Economics Bulletin

Volume 30, Issue 1

The relationship between the F-test and the Schwarz criterion: Implications for Granger-causality tests

Erdal Atukeren
ETH Zurich - KOF Swiss Economic Institute

Abstract

In applied research, the Schwarz Bayesian Information Criterion (BIC) and the F-test might yield different inferences about the causal relationships being investigated. This paper examines the relationship between the BIC and the F-tests in the context of Granger-causality tests. We calculate the F-test significance levels as a function of the model dimensionality and the sample size that would lead to the same conclusion as the BIC. We illustrate that the BIC would reject the null hypothesis of no-causality less often compared to an F-test conducted at five percent significance level for sample sizes above 50 especially when the chosen model dimensionality is small. Putting the philosophical issues aside, we suggest that the decision to choose between the F-test and the BIC should be made in view of the sample size.

I would like to thank an anonymous referee and the Associate Editor (J.P. Conley) for their comments ans suggestions that led to many improvements. The usual disclaimer applies.
Citation: Erdal Atukeren, (2010) "The relationship between the F-test and the Schwarz criterion: Implications for Granger-causality tests", *Economics Bulletin*, Vol. 30 no.1 pp. 494-499.
Submitted: Jan 12 2010. **Published:** February 08, 2010.

1. Introduction

According to Granger (1969), a (weakly) stationary stochastic variable X can be said to cause another (weakly) stationary stochastic variable Y if and only if the information contained in the history of X helps improve the prediction of Y when the prediction model already contains the history of Y and all other relevant information. Clearly, Granger's definition of causality is a pragmatic one – defined in terms of predictability.[1] In the bivariate case, Granger-causality from X to Y (both as defined above) can be operationalised and tested as follows.

$$y_t = \alpha_1 + \sum_{j=1}^{p} \beta_j y_{t-j} + \sum_{j=1}^{q} \gamma_j x_{t-j} + \varepsilon_t \qquad (1)$$

where: α is the constant term; β's and γ's are parameters to be estimated; p and q are lag-lengths; and ε_t is a well-behaved error term. X does not Granger-cause Y if $\gamma_1 = \ldots = \gamma_q = 0$.

In practice, the inference about whether X Granger-causes Y reduces to a model selection problem. This, in turn, is tightly linked to the selection of optimal lag-lengths in autoregressions and transfer functions, i.e., determining the p and q in equation (1). As a consequence, the results from Granger-causality tests are generally sensitive to the specification of the test equation. It should be noted that Granger (1969) employed fixed lag-lengths where p=q, while a flexible lag-lengths version (where p and q are allowed to differ) was developed by Hsiao (1982).

In order to identify what the "best" model is, various model selection criteria have been developed. The "best" model is taken as the "true" or the population model in the classical frequentist approaches based on significance testing. Another widely adopted approach is the use of a statistical model selection criterion. Among many statistical model selection criteria, the information criterion developed by Schwarz (1978) is grounded in Bayesian principles. As such, the Schwarz Bayesian Information Criterion (BIC) attempts to identify *a posteriori* what the "most probable" model is. When Gaussian errors are assumed, the order of the most probable model for a univariate autoregressive process, AR(p), is obtained by minimising $BIC = (RSS_p / T)\, T^{(p+1)/T}$, where T is the sample size, RSS_p is the residual-sum-of-squares when $Y_{t-1},\ldots,Y_{t-p}$ are used as regressors, and p+1 is the number of estimated parameters – including the constant term.[2]

The small- and large-sample properties of the BIC are well-researched in the literature. Mills and Prasad (1992), for instance, conduct extensive Monte-Carlo experiments and conclude that the BIC should probably be the first choice of the applied researchers. Among others, Lütkepohl (1985), Nickelsburg (1985), Yi and Judge (1988), Granger and Jeon (2004), and Raffalovich et al. (2008) also document evidence in favour of the BIC over other model selection criteria.

Testing for Granger-causality by means of the BIC takes the following form. First, the optimal order of p in equation (1) is found by minimising the BIC. This amounts to fitting an AR(p) model for the Y variable, with a calculated BIC_{AR} value. Then, the lags of X are introduced given the best specification for Y, and the order of X with the minimum BIC gives

[1] See Atukeren (2008) for a further discussion of the issues involved in testing for Granger causality. The notion of Granger-causality also received attention in the philosophy of science literature. For instance, James Woodward (2008: 234) states that:

> "Roughly speaking, X Granger-causes Y if X is temporally prior to Y and information about X improves our ability (relative to some baseline) to predict whether Y will occur. Interestingly, Granger-causation turns out to be a different notion of cause (and hence to be associated with a different notion of causal correctness) than the interventionist notion. X can be a Granger-cause of Y even though it is not a cause in the interventionist sense. It is thus a live question whether we should adopt this notion of cause instead of the interventionist notion."

