Data sharpening in practice: an application to nonlinear Granger causality

Wolski, M.; Diks, C.

Publication date
2013

Document Version
Final published version

Published in
Aenorm

Citation for published version (APA):
Wolski, M., & Diks, C. (2013). Data sharpening in practice: an application to nonlinear Granger causality. *Aenorm*, *21*(79), 19-21. http://www.aenorm.nl/files/nlaenorm2012/file/article_pdfs/sne8c_Marcin%20online.pdf

References

Bai, J., and S. Ng (2008): "Large Dimensional Factor Analysis," *Foundations and Trends in Econometrics*, 3(2).

Beck, G., K. Hubrich, and M. Marcellino (2011): "On the importance of sectoral and regional shocks for price-setting," *CEPR Discussion Papers* 8357.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120(1), 387-422.

Geweke, J. (1977): "The Dynamic Factor Analysis of Economic Time Series Models," in *Latent Variable in Socioeconomic Models*, ed. by D. J. Aigner, and A. S. Goldberger, pp. 365-387. North-Holland, Amsterdam.

Knight, K., and W. Fu (2000): "Asymptotics for lasso-type estimators," *Annals of Statistics*, 28(5).

Marcellino, M., J. H. Stock, and M. W. Watson (2006): "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, 135(1-2), 499-526.

Stock, J. H., and M. W. Watson (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20(2).

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*. Series B (Methodological), 58(1).

Watson, M. W., and J. H. Stock (2004): "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, 23(6), 405-430.

Data Sharpening in Practice: An Application to Nonlinear Granger Causality

by: Marcin Wolski and Cees Diks

Introduction

Since the introduction of Granger causality over four decades ago (Granger, 1969), the body of literature on (mainly linear) Granger causality has grown substantially, becoming a standard methodology not only among economists and econometricians, but also finding followers in physics or even biology.

To put it formally, imagine a strictly stationary bivariate process $\{X_t, Y_t\}$, $t \in Z$. We say that $\{X_t\}$ is a Granger cause of $\{Y_t\}$ if past and current values of X contain additional information on future values of Y that is not contained in past and current Y-values alone. If we denote the information contained in past observations $\{X_s\}$ and $\{Y_s\}$, $s \le t$, by $F_{X,t}$ and $F_{Y,t}$, respectively, and let '$\sim$' denote equivalence in distribution, the property might be stated in the following definition.

Definition 1. For a strictly stationary bivariate time series process $\{X_t, Y_t\}$, $t \in N$, X_t is a Granger cause of Y_t if, for some $k \ge 1$

$$(Y_{t+1}, \ldots, Y_{t+k}) \,|\, (F_{X,t}, F_{Y,t}) \nsim (Y_{t+1}, \ldots, Y_{t+k}) \,|\, F_{Y,t}.$$

Here it is worth pointing out the difference between the bivariate and multivariate setting. Throughout this paper, we will refer to a multivariate setting by a situation where vectors $\{X_t\}$ and $\{Y_t\}$ are allowed to be multidimensional, i.e. $\{X_t\} = \{X_{1,t}, X_{2,t}, \ldots, X_{m,t}\}$ and $\{Y_t\} = \{Y_{1,t}, Y_{2,t}, \ldots, Y_{m,t}\}$, $m > 1$. Similarly, by a bivariate setting we will refer to a situation where both vectors are univariate ($m = 1$).

The most commonly used nonparametric test for the above hypothesis testing (Def.1) is that proposed by Hiemstra and Jones (1994). Its main advantage lies in a very clear and intuitive reasoning together with a strong asymptotic theory, derived even for a multivariate setting. Nevertheless, the test can severely over-reject if the null is satisfied (Diks and Panchenko, 2005). Therefore, Diks and Panchenko (2006) (hereafter DP) proposed a new test statistic, which corrects for this shortcoming but, as it turns out, because of the relatively large kernel estimator bias compared to the variance the DP test can no longer be shown to be asymptotically standard normally distributed in the multivariate setting.

