Special Feature C

Forecasting Singapore GDP Using SPF Data

Tian Xie and Jun Yu[1]

In this Special Feature, we use econometric and machine learning (ML) methods, as well as 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 MAS. It is shown that the relationship between the actual GDP growth rates and the forecasts from individual professionals is highly non-linear and non-additive, making it difficult 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 by about 50% relative to that of the sample median.

1 Introduction

A very large body of applied work in economics has tried to forecast key macroeconomic indicators, including GDP growth rates, unemployment rates, and inflation rates, reflecting the vital importance of these macroeconomic variables to many decision makers in the economy. In this Special Feature, we focus our attention on predicting the GDP growth rate in Singapore with both conventional econometric and machine learning (ML) methods, using responses to MAS' Survey of Professional Forecasters (SPF).[2]

The SPF is a leading survey of macroeconomic forecast consensus in Singapore, which has been conducted by MAS since Q4 1999.[3] The survey is conducted quarterly following the release of economic data for the previous quarter by the Ministry of Trade and Industry (MTI) and contains forecasts for 15 key economic indicators, including the y-o-y real GDP growth rate. It should be noted that the SPF results do not represent MAS' own views or forecasts.

Every quarter, MAS reports the sample median and the empirical density of the forecasts from respondents. In this Special Feature, we denote the sample median as the benchmark forecast whereas Genre *et al.* (2013) employ the sample mean as the benchmark in another strand of the literature. We find that the difference between the sample median and the sample mean is negligible in the SPF.

[1] Tian Xie is an Associate Professor in the College of Business, Shanghai University of Finance and Economics (SHUFE), Shanghai, China. Jun Yu is a Lee Kong Chian Professor of Economics and Finance at both the School of Economics and the Lee Kong Chian School of Business, Singapore Management University (SMU). Support for research on this topic from MAS under the AI and Data Analytics (AIDA) Grant programme (grant: RFP001) is gratefully acknowledged. The views in this article are solely those of the authors and should not be attributed to MAS, SHUFE or SMU.

[2] The SPF is made available to the public at https://www.mas.gov.sg/monetary-policy/MAS-Survey-of-Professional-Forecasters.

[3] 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 ECB for the Eurozone since the late 1990s.

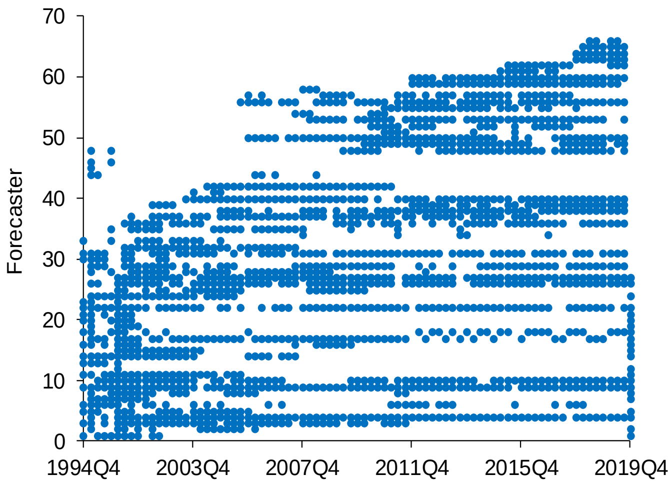
We first describe the data in Section 2. In Section 3, we introduce alternative methods for obtaining the forecasts and discuss the criteria used to evaluate those forecasts. Section 4 provides an empirical analysis to contrast the performance of alternative methods and the benchmark method. Section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

2 Data

In this Special Feature, we consider utilising individual forecasts from the SPF, denoted as $\{x_{1t},...,x_{pt}\}$, where $i=1,...,p$, to predict the real GDP growth rate, denoted as y_{T+1}. Here i represents the i^{th} forecaster, t represents the period t where $t=1,...,T$. From Q4 1999 to Q4 2019, the SPF collected one-month-ahead predictions of the quarterly real GDP growth rate from 66 different forecasters.[4] At period T, the sample median of $\{x_{1T},...,x_{pT}\}$, which is the "middle" number of these numbers when they are listed in ascending order, acts as the final forecast of y_{T+1}.

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. **Chart 1** describes the entries and exits of individual forecasters over the survey period. A blue dot represents a specific forecaster (labelled in the vertical axis) if he or she submitted a survey response and a blank space indicates otherwise.

