

An abridged version of this paper is forthcoming in
Computational Statistics and Data Analysis
Published online 5 Aug. 2015
DOI: 10.1016/j.csda.2015.07.003

**Moment Ratio Estimation of Autoregressive/Unit Root Parameters
and Autocorrelation-Consistent Standard Errors**
(Unabridged Version)

J. Huston McCulloch
Ohio State University (Emeritus)
April 6, 2015
Corrections Aug. 1, 2015

ABSTRACT

A *Moment Ratio* estimator is proposed for an $AR(p)$ model of the errors in an OLS regression, that provides standard errors with far less median bias and confidence intervals with far better coverage than conventional alternatives. A unit root, and therefore the absence of cointegration, does not necessarily mean that a correlation between the variables is spurious.

The estimator is applied to a quadratic trend model of real GDP. The rate of change of GDP growth is negative with finite standard error but is insignificant. The “output gap” has an infinite standard error and is therefore a statistical illusion.

Keywords: Method of Moments, Autoregressive processes, Unit Root Processes, Regression Errors, Consistent Covariance matrix, Real GDP growth

JEL codes: C13 (Estimation), C22 (Time Series Models), E17 (Macroeconomic Forecasting and Simulation)

URL to PDF of latest version of paper, with color graphs:
<http://www.econ.ohio-state.edu/jhm/papers/MomentRatioEstimator.pdf>

Corresponding author: J. Huston McCulloch, 145 W 67th St, New York, NY 10023.
Cell: (614) 460-1195. E-mail: mcculloch.2@osu.edu

1. Introduction

Serial correlation is a pervasive problem in time series models in econometrics, as well as in statistics in general. When, as is often the case, positive serial correlation is present in both the errors and the regressors, it has long been well known that the Ordinary Least Squares (OLS) estimates of the standard errors are generally too small, and hence the derived t-statistics too large.¹

If the form and parameters of the error serial correlation were known, it would be elementary to compute correct standard errors for OLS regression coefficients. However, observed regression residuals are typically much less persistent than the unobserved regression errors. Correlations estimated directly from the regression residuals therefore provide inadequate indication of the serial correlation that is actually present. This problem is particularly severe as the persistence in the errors approaches or even reaches a unit root.

The present paper proposes a *Moment Ratio* (MR) estimator for the parameters of an Autoregressive (AR) model of the errors in an OLS regression. Although it is computed from conventional correlation coefficients, it reduces their negative bias, and provides standard errors with far less median bias and confidence intervals with far better coverage than conventional alternatives. The MR estimator is in the spirit of the Median Unbiased estimator of Andrews (1993) and McCulloch (2008), but does not require laborious Monte Carlo simulation of the distribution of the sample autocorrelations.

By allowing the AR order to increase with the sample size by the same formula that is commonly used for Newey-West (1987) Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix, the MR estimator shares its consistency with respect to autocorrelation, while greatly mitigating its substantial finite-sample bias.

A unit root in the errors requires reposing the problem and greatly increases the variance of OLS slope coefficients, but otherwise presents no insurmountable difficulties. In particular, the presence of a unit root in the errors does not by itself indicate that an OLS regression is spurious. A unit root test similar in spirit to that of Andrews and Chen (1994) is developed that has only moderate size distortion in simulations using a trend line regression.

An MR-HAC estimator that generalizes the White (1980) Heteroskedasticity Consistent Covariance (HCC) matrix is proposed to capture regressor-conditional heteroskedasticity. It is found to have less finite-sample bias and better coverage than the Newey-West HAC estimator in simulated regressions.

The Moment Ratio estimator is applied to a quadratic trend model of US real GDP. OLS, HAC, and even AR(4) standard errors are found to greatly overstate the precision of the coefficients, and all give the misleading impressions that the terminal output gap of -4.80% in 2014Q2 is significantly negative and that the rate of change of

¹ This understanding goes back at least to Bartlett (1935) and Quenouille (1952).

GDP growth is significantly negative. The Moment Ratio persistence coefficient estimate is quite close to unity, and a unit root in the errors cannot be rejected. Using the preferred “blended” covariance matrix, the “output gap” has infinite variance and therefore is illusory, and the estimated secular rate of decline of GDP growth of -2.82% per annum per century is statistically insignificant. Nevertheless, expected GDP growth in 2014Q2 of +2.30% per annum remains strongly significant and positive despite the absence of cointegration.

Section 2 below reviews basic theory. Section 3 develops the “MR(1)” model, i.e. the MR estimator in the case of stationary AR(1) errors, and compares the derived standard errors to conventional alternatives. Section 4 discusses the Newey-West HAC estimator. Section 5 develops the modifications required for the MR(1) estimator in the unit root case. Section 6 investigates the bias in MR(1) standard errors using Monte Carlo simulations. Section 7 develops an exact unit root test for the case of AR(1) errors. Section 8 develops the MR(p) estimator, which extends the MR(1) estimator to a more general AR(p) error process. Section 9 uses simulations to compare the bias and size distortion of the MR(p) estimator to that of OLS, HAC, and asymptotic Method of Moments. Section 10 develops an approximate unit root test for the MR(p) case, and finds that it has only moderate size distortion. Section 11 uses the p-value of the unit root test to construct a blended covariance matrix that greatly reduces the remaining size distortion of the MR(p) estimator. Section 12 proposes an MR-HAC estimator and compares it to the non-parametric estimator of Kiefer, Vogelsang and Bunzel (2000). Section 13 applies the MR estimator to a quadratic trend model of US real GDP. Section 14 concludes.

2. Basics

Consider a time-series linear regression of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{X} is an $n \times k$ matrix of exogenous regressors whose first column is a vector of units so that β_1 is the intercept. We assume that the $n \times 1$ error vector $\boldsymbol{\varepsilon}$ has mean $\mathbf{0}$, is independent of \mathbf{X} , and, if stationary, has a time-invariant autocovariation structure,

$$\boldsymbol{\Gamma} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = (\gamma_{|i-j|})^2.$$

The OLS estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}$$

then has covariance matrix³

² Although this independence assumption is much more restrictive than one of weak exogeneity, it is required by the factorizations below that motivate the Moment Ratio estimator. As Peron and Yabu (2009) point out, deterministic linear trend regressions are often of interest, both to macroeconomists and climate scientists, in which case independence is a not unreasonable assumption. Likewise, the quadratic trendline illustration in Section 13 below reasonably satisfies independence. The author is optimistic that the proposed Moment Ratio estimator will be beneficial even with the weakly exogenous regressors that are often encountered in econometrics. Section 12 below modifies the estimator to allow for regressor-conditional heteroskedasticity.

³ See, e.g., Greene (2003: 193).

$$\begin{aligned}
\mathbf{C} &= \text{Cov}(\hat{\boldsymbol{\beta}}) \\
&= E\left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Gamma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \gamma_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{R}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},
\end{aligned} \tag{2}$$

where \mathbf{R} is the population autocorrelation matrix,

$$\mathbf{R} = (\rho_{|i-j|}) = \boldsymbol{\Gamma} / \gamma_0. \tag{3}$$

The vector of observed OLS residuals equals the “annihilator matrix” \mathbf{M} times the vector of unobserved errors:

$$\mathbf{e} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\varepsilon} = \mathbf{M}\boldsymbol{\varepsilon}. \tag{4}$$

Define

$$s_j = \text{tr}_j(\mathbf{e}\mathbf{e}'), \quad j = 0, \dots, n-1,$$

where we define the j -th order trace operator $\text{tr}_j(\cdot)$ for an $n \times n$ matrix $\mathbf{A} = (a_{i,j})$ by

$$\text{tr}_j(\mathbf{A}) = \sum_{i=1}^{n-j} a_{i,i+j}.$$

The residual autocorrelations are then customarily⁴ computed from the residuals as

$$r_j = s_j / s_0, \quad j = 0, \dots, n-1. \tag{5}$$

In general,

$$E\mathbf{e}\mathbf{e}' = E(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{M}) = \mathbf{M}\boldsymbol{\Gamma}\mathbf{M}, \tag{6}$$

so that

$$E s_j = \text{tr}_j(\mathbf{M}\boldsymbol{\Gamma}\mathbf{M}) = \gamma_0 \text{tr}_j(\mathbf{M}\mathbf{R}\mathbf{M}). \tag{7}$$

