International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Comparing the effectiveness of traditional vs. mechanized identification methods in post-sample forecasting for a macroeconomic Granger causality analysis

Haichun Ye [a,*,1], Richard Ashley [b], John Guerard [c]

[a] School of Economics, Shanghai University of Finance and Economics, Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, 111 Wuchuan Road, Shanghai 200434, China

[b] Department of Economics (0316), Virginia Tech, Blacksburg, VA 24061, United States

[c] McKinley Capital Management, LLC, 3301 C Street, Suite 500, Anchorage, AK 99503, United States

ARTICLE INFO

Keywords:
Post-sample forecasting
Post-sample Granger causality
Identification methods

ABSTRACT

We identify forecasting models using both a traditional, partially judgmental method and the mechanized Autometrics method. We then compare the effectiveness of these two different identification methods for post-sample forecasting, in the context of a relatively large-scale exemplar of macroeconomic post-sample Granger causality testing. This example examines the Granger causal relationships among four macroeconomically important endogenous variables – monthly measures of aggregate income, consumption, consumer prices, and the unemployment rate – embedded in a six-dimensional information set which also includes two interest rates, both of which are taken to be weakly exogenous in this context. We find that models indentified by the traditional method tend to have better post-sample forecasting abilities than analogous models identified using the mechanized method, and that the analysis done using the traditional identification method generates stronger evidence for post-sample Granger causality among the four endogenous variables.

1. Introduction

In-sample Granger causality analysis is typically based on an F-test of the null hypothesis that the coefficients on the putatively-causing variates in a particular VAR model equation are all zero. It has long been known that such tests are so routinely misleading as to be of doubtful usefulness. As was discussed by Racine and Parmeter (2013, Section 1) and Efron (1982, Chapter 7), this is an inevitable consequence of the fact that these in-sample F tests are inherently based on model fitting errors. These fitting errors – the magnitudes of which are, by definition, being minimized by the estimation process itself – correspond to what Efron calls 'apparent' rather than 'true' errors. Consequently, a comparison of the post-sample forecasting effectiveness over varying information sets has long been the methodology of choice in this area, albeit implemented in a variety of ways: see Ashley (2003), Ashley, Granger, and Schmalensee (1980), Guerard (1985), and Thomakos and Guerard (2004). The reader is referred to Ashley and Tsang (2014) and Ashley and Ye (2012) for a review of this literature.[2]

* Corresponding author.
E-mail address: haichunye@gmail.com (H. Ye).

1 Haichun Ye appreciates the financial support of the Shanghai Pujiang Program.

2 Notably, these papers discuss recent criticisms of the post-sample forecasting testing framework, including the developing realization that

http://dx.doi.org/10.1016/j.ijforecast.2014.08.004
0169-2070/

While it is well-known that a key step in post-sample forecasting is to identify relevant time series models over both the full and restricted information sets, very little is known about the effectiveness of different model identification methods in post-sample forecasting. In this study, we address this issue by identifying models in two interestingly distinct ways and then comparing the effectiveness of the two model identification approaches. Specifically, as per Ashley and Ye (2012), the models (over both the full and restricted information sets) are first identified in the somewhat *ad hoc* "large-to-small" manner commonly identified with David Hendry: one starts with as complicated a model as the data set will support (i.e., a vector autoregression in each included variable, utilizing all lags out to at least the seasonal lag), then pares down this formulation by eliminating statistically insignificant terms, starting with the largest, least plausible, lags.[3] It is common (and sensible) to use some judgment in this process, so we will identify this below as the "partially judgmental" identification procedure. For example, an isolated statistically significant lag structure term at lag twelve is likely to be worth retaining in a model for monthly data, whereas such a term at a lag of eight or eleven is not.[4] Alternatively, analogous models (over both the full and restricted information sets) are also identified and estimated using the "Autometrics" mechanized model specification procedure introduced by Doornik and Hendry (2007) and currently implemented in the *Oxmetrics* software program. Both of these model identification algorithms – along with their sample fits to the data considered here – are described at greater length in Section 2 below. The relative effectiveness of these two identification algorithms in post-sample forecasting is then examined in Section 3, in the context of a new, relatively large-scale exemplar of Granger causality testing.

