Bayes multivariate signification tests and Granger causality

Ciuiu, Daniel

Technical University of Civil Engineering, Bucharest, Institute for Economic Forecasting

25 September 2011

Bayes Multi-Variate Signification Tests and Granger Causality

Daniel Ciuiu *,**

Abstract

The Granger causality test is reduced, after co-integration, to the test of the fact that some coefficients of linear regressions are equal to zero or not.

In this paper we will build multi-variate Bayes tests for the signification of the parameters of linear regression provided by the above Granger causality, instead of using the classical F statistics. We will consider the cases of known variance, respectively unknown variance.

Because we replace in practice the Student tests by the Z tests if the involved number of degrees of freedom is at least 30, we can replace in our paper the case of unknown variance with that of known variance, if the above number of degrees of freedom is at least 30.
Keywords: Bayes multi-variate test, Granger causality
JEL Classification: .

Aknowledgement: This paper is suported by the Sectorial Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number SOP HRD/89/1.5/S/62988.

1 Introduction

Consider X a random variable whose distribution depends on the parameter $\theta \in \Theta$. If X is discrete, we denote by $p\left(x|\,\theta_0\right) = P\left(X = x|\,\theta = \theta_0\right)$, and, if X is continuous, by $f\left(x|\,\theta\right)$ the pdf. of X conditioned on the fact that $\theta = \theta_0$.

In the Bayesian inference the parameter θ is a random variable with the pdf $g\left(\theta\right)$, called prior pdf. of θ. To determinate the posterior pdf. of θ we use the formula of Bayes, and we obtain (Preda, 1992; Liu, 1996; Lo and Cabrera, 1987)

*Technical University of Civil Engineering, Bucharest. Bd. Lacul Tei, no. 124, Bucharest, Romania.
**Romanian Institute for Economic Forecasting, Casa Academiei, Calea 13 Septembrie no. 13 , Bucharest, Romania.
e-mail: dciuiu@yahoo.com

$$g(\theta_0|x) = \frac{p(x|\theta_0) \cdot g(\theta_0)}{\int\limits_{\Theta} p(x|\theta) \cdot g(\theta) d\theta} \tag{1}$$

in the discrete case, respectively

$$g(\theta_0|x) = \frac{f(x|\theta_0) \cdot g(\theta_0)}{\int\limits_{\Theta} f(x|\theta) \cdot g(\theta) d\theta} \tag{1'}$$

in the continuous case.

Definition 1 (Preda, 1992) *A family of prior distribution $\mathcal{P}$ is called conjugated prior distribution for the pdfs family $\mathcal{F} = \{ f(x|\theta) | \theta \in \Theta\}$ if $(\forall g \in \mathcal{P})$ $(\forall f \in \mathcal{F})$ we have $g(\theta|x) \in \mathcal{P}$.*

A special case of conjugated family of prior distribution is that when $\mathcal{P}$ and $\mathcal{F}$ are identical to the normal family (Preda, 1992): if $X \sim N(\theta, \sigma^2)$ with known σ^2, and $\theta \sim N(\mu, \tau^2)$ then

$$\theta|X \sim N\left(\frac{\tau^2 \cdot X + \sigma^2 \cdot \mu}{\sigma^2 + \tau^2}, \frac{\sigma^2 \cdot \tau^2}{\sigma^2 + \tau^2}\right). \tag{2}$$

For Bayes estimators we have to define first some loss function, like $U(\theta_1, \theta_0) = (\theta_1 - \theta_0)^2$, or $U(\theta_1, \theta_0) = |\theta_1 - \theta_0|$ (Preda, 1992). In these formulae $U(\theta_1, \theta_0)$ is the loss function if we decide that $\theta = \theta_1$ if in fact $\theta = \theta_0$. Other loss function, as weighted quadratic one is used by Ciumara (2005).