[2] For a Gaussian process, an alternative way of representing optimal model order chosen by the BIC criterion is that it results from minimising $\ln(\sigma^2_p)+p\ln(T)/T$, where σ^2_p is the maximum likelihood error variance for the AR(p) model, ln is the natural logarithm operator, and T is the sample size.

is chosen – resulting in a BIC_{TF} value. This step amounts to fitting a transfer function. Next, the BIC values from the two steps are compared. If $BIC_{TF} < BIC_{AR}$, X is said to Granger-cause Y.

In this paper, we investigate the relationship between the F-test and the BIC in the context of Granger-causality tests. We first calculate the required significance levels for conducting an F-test that would yield equivalent results as the application of the BIC. Then, we examine the behaviour of the required significance levels of the F-test as a function of the sample size and the model order. Conclusions follow.

2. Relationship between the F-test and the BIC in the context of Granger-causality tests

In the context of equation (1), the BIC for the AR(p) model for Y (BIC_{AR}) and the BIC for the transfer function involving the X variable (BIC_{TF}) can be calculated from:

$$BIC_{AR} = (RRSS / T)\, T^{((p+1)/T)} \quad (2)$$

$$BIC_{TF} = (URSS / T)\, T^{((p+q+1)/T)} \quad (3)$$

where RRSS stands for the restricted residual sum of squares (i.e., q = 0), URSS stands for the unrestricted residual sum of squares (q > 0), and T is the sample size. Note that, including the constant term, we have (p+1) and (p+q+1) parameters to estimate in BIC_{AR} and in BIC_{TF}, respectively. In other words, the order q gives the number of restrictions to be tested.

Let us denote (p+1) with m and (p+q+1) with k. Then, in a conventional F-test, the statistical validity of the restrictions would be tested by comparing the computed F-statistic with the table value – given the number of observations, the number of restrictions, and the degrees of freedom.

$$F = \frac{(RRSS - URSS)/(k-m)}{URSS/(T-k)} = \left(\frac{RRSS}{URSS} - 1\right)\left(\frac{T-k}{k-m}\right) \quad (4)$$

As stated above, if $BIC_{TF} < BIC_{AR}$, then "X can be said to Granger-cause Y". If $BIC_{TF} \geq BIC_{AR}$, then "X does not Granger-cause Y". Let us reconsider the borderline between the rejection and the non-rejection of Granger-causality, i.e., when $BIC_{TF} = BIC_{AR}$. Setting equations (2) and (3) equal and after manipulation, we get:

$$(RRSS / URSS) = T^{(k-m)/T} \quad (5)$$

Further substituting (5) into (4), we obtain:

$$F^* = (T^{(k-m)/T} - 1)\,[(T-k) / (k-m)] \quad (6)$$

The F* value is the critical F-value beyond which the BIC rejects non-causality. In other words, $BIC_{TF} < BIC_{AR}$ implies F > F*. As a numerical example, let us assume that Y is modelled as an AR(2) process (p=2) and X appears with one lag (q=1) in the transfer function. Hence, we have k = 4 and m=3. Let us further assume that T = 100.

According to (6), the critical F-value is: $F^*_{1,96} = 4.5243$. This results in a p-value of 0.0360. That is, rejecting the null hypothesis "X does not Granger-cause Y" if $BIC_{TF} < BIC_{AR}$ corresponds to a conventional F-test with a statistical significance level of 3.6 per cent. Note that this is a stronger requirement than the conventional 5 per cent statistical significance yardstick. Hence, a conventional F-test would reject the null hypothesis at the 5 per cent level whereas the BIC would not.

Let us take another example. Let k = 5, m = 3 and T = 100. That is, we again have an AR(2) process for Y, but we now include two lags of X in the transfer function. Then, $F^*_{2,95}$ = 4.5827, with the p-value of only 0.0126, or 1.26 per cent statistical significance level. For the same sample size and the same AR order for Y but with just one more additional lag of the X variable, the rejection of the null hypothesis of non-causality on the basis of the BIC-criterion corresponds to a much lower (stricter) significance level in terms of a conventional F-test.

In general, the corresponding F-test significance levels (p-values) to the BIC thresholds can be calculated for different sample sizes and model dimensions. In Table 1, we tabulate the statistical significance levels of the F-tests that would be required to reject the null hypothesis of no-causality from X to Y in the context of equation (1). Note that the maximum AR order for Y and the maximum lags of the X variable in the transfer function specification are taken as eight, i.e., p^{max}=8 and q^{max}=8, and the considered sample sizes are 25, 50, 75, 100, and 200. In principle, the corresponding significance levels can be calculated for any sample size and the model order using the equation (6) and a p-value calculator.

