2), most machine learning techniques try to find a good approximation to $f(\mathbf{X}_t)$ so that the approximation leads to an accurate forecast of y_{t+1}.

The support vector regression (SVR) of Drucker et al. (1996) approximates $f(\mathbf{X}_t)$ by a set of basis functions $\{h_s(\mathbf{X}_t)\}_{s=1}^{S}$ that can be of infinite-dimensional. Equation (1) can thus be rewritten in the following form

$$y_{t+1} = f(\mathbf{X}_t) + \varepsilon_t \approx \sum\nolimits_{s=1}^{S} \beta_s h_s(\mathbf{X}_t) + \varepsilon_t. \tag{5}$$

To estimate $\boldsymbol{\beta} = [\beta_1, \dots, \beta_S]'$, we minimize

$$H(\boldsymbol{\beta}) = \sum_{t=1}^{T-1} V_e\left(y_{t+1} - f(\mathbf{X}_t)\right) + \lambda \sum_{s=1}^{S} \beta_s^2, \tag{6}$$

where $V_e(\cdot)$ is the loss function given by

$$V_e(x) = \begin{cases} 0, & if\ |x| < e \\ |x| - e, & if\ |x| \geq e \end{cases}. \tag{7}$$

Suykens and Vandewalle (1999) modify the SVR algorithm which results in solving a set of linear equations under a squared loss function. This method, also known as least squares SVR (LSSVR), minimizes

$$H(\boldsymbol{\beta}) = \sum_{t=1}^{T-1} \left(y_{t+1} - f(\mathbf{X}_t)\right)^2 + \lambda \sum_{s=1}^{S} \beta_s^2, \tag{8}$$

where the loss function is specified to be a squared loss function in LSSVR. To minimize the quantity in Equation (8), the Lagrangian equation may be set up so that we have the following expression for the optimal solution,

$$\hat{f}(\mathbf{X}_t) = \sum_{t=1}^{T-1} \hat{\alpha}_t K(\boldsymbol{x}, \mathbf{X}_t), \tag{9}$$

where $\boldsymbol{x}$ is any given values for predictors, $\{\hat{\alpha}_t\}_{t=1}^{T}$ are the estimated Lagrangian multipliers, and $K(\cdot,\cdot)$ is the predetermined kernel function. In this article, we consider the Gaussian kernel function given by

$$K(\boldsymbol{x},\boldsymbol{X}) = e^{-(\|\boldsymbol{x}-\boldsymbol{X}\|)/(2\sigma_x^2)}, \tag{10}$$

where σ_x^2 is a hyperparameter that users specify in advance.

3.4 LSSVR$^{\text{MA}}$

Most machine learning methods, including LSSVR, do not account for model uncertainty. While the CSR method accounts for model uncertainty, it assumes that the relationship between y_{t+1} and each x_{it} is linear. If the relationship between y_{t+1} and some x_{it} is nonlinear and hence model uncertainty needs to be accounted for, then a reasonable approach is to apply the idea of forecast combinations to a set of machine learning strategies, as suggested in Qiu et al. (2020). In this article, following Qiu et al. (2020), we blend the idea of forecast combination with the LSSVR method. The new method is denoted LSSVR$^{\text{MA}}$, where the superscript MA indicates model averaging.

Let $\boldsymbol{y} = [y_2, \dots, y_T]'$. Suppose the mth LSSVR strategy uses $\mathbf{X}_t^{(\boldsymbol{m})}$, which is a subset of $\mathbf{X}_t$, to forecast y_{T+1} with $m = 1, \dots, M$. That is, in total there are M strategies. Denote $\hat{y}_{T+1}(m)$ the forecast of y_{T+1} under the mth LSSVR strategy. Qiu et al. (2020) show that LSSVR leads to $\hat{f}\left(\mathbf{X}_t^{(\boldsymbol{m})}\right) = \boldsymbol{P}_{(m)}\boldsymbol{y} := \boldsymbol{P}\left(\mathbf{X}_{(\boldsymbol{m})}, \mathbf{X}_{(\boldsymbol{m})}\right)\boldsymbol{y}$ where $\mathbf{X}_{(\boldsymbol{m})} = \left[\mathbf{X}_1^{(\boldsymbol{m})}, \dots, \mathbf{X}_{T-1}^{(\boldsymbol{m})}\right]$ for any $m = 1, \dots, M$. Let the weighted average forecast of y_{T+1} be

$$\hat{f}(\boldsymbol{w}) = \sum_{m=1}^{M} w_{(m)}\hat{f}\left(\mathbf{X}_t^{(\boldsymbol{m})}\right) = \sum_{m=1}^{M} w_{(m)}\boldsymbol{P}_{(m)}\boldsymbol{y} = \boldsymbol{P}(\boldsymbol{w})\boldsymbol{y}, \tag{11}$$

where $\boldsymbol{P}(\boldsymbol{w}) \coloneqq \sum_{m=1}^{M} w_{(m)}\boldsymbol{P}_{(m)}$ and the weight vector $\boldsymbol{w} \in \mathcal{H}$ with $\mathcal{H}$ being a M-dimensional simplex.

