WDS'09 Proceedings of Contributed Papers, Part I, 143–147, 2009. ISBN 978-80-7378-101-9 © MATFYZPRESS

Robust Estimation of the VAR Model

P. Jonáš

Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic.

Abstract. Vector autoregressive model is a very popular tool in multiple time series analysis. Its parameters are usually estimated by the least squares procedure which is very sensitive to the presence of errors in data, e.g. outliers. If outliers were present, the estimation results would become unreliable. Therefore in the presented paper we will propose a new procedure for estimating multivariate regression model. This method is a multivariate generalization of the least weighted squares (LWS) of residuals and we will use it for estimating the coefficients of vector autoregressive model.

Introduction

In many situations one does not observe just a single time series, but several series, possibly interacting with each other. The aim of mutliple time series analysis is then statistically describe the data (including the relationship among variables), suggest a model best fitting our data and estimate a future development (forecasting). For these multiple time series the vector autoregressive model became very popular.

The year 1980, when Christopher Sims in his article Sims, C. A. [1980] advocated vector autoregressive (VAR) model as an alternative to simultaneous equation model, used to be refered to as a milestone in their development. Because VAR models are linear models it is relatively easy to deal with them in the theoretical and practical way. The ease of computation and quality forecasts were the main reasons of their wide usage. Nowadays is VAR described in standard textbooks on time series (e.g. Lütkepohl [2005] and Hamilton, J. D. [1994]) and econometrics (Greene, W. H. [2002]).

Let $\{Y_t | t \in \{-p+1, \ldots, T\}\}$ be a $K-$dimensional stationary time series. The object of interest is the *vector autoregressive model of order p* (VAR(p)), given by

$$Y_t = \nu + B_1 Y_{t-1} + B_2 Y_{t-2} + \ldots + B_p Y_{t-p} + U_t, \tag{1}$$

where $B_1, \ldots, B_p (B_p \neq 0)$ are $(K \times K)$ coefficient matrices, ν is $(K \times 1)$ vector of intercept term and the $K-$dimensional error terms $\{U_t, t \in \mathrm{Z}\}$ are supposed to be independently and identically distributed with a density of the form

$$f_{U_t}(x) = \frac{g(x^{'} \Sigma^{-1} x)}{(\det(\Sigma))^{1/2}}, \tag{2}$$

where Σ and g are a positive definite matrix called the *scatter matrix* and a positive function. When the second moments of U_t exist, Σ will be proportional to the covariance matrix of the error term. Throughout this paper $A^{'}$ will stand for the transpose of a matrix A.

This model is usually estimated by LS procedure (c.f. Lütkepohl [2005]) which is extremely sensitive to outliers, therefore it is important to investigate robust multivariate regression methods which could be used for estimating the VAR model.

Suppose we have observations of time series Y_t for $t = -p+1, \ldots, T$. When we use following notation: for $t = 1, \ldots, T$ denote $X_t = (1, Y_{t-1}^{'}, \ldots, Y_{t-p}^{'})^{'}, B = (\nu, B_1, \ldots, B_p)^{'}$, we can rewrite the equation (1) as the multivariate regression model

$$Y_t = B^{'} X_t + U_t. \tag{3}$$

Using matrix notation $X = (X_1, \ldots, X_T)'$ and $Y = (Y_1, \ldots, Y_T)'$

$$Y = XB + U. \tag{4}$$

The LS estimator of B is given by a well known formula

$$\hat{B}_{LS} = (X'X)^{-1}X'Y, \tag{5}$$

and the scatter matrix Σ is estimated by

$$\hat{\Sigma}_{LS} = \frac{1}{T-K}(Y - X\hat{B}_{LS})'(Y - X\hat{B}_{LS}). \tag{6}$$

The LS estimates of B and Σ could be very negatively influenced by outliers and then we can get confusing estimation results and wrong forecasts. Outliers in multiple time series can be of different kinds. The most well known ones being innovational outliers and additive outliers. In the context of VAR model an observation Y_t is said to be an *innovational outlier* if the error term U_t in (1) is contamined. Due to the dynamic structure, innovational outlier will have an effect on the next observations as well. On the other hand we call an observation Y_t the *additive outlier* if only its value has been affected by contamination. Althought additive outlier has an isolated effect on the time series it can still cause a serious errors in our parameter estimates.

One way, commonly used in practice, how to handle with outliers in multiple time series analysis is to first apply a robust technique on univariate components of the series, then identify and remove outliers and model the multiple time series as outlier free using LS. This method has some difficulties. First, the contamination in one series may be caused by the contamination in other components. Secondly, due to the masking effect, multivariate outlier cannot always be detected by looking at the univariate components separately. Therefore it is better to use some multivariate technique. In the following parts of this paper we will present some multivariate robust regression methods which could be used for this purpose.