To illustrate this phenomenon, let us first consider the standard, i.e. bivariate, the DP test in a case when $k = 1$. Let us further denote $\{X_t\} = X, \{Y_t\} = Y$ and $\{Y_{t+1}\} = Z$. The null hypothesis might be then simplified to

$$E\left[f_{X,Y,Z}(X,Y,Z) f_Y(Y) - f_{X,Y}(X,Y) f_{Y,Z}(Y,Z)\right] = 0,$$

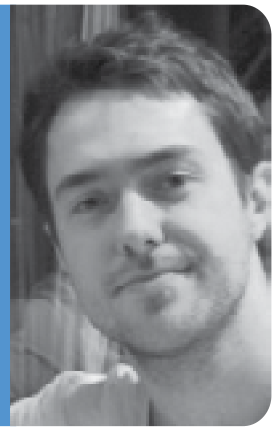
and the test statistic becomes

$$T_n(\varepsilon_n) = \frac{n-1}{n(n-2)} \sum_{i=1}^{n} E\left[\hat{f}_{X,Y,Z}(X_i,Y_i,Z_i)\hat{f}_Y(Y_i) - \hat{f}_{X,Y}(X_i,Y_i)\hat{f}_{Y,Z}(Y_i,Z_i)\right],$$

with $\hat{f}(.)$ being the ordinary kernel density estimator (for an example see the next section) with bandwidth being given by $\varepsilon_n \sim n^{-\beta}$. For clarity, let $W = (X,Y,Z)$. Following the methodology developed in DP, the test statistics has a corresponding U-statistics representation, with a mean squared error being dependent on the true variance ($Var[r_0(W)]$), the bandwidth, kernel estimator bias (here α) and the dimensionality of the multivariate

Marcin Wolski

Marcin is a second year PhD student, pursuing the European Doctorate in Economics - Erasmus Mundus, at the University of Amsterdam and Bielefeld University. His core topics comprise nonlinear dynamics and econometrics with applications to the macro models. Currently, he explores his interests in collaboration with the International Monetary Fund.



Cees Diks

Cees Diks is professor of Data-analysis and Economic Statistics at the University of Amsterdam. He is an expert on non-linear time series analysis, a new field at the border between non-linear dynamical systems and statistical data-analysis. Within his research in the Center for Nonlinear Dynamics in Economics and Finance (CeNDEF) he focuses on detecting and modelling nonlinearities, which often arise in systems with boundedly rational economic agents.

vector at hand ($\gamma = d_X + d_Y + d_Z$ and $\delta = d_X + 2d_Y + d_Z$), i.e.

$$MSE[T_n] = q_1\varepsilon_n^{2\alpha} + \frac{9}{n}C_0\varepsilon_n^{\alpha} + \frac{9}{n}Var[r_0(W_i)] + \frac{18}{n^2}q_2\varepsilon_n^{-\gamma} + \frac{5}{n^3}q_3\varepsilon_n^{-\delta} + o\left(\frac{1}{n^2}\right).$$

Assuming ordinary kernel density estimator (with the bias of order 2), it is clear that for the simplest bivariate case, there is a range for parameter, i.e. $\beta \in (1/4, 1/3)$, where the mean squared error is dominated by the true variance in the limit. This eventually guarantees that the test statistic is asymptotically normally distributed, i.e.

$$\sqrt{n}\left((T_n(\varepsilon_n) - \tau)/\left(3\sqrt{Var(r_0(W_i))}\right)\right) \xrightarrow{d} N(0,1),$$

and the properties of the test might be derived analytically. However, if we increase the system dimensionality by 1, without changing the kernel specification, in order to guarantee asymptotic normality β parameter would have to be between $1/(2\alpha)$ and $1/(d_x + d_y + d_z + 1)$, which clearly gives an empty set for the possible β-values.

The solution to the above mentioned problem is the bias reduction in the kernel density estimator. Given that it is lower (so that α-values are larger) there would be enough space for β to endow the test statistics with asymptotic normality.

In this study we apply the Data Sharpening methodology, as it seems to bring extraordinary advantages over the standard bias reduction methods. Firstly, it is simple and straightforward to implement. Secondly, it does not affect the derivation of the test properties, i.e. it reduces the bias without having any other impact on the test characteristics, which is clearly very practical in such a complex environment.