Chart 1 An illustration of the entries and exits of individual forecasters



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 remove irregular respondents if he or she misses more than 50% of the observations. In the end, we narrowed down to $p=15$ qualified forecasters. Then the missing observations for each forecaster are filled using the approach suggested in Genre *et al.* (2013).

4 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.

3 Methods

Let $\boldsymbol{X}_{\mathrm{t}} = [1, x_{1t}, ..., x_{pt}]'$. If all the p forecasters are used in the prediction model, 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:

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

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 the Least Squares method.

In practice, we do not know if all forecasters improve model projections. If most of the variables in $\boldsymbol{X}_t$ 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 $\boldsymbol{X}_t$ should be linear and additive. Although it is theoretically possible to specify a general functional form to relate y_{t+1} and $\boldsymbol{X}_t$ as follows:

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

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

In this section, we review four methods to forecast Singapore's GDP based on SPF survey outcomes. Other than the benchmark method of the sample median, we also use the complete subset regression (CSR) of Elliott *et al.* (2013), the Elastic Net (EN) method of Zou and Hastie (2005), the Least Squares Support Vector Regression (LSSVR) method of Suykens and Vandewalle (1999) and 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 ML technique. The last method combines an econometric method with a ML method. Most econometric methods impose prior assumptions on the data generating process (DGP), which is necessary for deriving useful statistical properties. On the other hand, many ML methods are data-driven and do not require assumptions on the DGP. For these methods, statistical properties are not the primary concern. A more extensive survey of both econometric and ML methods for a forecasting purpose can be found in Xie *et al.* (2020).

3.1 Complete Subset Regression (CSR)

The CSR of Elliott *et al.* (2013) is a method for mixing forecasts from all possible linear regression models, each of which uses only a subset of predictors. Specifically, each model 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 one, although we use five predictor variables in our empirical study, implying that there are p subsets of predictors, and thus p possible linear regression models. In this case, the equally weighted forecast of y_{T+1} is given by:

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

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

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

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

3.2 Elastic Net (EN)

When the number of predictors p is large and a significant subset of predictors is not informative in predicting y_{t+1}, Equation (1) and the Least Squares method do not perform well out-of-sample. Many penalised regressions have been proposed to select predictors which can improve predictive precision. One of the successful methods is the EN of Zou and Hastie (2005). The idea of the EN is to shrink the slope parameter towards zero if the associated predictor is not significant. An insignificant predictor provides little explanatory power on y_{t+1} but may introduce a large variation on prediction outcomes. By shrinking the magnitude of the slope parameter, we reduce the prediction variance and therefore improve the prediction accuracy.

The EN imposes a constraint on the sum of squared coefficients excluding the intercept. That is,

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

where $\boldsymbol{\beta}^* = [\beta_0, \beta_1, ..., \beta_p]'$, the second term in the curly 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 the 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 (LSSVR)

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

The Support Vector Regression (SVR) of Drucker *et al.* (1996) approximates $f(\boldsymbol{X}_t)$ by a set of basic elements, in which the ensemble of these elements mimics $f(\boldsymbol{X}_t)$. In mathematics, we call these basic elements basis functions. Denote $\{h_s(X_t)\}_{s=1}^{S}$ as basic functions that can be of infinite dimensions. Equation (1) can thus be rewritten in the following form:

$$y_{t+1} = f(\boldsymbol{X}_t) + \varepsilon_t \approx \sum_{s=1}^{S} \beta_s h_s(\boldsymbol{X}_t) + \varepsilon_t \tag{5}$$

To estimate $\boldsymbol{\beta} = [\beta_1, ..., \beta_s]'$, we minimise the following criterion:

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

where $V_e(.)$ is the loss function. If $|.| < e$, the loss takes a value zero as if its loss is "tolerated" by the method. If $|.| \geq e$, the loss is defined to be $|.| - e$.

