

Improvement of Regression Forecasting Models

Vasily Aleksandrovich Zubakin¹, Oleg Anatolyevich Kosorukov¹ & Nikita Aleksandrovich Moiseev¹

¹ Plekhanov Russian University of Economics, Russian Federation

Correspondence: Vasily Aleksandrovich Zubakin, Plekhanov Russian University of Economics, 117997, Moscow, Stremyanny per., 36, Russian Federation.

Received: December 15, 2014 Accepted: January 18, 2015 Online Published: May 30, 2015

doi:10.5539/mas.v9n6p344

URL: <http://dx.doi.org/10.5539/mas.v9n6p344>

Abstract

In this paper authors propose the technique, which decreases average forecast error of regression based models. The main idea of the method is to use the weighted sum of several regression equations, which satisfy Ordinary Least Squares prerequisites and have independent residuals, instead of only one. It is shown that if all method requirements are met, it is possible to decrease Mean Squared Error almost by half, using just three equations. This technique allows deriving equations which contain more predictors than the number of observations. Additionally, this method proves to be more consistent in time than any of regressions, used in it, separately. It is also illustrated, that the proposed method outperforms the regression equation, computed with the same independent variables, and, thus, it gives more accurate estimators of regression coefficients. Empirical results are provided as well.

Keywords: regression model, forecast error, coefficients estimation.

1. Introduction

When modeling a social or economic process a researcher nearly always encounters an uncertainty whether created model will work in the future with the same effectiveness, what is partly highlighted by James Stock and Mark Watson (2007), Alhamzawi, R and Yu, K. (2012), Clark, T.E., McCracken, M.W. (2009), Gneiting, T. (2011) and Giordani, P., Kohn, R., van Dijk, D. (2007). In other words, even if a linear model fully satisfies all the OLS (Ordinary Least Squares) prerequisites and has some forecast error magnitude, one cannot assert that while using it for prediction this very magnitude stays at least within some acceptable area. Sometimes a forecast error may be several times as big as the original one and in this case such model will not be of much use. That happens due to the following reasons. Either unaccounted factors changed their values so, that coefficients estimators become biased, or accounted factors change the extent of their impact on the output variable. Actually, it can also be a combination of both, see for example Orphanides, A. and S. van Norden (2005) and Primiceri, G. (2005). In order to decrease those errors a researcher can work out a model which takes into account structural breaks and coefficients variability in regression equation, refer for example to Jan J.J. Groen, Richard Paap and Francesco Ravazzolo (2009) and Sensier, M., Van Dijk, D. (2004), Clark, T.E. (2011), Jore, A.S., J. Mitchell and S. P. Vahey (2010) and Koop, G. and S. Potter (2007). However, the following problem, to authors' personal opinion, is still not fully solved. While specifying a regression equation and trying to adjust it to satisfy all the OLS prerequisites, one may skip a lot of data which actually impacts the model output. That is why we devote this paper to development of the method that could grasp more explanatory variables and thus significantly decrease forecast error without violating the classical way of regression model specification.

2. Empirical Background of the Method

Let us say we are creating an inflation forecasting model for the USA economy. Quarterly CPI (Consumer Price Index) is selected as the output variable y_t . As possible independent variables we test three lags of the dependent variable y_t and of each quarterly index for the following macroeconomic indicators: GDP (Gross Domestic Product), GDI (Gross Domestic Income), Federal Funds Rate, Non-Farm Payrolls, Brent Oil Price, Dow Jones Industrial Average, RGDP (Real Gross Domestic Product), Monetary Base M2, Total Export, Total Import, Money Velocity (calculated as a ratio of GDP and M2), Employment Rate. Thus, the regression equation can be written as follows:

$$y_{t+1} = a_0 + \sum_{j=1}^3 a_j y_{t-j} + \sum_{i=1}^n \sum_{j=1}^3 b_{ij} x_{i(t-j)} \tag{1}$$

where n is the number of explanatory variables, a_j and b_{ij} stand for coefficients for the lagged CPI and lagged explanatory variables respectively, a_0 – constant term.

