CESIS Electronic Working Paper Series

Paper No. 213

The Effectiveness of Information Criteria in Determining Unit Root and Trend Stratus

R. Scott Hacker

Jönköping International Business School

February 2010

The Royal Institute of Technology
Centre of Excellence for Science and Innovation Studies (CESIS)
http://www.cesis.se

The Effectiveness of Information Criteria in Determining Unit Root and Trend Status

R. Scott Hacker

Jönköping International Business School
Jönköping University

P.O. Box 1026, SE-551 11, Jönköping, Sweden
Scott.Hacker@ihh.hj.se

Telephone: +46 36 10 17 50, Fax + 46 36 12 18 32

Abstract

This paper compares the performance of using an information criterion, such as the Akaike information criterion or the Schwarz (Bayesian) information criterion, rather than hypothesis testing in consideration of the presence of a unit root for a variable and, if unknown, the presence of a trend in that variable. The investigation is performed through Monte Carlo simulations. Properties considered are frequency of choosing the unit root status correctly, predictive performance, and frequency of choosing an appropriate subsequent action on the examined variable (first differencing, detrending, or doing nothing). Relative performance is considered in a minimax regret framework. The results indicate that use of an information criterion for determining unit root status and (if necessary) trend status of a variable can be competitive to alternative hypothesis testing strategies.

Key words: Unit Root, Stationarity, Model Selection, Minimax regret, Information Criteria

JEL classification: C22

Introduction[1]

One of the central issues in applied research using time series data is to determine whether the underlying variables are stationary or contain unit roots. Applied researchers in economics are sometimes interested in the presence or not of the unit root itself due to its economic implications, for example in consideration of purchasing power parity or the efficient markets hypothesis. Unit root testing is also often a first step in time series analysis of economic variables. If the variables in a regression contain unit roots then there is risk for the regression to result in spurious regression relations as was pointed out by Granger and Newbold (1974) through Monte Carlo simulations and later proved analytically by Phillips (1986).

The earliest and one of the simplest unit root tests used is the Dickey and Fuller (1979) test. The general equation for Dickey-Fuller unit-root testing for a variable y is

$$\Delta y_t = a + b y_{t-1} + ct + \varepsilon_t \tag{1}$$

where t is the time-period variable, Δ is the first difference operator, ε_t is white noise, and a, b, and c are constant coefficients.[2] Stationary processes are those for which $-2 < b < 0$. For a nonoscillatory y, a unit root is present if $b = 0$, so testing under such circumstances is focused on whether or not that hypothesis can be rejected, with $b < 0$ being the alternative hypothesis. Such testing, however, is very sensitive to whether nonzero values exist for a and c. There are in fact three different Dickey-Fuller tests for unit roots one can do with estimates of equation (1) based on three possible prior restriction situations:

i) no prior restrictions on a or c,

[1] I thank Professor Abdulnasser Hatemi-J for his commentary on the structure and content of this paper, and Pär Sjölander for discussions about unit root testing strategies.

[2] Note that if ε_t is autocorrelated then the model should be augmented by lags of Δy according to Said and Dickey (1984).

ii) $c = 0$,

iii) $a = 0$ and $c = 0$.

The distribution of the unit-root test statistic (the conventionally computed t-statistic associated with b) does not follow any standard distribution when $b = 0$, and it is affected by which of the above situations match the true model. Fortunately critical values for this test statistic have been provided by Dickey and Fuller (1979), and MacKinnon (1991) has produced a formula and tables to calculate more comprehensive critical values, covering more sample sizes. As applied researchers are often not sure about which of the three scenarios is appropriate for testing, various authors such as Enders (2004), Dolado, Jekinson, and Sosvilla-Rivero (1990), and Ayat and Burridge (2000) have suggested sequential testing strategies for determining which scenario should be used for the unit root test.[3] In such sequential testing strategies, possible repeated testing of the unit root is done based on the parameter restriction situations (i), (ii), (iii), in sequence, continuing as long as the associated parameter restrictions are justified by the data (through other tests) and until the unit root is rejected or further parameter restriction is impossible. Various additional tests for the unit root at different points in this sequence may also be included in such a strategy.[4]

Repeated testing of the unit root can lead to serious mass significance problems, as demonstrated in simulations by Hacker and Hatemi-J (2008) of the Enders (2004) sequential testing strategy. Elder and Kennedy (2001) suggest an alternative approach to this sequential testing strategy in which unrealistic outcomes are removed as possibilities and prior knowledge about the trend status of y is used if available. Their approach has testing for the

[3] Similar suggestions have been made by Perron (1988), Holden and Perman (1994), and Ayat and Burridge (2000), as noted in Elder and Kennedy (2001).
[4] A sequential testing strategy for unit roots can include augmentation of the regression equation (1) with various lags of Δy included as necessary to deal with possible error-term autocorrelation. For simplicity, this paper does not include these augmentations, and does not deal with the autocorrelation issue.

unit root done only once, resulting in the mass significance problem vanishing for unit-root determination.

An interesting aspect of these various strategies is that they are closely associated with model selection. One may consider that there are six models possible in this Dickey-Fuller environment, given that in correspondence with the prior knowledge situations (i), (ii), and (iii) there are three possible equations, each of which having two unit root statuses (there is a unit root or there is not). The sequential testing strategies go down the path of selection among these six models, although incompletely in some circumstances (e.g. if a unit root is rejected when using equation (1) with no parameter restrictions, then no effort may be made to determine whether $a = 0$ or $c = 0$). The Elder and Kennedy (2001) strategy when there is no knowledge about whether y has a trend reduces the choice to be from among four models rather than six on the basis that the excluded two models are unlikely or unrealistic in most applications.

This paper investigates the relative performance of using formal model selection techniques, in the form of information criterion minimization, rather than hypothesis testing for choosing among the four models of Elder and Kennedy (2001). Such techniques provide some optimization of the trade-off between bias and inefficiency in considering inclusion of explanatory variables. Using a formal model-selection technique rather than hypothesis testing has a weakness of losing control over the probability of making the error of rejecting a model with parameter restrictions when that model is true (type I error in hypothesis testing). In unit root testing this may be a relevant issue if the rejection of one model has stronger consequences than non-rejection of it. If the researcher considers erroneous rejection of no unit root just as bad as erroneous rejection of the unit root, then a formal model selection technique may very well be more desirable to use.

Performance in this paper is based on the following criteria: (1) correct choice of unit root status, (2) predictive capability of chosen model, and (3) the induced action based on model choice (taking a first difference, detrend, or do nothing with respect to the variable examined). Simulations are performed to consider which procedure minimizes the maximum regret over various possible true parameter permutations, with regret based upon how badly a strategy performs according to one of the above-mention criteria in comparison to another strategy.

The rest of this paper is organised in the following way. The next section describes the six models in the Dickey-Fuller environment. It also explains the Elder and Kennedy strategy for choosing from among four of the six models when no prior knowledge about the trend status of y is available, or from two of the six models when such prior knowledge does exist. That section furthermore describes the information criteria considered in this paper. Section 3 presents some examples of performance of the various methods as the parameter b varies. Section 4 examines the minimax regret properties of using information criteria or hypothesis testing in this environment. Section 5 provides the conclusion.