Based on a Mallows-type criterion, Qiu et al. (2020) propose the following method to choose the weights,

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}\in H}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{P}(\boldsymbol{w})\boldsymbol{y}\|^2 + 2\hat{\sigma}^2(\boldsymbol{w}) \sum_{t=1}^{T} \boldsymbol{P}_{tt}(\boldsymbol{w}), \tag{12}$$

where $\boldsymbol{P}_{tt}(\boldsymbol{w})$ is the tth diagonal term in $\boldsymbol{P}(\boldsymbol{w})$.

4. Empirical Results

We conduct forecasting exercises using the data described in Section 2. We list the 5 forecasting methods, the tuning parameters, and the model settings in Table 1.[4]

Table 1: Summary of the 5 methods to forecast the Singapore GDP growth

Method	Parameter
Median	Median of all available forecasts
CSR	5 predictors, 1000 models, equal weight
Elastic Net	$\lambda = 0.5, \alpha = 0.5$
LSSVG	Gaussian kernel, $\sigma_x = 10$
$LSSVR^{MA}$	Gaussian kernel, $\sigma_x = 10$, Full combination

A rolling window approach is implemented to obtain a one-quarter-ahead forecast of the Singapore GDP growth. The initial period for making the forecast is the last quarter of 2009. The window length is set to 40. The out-of-sample performance of the 5 methods is evaluated by mean squared forecast error (MSFE) and mean absolute forecast error (MAFE) as defined by

$$\text{MSFE} = \frac{1}{K}\sum_{k=1}^{K}(y_{T+k} - \hat{y}_{T+k})^2, \tag{13}$$

$$\text{MAFE} = \frac{1}{K}\sum_{k=1}^{K}|y_{T+k} - \hat{y}_{T+k}|, \tag{14}$$

where K is the total number of quarters when we forecast the GDP growth, $\hat{y}_{T+k}$ is the one-step-ahead forecasted value of y_{T+k} at period $T + k$ by one of the 5 methods.

The values of MSFE and MAFE and their associated ranking for all the 5 models are reported in Tables 2. The lowest MSFE and MAFE are presented in boldface.

Table 2. Out-of-sample forecasting comparison of 5 methods

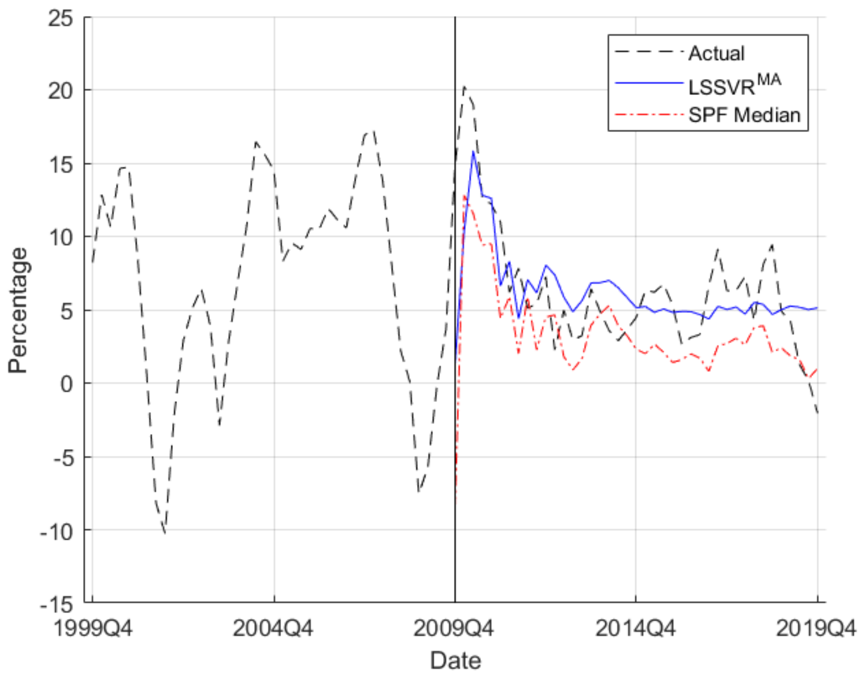
Methods	MSFE		MAFE	
	value	ranking	value	ranking
Median	26.7336	4	3.5439	5
CSR	28.8199	5	3.5042	4
Elastic Net	25.7032	3	3.2725	3
LSSVR	14.2397	2	2.7383	2
$LSSVR^{MA}$	**13.9567**	1	**2.6861**	1

[4] We also consider alternative settings of tuning parameters. The results are qualitatively intact.