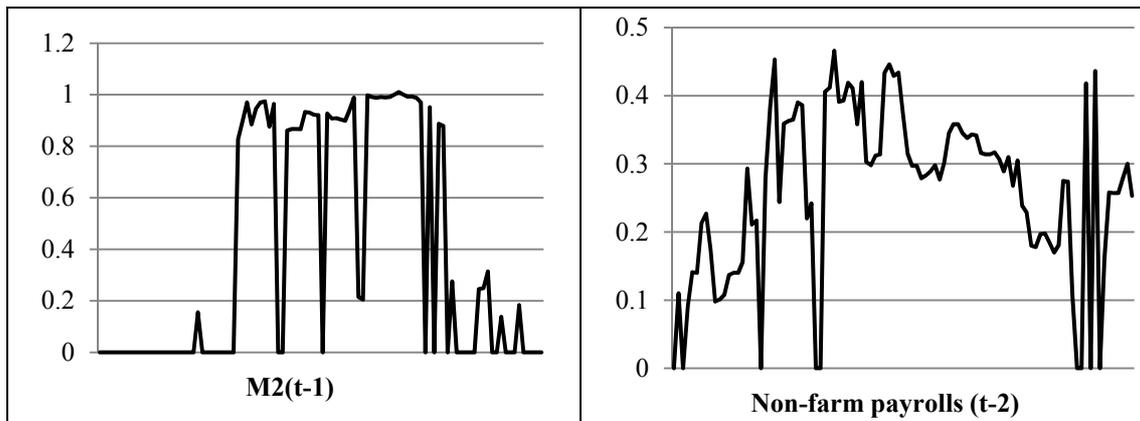
Here we don't take into account zero lags, because all the data mentioned above are not available at the very beginning of the next time period and are issued only during the quarter. Therefore one cannot make a forecast before another quarter starts. For building the model we take 95 observations, beginning from the 1st quarter of 1960. After the first equation is set up, the data frame is shifted forward by one observation and again new equation is computed. This procedure goes on for 70 times. To specify the regression equation the following optimization problem with restrictions is used:

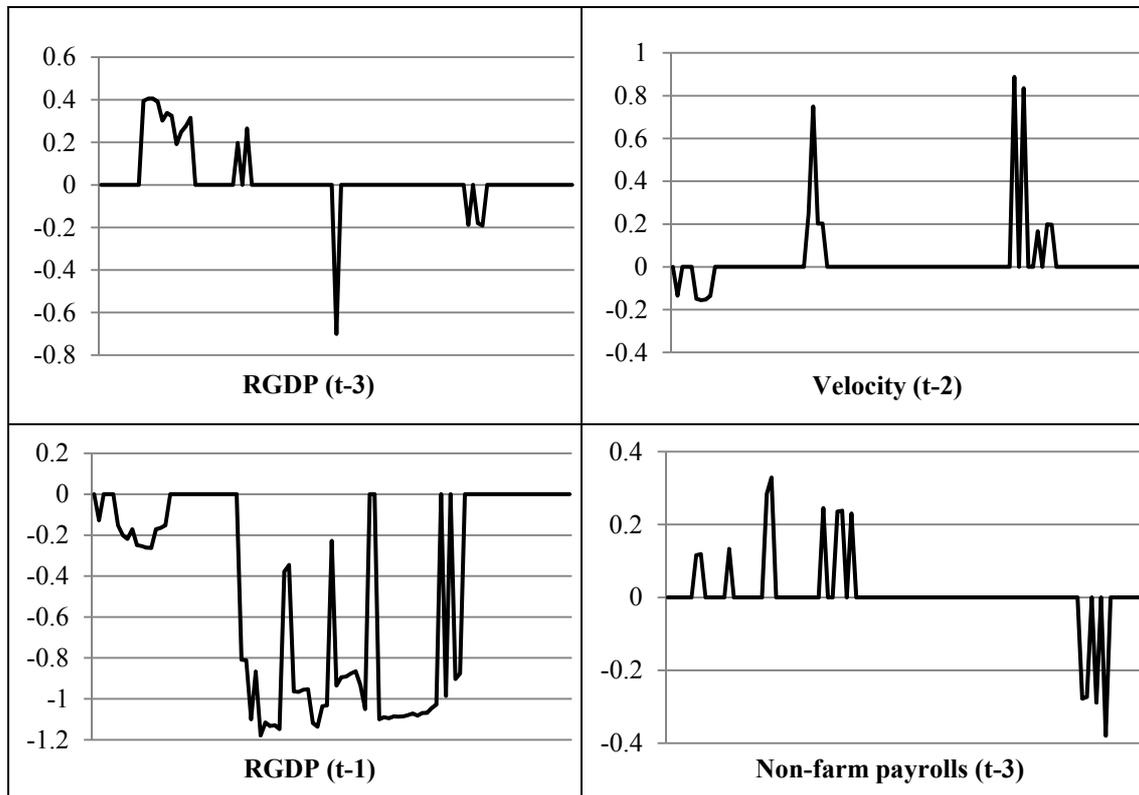
$$\begin{aligned}
 &MSE \rightarrow \min \\
 &\left\{ \begin{aligned}
 &a_k < 0,05; \\
 &m \leq 10; \\
 &VIF_k \leq 7; \\
 &CN \leq 6; \\
 &1,5 < DW < 2,5
 \end{aligned} \right. \tag{2}
 \end{aligned}$$

where MSE – Mean Squared Error, α_k – significance level for k -th predictor, $k \in [1..m]$, m – number of selected predictors, VIF_k - Variance Inflation Factor for k -th predictor, CN – Condition Number.