I. Models, the Elder and Kennedy strategy, and information criteria

Table 1 presents the six different models associated with all of the permutations of the models implied by the prior-restriction situations (i), (ii), and (iii) mentioned in the introduction and two possibilities for b ($b = 0$ and $b < 0$). Due to model (5)'s explosiveness and model (2)'s unlikeliness with a stationary process around an exactly-zero equilibrium, Elder and Kennedy (2001) suggest these two models be removed as selectable models, so that only models (1), (3), (4), and (6) be allowed as possible choices. There are two possible trend statuses for y: no trend as in models (1) and (4) and a trend as in models (3) and (6). The trend arising in model (6) is referred to as a *deterministic trend*, since it arises from the nonzero c term, whereas the trend arising in model (3) is referred to as a *stochastic trend*, since it arises from the nonzero drift term a.

Table 1. Definitions of models using the general equation $\Delta y_t = a + by_{t-1} + ct + \varepsilon_t$

Model (1)	$a = 0, b = 0, c = 0$	Unit root, no trend
Model (2)	$a= 0, b < 0, c = 0$	Stationary around zero equilibrium
Model (3)	$a\neq 0, b = 0, c = 0$	Unit root with drift
Model (4)	$a\neq 0, b < 0, c = 0$	Stationary around constant nonzero equilibrium
Model (5)	$a \neq 0, b = 0, c \neq 0$	Unit root around a deterministic time trend
Model (6)	$a \neq 0, b < 0, c \neq 0$	Trend stationary

The Elder and Kennedy strategy starts by further limiting the allowable models based on *a priori* knowledge about the trend status of y if possible. If there is known to be no trend in y, then only models (1) and (4) are allowable, and otherwise if there is known to be a trend in y, then only models (3) and (6) are allowable. Under these circumstances of *a priori* knowledge of the trend status for y, the choice between the resulting two models is then determined by a Dickey-fuller unit root test. The use of *a priori* information in this way improves the likelihood of choosing the correct model and, if there is known to be no trend in y, improves the power of the unit root test.[5] If instead no *a piori* information on the trend status of y is available, then the Elder and Kennedy strategy consists of two tests, the first to determine whether are not y has a unit root, and the second to determine whether or not a trend in y is involved. The Elder and Kennedy strategy is displayed in detail in Figure 1.

[5] These claimed improvements are particularly true if the prior knowledge of trend status is based on theoretical grounds, but if instead the prior knowledge of trend is based on what the data or prior similar data looks like, then it is arguable that this prior knowledge has been gained through an informal "eyeball test" that could affect the size of subsequent testing.

How to proceed with hypothesis testing in this environment when the trend status is unknown is a controversial subject. The strength of the Elder and Kennedy (2001) strategy under these circumstances is that it forces the size of the unit root test to be the nominal size when the true model is model (1) or (3), and due to its simplicity and wide dissemination of its tests in econometrics textbooks and software, it is possibly closer to what many practitioners do than other strategies that use tests that are not as widely disseminated in textbooks or software. Some may argue that performing first a trend status test that is robust to the unit root status (as in Vogelsang, 1998, and Bunzel and Vogelsang, 2005), and then doing the unit root test contingent upon the trend status found in the first step is more appropriate. Of course the pretesting for the trend can still affect the size of a subsequent unit root test. Others may argue for a sequential testing strategy for the unit root, as noted in the introduction, to increase the likelihood of choosing stationarity given the unit root test tends to have low power (the probability of not accepting the alternative hypothesis of stationarity when it is true is low). However this strategy also affects the size of the unit root test.[6]

[6] Interested readers may wish to read Hacker and Hatemi-J (2008) for a comparison of performance between the Elder and Kennedy strategy and the sequential testing strategies presented in the Enders (2004) textbook.

$\left\{\begin{array}{c}\text{Is there known to}\\ \text{be no trend in } y?\end{array}\right\} \xrightarrow{\text{Yes}} \left\{\begin{array}{c}\text{Estimate}\\ \Delta y_t = a + by_{t-1} + \varepsilon_t\end{array}\right\}$

$\downarrow$ No

$\downarrow$

$\left\{\begin{array}{c}\text{Can } b = 0 \text{ be rejected}\\ \text{using Dickey - Fuller}\\ \text{critical value?}\end{array}\right\} \begin{array}{l}\to \text{No} \to \{\text{Decide Model 1: unit root, no trend}\}\\ \to \text{Yes} \to \{\text{Decide Model 4: stationary, nonzero equilibrium}\}\end{array}$

$\left\{\begin{array}{c}\text{Is there known}\\ \text{to be a nonzero}\\ \text{trend in } y?\end{array}\right\} \xrightarrow{\text{Yes}} \left\{\begin{array}{c}\text{Estimate}\\ \Delta y_t = a + by_{t-1} + ct + \varepsilon_t\end{array}\right\}$

$\downarrow$ No

$\downarrow$

$\left\{\begin{array}{c}\text{Can } b = 0 \text{ be rejected}\\ \text{using Dickey - Fuller}\\ \text{critical value?}\end{array}\right\} \begin{array}{l}\to \text{No} \to \{\text{Decide Model 3: unit root with drift}\}\\ \to \text{Yes} \to \{\text{Decide Model 6: trend stationary}\}\end{array}$

(unknown trend status)

$\downarrow$

$\left\{\begin{array}{c}\text{Estimate}\\ \Delta y_t = a + by_{t-1} + ct + \varepsilon_t\end{array}\right\} \to \left\{\begin{array}{c}\text{Can } b = 0 \text{ be rejected}\\ \text{using Dickey - Fuller}\\ \text{critical value?}\end{array}\right\} \to \text{Yes} \to \left\{\begin{array}{c}\text{Can } c = 0 \text{ be rejected}\\ \text{using } t \text{ - statistic for } c \text{ estimate}\\ \text{and the usual } t \text{ distribution?}\end{array}\right\}$

$\downarrow$No $\qquad \downarrow$No $\qquad \downarrow$Yes

$\left\{\begin{array}{c}\text{Estimate}\\ \Delta y_t = a + \varepsilon_t\end{array}\right\} \qquad \left\{\begin{array}{c}\text{Decide Model 4:}\\ \text{stationary, non -}\\ \text{zero equilibrium}\end{array}\right\} \qquad \left\{\begin{array}{c}\text{Decide Model 6:}\\ \text{trend stationary}\end{array}\right\}$

$\downarrow$

$\left\{\begin{array}{c}\text{Can } a = 0 \text{ in last equation be}\\ \text{rejected using } t \text{ - statistic for } a\\ \text{estimate and the usual } t \text{ distribution?}\end{array}\right\} \begin{array}{l}\to \text{No} \to \{\text{Decide Model 1: unit root, no trend}\}\\ \to \text{Yes} \to \{\text{Decide Model 3: unit root with drift}\}\end{array}$

Figure 1. The Elder and Kennedy Strategy

Despite the difficulties of these alternative hypothesis testing strategies, they may have superior model-selection qualities compared to the strategy suggested by Elder and Kennedy. The simulation results presented later comparing the performance of information criterion strategies to that of the Elder and Kennedy strategy under these circumstances naturally cannot be generalized to how information criterion strategies perform relative to all hypothesis testing strategies. Readers who dislike the Elder and Kennedy strategy when there is no *a priori* knowledge of the trend status may nevertheless be still interested in the simulations later in this paper that assume *a priori* knowledge of trend status.