Under the classic OLS assumption $\boldsymbol{\Gamma} = \gamma_0 \mathbf{I}$, (2) becomes

$$\mathbf{C}^{OLS} = \gamma_0 (\mathbf{X}'\mathbf{X})^{-1}. \tag{8}$$

In this case,

$$s^2 = \frac{s_0}{n-k} \tag{9}$$

is an unbiased estimator of γ_0 . Furthermore,

$$\hat{\mathbf{C}}^{OLS} = s^2 (\mathbf{X}'\mathbf{X})^{-1} \tag{10}$$

⁴ This definition follows Hayashi (2000: 408), Greene (2003: 268), Paolella (2003) and others, by taking the ratio of the sum of $n-j$ terms to that of n terms, to obtain what might be called the *weak* autocorrelations. The *strong* autocorrelations, defined as the ratio of the *average* value of $e_t e_{t-j}$ to the *average* value of e_j^2 , are larger by a factor of $n/(n-j)$. Percival (1993) notes that time series texts commonly refer to these as the “biased” and “unbiased” estimators, respectively, although both in fact can be biased. The Moment Ratio estimator could as easily be based on the strong autocorrelations, with identical results, but this would require adding the additional factor to the formulas.

is an unbiased estimator of \mathbf{C} . However, when, as is often the case, the errors and regressor(s) are both positively serially correlated, s^2 is no longer unbiased and $\hat{\mathbf{C}}^{OLS}$ will underestimate the variances of the $\hat{\beta}_j$.

If $\mathbf{\Gamma}$, or even \mathbf{R} , were known, Generalized Least Squares (GLS) would provide the efficient estimator of β , along with an unbiased estimate of its variance. Thus, for example, Canjels and Watson (1997) and Choi, Hu and Ogaki (2008) recommend using Feasible GLS (FGLS), based on an estimate of the covariance structure. However, this covariance estimate must be based on a preliminary OLS regression, and it is recommended here that the Moment Ratio covariance matrix described below be used for this purpose. Perron and Yabu (2009) note that even in the presence of a unit root, the OLS estimate of the slope of a trendline may have smaller variance than the first difference estimate of the growth rate in finite samples, if the first differences themselves exhibit positive serial correlation.

3. Stationary AR(1) errors

The serial correlation in econometric time series regressions is often approximately AR(1) in structure:

$$\varepsilon_t = \varphi \varepsilon_{t-1} + u_t, \quad (11)$$

where the innovations u_t are iid with mean 0 and finite variance σ^2 . Under (11), the population autocorrelations are

$$\rho_j = \varphi^j. \quad (12)$$

For the moment, we assume $|\varphi| < 1$. The unit root case $\varphi = 1$ requires reposing the problem somewhat, and is discussed in Section 5 below.

Figure 1 illustrates the autocovariance function for a simple trend line regression

$$y_t = \beta_1 + \beta_2 t + \varepsilon_t, \quad (13)$$

whose errors are an AR(1) process with $\varphi = 0.9$ and innovation variance $\sigma^2 = 1$. The blue line shows the true autocovariances $\gamma_0, \dots, \gamma_{n-1}$. These decay geometrically from $\gamma_0 = \sigma^2 / (1 - \varphi^2) = 5.26$ by factors of φ and are all positive. A pseudo-random draw from this process was constructed with Gaussian innovations, and was fit by OLS to a linear trend with an intercept. The irregular green line plots the ordinary estimated autocovariances⁵ of the residuals,

$$\hat{\gamma}_j = s_j / n.$$

These estimated autocovariances initially behave in a qualitatively similar fashion to the true autocovariances. However, they exhibit several typical problems when serial correlation is present: First, $\hat{\gamma}_0$ typically understates the true unconditional variance γ_0 . Second, they typically decay faster initially than the true rate φ . Third, as noted by Percival (1993),

⁵ These are “weak” autocovariances” corresponding to the “weak” autocorrelations in the sense defined in footnote 3.

$$\sum_{j=1}^{n-1} \hat{\gamma}_j = -\hat{\gamma}_0 / 2$$

whenever a constant term is included in the regression, so that the sample autocorrelations of order 1 and higher are necessarily negative on average even if the true autocorrelations are all positive. And fourth, when serial correlation is high they tend to exhibit the spurious oscillations that characterize trend line residuals with unit root or near unit root errors (Nelson and Kang 1981). Because of these qualitatively misleading properties, the highest order sample autocorrelations should be ignored altogether.

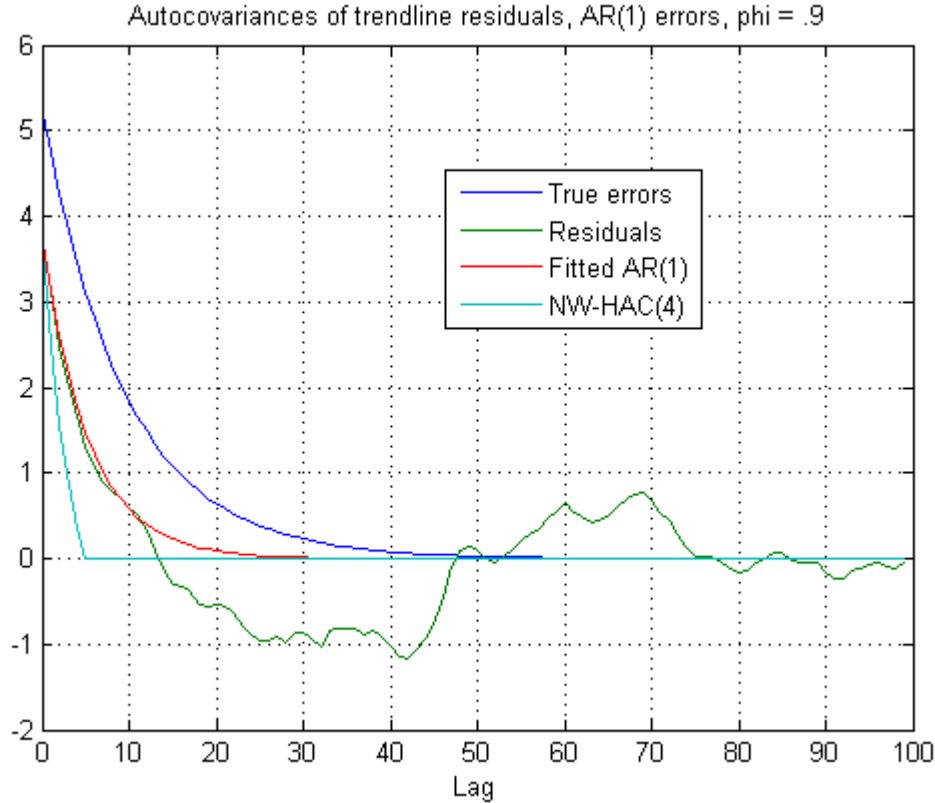


Figure 1

Illustrative simulated autocovariances of regression residuals (green), compared to true autocovariances of AR(1) errors (blue) for a trend line regression with sample size $n = 100$, true autoregressive parameter $\varphi = 0.9$, and unit innovation variance. Red line gives the autocovariances of an AR(1) process fit to the lag 0 and 1 autocovariances of the residuals. Cyan line shows the effective autocovariances employed by the Newey-West HAC estimator with bandwidth $m = 4$, as discussed in Section 4.

The red line in Fig. 1 is the autocovariance function of an AR(1) process fit by equating γ_0 and γ_1 directly to $\hat{\gamma}_0$ and $\hat{\gamma}_1$. This qualitatively captures the geometric shape of the true autocovariances, while it ignores the misleading behavior of the estimated

autocovariances at the long end. However, it starts with too low a value ($\hat{\gamma}_0 = 3.77$), and then decays at too fast a rate ($r_1 = 0.83$). Hence it will tend to give too small a value to the estimated standard errors of the regression parameters when substituted into equation (2).⁶

Figure 2 illustrates

$$E s^2 / \gamma_0 = \text{tr}_0(\mathbf{MRM}) / (n - k), \quad (14)$$

the expectation of the “unbiased” estimator s^2 relative to the true variance γ_0 of the regression errors, as a function of the true AR(1) parameter φ , again in the case of the simple trend line model with sample size $n = 100$. For $\varphi > 0$ this ratio is less than unity, and hence s^2 is downward biased, despite its “bias correction.”⁷ This bias depends on the observed regressor matrix \mathbf{X} by way of \mathbf{M} , but is also a function of the in-practice unknown parameter φ via the correlation matrix \mathbf{R} . As $\varphi = 1$ is approached, this bias actually becomes -100% , since γ_0 is infinite while the expectation of s^2 is necessarily finite when an intercept is included in the regression.