Here it is actually not so important which variable selection method we use, whether it is Bayesian variable selection, Gibbs sampling or classical approach, refer to De Mol C., D. Giannone and L. Reichlin (2008) and Wright, J.H. (2009). Let us say we just choose the best fitting regression shape according to some algorithm. To illustrate the regression structure inconsistency we provide some coefficients magnitude dynamics throughout these 70 equations. The results are shown in the Table 1, where Y-axis denotes coefficient magnitude and X-axis – the equation number. If a predictor was not included in equation, its coefficient takes up zero.

Table 1. Coefficients value dynamics





From the table one can see, that there are either some more or less consistent predictors, such as M2(t-1), Non-farm Payrolls (t-2) and RGDP (t-1), or not consistent at all (RGDP (t-3), Velocity (t-2) and Non-farm Payrolls (t-3)). The experiment revealed a huge amount of structural breaks while calculating the equations. Nearly every shift of the data frame a better regression could be built. Moreover, some coefficients even change their magnitude from positive to negative and vice versa. Therefore, we dare to assume, that when skipping some factors which are actually important, coefficients of accounted predictors bear some part of their impact on regression output, see for example Pesaran, M. H., D. Pettenuzzo and A. Timmermann (2006) and Gneiting, T., Raftery, A.E. (2007). For instance, M2(t-1) often performs as quite a significant predictor (coefficient is relatively high). And when M2(t-1) is not accounted in an equation due to multicollinearity reasons doesn't mean it stopped its influence on CPI. Its impact is just redistributed among other predictors coefficients, refer to Justiniano, A. and G. E. Primiceri (2008), Giordani, P. and M. Villani (2010) and Jore, A.S., Mitchell, J., Vahey, S.P. (2010). And if during forecasting M2(t-1) starts to display greater volatility, our equation will not be any more consistent as it doesn't comprise M2(t-1) as an explanatory variable. That's why a researcher should aspire to include as many dependent variables into a model as possible. But at the same time the more predictors one takes the higher risk of multicollinearity and greater errors in coefficients estimators is (Primiceri, G.E. (2005) and P. Newbold (1997)). Thus, to find the optimal balance is very important. Moving forward, we noticed that within one data frame several regression equations can be specified which would satisfy inequality set (2). And the problem of selecting one of them converges just to picking an equation with the minimum MSE. However, using the latter for forecasting doesn't guarantee us the best prediction compared to other possible regressions which were discarded, especially if MSE differ insignificantly. For example, let us consider three possible regression equations R_1 , R_2 and R_3 for the very first data frame. We provide general information about each of them in Table 2.

Table 2. Summary for R_1 , R_2 and R_3

Indicator	R_1	R_2	R_3
m	8	10	7
R	0.943	0.947	0.922
F-statistic sig.	0.000	0.000	0.000
DW	1.531	1.763	1.757
α	CPI (t-2) - 0.001	NFPR (t-1) - 0.000	CPI (t-1) - 0.000

	CPI (t-3) – 0.000	Brent (t-1) – 0.005	GDP (t-2) – 0.000
	NFPR (t-2) – 0.000	M2 (t-1) – 0.000	NFPR (t-2) – 0.000
	Brent (t-1) – 0.021	DJIA (t-1) – 0.004	Brent (t-2) – 0.021
	Import (t-2) – 0.020	RGDP (t-1) – 0.000	DJIA (t-1) – 0.003
	Export (t-1) – 0.000	Import (t-2) – 0.037	RGDP (t-1) – 0.022
	Export (t-2) – 0.021	Export (t-2) – 0.005	Export (t-1) – 0.012
	Emp-ment (t-1) – 0.010	Velocity (t-1) – 0.000	
		Emp-ment (t-2) – 0.000	
		Emp-ment (t-3) – 0.000	
VIF	CPI (t-2) – 6.882	NFPR (t-1) – 2.993	CPI (t-1) – 3.179
	CPI (t-3) – 5.653	Brent (t-1) – 1.170	GDP (t-2) – 4.373
	NFPR (t-2) – 2.054	M2 (t-1) – 4.242	NFPR (t-2) – 1.753
	Brent (t-1) – 1.183	DJIA (t-1) – 1.679	Brent (t-2) – 1.142
	Import (t-2) – 2.882	RGDP (t-1) – 5.331	DJIA (t-1) – 1.179
	Export (t-1) – 2.303	Import (t-2) – 3.228	RGDP (t-1) – 1.817
	Export (t-2) – 2.377	Export (t-2) – 2.731	Export (t-1) – 1.723
	Emp-ment (t-1) – 2.638	Velocity (t-1) – 3.003	
		Emp-ment (t-2) – 2.437	
		Emp-ment (t-3) – 1.527	
CN	5.887	5.471	4.225
MSE	0.53083	0.53548	0.63767