Definition 2 (Preda, 1992) *Consider $g(\theta|x)$ the posterior pdf. of the parameter θ and $U(\theta_1, \theta_0)$ a loss function as above. The Bayes risk is $E(U(\theta, \theta_0))$, where θ_0 is the decided (chosen) value of θ, and the pdf. of the random variable θ is $g(\theta|x)$.*

Definition 3 (Preda, 1992) *The Bayes estimator of the parameter θ is the value θ_0 chosen such that the Bayes risk is minimum.*

If $U(\theta_1, \theta_0) = (\theta_1 - \theta_0)^2$ the Bayes estimator is the average, and if $U(\theta_1, \theta_0) = |\theta_1 - \theta_0|$ the Bayes estimator is the median.

In Preda (1992) there is presented a bilateral Bayes test with the first degree error (the treshold) ε. There are considered the conditioned pdf. $f(x|\theta)$, the prior pdf. of θ, $g(\theta)$, and the number $p_0 \in (0, 1)$. Because p_0 is interpreted as the prior probability to have $\theta = \theta_0$, we usualy choose $p_0 = 0.5$ from the maximum entropy principle. The posterior probability to have $\theta = \theta_0$ is

$$P(\theta = \theta_0|X) = \frac{p_0 \cdot f(x|\theta_0)}{p_0 \cdot f(x|\theta_0) + (1 - p_0) \int\limits_{\Theta} f(x|\theta) \cdot g(\theta) d\theta}. \tag{3}$$

We accept the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta \neq \theta_0$ with the treshold ε if and only if $P\left(\theta = \theta_0 \middle| X\right) > 1 - \varepsilon$.

The computation of the integral from the denominator is done, and the example of the Bayes bilateral test in the case of the normal distribution with known variance is presented (Preda, 1992).

Consider n points in $\mathbb{R}^{k+1}$, $X^{(1)}, ..., X^{(n)}$, where $X^{(i)} = \left(X_1^{(i)}, X_2^{(i)}, ..., X_k^{(i)}, Y_i\right)$. The regression hyper-plane is

$$H : Y = A_0 + \sum_{i=1}^{k} A_i X_i \text{ such that} \tag{4}$$

$$\sum_{i=1}^{n} u_i^2 \text{ is minimum,} \tag{4'}$$

where the residues u_i have the formula

$$u_i = Y_i - A_0 - \sum_{j=1}^{k} A_j X_j^{(i)}. \tag{4''}$$

For the computation of A_i from (4) we have to solve the system (Saporta, 1990)

$$\sum_{j=0}^{k} \overline{X_i \cdot X_j} \cdot A_j = \overline{X_i \cdot Y},\ i = \overline{0, k}, \tag{5}$$

where $\overline{X_0 \cdot X_i} = \overline{X_i}$ and $\overline{X_0^2} = 1$.

The polynomial model is in fact the multilinear model (4) with the explanatory variables $X_1 = X$, $X_2 = X^2$ and so on (Ciuiu, 2010).

For the obtained estimators of A_i using (5) and for the residues u_i we have the following hypotheses (Jula, 2003; Voineagu et al., 2007):

1. The estimators of A_i are linear.
2. The estimators of u_i have the expectation 0 and the same variance (homoskedasticity).
3. The estimators of u_i are normal.
4. The random variables u_i are independent.

From the above hypotheses and from Gauss—Markov theorem we obtain the following properties (Jula, 2003; Voineagu et al., 2007):

1. The estimators of A_i are consistent.
2. The estimators of A_i are unbiased.

3. The estimators of A_i have the minimum variance.

4. The estimators of A_i have the maximum likelihood.

If we denote by $\widehat{A}$ the vector of the estimated parameters A_i, the variance-covariance matrix of $\widehat{A}$ is (Jula, 2003)

$$Var\left(\widehat{A}\right) = \sigma_e^2 \left(XX'\right)^{-1}. \tag{6}$$

If we denote by d_{ij} the value from the row i and column j in the matrix $\left(XX'\right)^{-1}$, we obtain the estimator of $\sigma^2_{A_i}$

$$S^2_{\widehat{A}_i} = \sigma_u^2 \cdot d_{ii}, \tag{7}$$

where σ_u^2 is the estimated variance of the residues.