A number of conclusions can be drawn from the results presented in Table 1. In small samples, the BIC may lead to the conclusion that "X Granger-causes Y" more often than an F-test conducted at the five per cent significance level. This is in line with the BIC philosophy. It is harder to detect the most probable model in small samples. In large samples, however, the BIC will tend to reject causality from X to Y more often than the F-test with a conventional uniform significance level (e.g., five per cent).

However, one also has to consider whether it is the F-test that rejects the null hypothesis of no-causality too often or whether it is the BIC that is too conservative in making Granger-causal inferences. For instance, the BIC is known to select more parsimonious models than chosen by other criteria. In our context, the question becomes whether selecting a small model dimension changes the qualitative conclusions reached by the BIC compared to the F-test. In this respect, Table 1 illustrates that the BIC does (not) lead to "X Granger-causes Y" conclusions more frequently than the F-test conducted at five per cent significance level for the case of p=1 and q=1 for sample sizes below (above) 50. Hence, putting the philosophical issues aside, the decision to choose between the F-test and the BIC should be made in view of the sample size.

3. Discussion

This study investigates the implications of using the F-test and the Schwarz Bayesian Information Criterion in testing for Granger non-causality. We demonstrate that the application of a uniform statistical significance level that does not vary with the sample size might lead to different causal inferences than those obtained by the BIC. This is often the case in applied research: conflicting results about Granger-causality from X to Y can arise if the BIC is used as the decision criterion rather than the F-test. This paper calculates the threshold values of the F-statistic and their associated significance levels, for a given sample size and model dimension, that would be required to reach the same conclusion as the BIC.

Overall, we show that for small sample sizes and large model dimensions, the BIC might conclude in favour of Granger-causality from X to Y more frequently than the F-test conducted at five per cent level. In samples with more than 50 observations, however, the application of the F-test at the five per cent significance level for Granger-causal inference would lead to the less frequent rejections of the null hypothesis of no causality compared to the BIC. In this sense, the BIC behaves more conservatively compared to an F-test conducted at the five per cent significance level.

Of course, the choice between the F-test and the BIC is a philosophical issue. However, our results illustrate the connections between the two approaches and provide the

threshold significance levels (in the frequentist sense) that correspond to the application of a Bayesian model selection criterion in view of the sample size and model dimensionality.

References

Atukeren, E. (2008) "Christmas cards, Easter bunnies, and Granger-causality" *Quality & Quantity* **42**(6), 835-844.

Granger, C.W.J. (1969) "Investigating causal relationships by econometric models and cross-spectral methods" *Econometrica* **36**, 424-438.

Granger, C.W.J and Y. Jeon (2004) "Forecasting performance of information criteria with many macro weries" *Journal of Applied Statistics* **31**(10), 1227-1240.

Hsiao, C. (1982) "Autoregressive modelling and causal ordering of economic variables" *Journal of Economic Dynamics and Control* **4**, 243-259.

Lütkepohl H. (1985) "Comparing of criteria for estimating the order of a vector autoregressive process" *Journal of Time Series Analysis* **6**(1), 35-52.

Mills, J.A. and K. Prasad (1992) "A comparison of model selection criteria" *Econometric Reviews* **11**, 201–33.

Nickelsburg, G. (1985) "Small-sample properties of dimensionality statistics for fitting VAR models to aggregate economic data: a Monte-Carlo study" *Journal of Econometrics* **28**, 183–92.

Poskitt, D.S. and A.R. Tremayne (1987) "Determining a portfolio of linear time series models" *Biometrika* **74**, 125–37.

Raffalovich, L.E., G.D. Deane, D. Armstrong, and H-S. Tsao (2008) "Model selection procedures in social research: Monte-Carlo simulation results" *Journal of Applied Statistics* **35**(10), 1093-1114.

Schwarz, G. (1978) "Estimating the dimension of a model" *Annals of Statistics* **6**, 461-464.

Yi, G. and G. Judge (1988) "Statistical model selection criteria" *Economics Letters* **28**, 47–51.

Woodward, J. (2008) "Invariance, Modularity, and All That: Cartwright on Causation", in *Nancy Cartwright's Philosophy of Science*, L. Bovens, C. Hoefer, and S. Hartmann (Eds.) Routledge Studies in the Philosophy of Science. Routledge. UK.