With the data being used to choose between two models or among four models, and with some of these models being non-nested, there becomes a consideration of whether the use of a formal model selection technique, such as the Akaike information criterion or the Schwarz information criterion, could be useful in deciding which of the models is best supported by the data. Each information criterion has its own goals in optimizing the trade-off between bias (by inappropriately omitting a variable) and inefficiency (by including an irrelevant variable). Considered for analysis in this paper are four information criteria usable with regression equations in which the error terms are assumed to be normally distributed with constant variance. First is the Akaike (1973, 1974) information criterion AIC, defined as

$$\mathrm{AIC} = T\ln\left(\frac{RSS}{T}\right) + 2(k+1)\,, \tag{2}$$

where RSS is the residual sum of squares, T is the number of observations, and k is the number of coefficients to estimate in the regression equation including the intercept term (k +1 is the number of parameters to be estimated including the error's variance). Second is the corrected Akaike information criterion, AICc, defined as

$$\mathrm{AICc} = T\ln\left(\frac{RSS}{T}\right) + \frac{2T(k+1)}{T-k-2}\,. \tag{3}$$

The corrected Akaike information criterion was introduced by Sugiura (1978) and Hurvich and Tsai (1989) and is meant to correct for some small-sample overfitting problems that the original AIC has, but it is constructed such that it is asymptotically equivalent to AIC

(McQuarrie and Tsai, 1998). Third is the corrected Akaike information criterion using an unbiased variance estimate, AICu:

$$\mathrm{AICu} = T\ln\left(\frac{RSS}{T-k}\right) + \frac{2T(k+1)}{T-k-2}. \tag{4}$$

This information criterion was suggested by McQuarrie, Shumway, and Tsai (1997) to deal further with the overfitting problems of AIC and AICc asymptotically, but in doing so it loses asymptotic equivalence with those other two criteria. The fourth criterion considered is the Schwarz (Bayesian) information criterion, SIC, defined as

$$\mathrm{SIC} = T\ln\left(\frac{RSS}{T}\right) + (\ln T)k\,, \tag{5}$$

which was introduced by Schwarz (1978) and an equivalent criterion was introduced by Akaike (1978).

The aim of AIC, AICc and AICu is optimal predictive performance in the original sample's domain. The original Akaike information criterion benefits from asymptotically efficiency (Shibata, 1980), in the sense that it gives a consistent estimator of predictive accuracy (Forster, 2001). Since AICc is asymptotically equivalent to AIC, AICc shares this property, but AICu does not due to its lack of asymptotic equivalence to these other two measures. The Schwarz information criterion instead is consistent in finding the true model if it is included in the set of choosable models, and in its Bayesian construction it is meant maximize the posterior probability of selecting the true model if that model is available to be chosen.

As noted earlier, with standard Dickey-Fuller unit root testing, there has often been a concern about the power of the test. Standard hypothesis testing in the frequentist tradition naturally can lead to low power, as only the probability of erroneously rejecting the null hypothesis is controlled for. Kwiatkowski, Phillips, Schmidt, and Shin (1992) introduced the KPSS test that reverses the unit root test so that stationarity is the null hypothesis and the unit root is the alternative hypothesis. Some have suggested doing unit-root testing both ways on the variable under investigation to build confidence about its unit-root status. Of course, doing so may also result in the tests indicating different hypotheses as being most appropriate, e.g. it may be the case that the Dickey-Fuller test does not reject the unit root while the KPSS test does not reject stationarity.

Minimizing an information criterion to decide whether or not there is a unit root avoids such conflict, but at the possible cost of losing information about confidence of one's finding. If the Dickey-Fuller test rejects the unit root at the 5% significance level for example, then when choosing stationarity the user can feel reasonably confident that he or she has not rejected the unit root wrongly. Choosing whether or not there is a unit root based on minimizing an information criterion cannot lead to such strong confidence in one's choice, but instead it may be considered in some ways as choosing the most likely model (unit root or not) given the evidence.[7] The "most likely" model, however, may be only slightly more likely. Whereas the legal analogy for hypothesis testing is determining whether somebody is guilty "beyond a reasonable doubt" (see Leamer, 1978), the legal analogy for model selection based on minimizing information criterion is determining whether somebody is guilty or not based on a "preponderance of the evidence".

---

[7] Here the phrase "most likely" needs to be interpreted loosely, as a model with $b < 0$ is actually a composite of an infinite number models, each with a different b less than zero, and probabilistic statements about the likelihood of the composite model become difficult. However, if one considers the "most likely" model to be the one that will most likely give good predictions or is more likely to be good in some other way, then the phrasing seems acceptable. Also making a model choice determination using information criterion does not necessarily mean that the degree of confidence in the choice must be absent. A possible measure of the strength of evidence of one model versus another using an information criterion may be based on the difference between the information criterion value for a particular model and that for which the information criterion is minimized. See for example Akaike (1978, 1983) and Burnham and Anderson (2002).

II. Examples of relative performance

This section presents some simulation results on the relative performance of using an information criterion rather than hypothesis testing when choosing among the four selectable models of Elder and Kennedy (2001). All the simulations are performed using GAUSS and are based on the results from 10,000 simulations on 50 observations with the variance of the error term in equation (1) drawn from a standard normal distribution. The small-sample situation of 50 observations is investigated as it may be considered reflective of a typical environment facing a macroeconomist dealing with yearly or quarterly data.

To avoid start-up values having a strong influence, there are also 50 unused observations on the variable *y* generated as prior observations to each used set of 50 observations. The first observation for the lagged *y* in the unused observations is 0, and the first observation on lagged *y* in the used observations is the last *y* in the unused prior observations on that variable. The deterministic time trend *t* moves in increments of one from one observation to the next.

First presented, in Figure 2, are some response curves representing the frequency of choosing the $b < 0$ hypothesis (stationarity) in various scenarios, comparing use of four information criteria (AIC, AICc, AICu, and SIC) and the Elder and Kennedy strategy in which each hypothesis test—the unit root test and, if necessary, the test for a trend—is done at the same significance level of 10% (referred to as EK10) or at the same significance level of 5% (referred to as EK05). The data generating process in Figure 2 is $\Delta y_t = by_{t-1} + \varepsilon_t$ with b varying between -0.5 and 0.[8] The figure has two parts: part (a) shows what happens when it is known that there is no trend in *y* (so the selectable models are only models (1) and (4)), and

[8] Even though $b < 0$ results in a true model (model (2)) that is not among those selectable, this figure is useful in considering unit root performance when a is sufficiently close to zero.

part (b) shows what happens when the trend status is unknown (so models (1), (3), (4), and (6) are selectable). For EK10 and EK05 what is being mapped out in both parts are power functions on the unit root test in the Elder and Kennedy strategy. The response curves associated with the information criteria cannot be called power functions, as these model selection techniques do not have declared null and alternative hypotheses. For values of b < 0, the higher the curve, the better the strategy is at choosing the correct model (as it indicates more frequent acceptance of $b < 0$ when it true). However, at $b = 0$, the lower the procedure's frequency of choosing $b < 0$, the better the strategy is at choosing the correct model, since choosing $b < 0$ under such circumstances is an error.