A few conclusions can be drawn from Table 2. First and foremost, LSSVR^{MA} always performs the best followed by LSSRV. The sound performance of LSSVR^{MA} relative to LSSVR suggests that there exists model uncertainty. Second, these two methods perform much better than the other three methods, implying a nonlinear dependence between y_{t+1} and x_{it}s. For example, compared to the benchmark method, LSSVR^{MA} gains at reducing the MSFE value by almost 50%. If we fit a partially linear model, one could see a strong nonlinear relationship between y_{t+1} and individual x_{it}. To save space, we do not report empirical results for the partially linear model. Third, the fact that the elastic net slightly outperforms CSR and the sample median indicates that there is no strong evidence of sparsity in x_{it}s.

To visually compare the forecast accuracy of the benchmark method and the LSSVR^{MA} method, we plot two forecasted series of the above two methods against the actual data in Figure 2. It is apparent that the median forecast often underestimates the actual values, especially for the recent 5 years. Although flatter than the actual values, the forecasts by the LSSVR^{MA} method captures the level and the trend reasonably well.

Figure 2: A comparison of two forecasts



To examine if the improvement in forecast accuracy is significant, we perform the Giacomini-White (GW) test of the null hypothesis that the column method performs equally well as the row method in terms of absolute forecast errors (Giacomini and White, 2006). Table 3 reports the p-values of the GW test in all pair-wise comparisons. The 5 methods can be divided into 2 groups. The sample median, CSR, and the elastic net form the first group.

There is no statistically significant difference in the forecasting performance of the methods in this group. LSSVR and $LSSVR^{MA}$ form the second group. There is no statistically significant difference in the forecasting performance of the methods in the second group. However, the methods in the second group statistically significantly outperform the methods in the first group at either the 5% level or the 10% level.

Table 3: the p-values of the GW test in all pair-wise comparisons

Methods	Median	CSR	Elastic Net	LSSVR
Median	-	-	-	-
CSR	0.4345	-	-	-
Elastic Net	0.4931	0.6245	-	-
LSSVR	0.0345	0.0508	0.0870	-
$LSSVR^{MA}$	0.0325	0.0589	0.0929	0.6881

5. Conclusion

We have considered five methods, including two conventional econometric methods, a variable selection method, a machine learning method, and a hybrid method, to forecast the GDP growth rate in Singapore based on the SPF. In particular, the performance of these methods is compared to the sample median used by the MAS. It is demonstrated that the hybrid method performs the best, reducing MSFE by about 50% over that of the sample median. The gain is verified to be statistically significant at the 5% level.

Our exercise suggests that it is possible to produce more accurate forecasts of the Singapore GDP growth rates than the median forecast of the SPF. Since forecasts of most, if not all, of the professional forecasters contain useful information about the next-quarter Singapore GDP growth rate, they should not be given a zero weight. Since the relationship between macroeconomic variables is nonlinear and complicated, a machine learning method is helpful in this case. Moreover, since the hybrid method can accommodate model uncertainty, it leads to the most accurate forecasts.

References

DRUCKER, H., C. J. C. B URGES, L. K AUFMAN, A. J. SMOLA, AND V. VAPNIK (1996): “Support Vector Regression Machines,” in Advances in Neural Information Processing Systems 9, ed. by M. C. Mozer, M. I. Jordan, and T. Petsche, MIT Press, 155–161.

ELLIOTT, G., A. GARGANO, AND A. TIMMERMANN (2013): “Complete subset regressions,” Journal of Econometrics, 177, 357 – 373.

GENRE, V., G. KENNY, A. MEYLER, AND A. TIMMERMANN (2013): “Combining expert forecasts: Can anything beat the simple average?” International Journal of Forecasting, 29, 108 – 121.

GIACOMINI, R. AND H. WHITE (2006): "Tests of Conditional Predictive Ability," Econometrica, 74, 1545–1578.

MAS: "Survey of Professional Forecasters: Dataset Documentation," https://www.mas.gov.sg/monetary-policy/MAS-Survey-of-Professional-Forecasters.

QIU, Y., XIE, T. AND J. YU (2020): "Forecast combinations in machine learning," Working Paper, Singapore Management University.

RAPACH, D. E., J. K. S TRAUSS, AND G. ZHOU (2009): "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy," The Review of Financial Studies, 23, 821–862.

SUYKENS, J. AND J. V ANDEWALLE (1999): "Least Squares Support Vector Machine Classifiers," Neural Processing Letters, 9.

XIE, T., YU, J., AND T. ZENG (2020): "Econometric Methods and Data Science Techniques: A Review of Two Strands of Literature and an Introduction to Hybrid Methods," Working Paper, Singapore Management University.

ZOU, H. AND T. HASTIE (2005): "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67, 301–320.