For testing the null hypothesis $H_0 : A_i = 0$ against the alternative hypothesis $H_1 : A_i > 0$ we use the Student test (Jula, 2003). We compute

$$T_i = \frac{\widehat{A}_i}{S_{\widehat{A}_i}}, \tag{8}$$

and we accept the null hypothesis (the parameter is not significant) if and only if $T_i < t_{n-k-1;\varepsilon}$, where n is the data size, k is the number of explanatory variables, and $t_{n-k-1;\varepsilon}$ is the centil of the error ε of the Student distribution with $n-k-1$ degrees of freedom.

The Granger causality test is based uppon the following regressions (Jula, 2003)

$$\begin{cases} y_t = \beta_0 + \sum\limits_{k=1}^{N} \beta_k \cdot y_{t-k} + \sum\limits_{l=1}^{N} \alpha_l \cdot x_{t-l} + u_t \\ x_t = \gamma_0 + \sum\limits_{k=1}^{N} \delta_k \cdot y_{t-k} + \sum\limits_{l=1}^{N} \gamma_l \cdot x_{t-l} + v_t \end{cases}, \tag{9}$$

where N is the common number of lags, u_t is the residue of the regression equation of y in term of x, and v_t is the residue of the regression equation of x in term of y.

The test consists in testing the null hypothesis H_0: $\alpha_l = 0$, respectively $\delta_k = 0$ for any $l = \overline{1,N}$ ($k = \overline{1,N}$) against the alternative hypothesis H_1: there exist $l = \overline{1,N}$ (or $k = \overline{1,N}$) such that $\alpha_l \neq 0$ (or $\delta_k \neq 0$), with a given error ε. If in both cases it is accepted H_0 we have no causality between x and y. If at least one α_l is statisticaly significant and all $\delta_k = 0$, then x is a cause for y (y can be explained by x). In the reverse way, if at least one δ_k is statisticaly significant and all $\alpha_l = 0$, then y is a cause for x (x can be explained by y). If the above existence is fulfilled bot for at least one l and at least one k, then the causality is in both senses.

For each of the equations we use the statistics (Jula, 2003)

$$F = \frac{RSS_r - RSS_n}{RSS_n} \cdot \frac{T - 2N - 1}{N}, \tag{10}$$

where T is the number of observation, N is the above number of considered lags, RSS_r is the sum of square of the residues in the case of restricted equation (all the coefficients zero), and RSS_n is the same sum of squares in the case of non-restricted equation. We notice that the above F statistics is well built, taking into account that for the non-restricted equation we have $T-2N-1$ (the number of observation minus the number of parameters), and for the restricted one we have to subtract the number of parameters we enforce to be zero, i.e. N.

2 The Bayes Test for Multi-Variate Normal Distribution

2.1 Case of Known Variance

If we expand the fraction of (3) by $\frac{g(\theta_0)}{\int\limits_{\Theta} f(x|\theta)\cdot g(\theta)d\theta}$ we obtain the posterior probability to have $\theta=\theta_0$ in the case of Bayes bilateral test

$$P\left(\theta=\theta_0\right|X)=\frac{p_0\cdot g\left(\theta_0\right|X)}{p_0\cdot g\left(\theta_0\right|X)+(1-p_0)\cdot g\left(\theta_0\right)}. \tag{11}$$

Therefore for testing if $\theta=\theta_0$ with a given error using Bayes test is enough to know the prior and posterior densities of θ. Therefore it does not matter if we have $\theta\in\mathbb{R}$ or $\theta\in\mathbb{R}^d$ with $d>1$. The case of the known variance is easy generalized, because the marginals of the normal distribution (the distribution of the coefficients that must be zero according to the null hypothesis) are also normal.