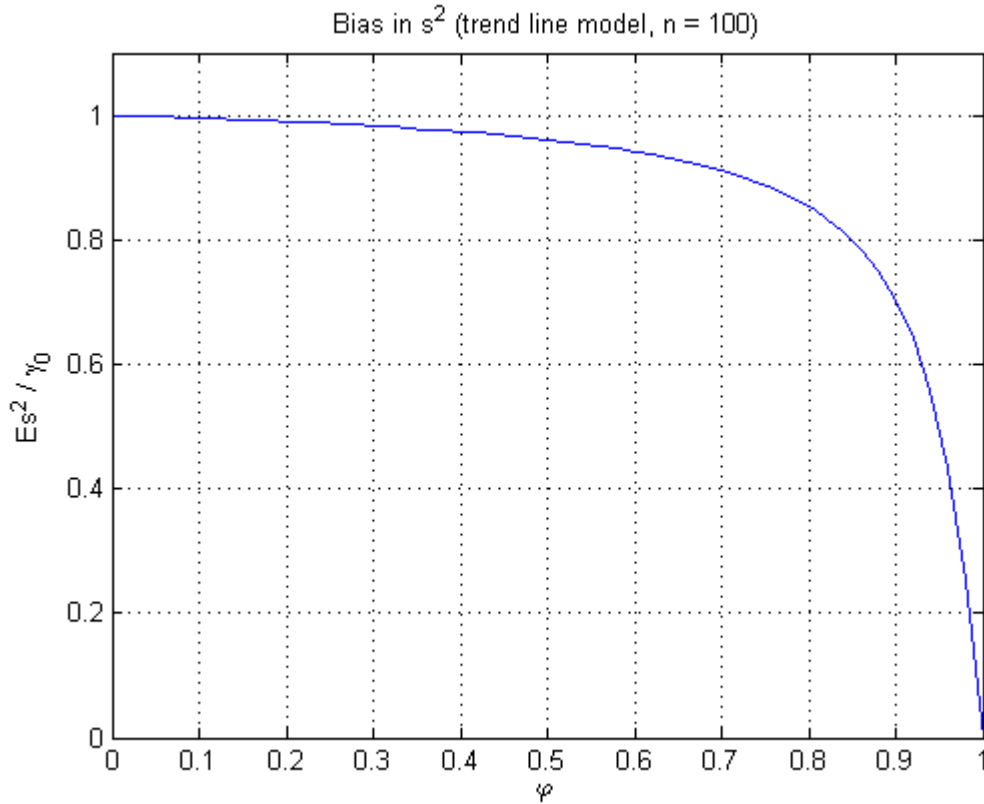


Figure 2

⁶ The cyan line in Figure 1 identified as HAC(4) is discussed in Section 4 below.

⁷ $s^2 = \hat{\gamma}_0 n / (n - k)$ is unbiased only in the absence of serial correlation.

Bias in s^2 relative to the true variance γ_0 of the regression errors, in a trend line regression with AR(1) errors, as a function of the true AR(1) parameter φ .

In order to correct these deficiencies in the AR(1) case, we define the *Moment Ratio function* for r_1 as the ratio of the population moments whose sample counterparts define r_1 per (5) and (7):

$$\psi(\varphi; \mathbf{X}) = \text{tr}_1(\mathbf{MRM}) / \text{tr}_0(\mathbf{MRM}). \quad (15)$$

The value of $\psi(\cdot)$ depends on φ through the correlation matrix \mathbf{R} , and also on \mathbf{X} via \mathbf{M} . Fortunately, however, it does *not* depend on the unknown coefficient vector $\boldsymbol{\beta}$ or the error variance γ_0 . Figure 3 illustrates this value as a function of φ , for the special case of a trend line regression with $n = 100$. A 45 degree line representing the true value of φ is also plotted.

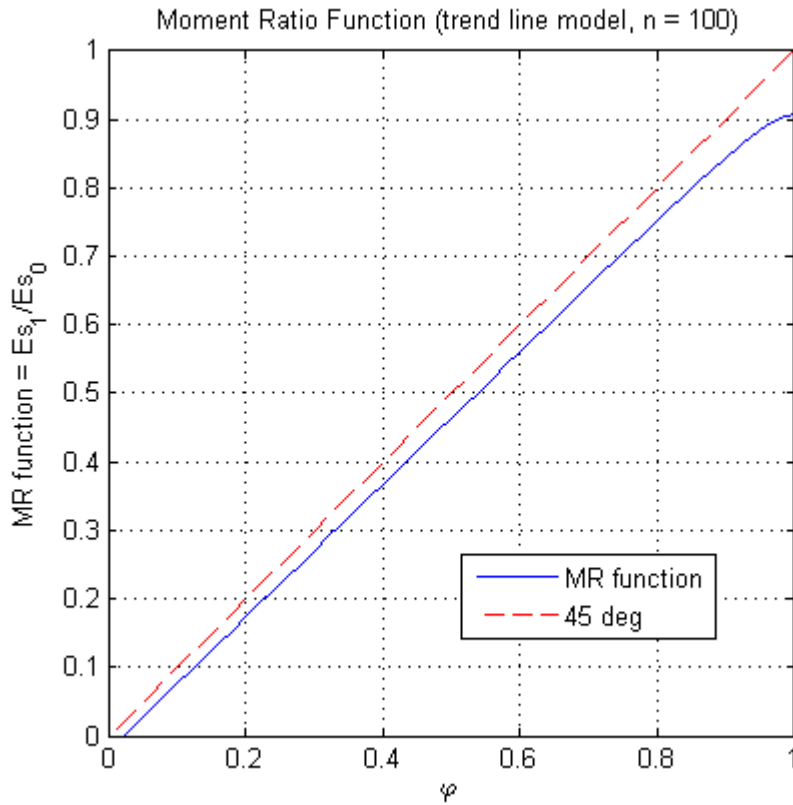


Figure 3

The Moment Ratio function for r_1 as a function of the true AR(1) autoregressive parameter φ for a trend line regression with sample size $n = 100$ (blue). This function lies entirely below the 45 degree line (dashed red).

It may be seen that r_1 already has a small downward bias (in the Moment Ratio sense) when $\varphi = 0$, and that this downward bias increases as φ increases to 1. The bias

may also be computed for $\varphi < 0$, but was found to vanish as $\varphi \downarrow -1$. Hence, only the more commonly encountered case $\varphi \geq 0$ is illustrated.

This bias in r_1 as an estimator of φ may be eliminated, simply by numerically evaluating the inverse of the function $\psi(\varphi; \mathbf{X})$, with respect to its first argument, at the empirical r_1 , when possible. If r_1 exceeds $\psi(1; \mathbf{X})$ (0.91 in the example), we instead use the point of nearest approach, namely $\varphi = 1$. Formally, the MR(1) estimator is defined by

$$\hat{\varphi}^{MR} = \arg \min_{\varphi \in (-1, 1]} |r_1 - \psi(\varphi; \mathbf{X})|. \quad (16)$$

This inverse function and its extension are illustrated in Figure 4 below.

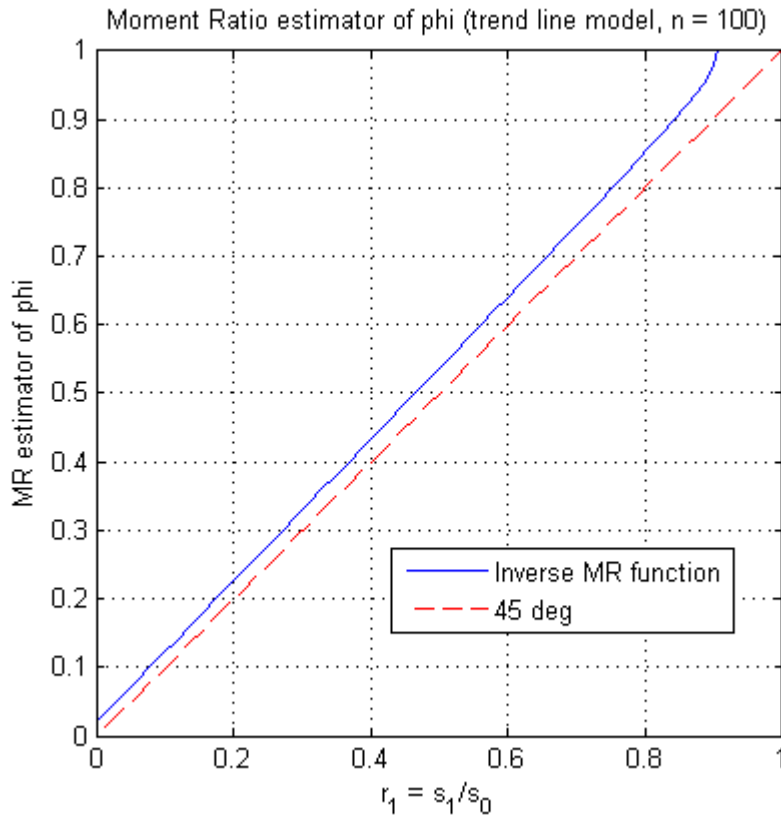


Figure 4

The inverse Moment Ratio function (blue) gives $\hat{\varphi}^{MR}$ as a function of the residual serial correlation coefficient r_1 . This inverse function lies entirely above the 45 degree line (dashed red).

