Journal of Econometrics 187 (2015) 358–375

Contents lists available at ScienceDirect

Journal of Econometrics

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

Instrumental variable and variable addition based inference in predictive regressions[☆]

Jörg Breitung[a], Matei Demetrescu[b,*]

[a] University of Cologne, Institute of Econometrics and Statistics, Albertus-Magnus-Platz D-50923 Köln, Germany

[b] Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany

ARTICLE INFO

Article history:
Received 22 February 2012
Received in revised form
28 October 2013
Accepted 29 October 2013
Available online 14 March 2015

JEL classification:
C12
C32

Keywords:
Causality test
Persistence
Integration
Long memory
IV estimation

ABSTRACT

Valid inference in predictive regressions depends in a crucial manner on the degree of persistence of the predictor variables. The paper studies test procedures that are robust in the sense that their asymptotic null distributions are invariant to the persistence of the predictor, that is, the limiting distribution is the same irrespective of whether the regressors are stationary or (nearly) integrated. Existing procedures are often conservative (e.g. tests based on Bonferroni bounds), are based on highly restrictive assumptions (such as homoskedasticity or assuming an AR(1) process for the regressor) or fail to have power against alternatives in a $1/T$ neighborhood of the null hypothesis. We first propose a refinement of the variable addition method with improved asymptotic power approaching the optimal rate. Second, inference based on instrumental variables may further improve the (local) power of the test and even achieve local power under the optimal $1/T$ rate. We give high-level conditions under which the suggested variable addition and instrumental variable procedures are valid no matter whether the predictor is stationary, near-integrated or integrated, or exhibits time-varying volatility. All test statistics possess a standard limiting distribution. Monte Carlo experiments suggest that tests based on simple combinations of instruments perform most promising relative to existing tests. An application to quarterly US stock returns illustrates the need for robust inference.

1. Introduction

Predictive regressions play an important role in empirical economics. Granger causality implies that a variable does not cause another if the former is not able to predict the latter. Also, in financial economics, it is of interest whether variables like dividend yields or interest spreads contain information about future stock price returns. One important practical problem with performing such predictive regressions is that the regressor is highly persistent in many cases, whereas the dependent variable is close to white noise. For example, stock price returns or exchange rate changes are approximately uncorrelated, whereas predictors like dividend yields or interest rate differentials behave roughly like a random walk. As shown by Elliott and Stock (1994) the t-statistic may suffer from severe size distortions in such cases.

We start within the framework of Elliott and Stock and consider as a baseline model the dynamic system given by the triangular representation

$$y_t = \beta x_{t-1} + u_t \tag{1}$$

$$x_t = \varrho x_{t-1} + v_t, \tag{2}$$

$t = 2, \ldots, T$, with $\Sigma = \mathrm{E}\left(\begin{pmatrix} u_t \\ v_t \end{pmatrix} \begin{pmatrix} u_t & v_t \end{pmatrix}\right) = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$. Note that the regressor x_t is assumed to be weakly exogenous since $\mathrm{E}(x_{t-1}u_t) = 0$ but $\mathrm{E}(x_{t-1}u_{t-1}) \neq 0$ whenever $\sigma_{uv} \neq 0$. If $\sigma_{uv} = 0$, then the regressor is strictly exogenous. We first abstract from any deterministic component such as an intercept or linear trend to focus on the main issues without the extra notational burden. In Section 3.2 we expand our model accordingly and show that deterministic terms can easily be dealt with in the usual manner.

To model persistent regressors, the variable x_t is often assumed to be nearly integrated,

$$\varrho = 1 - \frac{c}{T} \tag{3}$$

☆ The authors would like to thank two anonymous referees, as well as Christoph Hanck, Uwe Hassler, Søren Johansen, Lynda Khalaf, Helmut Lütkepohl, Tassos Magdalinos, Massimiliano Marcellino, Dominik Poß, Paulo M.M. Rodrigues, Michalis Stamatogiannis, Rob Taylor and Martin Wagner for very helpful comments and suggestions.