Data Sharpening

Data Sharpening (DS), as formally introduced by Choi and Hall in 1999, comprises a class of methods, which by a slight perturbation of the original dataset, improves the performance of relatively conventional statistical procedures without any severe consequences. The applications of DS include, inter alia, estimating a regression function subject to qualitative constraints, increasing the robustness of statistical procedures or reducing the variability of the bootstrap methods. However, the most widely cited practical adaptation of the DS method is bias reduction in ordinary kernel density estimators. To put it formally, let f denote the unknown distribution of some random variable x. Given a sample of observations $(X_1, X_2, \dots, X_n)$, we may estimate the density by the ordinary kernel estimator

$$\hat{f}(x) = (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where $K(.)$ is a smooth, bounded and symmetric probability density function and h is a bandwidth. Assuming the standard Gaussian kernel, the bias of the estimator is of order h^2 with variance of order $(nh)^{-1}$. As argued by Choi and Hall (1999), by applying a perturbation (or sharpening) function, Ψ, to the original dataset we may reduce the bias of the estimator, without affecting its variance. Imagine a sharpened form of the kernel density estimator

$$\hat{f}_s(x) = (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{x - \Phi(X_i)}{h}\right),$$

with the following sharpening function

$$\Phi(x) = x + \frac{\Sigma_i(X_i - x)K\left(\frac{x - X_i}{h_s}\right)}{\Sigma_i\left(\frac{x - X_i}{h_s}\right)},$$

where $h_s = 1/\sqrt{(2h)}$. Choi and Hall (1999) prove that the bias of the sharpened estimator is of order h^4 (hence much smaller than the original order h^2 with variance's order unchanged.

Although the range and the scope of DS methods seem to be unlimited, in practice the number of applications remains surprisingly low. For instance, the Web of Science returns only 39 papers which link directly to the original paper of Choi and Hall (1999). One of the reasons for that is fierce competition from the other bias reduction methods, including, for example, higher order kernel estimators, or relatively larger computing power required for more sophisticated problems. Nevertheless, due to its simplicity and straightforward application, we find DS a powerful tool which turned out to solve the majority of difficulties we were facing with the nonparametric Granger causality test in a multivariate setting.

Data Sharpening in Practice

Let us turn back to the original problem of the bias reduction in the nonlinear Granger causality, described in the introduction. Imagine now that instead of using the ordinary kernel estimator, we replace it by its sharpened form. The expected value of the estimator becomes

$$\begin{aligned}
E[\hat{f}_s(W_i)] &= \varepsilon_n^{-1}\int K\left(\frac{W_i - \Phi(W_i)}{\varepsilon_n}\right)dF(W) \\
&= \varepsilon_n^{-1}\int K\left(\frac{W_i - V}{\varepsilon_n}\right)\left(f(V) - \frac{1}{2}\kappa_2\varepsilon_n^2 f''(V) + o(\varepsilon_n^2)\right)dV \\
&= \int K(u)\left(f(W_i - \varepsilon_n u) - \frac{1}{2}\kappa_2\varepsilon_n^2 f''(W_i - \varepsilon_n u)\right)du + o(\varepsilon_n^2) = \\
&= \int K(u)f(W_i)du - \varepsilon_n\int uK(u)f'(W_i - \varepsilon_n u)du + o(\varepsilon_n^2) = f(W_i) + o(\varepsilon_n^2)
\end{aligned}$$

This guarantees that the bias α is of order 4 and, using the same reasoning as before, we may find that for $\beta \in (1/8, 1/4)$ the test statistics is asymptotically normally distributed. In principal, it is possible to reduce the kernel estimator bias to arbitrary low levels by applying more sophisticated sharpening functions; see Choi and Hall (1999).

Numerical results

In order to evaluate our methodology, we assess the test performance in a situation when the null is satisfied and when it is violated. We simulate 1000 multivariate time series where Granger causality is present (the null is violated) and where it is not. The results are presented in Figures 1 and 2.