All three equations satisfy inequality set (2) and according to the chosen algorithm we should pick R_1 for making forecast. However, applying R_1 to calculate the future magnitude of the dependent variable we will not always receive smaller forecast errors than if we used R_2 or R_3 . In order to illustrate this point, we take the data frame equal to 30 observations, starting from the first forecast value, and compute MSE for all three regressions considered. This procedure is made for 40 times shifting the frame by one observation forward. Thus, we take a look at how these regression equations would have worked if we had derived them back in 1984. The results are shown in Figure 1.

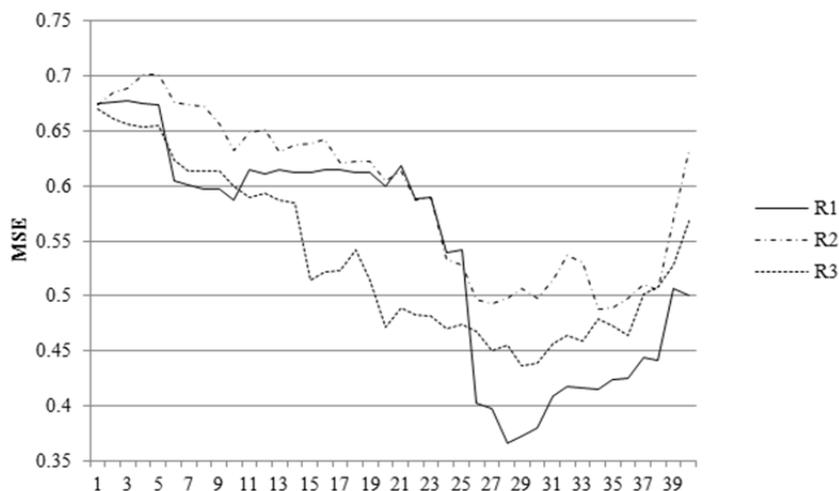


Figure 1. MSE dynamic for R_1 , R_2 and R_3

One can see that MSE of R_1 , R_2 and R_3 don't stay in the same order as in Table 2 and interweave throughout the time. It underlines the main idea of this section, that if the quality of several regression equations is more or less equal, then there is a high uncertainty level concerning which equation to choose for prediction. Therefore in the following section authors make an attempt to solve this problem.

3. The Method

Imagine we have a target variable y and a relatively large set of dependent variables $(x_1, x_2 \dots x_n)$. Let us also assume it is possible to derive l regression equations $(R_1, R_2 \dots R_l)$ within one observations frame and each R_i has $e_i \sim N(0; \sigma_i)$ and satisfies OLS prerequisites with significant predictors. Then we can create a new regression equation R , using the weighted sum of already computed ones.

$$\bar{y} = \sum_{i=1}^l \hat{y}_i f_i \tag{3}$$

where \hat{y}_i is an output of R_i , f_i - weighing function

For calculating weighing function f_i let us first consider standard deviation of \hat{y} , which denotes MSE of R . According to (3) $SD(\hat{y})$ is presented by the following expression.

$$SD(\hat{y}) = \sqrt{D\left(\sum_{i=1}^l e_i f_i\right) + D\left(\sum_{i=1}^l \hat{y}_i f_i\right)} \tag{4}$$

From formula 4 one can see that mean error of the combined model consists of the variance of errors and the variance of the regression line. Variance of errors is computed as follows.

$$D\left(\sum_{i=1}^l e_i f_i\right) = \sum_{i=1}^l \sum_{j=1}^l \text{cov}(e_i, e_j) f_i f_j \tag{5}$$

Variance of the regression line can be presented in the following form.

$$D\left(\sum_{i=1}^l \hat{y}_i f_i\right) = \sum_{i=1}^l \sum_{j=1}^l \text{cov}(\hat{y}_i, \hat{y}_j) f_i f_j \tag{6}$$

Hence our task converges to computation of covariance between two regressions included in the ultimate model.

$$\begin{aligned} \text{cov}(\hat{y}_i, \hat{y}_j) &= M((\hat{y}_i - M(\hat{y}_i))(\hat{y}_j - M(\hat{y}_j))) = M(\hat{X}_i^T (B_i - \beta_i)(B_j - \beta_j)^T \hat{X}_j) = \\ &= M(\hat{X}_i^T (X_i^T X_i)^{-1} X_i^T \epsilon_i \epsilon_j^T X_j (X_j^T X_j)^{-1} \hat{X}_j) = \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T M(\epsilon_i \epsilon_j^T) X_j (X_j^T X_j)^{-1} \hat{X}_j = \\ &= \text{cov}(e_i, e_j) \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T X_j (X_j^T X_j)^{-1} \hat{X}_j \end{aligned} \tag{7}$$

where \hat{X}_i is a column-vector of the i -th model predictors values used for forecasting