* Corresponding author.
E-mail address: mdeme@stat-econ.uni-kiel.de (M. Demetrescu).

http://dx.doi.org/10.1016/j.jeconom.2013.10.018
0304-4076/

for $c \geq 0$. We are interested in testing the null hypothesis $\beta = 0$ whatever the value of c may be. Under suitable regularity conditions (e.g. Elliott and Stock, 1994) the ordinary least squares [OLS] t-statistic for the null $\beta = 0$ in (1) is asymptotically distributed as

$$t_{ls} \xrightarrow{d} \omega \frac{\int_0^1 J_c(r)\, \mathrm{d}W_v(r)}{\sqrt{\int_0^1 J_c^2(r)\, \mathrm{d}r}} + \sqrt{1-\omega^2}\mathcal{Z}, \tag{4}$$

where $\omega = \sigma_{uv}/(\sigma_u\sigma_v)$, $J_c(r)$ represents an Ornstein–Uhlenbeck process such that $T^{-1/2}x_{[rT]} \Rightarrow \sigma_v J_c(r)$ with $J_c(r) = W_v(r) - c\int_0^r e^{-c(r-s)} W_v(s)\, \mathrm{d}s$, and $W_v(r)$ a standard Brownian motion obtained as $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \sigma_v W_v(r)$ (with "$\Rightarrow$" denoting weak convergence in a space of càdlàg functions on [0, 1] endowed with a suitable norm). The standard normal variate $\mathcal{Z}$ is independent of $W_v(r)$ (and thus of $J_c(r)$). Hence, the distribution of the ordinary t-statistic is nonstandard and depends on the parameter c if $\sigma_{uv} \neq 0$.

Should the driving process x_t be stationary, i.e. $-1 < \varrho < 1$ fixed, standard asymptotic theory applies. The problem in applied research is that the nature of x_t is typically unknown, and pre-testing to check whether $|\varrho| = 1$ has been shown to induce serious size-distortions when ϱ is close to unity (Elliott and Stock, 1994).

For the baseline model given by (1) and (2) there already exist a number of test procedures that are robust to the value of the autoregressive coefficient ϱ. Elliott and Stock (1994) proposed a Bayesian mixture procedure and Cavanagh et al. (1995) consider various tests based on conservative bounds (as refined by Campbell and Yogo, 2006); the work of Jansson and Moreira (2006) can be casted in a restricted likelihood framework. The asymptotics of these procedures is however confined to the near-integrated case.

Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) proposed testing strategies that allow for valid inference irrespective of the nature of the autoregressive roots of x_t (local to unity or stationary). The idea is to augment the testing equation with additional (redundant) variables such that the coefficients to be tested are attached to stationary variables. Bauer and Maynard (2012) show that variable addition [VA] also works in the context of VAR(∞) processes with unknown persistence. Although such a robust approach is appealing, we argue that such tests may suffer from a dramatic loss of power. Specifically, they only have power in $1/\sqrt{T}$ neighborhoods of the null hypothesis compared with the typical rate of $1/T$ for tests involving nearly integrated regressors.[1] The shortcoming is shared to some extent by the nonparametric approach of Maynard and Shimotsu (2009) with a local power characterized by the rate $1/T^{0.75}$ (see their Lemma 9). Gorodnichenko et al. (2012) propose a quasi-differencing procedure applicable, like the VA method of Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996), in general dynamic models; but, like VA, the procedure only has power in $1/\sqrt{T}$ neighborhoods of the null.[2] Finally, Phillips and Magdalinos (2009) propose an instrumental variable [IV] procedure with local power arbitrarily close to the optimal $1/T$ of the size-distorted OLS estimator.

We therefore study inference in the presence of regressors with unknown persistence such that the power of the resulting tests remains close to that of an optimal test, while the limiting null distributions do not change with the degree of persistence of the regressors. Specifically, we generalize VA and IV procedures and provide classes of tests which exhibit similarities with the IVX approach of Phillips and Magdalinos (2009).