Table 1. Statistical Significance Level Equivalency of the F-test to the BIC

TF-Order (↓)	AR order (→)							
T = 25	1	2	3	4	5	6	7	8
1	0.096066	0.104141	0.112965	0.122620	0.133199	0.144807	0.157566	0.171619
2	0.066947	0.076146	0.086610	0.098511	0.112047	0.127444	0.144956	0.164875
3	0.047945	0.057000	0.067706	0.080348	0.095254	0.112797	0.133405	0.157560
4	0.036185	0.044859	0.055508	0.068542	0.084445	0.103776	0.127177	0.155374
5	0.028849	0.037224	0.047882	0.061382	0.078393	0.099705	0.126227	0.158981
6	0.024290	0.032563	0.043457	0.057710	0.076220	0.100060	0.130469	0.168833
7	0.021587	0.030012	0.041471	0.056923	0.077553	0.104789	0.140280	0.185834
8	0.020237	0.029122	0.041578	0.058845	0.082479	0.114352	0.156612	0.211530
T = 50	AR order (→)							
TF-Order (↓)	1	2	3	4	5	6	7	8
1	0.056453	0.059173	0.062033	0.065041	0.068205	0.071534	0.075038	0.078726
2	0.027349	0.029575	0.031982	0.034585	0.037399	0.040443	0.043734	0.047294
3	0.013610	0.015162	0.016886	0.018804	0.020935	0.023302	0.025932	0.028853
4	0.007070	0.008107	0.009292	0.010646	0.012192	0.013957	0.015970	0.018264
5	0.003830	0.004518	0.005327	0.006276	0.007389	0.008694	0.010221	0.012007
6	0.002160	0.002620	0.003176	0.003846	0.004653	0.005624	0.006791	0.008190
7	0.001266	0.001579	0.001967	0.002448	0.003042	0.003775	0.004678	0.005790
8	0.000771	0.000988	0.001265	0.001616	0.002062	0.002626	0.003340	0.004239
T = 75	AR order (→)							
TF-Order (↓)	1	2	3	4	5	6	7	8
1	0.042473	0.043945	0.045470	0.047051	0.048691	0.050390	0.052153	0.053981
2	0.016786	0.017780	0.018834	0.019950	0.021132	0.022384	0.023711	0.025116
3	0.006812	0.007380	0.007996	0.008662	0.009382	0.010162	0.011006	0.011919
4	0.002872	0.003181	0.003524	0.003902	0.004320	0.004783	0.005293	0.005858
5	0.001255	0.001421	0.001609	0.001820	0.002060	0.002330	0.002634	0.002978
6	0.000567	0.000656	0.000759	0.000877	0.001014	0.001172	0.001353	0.001562
7	0.000264	0.000313	0.000369	0.000436	0.000515	0.000608	0.000717	0.000845
8	0.000127	0.000153	0.000185	0.000223	0.000269	0.000325	0.000391	0.000470
T = 100	AR order (→)							
TF-Order (↓)	1	2	3	4	5	6	7	8
1	0.035021	0.035979	0.036965	0.037979	0.039021	0.040094	0.041199	0.042335
2	0.012023	0.012589	0.013183	0.013804	0.014454	0.015136	0.015849	0.016596
3	0.004238	0.004520	0.004821	0.005142	0.005483	0.005847	0.006235	0.006649
4	0.001548	0.001681	0.001826	0.001983	0.002153	0.002337	0.002537	0.002754
5	0.000584	0.000646	0.000714	0.000789	0.000872	0.000964	0.001065	0.001177
6	0.000227	0.000255	0.000287	0.000323	0.000364	0.000409	0.000460	0.000517
7	0.000091	0.000104	0.000119	0.000136	0.000156	0.000178	0.000204	0.000234
8	0.000037	0.000043	0.000050	0.000059	0.000068	0.000080	0.000093	0.000108
T = 200	AR order (→)							
TF-Order (↓)	1	2	3	4	5	6	7	8
1	0.022514	0.022859	0.023211	0.023568	0.023930	0.024299	0.024673	0.025053
2	0.005559	0.005708	0.005861	0.006019	0.006180	0.006346	0.006517	0.006692
3	0.001411	0.001465	0.001521	0.001579	0.001639	0.001701	0.001766	0.001833
4	0.000369	0.000388	0.000407	0.000427	0.000448	0.000470	0.000493	0.000518
5	0.000099	0.000105	0.000112	0.000119	0.000126	0.000133	0.000141	0.000150
6	0.000027	0.000029	0.000031	0.000034	0.000036	0.000039	0.000041	0.000044
7	0.000008	0.000008	0.000009	0.000010	0.000011	0.000011	0.000012	0.000013
8	0.000002	0.000002	0.000003	0.000003	0.000003	0.000003	0.000004	0.000004