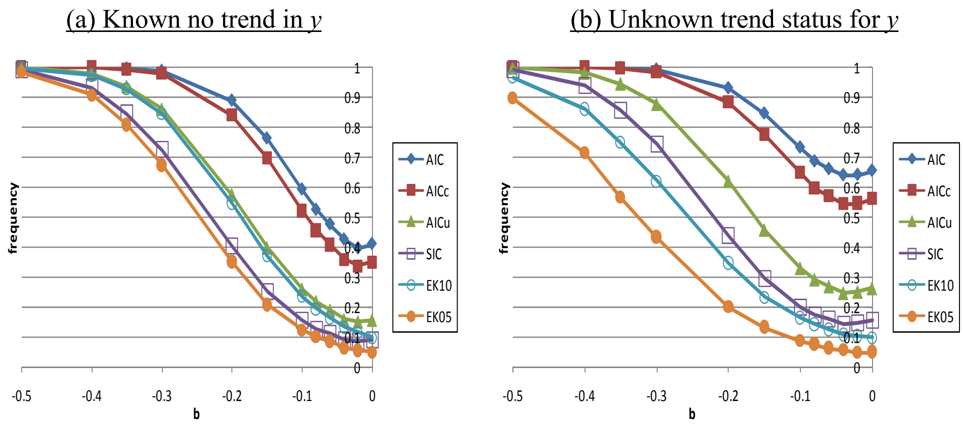

Figure 2. Frequency of choosing b < 0 with the data generating process $\Delta y_t = by_{t-1} + \varepsilon_t$

In Figure 2, we see that as b becomes larger in magnitude (measured leftward from the origin), each of the EK response curves increases, indicating higher likelihood of the technique discovering b's nonzero nature as b's magnitude increases. This pattern is what we would expect and it is repeated with the AIC and SIC curves except when b is very close to zero. Part (b) of this figure indicates that when there is neither a trend nor a nonzero intercept term a in the true model and the trend status of y is unknown, the information criteria do better than the techniques using the EK05 and EK10 at choosing the absence of a unit root when there is indeed no unit root, and the same can be said under the same circumstances when there is known to be no trend (as in part (a)), except EK10 finds $b < 0$ more frequently than SIC. However, when comparing any pair of strategies in the figure, the more frequent choosing of no unit root when none is present is at the clear cost of being less likely of choosing the unit root when a unit root exists.

In both parts (a) and (b) of Figure 2, the simulated frequency of not choosing the unit root when it exists is respectively 10% and 5% (rounded to the nearest whole number) for EK10 and EK05 respectively, so the size properties for the hypothesis testing strategies are very good as expected. The strategies using an information criterion tend to accept the unit root not as frequently when it exists: when the unit root exists, AICu and SIC result in respectively 15.7% and 8.9% frequencies on the choice of the unit root when the trend is known and respectively 26.3% and 15.4% frequencies on the choice of the unit root when the trend status is unknown. Compared to the other strategies, AIC and AICc have much more difficulty in finding the unit root when it exists, as can be seen in both parts (a) and (b) of the figure.

Figure 2 also displays an order to the performance of the strategies which we see repeated in later figures. This order is based on the parsimoniousness in accepting coefficients for estimation rather than setting them equal to zero. In general, the order of parsimoniousness found from lowest to highest is AIC, AICc, AICu, SIC, and EK05. EK10 is less parsimonious than EK05 and more parsimonious than AICu, and appears here and in later figures as sometimes more parsimonious than SIC, and at other times less parsimonious than SIC.

Figure 3 presents the same information as Figure 2 for the data generating process $\Delta y_t = 0.5 + by_{t-1} + \varepsilon_t$. With the inclusion of the substantially nonzero intercept term of 0.5 being the only change, the parts (a) and (b) in Figure 3 are almost identical to their counterparts in Figure 2, with the notable of exception of when b is close to zero. In both parts (a) and (b) of this figure there is a range near $b = 0$ for which as b gets closer to zero, the information criteria are increasing their frequency in which they correctly choose $b < 0$. The simulation with b close to zero is of course a bit unfair in part (a) of Figure 3, since it produces a process which is (or is close to) a stochastic trend, and a situation with a trend should not be under consideration given the assumed *a priori* information on that matter in this case.

(a) Known no trend in y (b) Unknown trend status for y

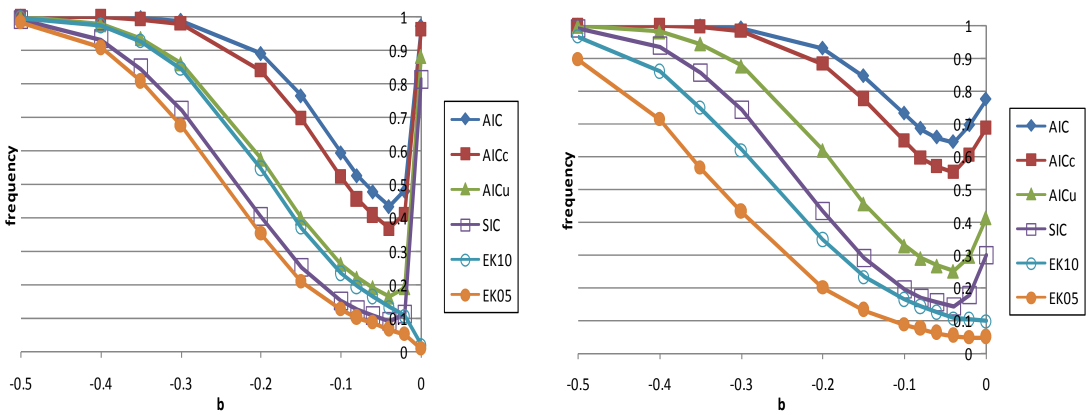

Figure 3. Frequency of choosing b < 0 with the data generating process $\Delta y_t = 0.5 + by_{t-1} + \varepsilon_t$

If the intercept term is higher, then the performance of the information criteria strategies can be substantially different. Figure 4 shows what happens when $\Delta y_t = 5 + by_{t-1} + \varepsilon_t$ and the trend status is unknown. In this case the information criteria strategies unusually have a local maximum in the frequency of choosing $b < 0$ when b is actually around -0.02.

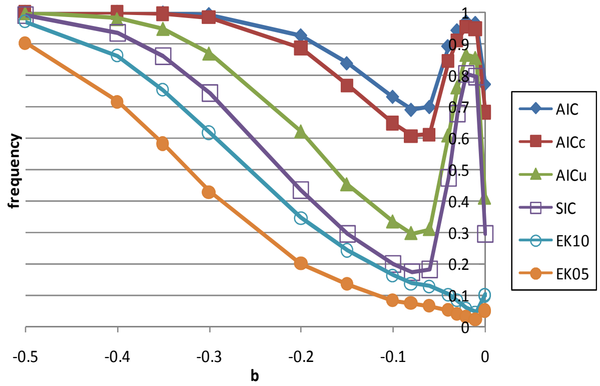

Figure 4. Frequency of choosing $b < 0$ with the data generating process $\Delta y_t = 5 + by_{t-1} + \varepsilon_t$, trend status unknown

Figure 5 presents the same information as Figure 4 for the data generating process $\Delta y_t = 0.5 + by_{t-1} + 0.25t + \varepsilon_t$, with there being *a priori* knowledge that there is a trend in y (thereby restricting the possible choices to models (3) and (6)). Without using this *a priori* knowledge the figure would not look substantially different, which is why such a figure is not presented. For $b \leq -0.08$, this figure has a pattern similar to previous figures. However, the frequency of choosing $b < 0$ when b is close to zero is exceptionally high for the information criteria strategies and when b is very close to zero that frequency goes to zero for EK05 and EK10. However, considering cases close to zero is somewhat unfair, since a unit root around a deterministic trend (model (5)) was removed *a priori* as a possibility for all strategies.[9]

[9] The fact that EK05 and EK10 do not converge to their nominal sizes in this case is not unexpected. An additional term such as t^2 would be needed to be included in the regression model for a legitimate test for the unit root under these circumstances (Harris and Sollis, 2003, p. 45).