Robust multivariate regression

To robustify the least squares regression estimator in multivariate case, two methods has been presented in Rousseeuw and Leroy [1987]. The Minimum Volume Ellipsoid (MVE) and the Minimum Covariance Determinant (MCD). The first method (MVE) is only $\sqrt[3]{n}$ consistent as n goes to infinity, which is extremely slow. The second one (MCD) converges with the usual rate $\sqrt{n}$ so we will deal only with MCD. Suppose we have n observations $x_i \in \mathrm{R}^p,\ y_i \in \mathrm{R}^q$ and denote the entire sample $Z_n = \{(x_i, y_i); i = 1, \ldots, n\}$. MCD estimator is the LS estimator computed from the h points for which the determinant of the residual covariance matrix is minimal. For this method also exists the fast and effective algorithm called *C-step* introduced in Rousseeuw and Van Driessen [1998]. Agulló et al. [2008] presented equivalent characterizations of MCD and they shown its connection to the Least Trimmed Squares (LTS) estimator. Let $\mathrm{H} = \{H \subset \{1, \ldots, n\} | \#H = h\}\}$ be the collection of all subsets of size h.

Definition 1 With the notation above *Multivariate Least Trimmed Squares (MLTS) estimator* is defined as

$$\hat{B}_{MLTS}(Z_n) = \hat{B}_{LS}(\hat{H}) \quad \text{where} \quad \hat{H} \in \underset{H \in \mathrm{H}}{\arg\min} \det \hat{\Sigma}_{LS}(H), \tag{7}$$

where $\hat{B}_{LS}(H)$ denotes the LS estimate based on H etc. As the corresponding estimate of scatter matrix of the error term we take

$$\hat{\Sigma}_{MLTS}(Z_n) = c_\alpha \hat{\Sigma}_{LS}(\hat{H}), \tag{8}$$

where c_α is a consistency factor. For details (see Croux and Haesbroeck [1999]). It has been shown in Agulló et al. [2008], that the MLTS estimator (and also MCD) can be equivallently defined as

$$\hat{B}_{MLTS} = \underset{B,\Sigma;|\Sigma|=1}{\arg\min} \sum_{s=1}^{h} d_{(s)}^2(B,\Sigma), \tag{9}$$

where $d_{(1)}(B,\Sigma) \leq d_{(2)}(B,\Sigma) \leq \ldots \leq d_{(n)}(B,\Sigma)$ is the ordered sequence of the residual Mahalanobis distances and $|\Sigma|$ denotes the determinant of the matrix Σ. Throughout this paper $|A|$ will stand for the determinant of a matrix A. The Mahalanobis distance is defined as

$$d_t(B,\Sigma) = ((Y_t - B^{'}X_t)^{'}\Sigma^{-1}(Y_t - B^{'}X_t))^{1/2}. \tag{10}$$

To generalize the LTS in the case of classical linear regression model Víšek [2000] proposed Least Weighted Squares (LWS)

$$\hat{\beta}_{LWS}^{n,w} := \underset{\beta \in \mathrm{R}}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta), \tag{11}$$

where $w : [0,1] \to [0,1]$ is given nonincreasing *weight function.* In the multivariate regression context we can generalize MLTS in the following way.

Definition 2 With the notation above *Multivariate Least Weighted Squares (MLWS) estimator* is defined as

$$\hat{B}_{MLWS}^{n,w} = \underset{B,\Sigma;|\Sigma|=1}{\arg\min} \sum_{s=1}^{n} w\left(\frac{s-1}{n}\right) d_{(s)}^2(B,\Sigma). \tag{12}$$

But there are some open questions about MLWS estimate. Under what conditions is this estimate consistent ($\sqrt{n}$ consistent)? What is its asymptotic distribution (variance)? How can we express this estimate as statistical functional? What is its influence function? And probably the most important question: What is the optimal choice of weights? For the LWS estimate defined in (11), this questions were answered in Mašíček [2004]. In the case of MLWS it will be the subject of future research.

Robust estimation of VAR model

As we have shown in equations (3) and (4) the VAR model can be rewritten as multivariate regression model and we can use estimation techniques described in the previous section. In this section we will show some results obtained from simulation study performed by Croux and Joossens [2008]. Our basic idea was to compare our results obtained from MLWS with their results, but unfortunately, there is not any fast and effective algorithm which could be used for computing MLWS. We can use iterative algorithm described in Mašíček [2004], but the computation takes too much time, so it is impossible to get comparable results in real time.

Since the efficiency of MLTS is rather low, many authors suggest to use a reweighted version to improve the performance.