Suykens and Vandewalle (1999) modified the SVR algorithm by replacing $V_e(\cdot)$ with a squared loss, which results in solving a set of linear equations. This method is known as LSSVR, which leads to the following expression for the optimal solution:

$$\hat{f}(\boldsymbol{X}_t) = \sum_{t=1}^{T-1} \widehat{\alpha}_t K(\boldsymbol{x}, \boldsymbol{X}_t) \tag{7}$$

where $\boldsymbol{x}$ is any given vector of values for predictors, $\{\widehat{\alpha}_t\}_{t=1}^{T}$ are the estimated Lagrangian multipliers in the optimisation problem, and $K(\cdot,\cdot)$ is the predetermined kernel function. We consider the Gaussian kernel function given by:

$$K(\boldsymbol{x}, \boldsymbol{X}) = e^{-(\|\boldsymbol{x}-\boldsymbol{X}\|)/(2\sigma_x^{\ 2})} \tag{8}$$

where $\sigma_x^{\ 2}$ is a hyperparameter that users specify in advance.

3.4 LSSVR^{MA}

Most ML 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 non-linear 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 ML strategies, as suggested in Qiu *et al.* (2020). Following Qiu *et al.* (2020), we blend the idea of forecast combination with the LSSVR method. The new method is denoted LSSVR^{MA}, where the superscript MA indicates model averaging.

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

$m = 1, \ldots, M$. Qiu *et al.* (2020) then construct the weighted average forecast of y_{T+1} and choose the weights using an information criterion.

4 Empirical Results

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

Table 1 Summary of the five methods to forecast Singapore's GDP growth

Method	Parameter
SPF median	Median of all available forecasts from SPF
CSR	5 predictors, 1000 models, equal weight
EN	$\lambda = 0.5$, $\alpha = 0.5$
LSSVR	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 Singapore's GDP growth. The initial period for making the forecast is Q4 2009. The window length is set to 40. The out-of-sample performance of the five methods is evaluated by mean squared forecast error (MSFE) and mean absolute forecast error (MAFE) as defined by:

$$\text{MSFE} = \frac{1}{K}\sum_{k=1}^{K}(y_{T+k} - \hat{y}_{T+k})^2 \tag{9}$$

$$\text{MAFE} = \frac{1}{K}\sum_{k=1}^{K}\left|y_{T+k} - \hat{y}_{T+k}\right| \tag{10}$$

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 five methods.

The values of MSFE and MAFE and their associated ranking for all the five methods are reported in **Table 2**. The lowest MSFE and MAFE are presented in boldface.

5 We also consider alternative settings of tuning parameters. The results are qualitatively intact.

Table 2 Out-of-sample forecasting comparison of five methods

Methods	MSFE		MAFE	
	Value	Ranking	Value	Ranking
SPF Median	26.73	4	3.54	5
CSR	28.82	5	3.50	4
EN	25.70	3	3.27	3
LSSVR	14.24	2	2.74	2
LSSVR^{MA}	**13.96**	**1**	**2.69**	**1**

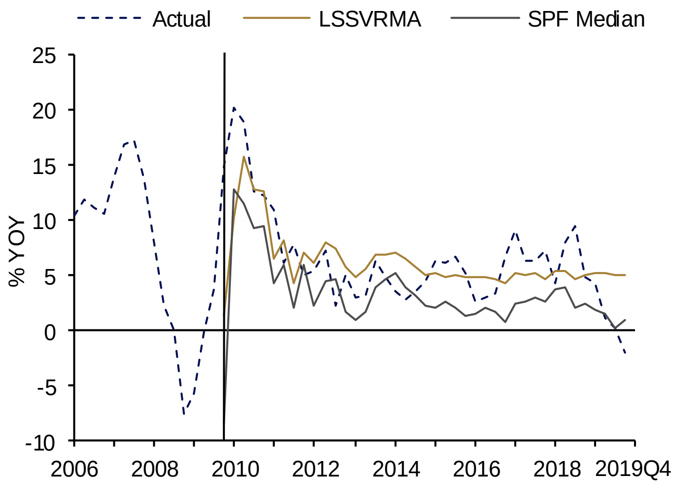
A few conclusions can be drawn from **Table 2**. First, it can be seen that all methods tested improve on the SPF median forecast with the exception of the CSR, which is worse than the SPF median under MSFE.

Second, LSSVR^{MA} always performs the best followed by LSSVR. LSSVR^{MA} acknowledges model uncertainty by combining forecasts from different candidate models, whereas LSSVR ignores model uncertainty and relies on one single model to deliver the forecasts. The sound performance of LSSVR^{MA} relative to LSSVR suggests that there exists model uncertainty. Note that this does not necessarily imply that all the models considered in LSSVR^{MA} help to improve the GDP growth prediction (at least not equally). For example, it is possible that an accurate combined forecast is due to two highly biased forecasters, one of which overpredicts and the other underpredicts.