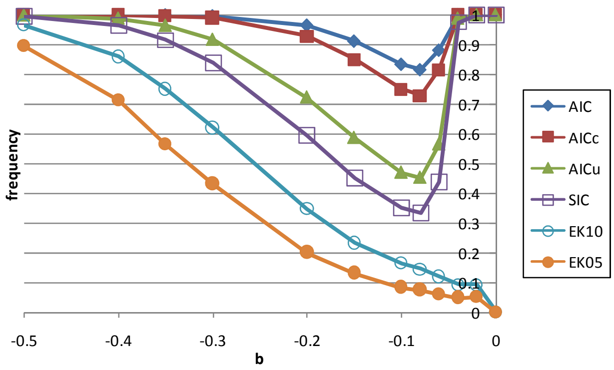

Figure 5: Frequency of choosing b < 0 with the data generating process $\Delta y_t = 0.5 + by_{t-1} + 0.25t + \varepsilon_t$, known existence of trend in y

One of the issues that the previous five figures masks is the relevance of choosing the incorrect model. If indeed there is a very slowly converging stationary process, it will tend to act very similarly to a unit root process except over a long time span. Also we have the issue in applied work that none of the models is the true model; any economic variable is in reality determined through a very complicated system. What we are often interested in instead is whether the model we choose for indicating the movement in an economic variable is sufficient in the sense that the variable can be treated *as if* it was being determined by that model. A reasonable way to consider whether the model is sufficient for such purposes is to look at its predictive capability.

For consideration of predictive capability, we can look at the L_2 distance measure. For the model $\Delta y_t = a + by_{t-1} + ct + \varepsilon_t$, the L_2 distance for Δy for simulations s is calculated as

$$L_{2s} = \frac{\sum_t \left((a + by_{t-1,s} + ct) - (\hat{a}_s + \hat{b}_s y_{t-1,s} + \hat{c}_s t)\right)^2}{T}, \tag{6}$$

where T is the number of observations, $\hat{a}_s$, $\hat{b}_s$, and $\hat{c}_s$ are the estimated or zero-restricted parameters associated with the estimation of a particular model in simulation s, and $y_{t-1,s}$ is the simulated lag value of y in simulation s. Therefore, L_{2s} is basically the mean of the squared differences between the expected value of the dependent variable and the estimated value of that variable (McQuarrie and Tsai, 1998) in simulation s, where the dependent variable here is the change in y. For any particular set of parameters, the L_2 distance presented in this paper is the mean over the various simulations for L_{2s}. The L_2 distances for the various techniques are shown in Figures 6, 7, and 8, respectively for same cases used in Figures 3, 4, and 5. Of course, the lower L_2 is, the better. All three of the new figures indicate superior performance of the information criteria (compared to the two hypothesis-testing strategies) for moderately large-magnitude in b, i.e. around b = -0.3, and superior performance of the EK05 strategy for b between -0.1 and -0.06.

In part (a) of Figure 6, the hypothesis testing strategies do horribly when b is very close to zero (both EK05 and EK10 meet the vertical axis near 0.25, whereas all the information criterion strategies do so below 0.09), but as noted previously, this is an unfair situation in which there is a stochastic trend, or close to one, and the assumed *a priori* knowledge is that there is no trend. Part (b) of Figure 6 indicates that when the same data generating process is present and no *a priori* knowledge of the trend status is used, EK05 has superior performance in comparison to the information criterion strategies for all values of b close to zero ($-0.1 \leq b \leq 0$).

(a) Known no trend in y (b) Unknown trend status for y

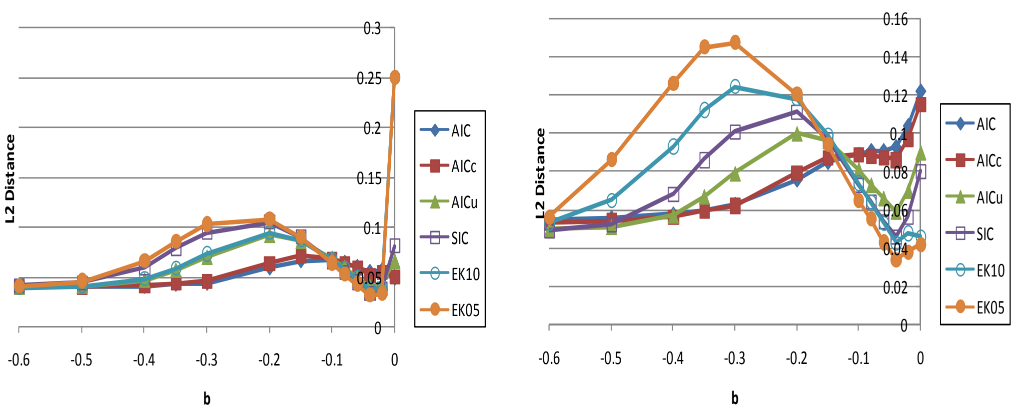

Figure 6. L_2 **distance with the data generating process** $\Delta y_t = 0.5 + by_{t-1} + \varepsilon_t$

In Figure 7, EK10 and EK05 have their L_2 distances jump up to around 0.14 when b gets close to -0.02. All of the information criterion strategies avoid this jump. This corresponds to what we see in Figure 4 with the information criterion strategies jumping in their frequency in choosing $b < 0$ when b gets close to -0.02. Interestingly, with Figures 4 and 7, a jump in one figure corresponds to no jump in the other.

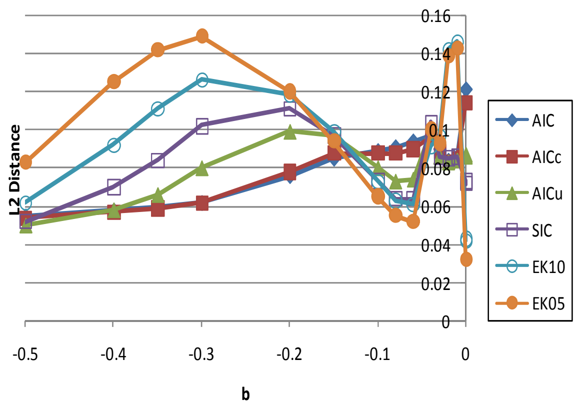

Figure 7. L_2 **distance with the data generating process** $\Delta y_t = 5 + by_{t-1} + \varepsilon_t$**, trend status unknown**

In Figure 8, for -0.06 < $b \leq 0$ the L_2 distance when using EK10 and EK05 shoots off to being very high (so high that the figure is truncated to have a maximum at 0.16, so other differences in the performances between the strategies can be seen). However, as noted with Figure 5, there is some unfairness in considering cases with b close to zero here since a unit root around a deterministic trend (model (5)) was removed *a priori* as a possibility for all strategies.

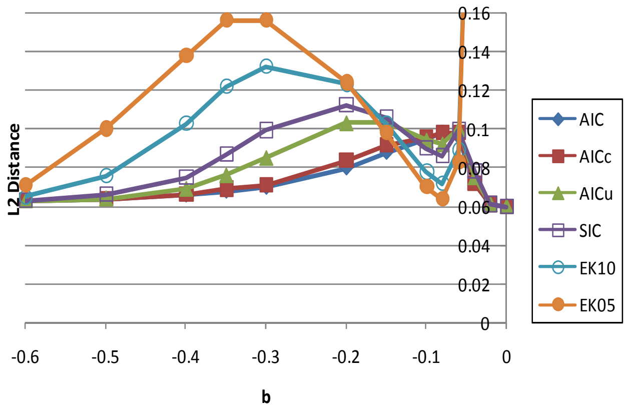

Figure 8. L_2 distance with the data generating process $\Delta y_t = 0.5 + by_{t-1} + 0.25t + \varepsilon_t$, **known existence of trend in *y***

III. Minimizing the maximum regret

This section considers many parameter permutations for the true model and considers the maximum regret of using one technique rather than another. How such regret is measured is done in a number ways and is discussed later, but first the set of various parameter permutations used for the true model in the simulations will be presented. The various permutations of the elements from following sets of numerical values are used for the true model based on the general equation $\Delta y_t = a + by_{t-1} + ct + \varepsilon_t$:

a = {0.0, 0.25, 0.35, 0.4, 0.45, 0.5, 1, 2, 2.5, 3, 5, 10},

b = {-1.0, -0.8, -0.6, -0.5, -0.4, -0.35, -0.3, -0.2 -0.15, -0.1, -0.08, -0.06, -0.04, -0.02, 0},

c = {0.0, 0.2, 0.4, 0.5, 1, 2, 5}.