Consider then the conditional distribution of the vector $\widehat{\theta}$ of θ (the k involved parameters) being normal $N\left(\theta,n\cdot\Sigma_1\right)$. Consider also the prior distribution of being $k-$variate normal distribution $N\left(\mu,\Sigma_2\right)$.

The posterior distribution of $\theta|\,\overline{X}$ is (Preda, 1992)

$$\theta|\,\overline{X}\sim N\left(\left(\Sigma_1^{-1}+\Sigma_2^{-1}\right)^{-1}\left(\Sigma_1^{-1}\mu+\Sigma_2^{-1}\overline{X}\right),\left(\Sigma_1^{-1}+\Sigma_2^{-1}\right)^{-1}\right). \tag{12}$$

The above formula is if we consider the bilateral case for each component, i.e. the alternative hypothesis is $\theta\neq 0$ (of course, as vector). If we consider for some components unilateral case (left if the estimator is less than zero, and right in the contrary case), the involved multiple integral is computed by the Monte Carlo method (Văduva, 2004): we generate 1000 vectors according to the prior distribution and we compute the average of a function on the generated vectors that takes the value zero if the generated vector is not in the right domain, otherwise the value of the conditional pdf.

In the case of the coefficients of linear regression, we replace Σ_1 with the estimated variance-covariance matrix of the corresponding coefficients.

2.2 Case of unknown variance

In this case the difficulty consists in the unknown variance-covariance matrix Σ_1. Nevertheless, tacking into account (6), the only extra parameter is σ_e^2.

Therefore we consider the same conditional distribution for $\theta|\,\overline{\theta}, \sigma_e^2$ as for previous $\theta|\,\overline{\theta}$ in the case of known variance-covariance matrix, but we replace Σ_1 by $\sigma_e^2 \cdot (XX')^{-1}$, with σ_e^2 unknown. For the corresponding coefficients we consider only these rows/ columns of the above matrix.

Denote by $\sigma^2 = \sigma_e^2$, by f_1 the conditional distribution of $\widehat{\theta}\Big|\,\theta, \sigma^2$, by f_2 the conditional distribution of $\widehat{\sigma^2}\Big|\,\sigma^2$, by g_1 the prior distribution of θ, and by g_2 the prior distribution of σ^2.

We compute for each value of σ^2 the posterior probability as in the case of known variance, and we apply the total probability formula.

The prior distribution g_2 is considered Γ having the expectation equal to the estimated σ^2, and the variance $\frac{1}{n}$.

When we apply the total probability formula we use the Monte Carlo method as follows. We generate first 1000 random variables having the above Gamma distribution (Văduva, 2004), and we compute the average of the posterior probabilities computed for the generated variances.

3 Results

Example 1 *Consider data on GDP per capita in PPS (Purchasing Power Standards) and on energy intensity of the economy (gross inland consumption of energy divided by GDP: kilograms of oil equivalent per 1000 Euro). The data are according to EUROSTAT, and there is considered the case of Romania in the period 1996-2009. We denote GDP by X, and energy intensity of the economy by Y. The data are in appendix A.*

If we consider two lags we obtain the regressions (9)

$$\begin{cases} y_t = 397.86263 + 0.67716 y_{t-1} - 0.01952 y_{t-2} - 9.22994 x_{t-1} + 4.6304 x_{t-2} \\ x_t = 31.57469 - 0.0008 y_{t-1} - 0.02319 y_{t-2} + 0.39113 x_{t-1} + 0.34242 x_{t-2} \end{cases}$$

The prior coefficients we want to test that are zero are those of x in the first equation, respectively y in the second one. The prior expectations are $(-9\ 5)^T$, respectively $(-0.001\ \ -0.025)^T$, and the prior variance-covariance matrices are $\begin{pmatrix} 41 & -40 \\ -40 & 45 \end{pmatrix}$ and $\begin{pmatrix} 0.25 & -0.22 \\ -0.22 & 0.22 \end{pmatrix}$. In fact the values are taken such that they are closed to the estimated values.