Andrews (1993) has noted that a pattern similar to that in Figure 3 emerges when the median of the Monte Carlo distribution of the OLS estimator of φ is plotted against φ . Andrews proposes that an *Exactly Median Unbiased* estimator of φ be computed by numerically inverting this median function. In a precursor to the present paper

(McCulloch 2008), the author implements and extends Andrews' method to find exactly median unbiased estimators of all the coefficients of an AR(p) model for regression errors. However, the Moment Ratio approach of the present paper gives very similar results, without the tedious and noisy Monte Carlo simulation at each step of the numerical search.

Once $\hat{\varphi}^{MR}$ has been found, the correlation matrix \mathbf{R} may be estimated by

$$\hat{\mathbf{R}}^{MR} = \left((\hat{\varphi}^{MR})^{|i-j|} \right). \quad (17)$$

A natural estimator of γ_0 , that would be unbiased by (7) if it were computed with the true value of φ , would be

$$s_0 / \text{tr}_0(\mathbf{M}\hat{\mathbf{R}}^{MR}\mathbf{M}). \quad (18)$$

However, this does not lead to a consistent estimate of the variance of the innovations in the limit as φ approaches unity, as discussed in section 5. This problem can nevertheless be avoided by instead basing the estimator of γ_0 on the *innovations*:

$$u_t = \varepsilon_t - \varphi \varepsilon_{t-1}, \quad t = 2, \dots, n$$

or

$$\mathbf{u} = \mathbf{D}_\varphi \boldsymbol{\varepsilon}$$

where the $(n-1) \times n$ quasi-differencing operator $\mathbf{D}_\varphi = (d_{ij})$ is defined by $d_{i,i+1} = 1$, $d_{i,i} = -\varphi$, and $d_{i,j} = 0$ otherwise. The consistently estimated last $n-1$ innovations are then

$$\hat{\mathbf{u}} = \mathbf{D}_\varphi \mathbf{e} = \mathbf{D}_\varphi \mathbf{M}\boldsymbol{\varepsilon}$$

so that

$$\mathbf{E} \hat{\mathbf{u}} \hat{\mathbf{u}}' = \mathbf{D}_\varphi \mathbf{M} \Gamma \mathbf{M} \mathbf{D}_\varphi' = \sigma^2 \mathbf{D}_\varphi \mathbf{M} \mathbf{G} \mathbf{M} \mathbf{D}_\varphi', \quad (19)$$

where $\mathbf{G} = (g_{|i-j|}) = \Gamma / \sigma^2 = g_0 \mathbf{R}$ is the covariance matrix of the errors, normalized to unit variance for the innovations. In the AR(1) case, $g_0 = 1/(1 - \varphi^2)$. Then

$$\mathbf{E} \hat{\mathbf{u}}' \hat{\mathbf{u}} = \sigma^2 \text{tr}_0(\mathbf{D}_\varphi \mathbf{M} \mathbf{G} \mathbf{M} \mathbf{D}_\varphi'),$$

so that

$$\hat{\sigma}_{MR}^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / \text{tr}_0(\mathbf{D}_\varphi \mathbf{M} \mathbf{G} \mathbf{M} \mathbf{D}_\varphi')$$

would be an unbiased and consistent estimator of the innovation variance if φ were known. When φ is estimated consistently by $\hat{\varphi}^{MR}$ and \mathbf{G} by

$$\begin{aligned} \hat{\mathbf{G}}^{MR} &= \hat{\mathbf{R}}^{MR} / (1 - (\hat{\varphi}^{MR})^2), \\ \hat{\sigma}_{MR}^2 &= \hat{\mathbf{u}}' \hat{\mathbf{u}} / \text{tr}_0(\mathbf{D}_{\hat{\varphi}^{MR}} \mathbf{M} \hat{\mathbf{G}}^{MR} \mathbf{M} \mathbf{D}_{\hat{\varphi}^{MR}}') \end{aligned} \quad (20)$$

becomes a consistent and at least approximately unbiased estimator. The Moment Ratio estimator of the covariance matrix \mathbf{C} of $\hat{\boldsymbol{\beta}}^{OLS}$ as given in (2) is then

$$\hat{\mathbf{C}}^{MR} = \hat{\sigma}_{MR}^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \hat{\mathbf{G}}^{MR} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1}. \quad (21)$$

If desired, the variance of the AR(1) errors may then be estimated by

$$\hat{\gamma}_0^{MR} = \hat{\sigma}_{MR}^2 / (1 - (\hat{\varphi}^{MR})^2). \quad (22)$$

4. The Newey-West HAC covariance matrix

The truncated-kernel Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix, introduced by Newey and West (1987), is now widely used by economists to “correct” the standard errors of OLS time series coefficients for serial correlation. Greene (2003: 201) reports that its use is now “standard in the econometrics literature.” Hayashi (2000: 409-12) mentions only it, the similar Quadratic Spectral HAC of Andrews and Monahan (1992), and the Vector Autoregression HAC (VAR-HAC) method of den Haan and Levin (1996) as appropriate methods for correcting OLS standard errors for serial correlation. The elementary text by Stock and Watson (2007) presents HAC as the *only* method worthy of consideration.

In the OLS case considered here, the Newey-West HAC estimator of C simplifies to

$$\hat{C}^{HAC} = (X'X)^{-1} X'FX(X'X)^{-1}, \quad (23)$$

where

$$F = (e_i e_j' K(|i - j|)) \quad (24)$$

and

$$K(l) = \max((m + 1 - l)/(m + 1), 0) \quad (25)$$

is the *truncated Bartlett Kernel function* at lag l for some *bandwidth* m . (Greene 2003: 200) Most econometric packages provide “automatic bandwidth selection” for HAC, using a formula such as the following:

$$m = \lfloor 4(n/100)^{2/9} \rfloor, \quad (26)$$

which just yields $m = 4$ (corresponding to 4 non-zero autocorrelation terms used) for $n = 100$.⁸

At least three types of heteroskedasticity are commonly encountered in econometrics: In *a priori* heteroskedasticity, the variance of the errors is known up to a common constant, and Weighted Least Squares (WLS) provides efficient coefficient estimates and correct standard errors. In *serially conditional* heteroskedasticity, episodes of high and low variance errors are clustered, in which case a GARCH model is often appropriate (McCulloch 1985, Bollerslev 1986). And finally, in what might be called *Regressor Conditional Heteroskedasticity* (RCH), the variance of the errors may depend in an unknown way on the equation’s regressors. The White (1980) HCC and Newey-West HAC estimators correct consistently for RCH, though not necessarily for the other two forms of heteroskedasticity unless they coincidentally coincide.

In the benchmark case of regressor-conditional homoskedasticity, the expectation of the HAC estimator becomes

⁸ EViews uses Equation (26), following a suggestion of Newey and West (1994). Stock and Watson (2007:607) suggest $m = \lfloor 0.75 n^{1/3} - 0.5 \rfloor$, which has very similar effect for n in the range 50 – 2000, and generally identical effect in the range 400 – 1000. Sun et al. (2008) note that while a power of 1/3 or 1/5 minimizes the asymptotic mean squared error of the long-run variance estimator for first and second order kernels, respectively, a power of 1/2 or 1/3 asymptotically minimizes a weighted average of Type I and Type II errors. The notation $\lfloor x \rfloor$ indicates the *floor function* of x .

$$E \hat{\mathbf{C}}^{HAC} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{H}^{HAC} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad (27)$$

where

$$\mathbf{H}^{HAC} = (h_{i,j} K(|i-j|)), \quad \text{and } \mathbf{H} = (h_{i,j}) = \mathbf{M}\mathbf{\Gamma}\mathbf{M}.$$

HAC thus effectively employs only the first m sample autocovariances, and replaces the others with zeros. At the same time it down-weights the autocovariances it does not discard entirely by the factor $(m+1-l)/(m+1)$. It also uses the regression residuals as if they were the errors themselves. For all three of these reasons, it tends in finite samples to underestimate the standard error for a regressor which is itself serially correlated. However, the amount by which it does this depends on both the ρ_j and the degree of serial correlation of the regressors themselves.