By focusing on values of *b* only in the range [-1, 0], only unit root and stationary processes that are nonoscillatory are dealt with. The 90 (=15 × 6) permutations from the above sets in which *c* > 0 and *a* = 0 are excluded as providing a legitimate model (none of the original six in Table 1 matches this case). Also, the parameter permutations that result in unacceptable models according to Elder and Kennedy (2001), i.e. models (2) and (5), are removed as possibilities for the true model, and parameter permutations that result in model (6) formally but with an equation that is really close to model (5) are removed, so there are overall 848 parameter permutations to be considered.[10] Of these, there are 693 which have a trend and 133 which have no trend and are not close to having a stochastic trend.[11]

The regret of using one procedure rather than the other is measured in various ways. One is how much higher the L_2 distance is using one strategy rather than another (the difference in L_2) for a particular parameter permutation. Table 2 shows the maximum regrets with regret measured in this way over all the parameter permutations we consider. Part (a) assumes *a priori* knowledge that there is no trend, so only the 133 parameter permutations with no trend and not close to a stochastic trend are used there. Part (b) assumes *a priori* knowledge of a existing trend, so the 693 parameter permutations with a trend are used there, and part (c) assumes no knowledge of the trend status of y, so all of the 848 considered parameter permutations are used in that part. The value in each cell represents the maximum regret of using the procedure noted in the corresponding row rather than the procedure in the corresponding column.

As an example of how to read the table, when using EK05 rather than AIC and there is known to be no trend (taking us to part (a) of the table), the maximum worsening of the L_2

[10] In particular, the 14 cases in which $b < 0$ and $a = c = 0$ are removed; the 264 (=11×4×6) cases in which $a \neq 0$, $-0.06 \geq b \geq 0$, and $c \neq 0$ are removed; and the 44 (=11×1×4) cases in which $a \neq 0$, $b = -0.08$ and $c \geq 0.5$ are removed. Therefore there are overall 12×15×7 – 90 – 14 – 264 – 44 = 848 parameter permutations to be considered.

[11] There are 22 (=11×2×1) cases with $a \neq 0$, $-0.04 \leq b < 0$, and $c = 0$ which are removed as possibilities as no-trend models even though they have no trend formally, since they are so close to having a stochastic trend.

distance over all the 133 parameter permutations is 0.060 units,[12] whereas when using AIC rather than EK05, the maximum worsening of the L_2 distance over all those permutations is 0.040 units. Based on the difference in L_2 distance, the maximum regret of using EK05 rather than AIC is therefore 0.060 units and the maximum regret of using AIC rather than EK05 is 0.040 units. Using the concept of minimax regret, AIC would then be preferable to use in comparison to EK05 since AIC has a lower maximum regret compared to EK05. The table indicates that when there is known to be no trend, AIC minimizes the maximum regret in comparison to each of the other procedures. The highlighted cells in Table 2(a) are the relevant ones for the comparisons needed for this statement: in AIC's comparison to AICc, 0.005 < 0.006; in its comparison to AICu, 0.024 < 0.034; in its comparison to SIC, 0.033 < 0.052; in its comparison to EK10, 0.032 < 0.036; and in its comparison to EK05, 0.040 < 0.060.

Table 2. Maximum regrets, predictive, based on the difference in L_2 distance

(a) Known there is no trend

	AIC*	AICc	AICu	SIC	EK10	EK05
AIC*	0	0.005	0.024	0.033	0.032	0.040
AICc	0.006	0	0.019	0.028	0.028	0.035
AICu	0.034	0.029	0	0.009	0.009	0.016
SIC	0.052	0.050	0.026	0	0.024	0.007
EK10	0.036	0.031	0.004	0.004	0	0.007
EK05	0.060	0.058	0.035	0.009	0.032	0

[12] For one permutation of the parameters L_2 was 0.103 for EK05 and 0.043 for AIC, leading to the worsening of the L_2 of 0.103 – 0.043 = 0.060 when using EK05 rather than AIC with that parameter permutation.

(b) Known there is a trend

	AIC	AICc	AICu*	SIC	EK10	EK05
AIC	0	0.008	0.037	0.052	0.080	0.089
AICc	0.005	0	0.030	0.044	0.072	0.082
AICu*	0.024	0.020	0	0.015	0.043	0.054
SIC	0.034	0.031	0.016	0	0.029	0.039
EK10	0.120	0.119	0.110	0.100	0	0.011
EK05	0.130	0.129	0.120	0.11	0.039	0

(c) Unknown trend status

	AIC	AICc	AICu*	SIC	EK10	EK05
AIC	0	0.010	0.05	0.07	0.08	0.089
AICc	0.005	0	0.04	0.059	0.073	0.082
AICu*	0.025	0.022	0	0.019	0.045	0.054
SIC	0.039	0.039	0.023	0	0.036	0.04
EK10	0.399	0.401	0.406	0.408	0	0.014
EK05	0.405	0.407	0.411	0.413	0.038	0

The value in each cell represents the maximum regret of using the procedure noted in the corresponding row rather than the procedure in the corresponding column. The strategy providing minimax regret compared to all other strategies has an asterisk noted next to it, and the highlighted cells are the ones used in the comparison to back that conclusion.

Table 2(b) indicates that when there is known to be a trend in *y*, AICu minimizes the maximum regret in comparison to each of the other procedures (in AICu's comparison to AIC, 0.024 < 0.037; in its comparison to AICc, 0.020 < 0.030, etc.). Table 2(c) indicates that with an unknown trend status, AICu is again the strategy that minimizes the maximum regret (in terms of the difference in L_2 distance) with respect to each of the other strategies.

All of the information criterion strategies minimize the maximum regret in terms of the difference of the L_2 distance compared to EK05 or EK10, regardless of whether *a priori* information on trend status is used, except when comparing AICu or SIC with EK10 when there is known to be no trend.

Table 3 provides the same information as Table 2, except now to measure the maximum regret on predictive ability the difference in the ln L_2 distance is used (this is equivalent to the ln of the ratio of the L_2 distances, so exponentiation of the numbers in the table provides L_2 ratios).[13] Here we see that when *a priori* knowledge of no trend in *y* is used, the Elder and Kennedy strategy does very well, with both EK10 and EK05 doing better in minimax regret than each of the information criterion strategies. EK05 has minimax regret compared to all other strategies including EK10. However, if there is a known to be a trend in the data or if the trend status is unknown, then SIC has minimax regret versus all other strategies. If the trend status is unknown, all the information criteria have minimax regret with respect to the two hypothesis testing strategies, but when knowledge of the existence of a trend is used, only AICu and SIC have minimax regret compared to the hypothesis testing strategies.

When considering minimax regret, using the difference in ln L_2 for regret appears to favor more parsimonious strategies than using the difference in L_2. One weakness of using the difference in ln L_2 however is that if model (1) is the true model and is choosable, then an extreme strategy that always chooses that model will have minimax regret against all others since the L_2 distance is 0 for such a strategy when that model is the true one, so in effect any other strategy that sometimes does not choose that model will result in regret with respect to the extreme strategy that is in effect infinitely high (the ratio of the L_2 distance for the alternative strategy compared the L_2 distance of the extreme strategy of always choosing model (1) would be undefined since it would involve a division by zero). This problem does not arise when using the difference in L_2 for determining regret.