This paper's contributions are as follows. We consider a model with conditional and unconditional heteroskedasticity as well as short-run dynamics of the predictor and show in Section 2 that the original VA test may suffer from severe loss of (asymptotic) power. Although we demonstrate that the power of the VA procedure can be substantially improved by employing certain transformations of the involved variables, some loss of power remains. We then develop alternative test procedures based on instrumental variables that share with the VA tests the invariance to the persistence of the predictor. At the same time, an appropriate choice of instruments yields tests with power against a sequence of alternatives converging to the null hypothesis at the optimal rate. Moreover, the instruments we propose do not require additional data. In Section 3, we study the possibility of improving inference in the IV setup by combining instruments. Our methods can easily be extended to deal with deterministic components and an arbitrary number of regressors. Section 4 compares the proposed methods with existing alternatives by means of Monte Carlo experiments, and Section 5 illustrates the proposed methods with US data.

2. Variable addition and instrument variable tests

We first extend the baseline model to allow for more general data characteristics.

Assumption 1. Let

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} \bar{u}_t \\ \sum_{j\geq 0} b_j \bar{v}_{t-j} \end{pmatrix}$$

where $\sum_{j\geq 0} j\,|b_j| < \infty$ and $\sum_{j\geq 0} b_j \neq 0$, the innovations $\bar{u}_t$ and $\bar{v}_t$ are a bivariate white noise sequence with a component structure,

$$\begin{pmatrix} \bar{u}_t \\ \bar{v}_t \end{pmatrix} = H\left(\frac{t}{T}\right)\begin{pmatrix} \widetilde{u}_t \\ \widetilde{v}_t \end{pmatrix}$$

where $H(r)$ is a matrix of piecewise Lipschitz functions, invertible for all $r \in [0, 1]$, and $(\widetilde{u}_t, \widetilde{v}_t)'$ is a martingale difference sequence with identity covariance matrix satisfying $\sup_t \left|\frac{1}{T}\sum_{j\geq 1}\sum_{k\geq 1} \mathrm{E}\left(\widetilde{v}_{t-j}\widetilde{v}_{t-k}\widetilde{v}_t^2\right)\right| < \infty$ and $\sup_t \mathrm{E}\left(\left\|(\widetilde{u}_t, \widetilde{v}_t)'\right\|^{4+\epsilon}\right) < \infty$ for some $\epsilon > 0$.

The assumption allows the increments of the predictor process x_t to be serially correlated. The so-called 1-summability condition for the moving average coefficients is standard in the literature on integrated series. The disturbances u_t are uncorrelated with the increments of x_t at all lags (i.e. x_t is weakly exogenous with respect to u_t). The martingale difference assumption for the innovations is natural for the empirical applications we have in mind and allows for conditional heteroskedasticity. The summability condition on the cross-product moments $\mathrm{E}\left(v_{t-j}v_{t-k}v_t^2\right)$ slightly restricts the serial dependence in the conditional variances and is fulfilled by independent sequences, for instance.

Unconditional time heteroskedasticity is captured by the matrix $H(r)$ since

$$\mathrm{E}\left(\begin{pmatrix} \bar{u}_t \\ \bar{v}_t \end{pmatrix}\begin{pmatrix} \bar{v}_t & \bar{v}_t \end{pmatrix}\right) = H\left(\frac{t}{T}\right) H\left(\frac{t}{T}\right)'.$$

Should $H(r)$ be a diagonal matrix for all $r \in (0, 1]$, the innovations $\bar{u}_t$ and $\bar{v}_t$ may have time-varying variance but are uncorrelated. In

[1] The VA approach of Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) is more general and may perform more favorable in other applications such as testing for causality in cointegrated systems.

[2] It should be noted, however, that these methods are designed to work against a wider range of alternatives than we consider in (1). As argued by Lettau and Van Nieuwerburgh (2008) and Maynard and Shimotsu (2009), it is the stationary component of the predictor that matters for forecasting series like stock price returns. Accordingly, the rate of the sequence of local alternatives may be a misleading guide for assessing the power against economically relevant alternatives.