Aside from its RCH adjustment, the NW HAC estimator in effect therefore replaces the already twice-downward-biased $\hat{\gamma}_j^{OLS}$ with the triply-biased $\hat{\gamma}_j^{OLS} K(j)$, as depicted by the cyan line in Figure 1 above. Although NW provides some improvement over the OLS standard errors, the already seriously deficient AR(1) standard errors are far superior.⁹

As the sample size n rises without bound, the bandwidth determined by (26) also rises without bound, so that eventually every lag is included and $K(l)$ becomes arbitrarily close to unity for each l . Furthermore, as long as $n^{1/2}/m$ also rises to infinity, each autocorrelation is consistently estimated (Newey and West 1994: 633). However, in finite samples the estimator can be highly misleading.

5. The Unit Root Case $\varphi = 1$

The unit root case $\varphi = 1$ poses no insurmountable problems, so long as there is a constant term and a trend or trending variable(s) in the original regression (1), and so long as the regression residuals are not used to estimate the variance of the innovations. In particular, a unit root in the errors does not in itself indicate a “spurious regression” in which OLS coefficient slope estimates become meaningless. However, it does require reposing the problem, so as to replace certain undefined mathematical expressions with their limiting values. We assume that $\varphi \in (-1, 1]$, so that a unit root is the only type of nonstationarity that will be encountered.

As is well known, OLS coefficient estimates are inconsistent when both the regressor(s) and the errors contain a unit root with no drift. (See e.g. Choi, Hu and Ogaki, 2008: 330). Conditional on the observed regressor(s) and the finite innovation variance, the limiting coefficient error is Gaussian with positive variance. Even so, there may be a problem of estimating the innovation variance from even the estimated innovations if the residuals themselves are not consistently estimated.

⁹ Politis (2011) finds that a flat-top kernel and a data-driven bandwidth somewhat alleviate the downward bias in the Newey-West estimator. In order to ensure that the resulting covariance matrix is positive semidefinite, he must rectify it by zeroing out any negative eigenvalues in its singular value decomposition. Nevertheless, he uses residuals as if they were errors, with no compensation for their biased variance..

Nevertheless, when the regression includes a constant and a time trend, the slope coefficient becomes \sqrt{n} consistent: Hayashi (2000: 570) shows that after scaling the horizontal axis by n and the vertical axis by \sqrt{n} , the slope coefficient is Gaussian about its true value with a finite variance. Without the scaling, the slope coefficient is therefore Gaussian with standard error proportional to $1/\sqrt{n}$. Likewise, if the regressor is unit root with non-zero drift, the drift will dominate the unit root noise for large n , and so the regressor will act as if it were a time trend. This section is therefore limited to the case in which the regressors include either a time trend or a trending variable.

As $\varphi \uparrow 1$, each element of the unconditional covariance matrix $\mathbf{\Gamma}$ becomes infinite, holding the variance σ^2 of the innovations in (11) constant. Furthermore, each element of the correlation matrix \mathbf{R} becomes unity in this limit. These matrices are therefore no longer useful or informative in this case, and the problem must be reposed without them.

Although a random walk has infinite *unconditional* variance and covariances, its variances and covariances are all finite *conditional on its value at any point in time*, say $t = 0$. If the errors $\boldsymbol{\varepsilon}$ are a random walk that arise by accumulating white noise \mathbf{u} , we have

$$\boldsymbol{\varepsilon} = \varepsilon_0 + \mathbf{N}\mathbf{u}$$

where

$$\mathbf{N} = (\mathbf{1}(i \leq j)) \quad (28)$$

is the $n \times n$ integrator matrix, and the logical indicator $\mathbf{1}(i \leq j)$ takes the value 1 when the expression is true and 0 when it is false. The *conditional* covariance matrix of the errors, taken conditional on ε_0 , is then

$$\mathbf{\Gamma}_0 = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | \varepsilon_0) = \sigma^2 \mathbf{W}_0,$$

where

$$\mathbf{W}_0 = \mathbf{N}\mathbf{N}' = (\min(i, j)).$$

Furthermore, so long as there is a constant term in the regression, the OLS residuals will sum to zero regardless of the value of the random walk at $t = 0$, so that the actual value of ε_0 does not matter for their properties. Equation (6) then becomes

$$E\mathbf{e}\mathbf{e}' = E(\mathbf{e}\mathbf{e}' | \varepsilon_0) = \mathbf{M}\mathbf{\Gamma}_0\mathbf{M} = \sigma^2 \mathbf{M}\mathbf{W}_0\mathbf{M}$$

so that

$$E s_j = \sigma^2 \text{tr}_j(\mathbf{M}\mathbf{W}_0\mathbf{M}). \quad (29)$$

The limiting value of $\psi(\varphi; \mathbf{X})$, as plotted in Figure 2, then becomes

$$\psi(1; \mathbf{X}) = \text{tr}_1(\mathbf{M}\mathbf{W}_0\mathbf{M}) / \text{tr}_0(\mathbf{M}\mathbf{W}_0\mathbf{M}). \quad (30)$$

Due to sampling error, the actual value of r_1 could lie above or below this value. Any value of r_1 above $\psi(1; \mathbf{X})$ should simply be identified with $\hat{\varphi}^{MR} = 1$ as in (16) and Figure 4.

The innovation variance σ^2 could be estimated without bias, using (29), by

$$s_0 / \text{tr}_0(\mathbf{M}\mathbf{W}_0\mathbf{M}).$$

Unfortunately, however, this unbiased estimator, which is the limiting value of (19) taken together with (22), is not consistent, even in a trend line regression (see, e.g. Hayashi 2000:570-71). Simulations with Gaussian errors indicate that for a model in which only a constant term is estimated, its distribution is approximately a scaled χ^2 with 2.5 degrees of freedom, using either $n = 100$ or $n = 1000$.¹⁰ In a linear trend line model, it is approximately scaled χ^2 with 5 degrees of freedom, using either $n = 100$ or $n = 1000$.

However, using the estimated innovations (which are in fact simply the first-differenced residuals in the random walk case being discussed here), equation (19) becomes

$$\mathbf{E} \mathbf{u} \mathbf{u}' = \mathbf{E}(\mathbf{u} \mathbf{u}' | \varepsilon_0) = \sigma^2 \mathbf{D}_1 \mathbf{M} \mathbf{W}_0 \mathbf{M} \mathbf{D}_1',$$

whence

$$\mathbf{E} \mathbf{u}' \mathbf{u} = \mathbf{E}(\mathbf{u}' \mathbf{u} | \varepsilon_0) = \sigma^2 \text{tr}_0(\mathbf{D}_1 \mathbf{M} \mathbf{W}_0 \mathbf{M} \mathbf{D}_1'),$$

so that

$$\hat{\sigma}_{MR}^2 = \mathbf{u}' \mathbf{u} / \text{tr}_0(\mathbf{D}_1 \mathbf{M} \mathbf{W}_0 \mathbf{M} \mathbf{D}_1') \quad (31)$$

is an unbiased estimator when it is known that there is a unit root. Furthermore, simulations with Gaussian errors indicate that the distribution of this estimator is approximately scaled χ^2 with $n-k$ degrees of freedom, for both the mean and trend line models, and for both $n = 100$ and 1000 . Inference on the coefficients with t or F statistics based on (31) and (32) should therefore be at least approximately valid.