Definition 3 The *Reweighted Multivariate Least Trimmed Squares (RMLTS) estimator* is defined as

$$\hat{B}_{RMLTS}(Z_n) = \hat{B}_{LS}(J) \quad \text{and} \quad \hat{\Sigma}_{RMLTS} = c_\alpha \hat{\Sigma}_{LT}(J), \tag{13}$$

where c_α is a consistency factor, $J = \{j \in \{1,\ldots,n\} | d_j^2(\hat{B}_{MLTS}, \hat{\Sigma}_{MLTS}) < q\}\}$ and q is a choosen constant.

There is a simple idea behind RMLTS. Outliers have large residuals with respect to robust MLTS estimate and therefore a large Mahalanobis distances. For details see Croux and Joossens [2008]. In the next two tables (1 and 2) we will show the comparison of LS and RMLTS based on simulations.

	LS		RMLTS	
m	Bias	MSE	Bias	MSE
0	0.00	0.020	0.00	0.022
1	0.08	0.030	0.02	0.023
2	0.14	0.045	0.03	0.024
3	0.18	0.063	0.04	0.026
4	0.22	0.079	0.04	0.027
5	0.25	0.096	0.05	0.029
10	0.38	0.193	0.07	0.039
15	0.51	0.319	0.11	0.057
20	0.64	0.478	0.17	0.080
25	0.76	0.659	0.25	0.104

Figure 1. Simulated Bias and Mean Squared Error for the LS, and the robust RMLTS estimator of a bivariate VAR(2) model, in presence of m additive outliers in a series of length 500. Source: Croux and Joossens [2008], Table 1, page 494

	LS		RMLTS	
m	Bias	MSE	Bias	MSE
0	0.00	0.021	0.00	0.022
1	0.02	0.022	0.00	0.021
2	0.04	0.023	0.01	0.020
3	0.06	0.025	0.01	0.019
4	0.08	0.029	0.01	0.018
5	0.10	0.033	0.01	0.018
10	0.20	0.068	0.01	0.017
15	0.30	0.123	0.01	0.016
20	0.40	0.198	0.01	0.016
25	0.49	0.289	0.01	0.016

Figure 2. Simulated Bias and Mean Squared Error for the LS, and the RMLTS estimator of a bivariate VAR(2) model, in presence of m innovational outliers in a series of length 500. Source: Croux and Joossens [2008], Table 2, page 495

As we can see in this two tables robust methods give better estimates than LS when the amount of outliers increase.

Conclusion

In the presented paper a new method for estimating multivariate regression model was introduced, namely Multivariate Least Weighted Squares (MLWS). It was already shown in the literature (e.g. Jurczyk [2008], Mašíček [2004]), that for the one dimensional case, LWS can give better results than LTS. We hope, that this could be also possible in the multivariate case. The estimation of VAR model is not only application of multivariate regression. We can also use it for example to robustify discriminant analysis, cluster analysis etc. and on many other fields.

Acknowledgments. The presented work was supported by the Grant Agency of Czech Republic under contract 402/09/0557. The author also thanks to his supervisor prof. RNDr. Jan Ámos Víšek, CSc. for the professional help.

References

Agulló, J., Croux, C., and Van Aelst, S., The multivaritate least-trimmed squares estimator, *Journal of Multivariate Analysis*, *99*, 311–338, 2008.

Croux, C. and Haesbroeck, G., Influence function and efficiency of the mcd-scatter estimator, *Journal of Multivariate Analysis*, *71*, 161–190, 1999.

Croux, C. and Joossens, K., Robust estimation of the vector autoregressive model by a least trimmed squares procedure, *COMPSTAT 2008: Proceedings in computational statistics*, pp. 489–501, 2008.

Greene, W. H., *Econometric Analysis*, Prentice Hall, 5 edn., 2002.

Hamilton, J. D., *Time Series Analysis*, Princeton University Press, 1994.

Jurczyk, T., High breakdown point estimation in regression, *WDS'08 Proceedings of Contributed Papers, Part I*, pp. 94–99, 2008.

Lütkepohl, H., *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, 2005.

Mašíček, L., *Diagnostika a senzitivita robustních modelů*, PhD. thesis,(unpublished), 2004.

Rousseeuw, P. J. and Leroy, A. M., *Robust Regression and Outlier Detection*, John Wiley & Sons, 1987.

Rousseeuw, P. J. and Van Driessen, K., A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, *41*, 212–223, 1998.

Sims, C. A., Macroeconomics and reality, *Econometrica*, *48*, 1–48, 1980.

Víšek, J. Á., Regression with high breakdown point, *ROBUST 2000*, pp. 324–356, 2000.