Thus variance of the combined model may be rewritten as in formula 8.

$$D(\hat{y}) = \sum_{i=1}^l \sum_{j=1}^l \text{cov}(e_i, e_j) f_i f_j + \sum_{i=1}^l \sum_{j=1}^l \text{cov}(e_i, e_j) \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T X_j (X_j^T X_j)^{-1} \hat{X}_j f_i f_j \tag{8}$$

For being able to get $D(\hat{y})$ from formula 8, we need an unbiased estimator of models residuals covariance, which will be computed as presented below.

$$\text{cov}(e_i, e_j) = \frac{\sum_{m=1}^n e_{im} e_{jm}}{\sqrt{(n - k_i)(n - k_j)}} \tag{9}$$

where k_i is a number of i -th model parameters

As the main goal of the method being elaborated is to create an equation, which would outperform all the other computed, we encounter a problem of $D(\hat{y})$ minimization, which can be easily solved by quadratic programming method. In this case optimization task will look as follows.

$$\left| \begin{aligned} D(\hat{y}) &\rightarrow \min \\ \sum_{i=1}^l f_i &= 1 \end{aligned} \right. \tag{10}$$

Let us also take into account the fact that if the set of e_i are i.i.d. variables and are subject to $N(0; \sigma_i)$ (what we

assume to be true for all our models residuals), then $\hat{\epsilon} \sim N(0; \sigma_{\hat{\epsilon}})$ as well. Thus, we can figure out the interval estimate for the combined model forecast.

We point out that it is common for confidence interval computation in case of linear regression to use T -distribution with $(n-k-1)$ degrees of freedom if a constant is present in the equation and with $(n-k)$ degrees of freedom in case of its absence.

$$e - T_{\alpha, n-k-1} \times s < \mathcal{E} < e + T_{\alpha, n-k-1} \times s \quad (11)$$

where s is an unbiased estimator of the true standard deviation

In case of newly computed regression model, formula 11 will look as follows.

$$\bar{y} - T_{\alpha, r} \times SD(\bar{y}) < \hat{y} < \bar{y} + T_{\alpha, r} \times SD(\bar{y}) \quad (12)$$

Now our task converges to the search of degrees of freedom number r . As

$$\frac{e - \mathcal{E}}{s} = \frac{\frac{e - \mathcal{E}}{\sigma}}{\sqrt{\frac{(n-k)s^2}{(n-k)\sigma^2}}} \sim T_{n-k} \quad (13)$$

Then we can infer, that

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi_{n-k}^2 \Rightarrow \frac{rs^2}{\sigma^2} \sim \chi_r^2 \quad (14)$$

Using the fact, that the variance of chi-square distribution is two times as great as the number of degrees of freedom, we have:

$$Var(\chi_r^2) = 2r \Rightarrow Var\left(\frac{rs^2}{\sigma^2}\right) = 2r \Rightarrow \frac{r^2}{\sigma^4} Var(s^2) = 2r \Rightarrow r = \frac{2\sigma^4}{Var(s^2)} \quad (15)$$

As $\bar{y} = \sum_{i=1}^l \hat{y}_i f_i$, then

$$\begin{aligned} Var(\bar{y}) &= Var\left(\sum_{i=1}^l \sum_{j=1}^l cov(e_i, e_j) f_i f_j \left(1 + \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T X_j (X_j^T X_j)^{-1} \hat{X}_j\right)\right) = \\ &= \sum_{i=1}^l \sum_{j=1}^l f_i^2 f_j^2 \left(1 + \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T X_j (X_j^T X_j)^{-1} \hat{X}_j\right)^2 Var(cov(e_i, e_j)) \end{aligned} \quad (16)$$

as we suppose there is no dependencies among covariances.

In order to compute the variance of residuals covariance let us refer to the Wishart distribution. Hence we give a quick reference on this distribution. Suppose X is an $n \times p$ matrix, each row of which is independently drawn from a p -variate normal distribution with zero mean.

$$X_{(i)} = (x_i^1, x_i^2, \dots, x_i^p) \sim N_p(0, V) \quad (17)$$

Then the Wishart distribution is the probability distribution of the $p \times p$ random matrix $S = X^T X$. One indicates that S has that probability distribution by writing

$$S \sim W_p(V, n) \quad (18)$$

The positive integer n is the number of degrees of freedom. If $p = V = 1$ then this distribution is a chi-squared distribution with n degrees of freedom.