[13] In the case of using EK05 rather than AIC given prior knowledge of there being no trend in *y*, for example, the maximum regret is given as 0.871, which means using EK05 rather than AIC results in L_2 being higher by about 139% (exp(0.871) = 2.39; 2.39-1=1.39=139%) for one parameter permutation, the worst situation when using EK05 rather than AIC.

Table 3. Maximum regrets, predictive, based on the difference in ln L_2 distance

(a) Known there is no trend

	AIC*	AICc	AICu	SIC	EK10	EK05*
AIC*	0	0.105	0.659	1.111	1.062	1.634
AICc	0.098	0	0.554	1.006	0.957	1.529
AICu	0.464	0.421	0	0.452	0.403	0.974
SIC	0.785	0.743	0.327	0	0.290	0.523
EK10	0.502	0.460	0.047	0.096	0	0.572
EK05*	0.871	0.829	0.430	0.112	0.402	0

(b) Known there is a trend

	AIC	AICc	AICu	SIC*	EK10	EK05
AIC	0	0.066	0.369	0.554	1.071	1.348
AICc	0.057	0	0.304	0.491	1.007	1.287
AICu	0.269	0.221	0	0.196	0.709	0.993
SIC*	0.372	0.357	0.169	0	0.526	0.807
EK10	0.969	0.953	0.841	0.731	0	0.288
EK05	1.017	1.003	0.891	0.781	0.321	0

(c) Unknown trend status

	AIC	AICc	AICu	SIC*	EK10	EK05
AIC	0	0.101	0.638	1.061	1.229	1.808
AICc	0.055	0	0.537	0.959	1.128	1.707
AICu	0.281	0.243	0	0.422	0.729	1.170
SIC*	0.484	0.484	0.259	0	0.551	0.804
EK10	2.010	2.040	2.119	2.151	0	0.579
EK05	2.022	2.052	2.131	2.163	0.326	0

The value in each cell represents the maximum regret of using the procedure noted in the corresponding row rather than the procedure in the corresponding column. The strategy providing minimax regret compared to all other strategies has an asterisk noted next to it, and the highlighted cells are the ones used in the comparison to back that conclusion.

It is perhaps illuminating to understand what type of parameter permutations lead to the worst predictive regret for the information criteria against EK05 and vice versa. When the trend status is unknown, the data generating process that leads to the maximum regrets (in L_2 difference or ln L_2 difference) for EK05 against any of the information criteria is

$\Delta y_t = 10 - 0.02 y_{t-1} + \varepsilon_t$. Thus an almost stochastic trend situation with a high-valued intercept, as seen to the far right of Figure 7, is giving EK05 the most problems. When the trend status is unknown, what gives the information criteria the most difficulty against EK05, in terms of maximum regrets based on the difference in L_2 distance, are the simplest unit root case $\Delta y_t = \varepsilon_t$ (for AIC) and the stochastic trend situations $\Delta y_t = 2 + \varepsilon_t$ (for AICc and AICu) and $\Delta y_t = 5 + \varepsilon_t$ (for SIC). When the trend status is unknown, what gives the information criteria the most difficulty against EK05, in terms of maximum regrets based on the difference in ln L_2 distance, is the simplest unit root case $\Delta y_t = \varepsilon_t$ (for AIC, AICc, and AICu) and the stochastic trend situation $\Delta y_t = 2.5 + \varepsilon_t$ (for SIC).

A third way considered for measuring regret is how much more frequently the unit root status is chosen incorrectly compared to another procedure. Table 4 indicates the maximum regrets measured in this way and is otherwise read the same way as the previous two tables. For example, when it is known that there is no trend in *y* (part (a) in Table 4), AIC is shown at worst to get the unit root status incorrect by an extra 36.1 percentage points in frequency compared to the Elder and Kennedy strategy with five-percent significance testing. When it is known that there exists a trend in *y*, AICc minimizes the maximum regret on frequency of erroneous unit root choice compared to all other strategies, and when it is known that no trend exists or when the trend status is unknown, AIC minimizes the maximum regret on frequency of erroneous unit root choice compared to all other strategies. This table also shows that the information criterion strategies minimize the maximum regret compared to the techniques based on the Elder and Kennedy strategy in all three situations of *a priori* trend status knowledge, except EK10 has the minimax regret with respect to SIC when there is known to be no trend.

Table 4. Maximum regrets, difference in frequency of erroneous unit root choice

(a) Known there is no trend

	AIC*	AICc	AICu	SIC	EK10	EK05
AIC*	0	0.062	0.249	0.321	0.308	0.361
AICc	0.076	0	0.187	0.258	0.246	0.299
AICu	0.365	0.300	0	0.071	0.059	0.112
SIC	0.508	0.449	0.176	0	0.149	0.04
EK10	0.475	0.410	0.138	0.012	0	0.053
EK05	0.570	0.505	0.233	0.106	0.201	0

(b) Known there is a trend

	AIC	AICc*	AICu	SIC	EK10	EK05
AIC	0	0.095	0.376	0.484	0.656	0.705
AICc*	0.092	0	0.282	0.393	0.562	0.611
AICu	0.377	0.292	0	0.115	0.285	0.337
SIC	0.503	0.421	0.145	0	0.175	0.225
EK10	0.728	0.721	0.661	0.603	0	0.054
EK05	0.836	0.830	0.770	0.713	0.202	0

(c) Unknown trend status

	AIC*	AICc	AICu	SIC	EK10	EK05
AIC*	0	0.096	0.392	0.506	0.675	0.728
AICc	0.097	0	0.296	0.416	0.593	0.644
AICu	0.410	0.326	0	0.124	0.318	0.367
SIC	0.549	0.481	0.186	0	0.201	0.253
EK10	0.961	0.961	0.960	0.960	0	0.052
EK05	0.983	0.983	0.982	0.982	0.202	0

The value in each cell represents the maximum regret of using the procedure noted in the corresponding row rather than the procedure in the corresponding column. The strategy providing minimax regret compared to all other strategies has an asterisk noted next to it, and the highlighted cells are the ones used in the comparison to back that conclusion.

A fourth way considered for measuring regret is how much more frequently an action for further analysis is chosen incorrectly compared to another procedure. The actions are those typically taken when including the variable as a dependent variable (or a transformation of it) in a subsequent regression. The actions typically available are (i) take a first difference (correct if unit root found), (ii) detrend or include a time trend in subsequent regressions (correct if a trend-stationary process found), or (iii) do nothing (correct if stationary with no trend found). Action (i) corresponds to choosing model (1) or model (3), action (ii) corresponds to choosing model (6), and action (iii) corresponds to choosing model (4). Table 5 indicates the maximum regrets measured in this way with an unknown trend status of y and is otherwise read the same way as the previous table (when the trend status is known, choice of action is based solely upon unit root choice, so the maximum regrets would be the same as those in parts (a) and (b) of Table 3). For example, at worst AICu is shown in Table 5 for one parameter permutation to get the chosen action incorrect by an extra 36.7 percentage points in frequency compared to the Elder and Kennedy strategy with five-percent significance testing.