The prior values of a are 15000, respectively 30, and the prior values of b are 0.001, respectively 0.05.

The posterior probabilities to have the coefficients equal to zero are 0.00367 in the first case (Y resulting variable), respectively 0.99966 in the first case (X resulting variable).

Therefore the coefficients of GDP per capita are significant for energy intensity of the economy, and it means that the last variable can be explained by GDP, i.e. GDP is Granger cause for energy intensity of the economy.

In the second case the conclusion is that GDP can not be explained by energy intensity of the economy, hence it is not Granger cause for GDP.

4 Conclusions

The case of unknown variance of residues can be reduced, as we have seen in this paper to the case of known variance. We consider also a prior distribution for the variance and we apply the total probability formula.

The idea of using Bayes test for Granger causality is that in fact we must have the same thing for non-causality: some coefficients of derived linear regressions are zero. That's why we have replaced the F test by the Bayes test.

An open problem is to find a possible explanation of sign change for GDP as Granger cause for energy intensity of the economy: the coefficient of x_{t-1} is -9.22994, and the coefficient of x_{t-2} is 4.6304.

References

[1] Ciuiu, D., 2007. "Bayes, Neyman and Neyman—Bayes Inference for Queueing Systems", *Mathematical Modelling in Civil Engineering (Scientific Bulletin of Technical University of Civil Engineering, Bucharest)*, **4**, pp. 46-57.

[2] Ciuiu, D., 2010. "Informational Criteria for the Homoskedasticity of the Errors", *Romanian Journal of Economic Forecasting*, **XIII(2)**, pp. 231-244.

[3] Ciumara, R., 2005. "A Bayesian Approach in the Collective Risk Model Involving Weighted Quadratic Loss Function", *Mathematical Reports*, **Vol. 7(57), No. 1**, pp. 21-37.

[4] Dumitrescu, M. and Panaite, V., 2003. "Complete Neyman—Bayes Estimation Procedures for the Mean of a Normal Distribution", *Analele Universităţii Bucureşti*, **1**, pp. 31-43.

[5] Gelman, A. at al., 2000. *Bayesian Data Analysys*, Chapman & Hall/CRC.

[6] Jula, D., 2003. *Introducere în econometrie*, Ed. Professional Consulting, Bucharest.

[7] Jun S. Liu, 1996. "Nonparametric hierarhical Bayes via sequential imputations", *The Annals of Statistics*, **24(3)**, pp. 911-930.

[8] Lo, A.Y and Cabrera, J., 1987. "Bayes procedures for rotationally symmetric models on the sphere", *The Annals of Statistics*, **15(3)**, pp. 1257-1268.

[9] Preda, V., 1992. *Teoria deciziilor statistice*, Ed. Academiei, Bucharest.

[10] Saporta, G., 1990. *Probabilités, Analyse des Donées et Statistique*, Editions Technip, Paris.

[11] Văduva, I., 2004. *Modele de Simulare*, Bucharest University Printing House.

[12] Voineagu, V. et al., 2006. *Econometric Theory and Practice*, Meteor Press, Bucharest (Romanian).

[13] *Indicators of Sustainable Developement: Guidelines and Methodologies*, United Nations, 2001, www.un.org/esa/sustdev/publications/indisd-mg2001.pdf.

[14] *GDP per capita in Purchasing Power Standards (PPS)*, http://epp/eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsieb010.

A The data for Romania according EUROSTAT, on GDP per capita and energy intensity of the economy

Year	GDP per capita	Energy intensity of the economy
An1996	33	1128.9
An1997	29	1116.17
An1998	27	1037.95
An1999	26	924.41
An2000	26	906.05
An2001	28	869.24
An2002	29	857.74
An2003	31	847.43
An2004	34	766.7
An2005	35	732.99
An2006	38	704.78
An2007	42	659.09
An2008	47	612.76
An2009	46	576.9