Thus we can write that

$$\frac{\text{cov}(e_i e_j) \sqrt{(n-k_i)(n-k_j)}}{\sigma_i \sigma_j} \sim W_1(1, \sqrt{(n-k_i)(n-k_j)}) \tag{19}$$

We also know that the variance of a matrix element which is subject to the Wishart distribution is computed as follows.

$$D(s_{ij}) = n(v_{ij}^2 + v_{ii}v_{jj}) \tag{20}$$

Applying formula 19 to the formula 20 we obtain

$$\text{Var}\left(\frac{\text{cov}(e_i e_j) \sqrt{(n-k_i)(n-k_j)}}{\sigma_i \sigma_j}\right) = \sqrt{(n-k_i)(n-k_j)}(v_{ij}^2 + v_{ii}v_{jj}) \tag{21}$$

Making some trivial calculations we get

$$\text{Var}(\text{cov}(e_i e_j)) = \frac{(v_{ij}^2 + v_{ii}v_{jj})\sigma_i^2\sigma_j^2}{\sqrt{(n-k_i)(n-k_j)}} = \frac{(\text{cov}^2(\varepsilon_i \varepsilon_j) + \sigma_i^2\sigma_j^2)}{\sqrt{(n-k_i)(n-k_j)}} \tag{22}$$

Integrating (19) into (18) and then applying it to formula 16 we obtain the number of degrees of freedom r .

$$r = \frac{2\bar{\sigma}^{-4}}{\sum_{i=1}^l \sum_{j=1}^l f_i^2 f_j^2 \left(1 + \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T X_j (X_j^T X_j)^{-1} \hat{X}_j\right)^2 \frac{(\text{cov}^2(\varepsilon_i \varepsilon_j) + \sigma_i^2\sigma_j^2)}{\sqrt{(n-k_i)(n-k_j)}}} \tag{23}$$

As we do not know the true variance of the ultimate model and true covariances and variances of models included in the ultimate one, we should substitute them with their unbiased point estimator. And in this case we get our final formula of degrees of freedom.

$$r = \frac{2\left(\sum_{i=1}^l \sum_{j=1}^l \text{cov}(e_i e_j) f_i f_j \left(1 + \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T X_j (X_j^T X_j)^{-1} \hat{X}_j\right)\right)^2}{\sum_{i=1}^l \sum_{j=1}^l f_i^2 f_j^2 \left(1 + \hat{X}_i^T (X_i^T X_i)^{-1} X_i^T X_j (X_j^T X_j)^{-1} \hat{X}_j\right)^2 \frac{(\text{cov}^2(e_i e_j) + s_i^2 s_j^2)}{\sqrt{(n-k_i)(n-k_j)}}} \tag{24}$$

Thus, not only does proposed method allow to significantly decrease MSE acceptance of the hypothesis about residuals independence, but also it reduces its confidence interval. To summarize calculations and inferences above, we provide a stepwise algorithm of regression R computation.

1. Calculate all possible regression equations satisfying OLS prerequisites with significant predictors;
2. Find f_i by solving the optimization problem (10);
3. Compute regression R by formula 3 using already obtained weighing functions f_i ;
4. Calculate a point forecast using regression R from the previous step;
5. Compute confidence interval for the point forecast by formula 12 with degrees of freedom computed by formula 24.

An ideal case would be several regression equations with equal MSE and independent residuals. Just three regression equations which meet all the requirements allow us to decrease MSE of the fittest equation by about 40%. It is also worth noticing that according to the algorithm above proposed method allows deriving an equation, which would comprise even more explanatory variables than there are observations. The prominent feature of the method is that it is fully automatic and can operate without any interference from a researcher.

4. Application to Real Data

In this section we use the same data as in Section 1 for testing the method. In Table 3 we provide pair correlation for residuals of chosen regressions R_1, R_2 and R_3 .

Table 3. Pearson’s correlation for residuals of R_1 , R_2 and R_3

	Res1	Res2	Res3
Res1	1	0.670	0.649
Sig. (2-sided)	-	0.000	0.000
N	95	95	95
Res2	0.670	1	0.703
Sig. (2-sided)	0.000	-	0.000
N	95	95	95
Res3	0.649	0.703	1
Sig. (2-sided)	0.000	0.000	-
N	95	95	95

As one can see positive correlation amongst these regression residuals is not very strong although significant. But as it will be shown further even with such equations we can substantially improve the accuracy of the forecast. We also provide in Table 4 Kolmogorov-Smirnov test for normality of the residuals in order to make sure, that we can use parametric interval estimators of forecast errors.