MSc Econometrics recommended for readers of MSc-level

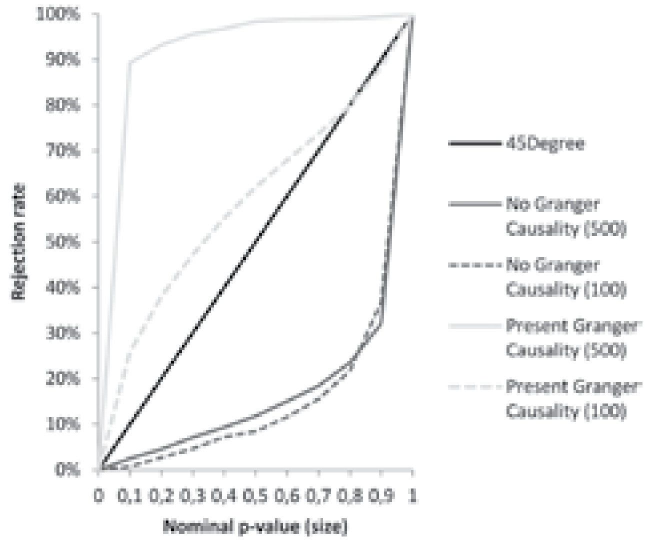


Figure 1. Size-size and power-size plot of the test performance. The length of time series is given in brackets. Source: own calculations.

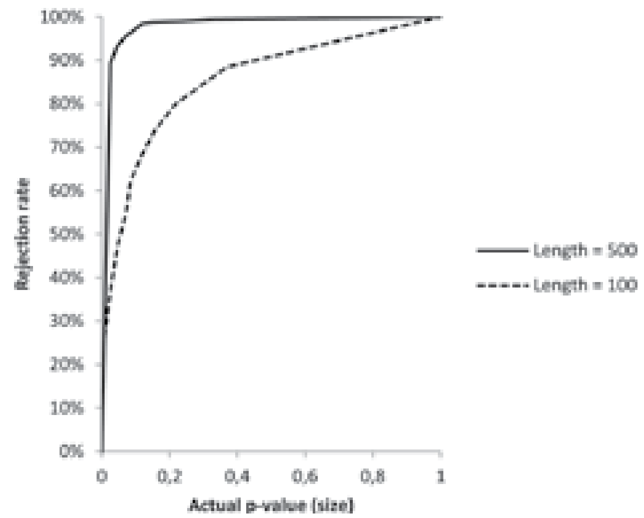


Figure 2. Size-corrected power plot of the test performance. Source: own calculations.

We observe that the test performs well in case of the presence of Granger causality. For a time series of length 500 we observe a very high rejection rate, as expected. When the time series are relatively short (100 observations), the test performs worse. If there is no Granger causality between the time series, the test under-rejects. However, the longer the time series the closer we are to the 45 degree line, which is the expected shape when the null is satisfied. In fact, the test might be viewed as conservative, in a sense that under-rejection under the null is better than over-rejection.

In practice, Granger causality is assessed on the time series of length much larger than 1000, which increases the power of the test. Nevertheless, it seems that the number of observations plays an important role when DS bias reduction is applied. The straightforward explanation for this phenomenon lies in the functional of the sharpening. In general it concentrates the points around high-density regions and makes it more dispersed where their density is low. In small samples it is difficult to determine the 'true' high and low density regions which biases the procedure. However, the more observations, the more likely it is that DS works well, which might be clearly observed in Figure 2 – correcting for the nominal size of the test, the power of the test is higher for longer time series.

Conclusions

In this study we show a practical application of the DS methodology in the nonparametric Granger causality testing. The problems with the asymptotic behavior of the test statistics, which arise as a consequence of the kernel estimator bias in a multivariate setting, might be effectively narrowed bringing back the desired properties.

The numerical results suggest that DS performs rather well in the large samples. Smaller samples might be suffering from the under-rejection bias when the null is satisfied. Nevertheless, practical applications of the Granger causality testing suggest that the size of the samples is much higher than 1000, narrowing the influence of the above-mentioned bias.

Because of the increasing popularity of the nonlinear methods, the DS might quickly find a broader range of applications as it seems that it serves as an ideal solution to possible issues with the nonparametric models.

References

Choi, E., and P. Hall. Data Sharpening as a Prelude to Density Estimation. *Biometrika* 86.4 (1999): 941-47.

Diks, Cees, and Valentyn Panchenko. A New Statistical and Practical Guidelines for Nonparametric Granger Causality Testing. *Journal of Economic Dynamics and Control* 30 (2006): 1647-669.

Diks, Cees, and Valentyn Panchenko. A Note on the Hiemstra-Jones Test for Granger Non-causality. *Studies in Nonlinear Dynamics & Econometrics* 9.4 (2005).

Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37.3 (1969): 424-38.