Having computed $\hat{\sigma}_{MR}^2$, we may estimate the covariance of the regression coefficients, conditional on the arbitrarily chosen reference point ε_0 , by

$$\hat{\mathbf{C}}_0^{MR} = \hat{\sigma}_{MR}^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}_0 \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1}. \quad (32)$$

If a reference point other than ε_0 , say ε_t , is chosen, the conditional covariance of the errors becomes

$$\mathbf{\Gamma}_t = \mathbf{E}(\mathbf{\varepsilon} \mathbf{\varepsilon}' | \varepsilon_t) = \sigma^2 \mathbf{W}_t,$$

where

$$\mathbf{W}_t = (w_{i,j}^t),$$

with

$$w_{i,j}^t = \begin{cases} \min(|i-t|, |j-t|), & (i > t \text{ and } j > t) \text{ or } (i < t \text{ and } j < t) \\ 0, & \text{else.} \end{cases}$$

Since $\mathbf{M} \mathbf{W}_t \mathbf{M}$ does not depend on the choice of t , exactly the same values of $\psi(1; \mathbf{X})$ and $\hat{\sigma}_{MR}^2$ will be obtained. Furthermore, except for its first row and column,

$$\hat{\mathbf{C}}_t^{MR} = \hat{\sigma}_{MR}^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}_t \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1}$$

does not depend on the choice of t . The estimated standard error of the constant term β_1 therefore does depend on the reference point defined by the choice of t , while those of the

¹⁰ Although s_0 can be expressed as a linear combination of squared $N(0,1)$ random variables, the weights are unequal so that its distribution is not exactly χ^2 . Nevertheless, the relation between the distribution's simulated mean and variance is the same as for the χ^2 with the indicated degrees of freedom.

slope coefficients β_2, \dots, β_k do not. In particular, the regression F statistic for the joint hypothesis $\beta_2 = 0, \dots, \beta_k = 0$ is at least approximately valid despite the unit root in the errors, and is invariant to the arbitrarily chosen reference point.

If the one-step-ahead forecast of y_{n+1} (conditional on the time $t = n+1$ values of the regressors) is of particular interest, the reference point $t = n+1$ may be useful. In this case,

$$\mathbf{W}_{n+1} = (\min(n+1-i, n+1-j)).$$

Another particularly natural reference point, one that does not single out any individual point in the sample period, is the mean error

$$\bar{\varepsilon} = \mathbf{1}'\boldsymbol{\varepsilon} / n,$$

where $\mathbf{1} = (1, \dots, 1)'$. Conditioning on this value, we have

$$\hat{\mathbf{C}}_{\bar{\varepsilon}}^{MR} = \hat{\sigma}_{MR}^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_{\bar{\varepsilon}} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where

$$\mathbf{W}_{\bar{\varepsilon}} = \mathbf{Z}\mathbf{W}_0\mathbf{Z}$$

and

$$\mathbf{Z} = \mathbf{I}_n - \mathbf{1}\mathbf{1}'/n. \quad (33)$$

The unconditional unit root covariance matrix $\hat{\mathbf{C}}^{UR}$ may be obtained from $\hat{\mathbf{C}}_0^{MR}$ or any of the other conditional unit root covariance matrices developed above, simply by replacing its (1,1) element by ∞ . With no loss of generality, the remaining elements of its first row and column may be set to 0, since the variance of any linear combination of the intercept and the regressors is infinite regardless of these covariances. The inverse of the unconditional unit root covariance matrix will be of use below when we compute a blended covariance matrix in the near-unit-root cases. It may be seen from the partitioned matrix inversion rule that the inverse of $\hat{\mathbf{C}}^{UR}$ is block-diagonal, with 0 in the (1,1) position, regardless of the covariances in the first row and column, so that again there is no loss of generality in setting them to 0. $\hat{\mathbf{C}}^{UR}$ may be similarly constructed from $\hat{\mathbf{C}}_0^{MR}$ in the MR(p) case considered below.

6. Monte Carlo Properties of Moment Ratio Estimator in AR(1) case

Figure 5 below shows the Moment Ratio Function $\psi(\varphi, \mathbf{X})$, along with the Monte Carlo mean, median, quartiles, and 5th and 95th percentiles of the distribution of r_1 , as a function of the true value of φ , again for the trend line model with $n = 100$, under the additional assumption of Gaussian errors.¹² The median line is essentially the same as the median function inverted by Andrews (1993) to determine his Median Unbiased

¹¹ The unit root covariance matrix could also be conditioned on ε_t for any t outside $[0, n+1]$, but this complicates the formula somewhat, and so is not treated here.

¹² This simulation was performed with $m = 99,999$ replications using MATLAB's `randn('state')` Gaussian random number generator, for φ in steps of 0.01. The same seed was used for each value of φ in order to make each simulated percentile a smooth function of φ .

estimator of the autoregressive parameter φ .¹³ The Moment Ratio Function is almost indistinguishable from the median for values of φ near 0, but rises slightly above the median as φ increases. Nevertheless, it remains well within the quartiles, and at least approximately corrects the median bias of r_1 when inverted. The downward skewness of the distribution of r_1 pulls the mean of the distribution down below the median by a comparable amount.

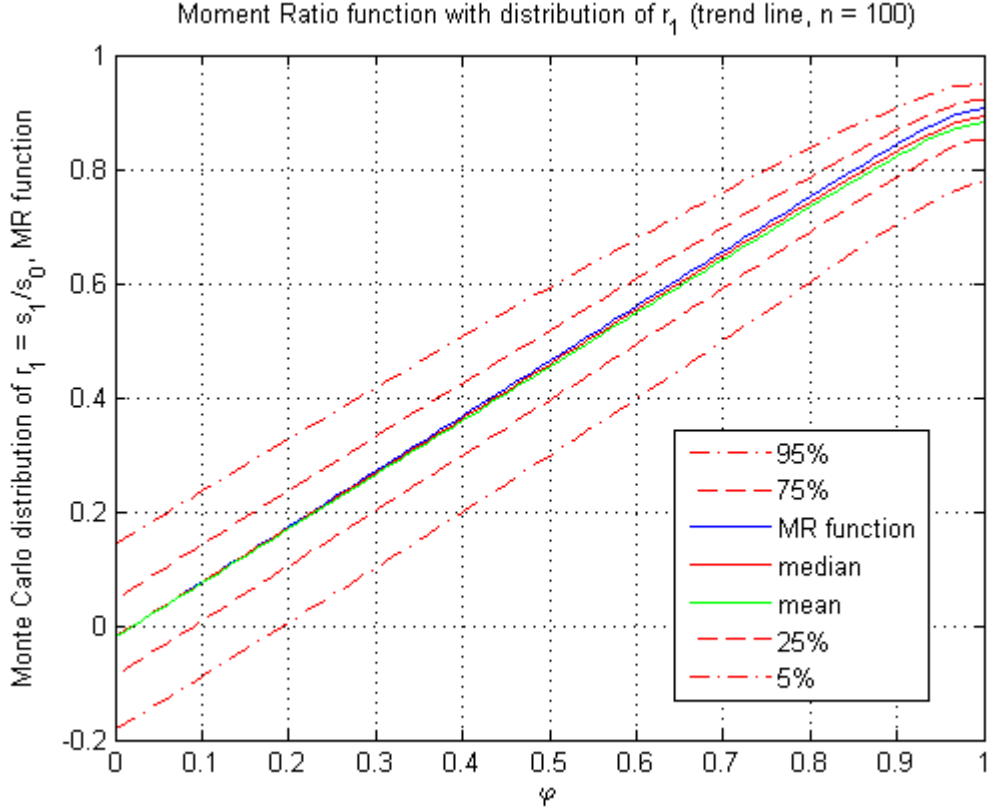


Figure 5

Monte Carlo distribution of the residual autocorrelation r_1 for a trend line regression with sample size $n = 100$, as a function of the true AR(1) parameter φ . The median is shown in solid red, the quartiles in dashed red, the 5% and 95% quantiles in dash-dot red, and the mean in solid green. The solid blue line shows the Moment Ratio function from Fig. 1 for comparison.

Because the Moment Ratio function lies above the median in Figure 5 for $\varphi > 0$, and both are increasing functions of φ , $\hat{\varphi}^{MR}$ is necessarily less than the Median Unbiased estimator computed by inverting the Monte Carlo median function at r_1 , and therefore

¹³ Andrews in fact bases his estimator on an OLS regression of y_t on y_{t-1} , a constant, and t , rather than on the first order autocorrelation of the residuals of a regression of y_t on a constant and t . The two approaches are nevertheless very similar. As demonstrated in McCulloch (2008), the latter approach is easily applied to the residuals of a general OLS regression with higher-order autoregressive errors.

provides a less conservative adjustment for serial correlation, at least for the trend line problem simulated.