Table 4. Kolmogorov-Smirnov test for normality

	Res1	Res2	Res3	Res
N	95	95	95	95
MSE	0.53083	0.53548	0.64099	0.50159
Kolmogorov-Smirnov statistics	0.730	0.467	0.794	0.452
Asymptotic sig. (2-sided)	0.661	0.981	0.554	0.987

From Table 4 we see, that residuals for regression R (Res) show even more normality than any of selected regressions separately (Res1, Res2, Res3). Thus, interval estimates will be also more reliable for regression R what is another doubtless positive feature of the method. In Figure 2 we compare three regression residuals, which were already shown in Figure 1, with regression R , computed according to proposed method.

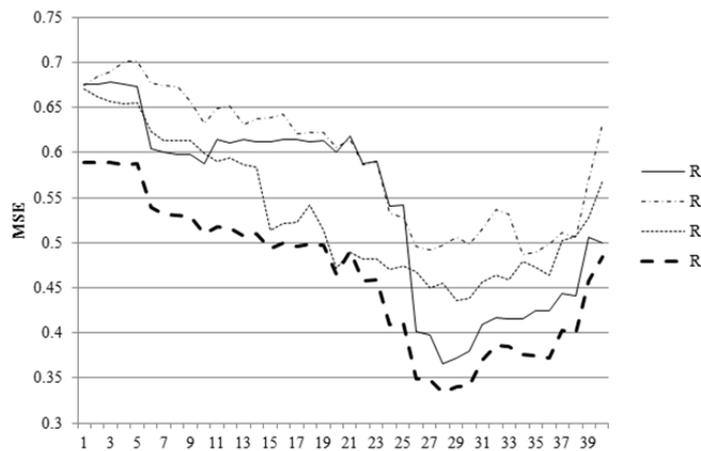


Figure 2. MSE dynamic for R_1 , R_2 , R_3 and R

It is easily seen, that R displays consistently smaller MSE throughout the data span of 70 observations than R_1 , R_2 and R_3 . we also want to draw our reader’s attention to the fact that during calculation of the equations Res was by 5.5% smaller than the minimum of Res1, Res2 and Res3 (0.50159 for Res against 0.53083 for Res1) and already during first 30 out-of-sample values Res became smaller than the same minimum by 12.2% (0.58913 for Res against 0.67063 for Res4). Regression R will stay more preferable unless there are some drastic shifts in MSE proportions of selected regressions. We can see that situation in Figure 2 around 19th -22nd data frame where Res3 was much lower than Res1 and Res2 what reflected on R negatively. Moving forward, in order to illustrate that the method allows us to get more accurate coefficients estimators we provide in Figure 3 MSE for regression (R_4) built using OLS, and which includes all the predictors used in R . We also provide the MSE for regression (R_5), which includes all predictors, claimed for possible inclusion into the model in Section 1.

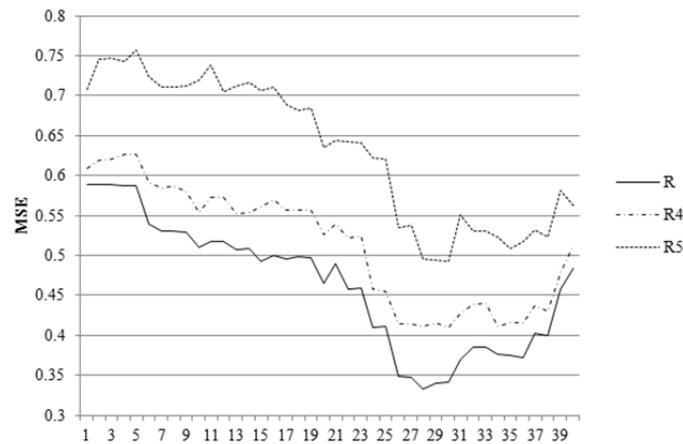


Figure 3. MSE dynamic for R_4 , R_5 and R

The reason why R is substantially better, than R_4 and R_5 is the following. Due to high multicollinearity there is a great risk of receiving wrong coefficients estimators, which actually don't properly reflect true dependencies amongst data. Of course within the data frame for deriving an equation the more predictors we use, the more accurate the model is. But as soon as we start using it for real forecast it returns much greater errors than an equation with less number of predictors but better satisfaction of OLS requirements. And dynamics of MSE for R_5 only proves this statement.