Table 5. Maximum regrets, difference in frequency of incorrect action due to model choice, unknown trend status

	AIC	AICc	AICu*	SIC	EK10	EK05
AIC	0	0.096	0.392	0.506	0.675	0.728
AICc	0.088	0	0.296	0.416	0.593	0.644
AICu*	0.369	0.29	0	0.124	0.318	0.367
SIC	0.491	0.41	0.147	0	0.201	0.253
EK10	0.726	0.771	0.858	0.885	0	0.056
EK05	0.787	0.771	0.858	0.885	0.202	0

The value in each cell represents the maximum regret of using the procedure noted in the corresponding row rather than the procedure in the corresponding column. The strategy providing minimax regret compared to all other strategies has an asterisk noted next to it, and the highlighted cells are the ones used in the comparison to back that conclusion.

Again, this table shows that when the trend status of y is unknown, the information criterion strategies minimize the maximum regret compared to the hypothesis testing techniques. AICu is now found to minimize the maximum regret compared to all the other strategies, in contrast with AIC being the minimax regret winner on frequency of erroneous unit root choice found in Table 3(c).

IV. Conclusions

This paper has examined the relative performance of using information criteria rather than hypothesis testing to choose among the various models of the classic Dickey-Fuller unit-root testing environment. The results indicate that utilizing information criteria for deciding whether or not there is a unit root and (if unknown) whether or not a trend exists is a competitive technique to traditional hypothesis testing techniques based on, for example, the Elder and Kennedy (2001) strategy. The information criterion strategies are competitive when the researcher is more interested in the model that is most supported by the data rather than finding sufficient evidence to confidently reject a particular hypothesis.

The simulations in this paper consider various situations of *a priori* knowledge on the trend status of the examined variable (known no trend, known existing trend, and trend status unknown) and various situations of performance measure used (minimax regret based on (1) the difference in L_2 distance, which is a predictive measure; (2) the difference in ln L_2 distance; (3) the difference in the frequency of choosing a unit root or not correctly; or (4) the difference in the frequency of correctly choosing the subsequent action—first differencing, detrending, or doing nothing when using the examined variable as a dependent variable in a subsequent regression). In all situations, the information criterion strategies perform well (in terms of minimax regret) compared to the Elder and Kennedy strategy with the following exceptions: (1) when prior knowledge of no trend in the examined variable is used, then the Elder and Kennedy strategy with 10% significance used for the unit root test performs better than AICu and SIC when considering the difference in L_2 distance and better than SIC when considering performance in unit-root choice and in choice of subsequent action, and (2) when considering the difference in ln L_2 distance, the Elder and Kennedy strategy, with either 10% or 5% used for the unit root test, performs better than AIC and AICc when prior knowledge of the existence of a trend is used and better than all the information criterion strategies when prior knowledge of no trend is used. When there is an unknown trend status or the information that a trend exists is used, then SIC and AICu always perform better than the two simulated hypothesis testing strategies, regardless of which of the four performance measures is used.

References

Akaike, H. (1973). 'Information theory and an extension of the maximum likelihood principle', in B.N. Petrov and F. Csaki eds. *2nd international symposium on Information Theory*, pp. 267–281. Budapest: Akademia Kiado.

Akaike, H. (1974). 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control AC*, Vol. 19, pp. 716–723.

Akaike, H. (1978). 'A Bayesian analysis of the minimum AIC procedure', *Annals of the Institute of Statistical Mathematics* Vol. 30, Part A, pp. 9–14.

Akaike H (1983). 'Information measures and model selection', *International Statistical Institute* Vol. 44, pp. 277–291.

Ayat, L. and Burridge, P. (2000): 'Unit root tests in the presence of uncertainty about the nonstochastic trend', *Journal of Econometrics* Vol. 95, pp. 71–96.

Burnham, K. P. and Anderson D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* 2nd edition. Springer-Verlag, New York.

Bunzel, H. and Vogelsang, T. J. (2005).' Powerful Trend Function Tests That Are Robust to Strong Serial Correlation, with an Application to the Prebisch-Singer Hypothesis', *Journal of Business and Economic Statistics* Vol. 23, pp. 381–394.

Dolado, J., Jenkinson, T. and Sosvilla-Rivero, S. (1990): 'Cointegration and unit roots', *Journal of Economic Surveys* Vol. 4, pp. 249–273.

Dickey, D. A., and Fuller, W. A. (1979): ‘Distribution of the Estimators for Autoregressive Time Series with a Unit Root’, *Journal of the American Statistical Association*, Vol. 74, pp. 427–431.

Elder J. and Kennedy P. E. (2001) ‘Testing for Unit Roots: What Should Students Be Taught?’, *Journal of Economic Education*, Vol. 32, pp. 137–146

Enders, W. (2004) *Applied Econometric Time Series*, Second Edition. John Wiley & Sons: United States.

Forster, M. (2001) ‘The New Science of Simplicity’, Chap. 5 in A. Zellner, H. A. Keuzenkamp, and M. McAleer eds. *Simplicity, Inference, and Modelling: Keeping It Sophisticatedly Simple*. Cambridge: Cambridge University Press.

Granger, C. and Newbold, P. (1974) ‘Spurious Regressions in Econometrics’, *Journal of Econometrics*, Vol. 2, 111–120.

Hacker, R.S. and Hatemi-J, A. (2008) ‘The Properties of Procedures Dealing with Uncertainty about Intercept and Deterministic Trend in Dickey-Fuller Unit Root Testing’. Working Paper.

Harris, R. and Sollis, R. (2003) *Applied Time Series Modelling and Forecasting*. John Wiley & Sons, Chichester, West Sussex, England.

Holden, D. and Perman, R. (1994) ‘Unit roots and cointegration for the economist’, in B. B. Rao, ed. *Cointegration for the Applied Economist*. 47–112. New York: St. Martin’s.

Hurvich, C. M. and Tsai, C. L. (1989). ‘Regression and Time Series Model Selection in Small Samples’, *Biometrika*, Vol. 76, pp. 297-307.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992), ‘Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?’, *Journal of Econometrics*, Vol. 54, pp. 159–178.

Leamer, E. (1978): *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. John Wiley & Sons, New York.

MacKinnon, J. G. (1991): 'Critical Values for Cointegration Tests', republished in *Long-run Economic Relationships, Readings in Cointegration,* Edited by Engle, R. F., and Granger, C. W. A., Oxford University Press.

McQuarrie, A. D. R., Shumway, R. H., and Tsai, C. L. (1997). 'The model selection criterion AICu. Statistics and Probability Letters', Vol. 34, pp. 285–292.

McQuarrie, A. D. R and Tsai, C. L. (1998) *Regression and Time Series Model Selection*. World Scientific Publishing, Singapore.

Phillips, P. (1986) 'Understanding Spurious Regressions in Econometrics', *Journal of Econometrics*, Vol. 33, pp. 311–340.

Perron, P. (1988) 'Trends and Random Walks in Macroeconomic Time Series', *Journal of Dynamics and Control*, Vol. 12, pp. 297–332.

Said, S. and Dickey, D. (1984) 'Testing for Unit Roots in Autoregressive-Moving Average Models with Unknown Order', *Biometrika*, Vol. 71, pp. 599–607.

Schwarz , G. (1978). 'Estimating the dimension of a model', *Annals of Statistics* Vol. 6, pp. 461–464.

Shibata, R. (1980). 'Asymptotic Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process', *Annals of Statistics*, Vol. 8, pp. 147–164.

Sugiura, N. (1978) 'Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections', *Communications in Statistics – Theory and Methods*, Vol. 7, pp. 13–26.

Vogelsang, T. J. (1998) 'Trend Function Hypothesis Testing in the Presence of Serial Correlation', *Econometrica*, Vol. 66, pp. 123–148.