Figures 6 through 17 are based on 10,000 Monte Carlo replications of a trend line regression with $n = 100$ equally spaced and demeaned values of the time trend variable, Gaussian AR(1) errors, and various true values of φ ranging from -0.2 to 1.0, inclusive. The “AR(1)” model is fit to the (weak) sample autocorrelation r_1 , in conjunction with an estimate of γ_0 that is unbiased conditional on $\varphi = r_1$ per (7). The “HAC” results are based on a bandwidth of $m = 4$, using formula (26).

Figure 6 shows the Monte Carlo median of the squared standard error of the slope coefficient, relative to the true variance of the coefficient, so that a value of unity reflects the target of zero median bias. As expected, OLS is right on the mark when the errors are uncorrelated, but then deteriorates rapidly as the degree of serial correlation increases. With random walk errors, the true variance of the slope coefficient still finite, but is 180 times larger than the OLS estimate, because OLS takes no account at all of serial correlation.

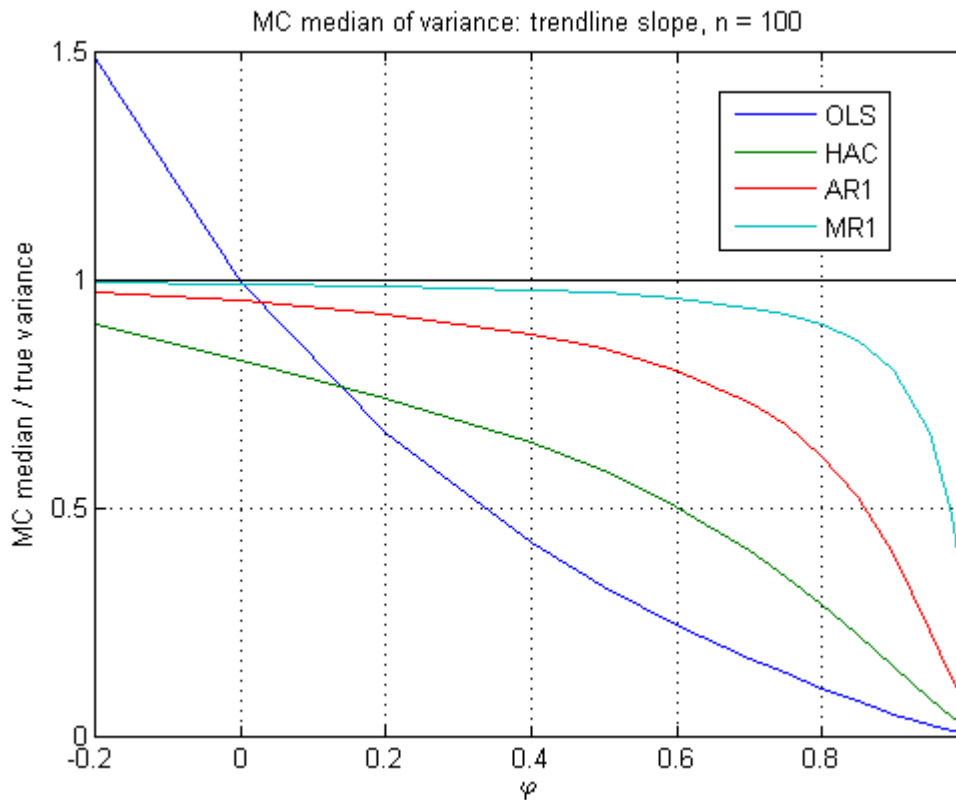


Figure 6
Monte Carlo median of estimated variance of slope coefficient, relative to true variance, versus true AR(1) parameter of error DGP, for a trend line regression with $n = 100$. OLS is shown in blue, NW-HAC in green, conventional AR(1) in red, and MR(1) in cyan.

NW HAC provides some improvement over OLS for φ greater than approximately 0.13. However, it already has substantial negative median bias (median/true = 0.90) even when the errors are white noise. This is because HAC takes no account of the distinction, important in finite samples, between the residuals and the errors. However, the AR(1) model that fits r_1 directly to its asymptotic counterpart is a big improvement over HAC.

Finally, the Moment Ratio estimator MR(1) is clearly the leader of the pack by this metric, except for OLS when φ is very near 0. Because of the upper bound imposed on $\hat{\varphi}^{MR}$, even MR(1) is inevitably somewhat biased downwards as φ approaches and reaches 1.

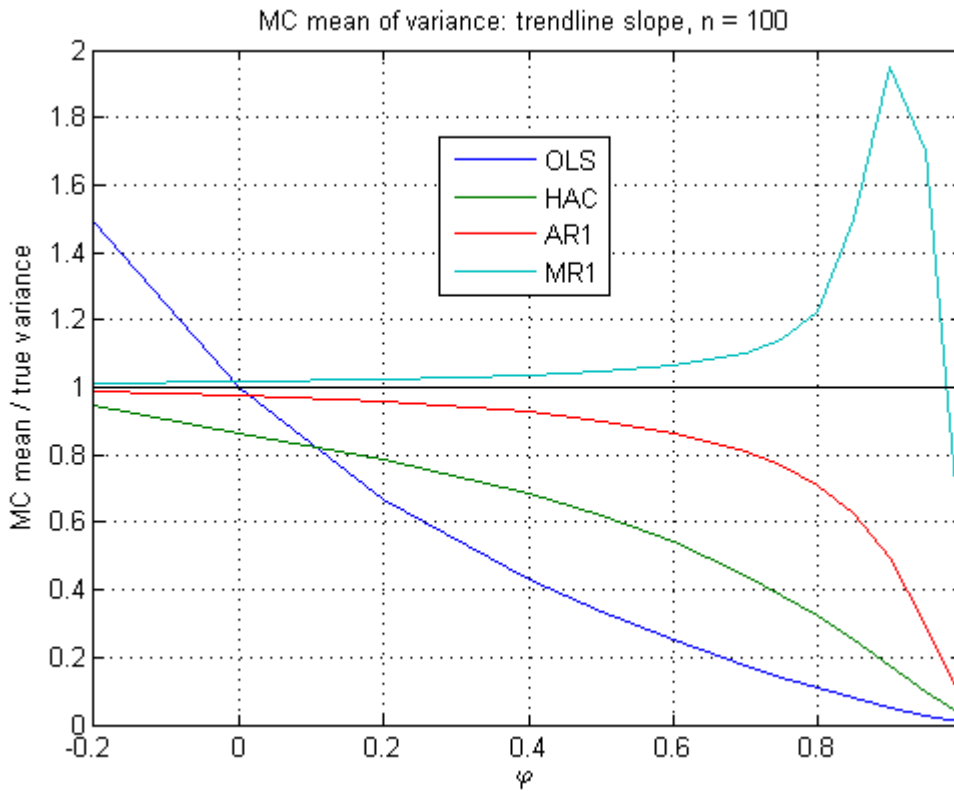


Figure 7

Monte Carlo mean of estimated variance of slope coefficient, relative to true variance, versus true AR(1) parameter of error DGP, for a trend line regression with $n = 100$. OLS is shown in blue, NW-HAC in green, conventional AR(1) in red, and MR(1) in cyan.

Figure 7 is similar to Figure 6, except that it gives the Monte Carlo *means* of the squared standard errors, relative to the true variance of the slope coefficient, rather than the medians. OLS, HAC and AR(1) all have mean biases qualitatively similar to the

median biases shown in Figure 6. However, MR(1) now has a *positive* mean bias for all but the highest values of φ . For $\varphi < 0.6$, this bias is quite small, but for larger values it grows quickly, reaching a peak of +95% at $\varphi = 0.95$ (corresponding to a RMSE of +40% for the standard error itself). This mean bias is due to the fact that the expected squared standard error is a highly convex function of φ , especially as $\varphi = 1$ is approached, so that overestimates of φ have a much bigger impact on the mean than do underestimates. The medians in Figure 6 are less affected by this factor, resulting in negative median biases throughout. For $\varphi > 0.95$, the upper bound $\hat{\varphi}^{MR} \leq 1$ becomes binding more frequently, with the result that under-estimates start to dominate, and the bias turns negative after 0.98. At $\varphi = 1$, the mean variance bias is -49% , corresponding to -30% in terms of the RMS standard error.

