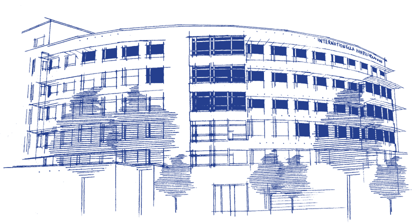
Issues of multicollinearity and conditional heteroscedasticity in time series econometrics

KRISTOFER MÅNSSON



Jönköping International Business School
Jönköping University

JIBS Dissertation Series No. 075 • 2012

KRISTOFER MÅNSSON **Issues of multicollinearity and conditional heteroscedasticity in time series econometrics** JIBS Dissertation Series No. 075 JIBS

Issues of multicollinearity and conditional heteroscedasticy in time series econometrics

KRISTOFER MÅNSSON

This doctoral thesis consists of four chapters all related to the field of time series econometrics. The main contribution is firstly the development of robust methods when testing for Granger causality in the presence of generalized autoregressive conditional heteroscedasticity (GARCH) and causality-in-variance (i.e. spillover) effects. The second contribution is the development of different shrinkage estimators for count data models which may be used when the explanatory variables are highly inter-correlated.

The first essay investigated the effect of spillover on some tests for causality in a Granger sense. As a remedy to the problem of over-rejection caused by the spillover effects White's heteroscedasticity consistent covariance matrix is proposed. In the second essay the effect of GARCH errors on the statistical tests for Granger causality is investigated. Here some wavelet denoising methods are proposed and by means of Monte Carlo simulations it is shown that the size properties of the tests based on wavelet filtered data is better than the ones based on raw data.

In the third and fourth essays ridge regression estimators for the Poisson and negative binomial (NB) regression models are investigated respectively. Then finally in the fifth essay a Liu type of estimator is proposed for the NB regression model. By using Monte Carlo simulations it is shown that the estimated MSE is lower for the ridge and Liu type of estimators than maximum likelihood (ML).

JÖNKÖPING INTERNATIONAL
BUSINESS SCHOOL
JÖNKÖPING UNIVERSITY

ISSN 1403-0470 ISBN 978-91-86345-27-3

Issues of multicollinearity and conditional heteroscedasticy in time series econometrics

KRISTOFER MÅNSSON

JÖNKÖPING INTERNATIONAL
BUSINESS SCHOOL
JÖNKÖPING UNIVERSITY

Jönköping International Business School
P.O. Box 1026
SE-551 11 Jönköping
Tel.: +46 36 10 10 00
E-mail: info@jibs.hj.se
www.jibs.se

*Issues of multicollinearity and conditional heteroscedasticy in time series econometrics*JIBS Dissertation Series No. 075

ISSN 1403-0470
ISBN 978-91-86345-27-3

Printed by ARK Tryckaren AB, 2012

Acknowledgments

To write a dissertation is not possible without the help and support of a great number of people. For this thesis my first gratitude must go to my head supervisor Professor Ghazi Shukur for his guidance and support for this dissertation work over the years. I cannot imagine a better supervisor. He has shown a great interest and constantly been reading drafts and provided me with suggestions on how to make improvements. Besides supervising my research he has also encouraged me to go to various conferences and encouraged me to participate in different research projects with researchers from many different universities. I will always be in debt to you, for the encouragement and for your willingness to share your research skills over these years. I would also like to thank my deputy supervisors Associate Professor Thomas Holgersson and Pär Sjölander for their support and guidance over the years. I am also grateful to Associate professor Abdullah Almasri and some anonymous referees for helpful comments and suggestions on earlier versions of the manuscript.

During the years at the Department of Economic, Finance and Statistics I have had the fortune of working with many stimulating people which has made the writing of the dissertation an enjoyable experience. I would first like to express my gratitude to Professor Åke E. Andersson for the valuable comments and advices during the internal seminars at the department. Then I would like to thank Professor Börje Johansson for his encouragement and valuable comments over the years. Then I would also like to express my gratitude to everyone that I have had the chance to co-author papers with. These are (in alphabetic order): Professor Martin Andersson, Professor Scott Hacker, Professor Thomas Holgersson, Professor Börje Johansson, Associate Professor Golam Kibria, Associate Professor Ghadban Khalaf, Hyunjoo Kim, Associate Professor Panagiotis Mantalos and Pär Sjölander. I would also like to thank the other phd candidates at Department of Economics and Statistics. A special thanks goes to Peter Karlsson (now a phd), Rashid Mansoor and Zangin Zeebari for the help and assistance during the work with the thesis and courses.