5. Conclusion

In the paper we try to make a small step on the way of improving existing methods of forecasting economic processes. We reckon that the main problem of all methods invented so far is inability to grasp larger number of predictors. Therefore a researcher may skip a lot of data which are significant and, thus, crucial for future predictions. Due to the peculiarity of economic systems, which is a dramatic lack of observations, it is quite a complicated task to derive a consistent forecasting method without enough statistical data. That is why we consider proposed method as extremely actual nowadays. Using it one can grasp more explanatory variables and calculate more consistent coefficients. It makes possible, for example, to include variables with high multicollinearity into a model and still get substantial coefficients, which would not be possible using just OLS. Additionally as it was mentioned before the method can be programmed and compute forecasts without any help from outside. The researcher's task is just to specify which variables to include for conducting selection, type of model and way of data transformation. Empirical results illustrate that even if method requirements are not fully met it is still possible to reach significantly better regression equation. Taking into account mentioned above, proposed method can be applied in different econometric models and specifically for government needs. Speaking of the future plans concerning further development of forecasting methods we plan to work out the way to apply the enunciated method for non-linear models. We also plan to work out the method of integration of equations computed on different data frames as an extension of the present work and believe that this procedure will help eliminate the problem of data frame selection.

Acknowledgements

This paper was completed under a research which was funded by Plekhanov Russian University of Economics. So we express our gratitude to the board of the University and especially to the rector Viktor Grishin, Doctor of Economics. We also thank Aleksander Kislitsyn, Lyubov Panfil, Nikita Mullin for their contribution to the research and valuable remarks on the present paper.

References

- Alhamzawi, R., & Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling, *Journal of Applied Statistics*, 39(4), 799-813. <http://dx.doi.org/10.1080/02664763.2011.620082>
- Clark, T. E., & McCracken, M. W. (2009). Tests of equal predictive ability with real-time data (2009) *Journal of Business and Economic Statistics*, 27(4), 441-454. <http://dx.doi.org/10.1198/jbes.2009.07204>
- Giannone, D., De Mol, C., & Reichlin, L. (2008). Forecasting with a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components? *Journal of Econometrics*, 146(5829), 318-328.

<http://dx.doi.org/10.1016/j.jeconom.2008.08.011>

- Giordani, P., & Villani, M. (2010). Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction. *International Journal of Forecasting*, 26, 312-325. <http://dx.doi.org/10.1016/j.ijforecast.2009.12.011>
- Giordani, P., Kohn, R., & van Dijk, D. (2006). A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, 137(1), 112-133. <http://dx.doi.org/10.1016/j.jeconom.2006.03.013>
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746-762. <http://dx.doi.org/10.1198/jasa.2011.r10138>
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359-378. <http://dx.doi.org/10.1198/01621450600001437>
- Jan, J. J., Groen, Richard, P., & Francesco, R. (2012). Real-Time Inflation Forecasting in a Changing World. Federal Reserve Bank of New York Staff Report no. 388. <http://dx.doi.org/10.1080/07350015.2012.727718>
- Jore, A. S., Mitchell, J., & Vahey, S. P. (2010). Combining Forecast Densities from VARs with Uncertain Instabilities. *Journal of Applied Econometrics* 25, 621-634. <http://dx.doi.org/10.1002/jae.1162>
- Jore, A. S., Mitchell, J., & Vahey, S. P. (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 25(4), 621-634. <http://dx.doi.org/10.1002/jae.1162>
- Justiniano, A., & Primiceri, G. E. (2008). The Time-Varying Volatility of Macroeconomic Fluctuations, *American Economic Review*, 98, 604-641. <http://dx.doi.org/10.1257/aer.98.3.604>
- Koop, G., & Potter, S. (2007). Estimation and forecasting in models with multiple breaks, *Review of Economic Studies*, 763-789. <http://dx.doi.org/10.1111/j.1467-937X.2007.00436.x>
- Orphanides, A., & van Norden, S. (2005). The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time, *Journal of Money, Credit and Banking*, 37, 583-601. <http://dx.doi.org/10.1353/mcb.2005.0033>
- Pesaran, M. H., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks, *Review of Economic Studies*, 73(4), 1057-1084. <http://dx.doi.org/10.1111/j.1467-937X.2006.00408.x>
- Primiceri, G. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies*, 72, 821-852. <http://dx.doi.org/10.1111/j.1467-937X.2005.00353.x>
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3), 821-852. <http://dx.doi.org/10.1111/j.1467-937X.2005.00353.x>
- Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, 29(3), 327-341. <http://dx.doi.org/10.1198/jbes.2010.09248>
- Sensier, M., Van Dijk, D. (2004). Testing for volatility changes in U.S. macroeconomic time series (2004) *Review of Economics and Statistics*, 86(3), 833-839. <http://dx.doi.org/10.1162/0034653041811752>
- Stock, J. H., & Watson, M. W. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking*, 39, 3-34. <http://dx.doi.org/10.1111/j.1538-4616.2007.00014.x>
- Wright, J. H. (2009). Forecasting US inflation by bayesian model averaging. *Journal of Forecasting*, 28(2), 131-144. <http://dx.doi.org/10.1002/for.1088>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).