Third, the two LSSVR-based methods perform much better than the other three methods, implying a non-linear dependence between y_{t+1} and x_{it}'s. For example, compared to the benchmark median method, LSSVR^{MA} gains at reducing the MSFE value by almost 50%. If we fit a partially linear model, one could see a strong non-linear relationship between y_{t+1} and individual x_{it}. In the interests of brevity, the empirical results for the partially linear model are not reported here. Fourth, the fact that the EN slightly outperforms CSR and the sample median indicates that there is no strong evidence of sparsity in x_{it}'s. We also note that even the best method yields a high MAFE value of 2.69%. This is due to the fact that our evaluation interval (Q4 2009 – Q4 2019) overlaps with periods of high macroeconomic volatility such as the recent trade war between the US and China, which makes forecasting GDP growth unusually difficult.

To visually compare the forecast accuracy of the benchmark method and the LSSVR^{MA} method, we plot two forecasted series of these two methods against the actual data in **Chart 2**. It is apparent that the SPF median forecast often underestimates the actual values, especially from 2015–18. Although flatter than the actual values, the forecasts by the LSSVR^{MA} method captures the level and the trend reasonably well.

Chart 2 A comparison of two forecasts of Singapore GDP growth



To examine if the improvement in forecast accuracy is significant, we perform the Giacomini-White (GW) test of the null hypothesis that the method listed in the columns of **Table 3** performs equally well as the method listed in the rows in terms of absolute forecast errors (Giacomini and White, 2006). The p-values of the GW test for all pair-wise comparisons are reported in the table. The five methods can be divided into two groups. The SPF median, CSR, and the EN 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	SPF Median	CSR	EN	LSSVR
SPF Median	-	-	-	-
CSR	0.4345	-	-	-
EN	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 ML method, and a hybrid method, to forecast the GDP growth rate in Singapore based on the SPF data. The performance of these methods is then compared to the sample median of SPF forecasts. 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. Our results also show that the forecasts of most, if not all, of the professional forecasters contain useful information about the next-quarter Singapore GDP growth rate. Therefore, they should not be given a zero weight in models for forecasting GDP. Since the relationship between SPF forecasts of GDP growth and GDP outturns is potentially non-linear and complex, a ML method is helpful in this case. Moreover, the hybrid method leads to the most accurate forecasts, likely because it can accommodate model uncertainty.

References

Drucker, H, Burges, C J C, Kaufman, L, Smola, A J and Vapnik, V (1996), "Support Vector Regression Machines", in Advances in Neural Information Processing Systems 9, *MIT Press,* pp 155–161.

Elliott, G, Gargano, A and Timmermann, A (2013), "Complete Subset Regressions", *Journal of Econometrics*, Vol. 177(2), pp. 357–373.

Genre, V, Kenny, G, Meyler, A and Timmermann, A (2013), "Combining Expert Forecasts: Can Anything Beat the Simple Average?", *International Journal of Forecasting*, Vol. 29(1), pp. 108–121.

Giacomini, R and White, H (2006), "Tests of Conditional Predictive Ability", *Econometrica*, Vol. 74(6), pp. 1545–1578.

MAS (n.d.), "Survey of Professional Forecasters" (URL: https://www.mas.gov.sg/monetary-policy/MAS-Survey-of-Professional-Forecasters).

Qiu, Y, Xie, T and Yu, J (2020), "Forecast Combinations in Machine Learning", *SMU Economics and Statistics Working Paper Series* No. 13-2020.

Rapach, D E, Strauss, J K and Zhou, G (2010), "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy", *Review of Financial Studies*, Vol. 23(2), pp. 821–862.

Suykens, J A K and Vandewalle, J (1999), "Least Squares Support Vector Machine Classifiers", *Neural Processing Letters*, Vol. 9, pp. 293–300.

Xie, T, Yu, J and Zeng, T (2020), "Econometric Methods and Data Science Techniques: A Review of Two Strands of Literature and an Introduction to Hybrid Methods", *SMU Economics and Statistics Working Paper Series* No. 16-2020.

Zou, H and Hastie, T (2005), "Regularization and Variable Selection via the Elastic Net", *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, Vol. 67(2), pp. 301–320.