Finally I would like to express my gratitude to my friends and my family. I always feel their support. In particular, the constant opportunity to come home to Malmö and do things completely different from research implies that I can forget the research a while and come back to work fresh minded.

The work on this thesis has been supported by Sparbankernas Forskningsstiftelse, for which I am most grateful.

Jönköping February 2012,

Kristofer Månsson

Abstract

This doctoral thesis consists of four chapters all related to the field of time series econometrics. The main contribution is firstly the development of robust methods when testing for Granger causality in the presence of generalized autoregressive conditional heteroscedasticity (GARCH) and causality-in-variance (i.e. spillover) effects. The second contribution is the development of different shrinkage estimators for count data models which may be used when the explanatory variables are highly inter-correlated.

The first essay investigated the effect of spillover on some tests for causality in a Granger sense. As a remedy to the problem of over-rejection caused by the spillover effects White's heteroscedasticity consistent covariance matrix is proposed. In the second essay the effect of GARCH errors on the statistical tests for Granger causality is investigated. Here some wavelet denoising methods are proposed and by means of Monte Carlo simulations it is shown that the size properties of the tests based on wavelet filtered data is better than the ones based on raw data.

In the third and fourth essays ridge regression estimators for the Poisson and negative binomial (NB) regression models are investigated respectively. Then finally in the fifth essay a Liu type of estimator is proposed for the NB regression model. By using Monte Carlo simulations it is shown that the estimated MSE is lower for the ridge and Liu type of estimators than maximum likelihood (ML).

Table of Content

1 Introduction

In time series econometrics the error term and/or the observations are almost always assumed to be identically and independently distributed (iid). However, in the real-life context the iid assumption is often too strong. For financial data it is nowadays well-known that the error term often follows an autoregressive conditional heteroscedasticity (ARCH) process introduced by Engle (1982) or a generalized ARCH (GARCH) process proposed by Bollerslev (1986). Further developments were then made by Granger (1986) where the author showed that causality-in-variance (i.e. spillover) often exists. When the error term is not iid then the commonly applied statistical tests often over-rejects the true null-hypothesis. This has been shown when testing for unit roots by for example Sjölander (2008) and when testing for causality in a Granger (1969) sense (i.e. Granger causality) by for example Mantalos et al. (2007). The problem of GARCH and spillover effects will be treated in this thesis and some solutions are proposed which are based on Whites heteroscedasticity consistent covariance matrix (HCCME) introduced by White (1980) and some wavelet denoising methods suggested by Donoho and Johnstone (1994).

Another example where the iid property is broken occurs when analyzing count data that follows a Poisson or negative binomial (NB) distribution. In this situation the observations are not iid and the relationship between the dependent variable and the explanatory variables are non-linear. Then the classical ordinary least square (OLS) method is not proper and the regression model is usually estimated by means of maximum likelihood (ML) by applying the iterative weighted least square (IWLS) algorithm. However, this estimation method is very sensitive to multicollinearity which is defined by Frisch (1934) as the situation where the explanatory variables are highly inter-correlated. This thesis treats this problem and suggests shrinkage estimators which are based on the classical ridge regression estimator proposed for linear regression by Hoerl and Kennard (1970a,b) and the Liu estimator suggested by Liu (1993).

This introduction chapter includes a discussion of the issues mentioned so far and a summary of the papers in the thesis.

2. Testing for Granger causality in the Presence of GARCH Effects

In this thesis some commonly applied statistical tests for causality in a Granger (1969, 1980) sense is investigated in the presence of GARCH and spillover effects. By causality in a Granger sense it is meant that the time series y_2 contains unique information about the development of time series y_1. Thus, the requirement set by Granger (1969, 1980) is that a cause in y_2 creates an effect in y_1 and this cause in y_2 happens before the effect in y_1. The causality test is most commonly performed in a vector autoregressive (VAR) framework by applying OLS. The effect of GARCH errors may be explained by considering the following VAR model:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \beta_{11}^1 & \beta_{12}^1 \\ \beta_{21}^1 & \beta_{22}^1 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \beta_{11}^p & \beta_{12}^p \\ \beta_{21}^p & \beta_{22}^p \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \tag{2.1}$$

where the error component u_{it} equals $h_{it}v_{it}$ for $i=1,2$, v_t is standard normal distributed and

$$\begin{bmatrix} h_{1t}^2 \\ h_{2t}^2 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} h_{1t-1}^2 \\ h_{2t-1}^2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} u_{1t-1}^2 \\ u_{2t-1}^2 \end{bmatrix}. \tag{2.2}$$