Ashley and Ye (2012) test for post-sample Granger causality between the median growth rate in these 31 sub-components of the US Consumer Price Index (i.e., the monthly CPI inflation rate) and the inter-quartile range of these 31 sub-components (i.e., the monthly dispersion in the inflation rates across the 31 categories), but this is only a bivariate analysis. Here we employ six, arguably more broadly interesting, US macroeconomic aggregates:

- Aggregate real income
  This variable is defined as the monthly growth rate of seasonally adjusted real disposable personal income, and is denoted "y_t" below.
- Aggregate real household consumption spending
  This variable is defined as the monthly growth rate of seasonally adjusted real personal consumption expenditures, and is denoted "c_t" below.
- CPI inflation rate
  This variable is defined as the monthly growth rate of the seasonally unadjusted consumer price index (CPI), and is denoted "π_t" below.
- Civilian unemployment rate
  This variable is defined as the monthly change in the seasonally unadjusted civilian unemployment rate, and is denoted "Δun_t" below.

These time series are taken to be endogenous, which is to say, potentially Granger-caused by each other and/or by the final two time series considered; lags in these last two time series are therefore taken to be weakly exogenous:[5]

- Short-term interest rate
  This variable is defined as the monthly change in the seasonally unadjusted 3-month Treasury bill rate, and is denoted "$\Delta tbill_t$" below.
- Long-term interest rate[6]
  This variable is defined as the monthly change in the seasonally unadjusted yield on 10-year Treasury bonds, and is denoted "$\Delta tbond_t$" below.

These data are all used in un-deseasonalized form whereever possible (i.e., for π_t, Δun_t, $\Delta tbill_t$, and $\Delta tbond_t$), as the Bureau of Economic Analysis' de-seasonalization method employs a two-sided filter which distorts causal inferences.

The data sources, summary statistics, time plots, and sample correlograms for these six time series are presented in Tables 1 and 2 and Fig. 1. The changes in Δun_t, $\Delta tbill_t$, and $\Delta tbond_t$ are used instead of their levels because these levels data are so highly persistent that a unit root in the levels time series cannot be rejected credibly on standard tests. The null hypothesis of a unit root is rejected at the 1% level for all six time series (as defined above) using both the *ADF* and *PP* tests; see Table 3.[7]

Consequently, we proceeded on the assumption that all six time series, as formulated above, are $I(0)$.

In this setting, we find that models identified by the "partially judgmental" data procedure tend not to fit the sample data as well, but produce smaller post-sample mean squared forecast errors (MSFE) than those identified by the Autometrics algorithm. The analysis based on the traditional, partially judgmental model specification approach yields stronger evidence for post-sample

particular care must be taken (as is done below) in choosing a statistical test for post-sample forecasting improvements in the context of nested models. Another problem with post-sample testing is the *ad hoc* nature of the data split between a model identification/estimation sub-period and a post-sample model evaluation sub-period. Ashley and Tsang (2014) and Racine and Parmeter (2013) have each developed model validation methods based on cross-validation which surmount this obstacle, for modest sample lengths and large sample lengths, respectively; a follow-on paper to the present work will apply the Racine–Parmeter cross-validation model validation procedure to the (large-sample) data set and models examined here.

3 If reasonably feasible, it is a good idea to exceed the seasonal lag at the outset, as a multiplicatively seasonal model can be expected to yield terms beyond the seasonal lag when one identifies an additive model.

4 See Ashley (2012, Section 14.4) for a discursive example.

5 We are by no means asserting that fluctuations in the other four variables do not Granger- cause fluctuations in these two interest rates, we are simply not testing for these causal links.

6 The yields used here as $tbill_t$ and $tbond_t$ are taken from the St. Louis Federal Reserve website as the secondary market rate for a three-month Treasury bill and the constant maturity rate for a ten-year Treasury bond. Measuring yields on such securities is a non-trivial endeavor, with the realized yields being likely to be slightly superior to those used here.

7 The absence of a strong negative sample autocorrelation at lag one in the correlograms for Δun_t, $\Delta tbill_t$, and $\Delta tbond_t$ confirms that they are not over-differenced. An *ARFIMA* model for the levels variables was not considered, for the reasons given, at length, by Ashley and Patterson (2010).

Please cite this article in press as: Ye, H., et al., Comparing the effectiveness of traditional vs. mechanized identification methods in post-sample forecasting for a macroeconomic Granger causality analysis. International Journal of Forecasting (2014), http://dx.doi.org/10.1016/j.ijforecast.2014.08.004