In this situation there exist spillover effects if the non-diagonal elements of the parameter matrices are non-zero. According to Granger and Newbold (1986) one can test for Granger causality by estimating each of the single linear equations in the VAR-model using ordinary least squares (OLS) and then evaluate a zero restriction in each of the estimated equations. In matrix notation this may be done by evaluating the following restriction:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{0},$$

where $\boldsymbol{\beta}$ is a $(2p+1)\times 1$ vector containing all of the coefficient parameters of one of the equations in the VAR(p) model, $\mathbf{R}$ is a fixed $p\times(2p+1)$ matrix with each row consisting of zeros except for the ones corresponding to an element in $\boldsymbol{\beta}$ being restricted to zero, and $\mathbf{0}$ is a $p\times 1$ vector. The tests for Granger causality used in this thesis are the following:

$$W = T\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right)'\left[\mathbf{R}\hat{\Sigma}_u\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right), \tag{2.3}$$

$$LR = T\ln\left(\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right)'\left[\mathbf{R}\hat{\Sigma}_u\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right)+1\right), \tag{2.4}$$

$$LM = T\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right)'\left[\mathbf{R}\hat{\Sigma}_r\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right), \tag{2.5}$$

$$F = \frac{\Delta}{q}\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right)'\left[\mathbf{R}\hat{\Sigma}_u\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}}\right), \tag{2.6}$$

where $\hat{\boldsymbol{\Sigma}}_u$ and $\hat{\boldsymbol{\Sigma}}_r$ are the estimated variances-covariance matrices of the unrestricted and restricted model respectively. Under an assumption of the error term being white noise the first three tests are all asymptotically $\chi^2(p)$ distributed and the final test is $F(q,\Delta)$ distributed where $\Delta = T - 2p - 1$. Under the white noise assumption we may estimate the variances-covariance matrices using the following formulas:

$$\hat{\mathbf{\Sigma}}_u = \hat{\sigma}_u^2 (\mathbf{X}'\mathbf{X})^{-1} \text{ and } \hat{\mathbf{\Sigma}}_r = \hat{\sigma}_r^2 (\mathbf{X}'\mathbf{X})^{-1}, \tag{2.7}$$

where $\hat{\sigma}_u^2 = \hat{\mathbf{u}}_u{}'\hat{\mathbf{u}}_u$ and $\hat{\sigma}_r^2 = \hat{\mathbf{u}}_r{}'\hat{\mathbf{u}}_r$ where $\hat{\mathbf{u}}_u$ and $\hat{\mathbf{u}}_r$ are the estimated residual series from the unrestricted and restricted model respectively. As a remedy to the problem of over-rejection of the true null hypothesis one may use the following White's HCCME:

$$T(\mathbf{X'X})^{-1}[\mathbf{X'SX}](\mathbf{X'X})^{-1}, \tag{2.8}$$

where

$$\mathbf{S} = diag\left(\hat{u}_1^2, \ldots, \hat{u}_t^2\right). \tag{2.9}$$

Another potential remedy is to use a wavelet decomposition of the time series. This method attempts to reproduce a Fourier analysis method but with functions (wavelets) that are better suited to capture the local behavior of time series. By using wavelet denoising proposed by Donoho and Johnstone (1994) one can divide a time series into signal and noise according to the following formula:

$$y_t = D_t + \varepsilon_t, \tag{2.10}$$

where D_t is the signal and ε_t is the noise assumed to be $N(0,\sigma^2)$. Based on wavelet methods one may then extract the signal from the noise by firstly conduct a wavelet decomposition of the time series using either discrete wavelet transform (DWT) or maximal overlap discrete wavelet transform (MODWT). The difference between these two methods is that DWT uses averages and differences over contiguous pairs of observations and is therefore limited to sample sizes that are elements of the didactic series (T = 2^j for some integer *j*). In contrary MODWT uses moving averages and differences and can thus handle any sample size. Therefore, in this thesis MODWT is chosen instead of DWT. Hence, the wavelet decomposition is made using weighted moving averages where the moving averages are denoted scaling coefficients and the weighted differences wavelet coefficients. Then the wavelet coefficients are denoised using the soft, mid and hard denoising methods where some of the coefficients are shrunk towards zero while some are kept. Finally a denoised time series is produced based on the scaling coefficients and the denoised wavelet coefficents. For this method the distribution of the tests are not assumed to be $\chi^2(p)$ distributed or $F(q,\Delta)$ distributed. Instead new critical values are generated using Monte Carlo simulations.

3. Estimating Count Data Models in the Presence of Multicollinearity

Count data models is commonly applied to model a dependent variable that is either Poisson or NB distributed. This type of models are usually estimated by ML using the following IWLS algorithm:

$$\hat{\boldsymbol{\beta}}_{\mathbf{ML}} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{W}}\hat{\mathbf{z}}, \tag{3.1}$$

where $\hat{\mathbf{W}}$ is equal to $diag\left[\hat{\mu}_i\right]$ and $diag\left[\frac{\hat{\mu}_i}{1+\hat{\alpha}\hat{\mu}_i}\right]$ for the Poisson and NB regression model respectively. Furthermore $\hat{\mu}_i = \exp\left(\mathbf{x_i}\hat{\boldsymbol{\beta}}\right)$ is an estimation of the conditional average of observation i based on the estimated parameters in the previous iteration while $\hat{\alpha}$ is an estimator of the overdispersion parameter. Finally $\hat{z}_i$ which is the ith row of $\hat{\mathbf{z}}$ equals $\log\left(\hat{\mu}_i\right)+\frac{y_i-\hat{\mu}_i}{\hat{\mu}_i}$ for both the Poisson and NB distributions. The ML estimator is asymptotically normally distributed and the MSE equals:

$$E\left(L_{ML}^2\right) = E\left(\hat{\boldsymbol{\beta}}_{\mathbf{ML}}-\boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}}_{\mathbf{ML}}-\boldsymbol{\beta}\right) = tr\left[\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\right] = \sum_{j=1}^{J}\frac{1}{\lambda_j}, \tag{3.2}$$

where λ_j is the jth eigenvalue of the $\mathbf{X'WX}$ matrix. When the explanatory variables are highly correlated the weighted matrix of cross-products, $\mathbf{X'WX}$, is ill-conditioned which leads to instability and high variance of the ML estimator. In that situation, it is very hard to interpret the estimated parameters since the vector of estimated coefficients is on average too long. As a solution to this problem this thesis proposes that one should use a shrinkage estimator instead of ML. We firstly propose the ridge type of estimator:

$$\hat{\boldsymbol{\beta}}_{\mathbf{RR}} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbf{ML}} = \mathbf{Z}\hat{\boldsymbol{\beta}}_{\mathbf{ML}}, \tag{3.3}$$

where k is a ridge parameter that may only take on values between zero and infinity. For the ridge estimator we have that when k is greater than zero then $\left\|\hat{\boldsymbol{\beta}}_{\mathbf{RR}}\right\| < \left\|\hat{\boldsymbol{\beta}}_{\mathbf{ML}}\right\|$. Since $\hat{\boldsymbol{\beta}}_{\mathbf{ML}}$ is, on average, too long in the presence of multicollinearity $\hat{\boldsymbol{\beta}}_{\mathbf{RR}}$ is assumed to perform better than $\hat{\boldsymbol{\beta}}_{\mathbf{ML}}$. Then a Liu type of estimator is considered:

$$\hat{\boldsymbol{\beta}}_{\mathbf{d}} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + \mathbf{I}\right)^{-1} \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + d\mathbf{I}\right)\hat{\boldsymbol{\beta}}_{\mathbf{ML}}, \tag{3.4}$$

where d is a shrinkage parameter that may only take on values between zero and one. For this estimator we have that when d is less than one $\left\|\hat{\boldsymbol{\beta}}_{\mathbf{d}}\right\| < \left\|\hat{\boldsymbol{\beta}}_{\mathbf{ML}}\right\|$. Hence, this estimator is also assumed to perform better than ML in the presence of multicollinearity.

4. Summary of the Papers

Article 1: Granger Causality Test in the Presence of Spillover Effects

This paper investigates by means of Monte Carlo simulations the effect of spillover on the reliability of eight different tests for Granger causality. As a remedy to the problem caused by spillover effects we propose that one could use Whites HCCME. The result from the simulation study show that the over-rejection of the true null-hypothesis in the presence of spillover effects decreases for all tests when Whites HCCME is used. Furthermore we may see that the best option, where the estimated size is closest to the nominal size, is to apply the LM test with Whites HCCME. Hence, this is the test we may recommend to practitioners.

Article 2: A Wavelet-Based Approach of testing for Granger Causality in the Presence of GARCH Effects

This paper proposes a new wavelet-based approach of testing for Granger causality. This method uses different type of wavelet denoising methods proposed by Donoho & Johnstone (1994) in order to obtain tests for Granger causality that are robust to GARCH effects. By means of Monte Carlo simulations it is shown that the commonly used causality tests tend to over-reject the true null hypothesis when the error term follows a GARCH process. Furthermore it is shown that the new wavelet-based approach improves the size properties of the Granger causality tests for all of the different situations evaluated. Based on the result from the simulation study we may recommend using the Haar filter with the soft thresholding rule.

Article 3: A Poisson Ridge Regression Estimator

This paper introduces a ridge regression estimator for the Poisson regression model. This new Poisson ridge regression estimator (PRR) is suggested since multicollinearity leads to a high variance and instability of the estimated coefficient vector when the traditional ML method is used. By means of Monte Carlo simulations it is shown that the new PRR estimator outperforms the ML method in the presence of multicollinearity. Based on the result from the simulation study we may also recommend some estimators of the ridge parameter which were first proposed by Muniz and Kibria (2009).

Article 4: On Ridge Estimators for the Negative Binomial Regression Model

In this paper a ridge regression estimator for the NB regression model is suggested. The NB regression model should be used instead of the Poisson regression model in the case of over-dispersion (i.e. the variance of the dependent variable exceeds the mean). To estimate the NB regression model one still has to use ML for which the estimated parameters have a high variance when the explanatory variables are highly inter-correlated. Using Monte Carlo simulations we show that the ridge regression estimator outperforms ML. Based on the result from the simulation study we may also recommend some estimators of the ridge parameter which were first proposed by Muniz and Kibria (2009).

Article 5: Developing a Liu Estimator for the Negative Binomial Regression Model: Method and Application

This paper introduces a Liu (1993) type of shrinkage estimator for the NB regression model. This shrinkage estimator is easier than the ridge regression estimator since the estimated parameters are a linear function of the shrinkage parameter. This paper also suggests some methods of estimating the shrinkage parameter and by using Monte Carlo simulations it is shown that the new Liu type of estimator outperforms the ML method in the presence of multicollinearity.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics,* 31: 307-327.

Bollerslev, T., Engle, R. F., Wooldridge, J. M. (1988). A Capital Asset Pricing Model with Time-varying Covariances. *The Journal of Political Economy,* 96: 116-131.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika,* 81: 425-455.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica,* 50: 987-1008.

Frisch, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems*. Publication 5 (Oslo: University Institute of Economics, 1934).

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric models and Cross-Spectral Methods. *Econometrica,* 37: 24-36.

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control,* 2: 329-352.

Granger, C. W. J. & Newbold, P. (1986). *Forecasting Economic Time Series,* Academic Press, New York.

Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: biased estimation for non-orthogonal Problems. *Technometrics,* 12: 55-67.

Liu, K (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods*, 22: 393-402.

Mantalos, P., Shukur, G. and Sjölander, P. (2007). An Examination of the Robustness of the Vector Autoregressive Granger-Causality Test in the Presence of GARCH and Shifts. *International Review of Business Research Papers,* 3: *280-296.*

Muniz, G., and Kibria, B. M. G. (2009). On some ridge regression estimators: An Empirical Comparisons. *Communications in Statistics-Simulation and Computation,* 38: 621- 630.

Sjölander, P. A new test for simultaneous estimation of unit roots and GARCH risk in the presence of stationary conditional heteroscedasticity disturbances. *Applied Financial Economics,* 18: 527-558.

White, H. (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica,* 48: 817-838.