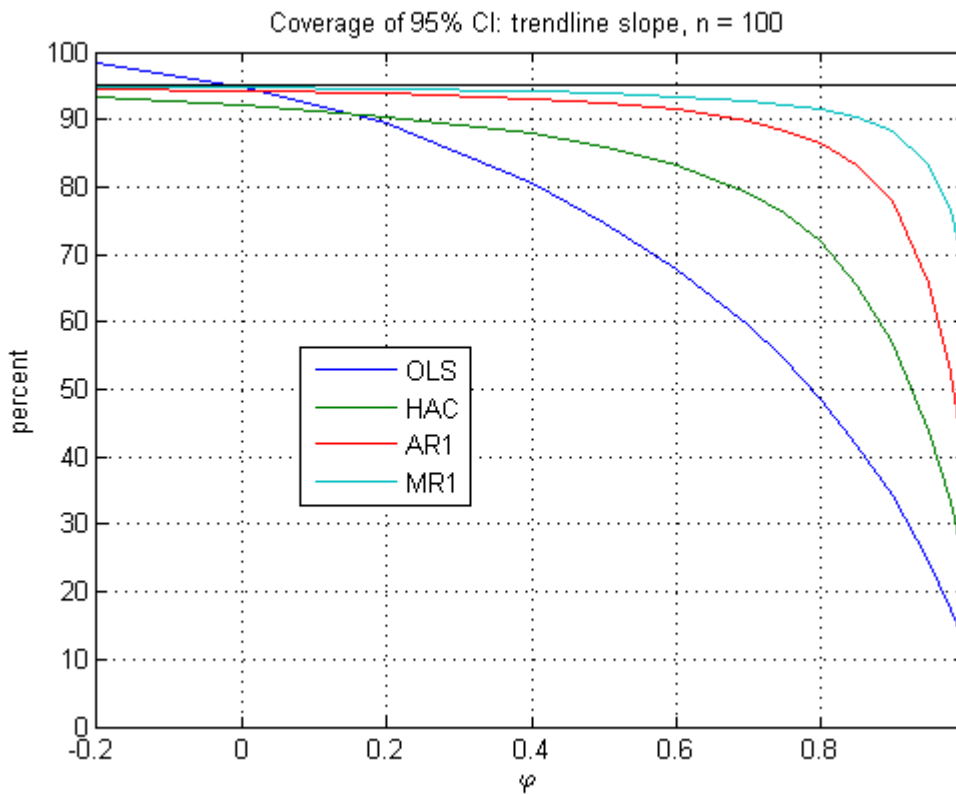


Figure 8

Monte Carlo coverage of 95% confidence interval for slope coefficient in a trend line regression with $n = 100$. OLS is shown in blue, NW-HAC in green, conventional AR(1) in red, and MR(1) in cyan.

For most practical purposes, the ultimately relevant criterion is not the mean or median bias of the corrected standard errors, but rather their ability to construct confidence intervals, and therefore hypothesis tests, with correct coverage.¹⁴ Figures 8

¹⁴ Peron and Yabu (2009) similarly evaluate the asymptotic behavior of estimators in terms of a weighted average of Type I and Type II errors, rather than the mean squared error of the estimated long-run variance itself.

and 9 below shows the simulated coverage of a 95% confidence interval for the slope in the same trend line regression with $n = 100$ and Gaussian innovations. Since 10,000 replications were again used, the sampling standard error of the coverage of a 95% confidence interval is 0.22%. The intercept and slope coefficients were, without loss of generality, both set to 0 for this simulation. Figure 8 shows the large-scale behavior as $\varphi \uparrow 1$, while Figure 9 is a detail of the same curves, allowing comparisons for φ near 0. In both figures, the horizontal black line at 95% is the target representing perfect coverage.

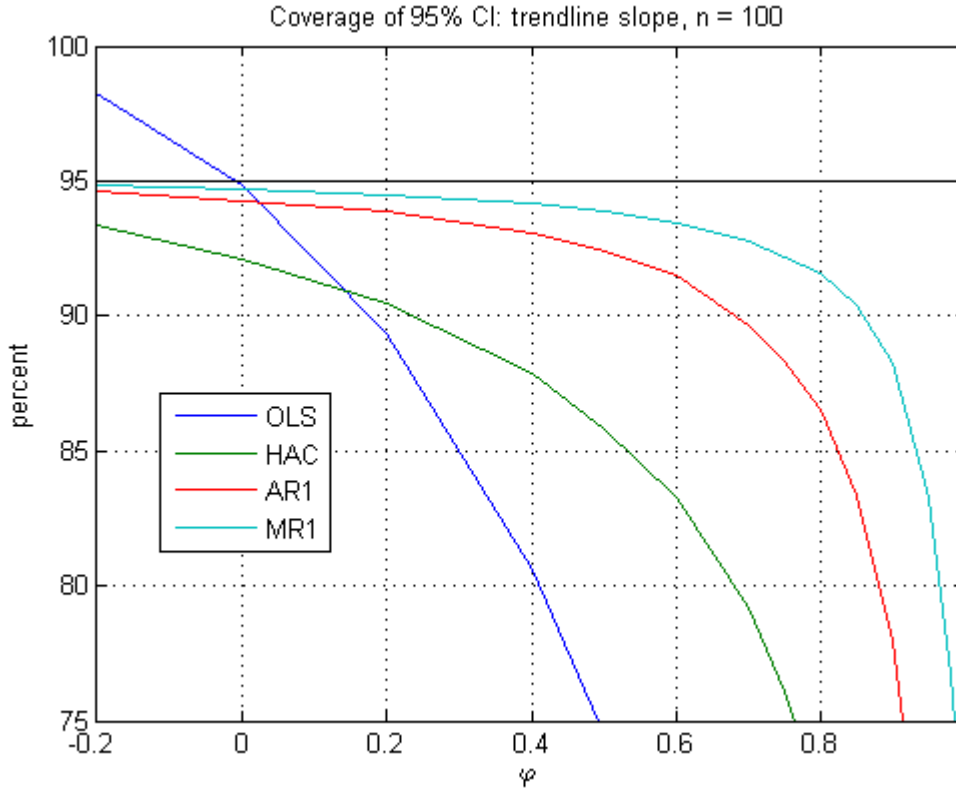


Figure 9
Detail of Figure 8.

Despite its upward mean bias for the slope variance, the MR(1) size distortion is always negative and becomes increasingly negative throughout the range $[-.2, 1]$. Although it does not have perfect coverage, the MR(1) estimator always has better performance than HAC or AR(1), and almost always dominates OLS, by this metric.

The pattern for the estimated intercept is qualitatively similar to that shown for the estimated slope in Figures 6 – 9, and hence is not shown here. However, since the autocorrelation of the errors interacts with the autocorrelation of the regressor(s), and a constant term's unit regressor is even more serially correlated than a trend line regressor, the biases and size distortions are slightly worse for the intercept than those depicted for the trend slope. A less serially correlated regressor would have less bias and size distortion.

7. AR(1) Unit Root Test

The Monte Carlo distribution of r_1 for $\varphi = 1$, summarized at the right edge of Figure 5 above, provides a simple test for a unit root, under the assumption of Gaussian errors, that is exact to within Monte Carlo sampling error: If r_1 is less than say the 5th percentile of this distribution (0.778 for the illustrated trend line regression with $n = 100$), a unit root can be rejected with a 5% one-tailed test size. The only practical way to perform this test is with a Monte Carlo simulation comparable to that required for Andrews' (1993) Median Unbiased estimator, but since the test is optional, and even then the simulation only needs to be performed under the null $\varphi = 1$, it is not nearly as computationally demanding as the Median Unbiased estimator.

8. Higher Order Autoregressive/Unit Root Models and Consistent Covariance Estimation

Although an AR(1) model is often a good first approximation to the autocovariation function of econometric time series regression errors, this model is unnecessarily restrictive. The true model may not even be a finite order autoregressive process. However, a finite order AR(p) model

$$\varepsilon_t = \sum_{j=1}^p \varphi_j \varepsilon_{t-j} + u_t \quad (34)$$

can approximate most covariance-stationary Gaussian processes to any desired precision, given a sufficiently high value of p .¹⁵

As in the AR(1) model, the unit root case is not an insurmountable obstacle. However, it does require separate treatment. Following Anderson and Chen (1993), it is useful to rewrite (34) in the equivalent Augmented Dickey-Fuller (ADF) form

$$\varepsilon_t = \alpha \varepsilon_{t-1} + \sum_{j=1}^{p-1} \theta_j (\varepsilon_{t-j} - \varepsilon_{t-j-1}) + u_t,$$

where

$$\begin{aligned} \alpha &= \sum_{j=1}^p \varphi_j, \\ \theta_j &= - \sum_{h=j+1}^p \varphi_h, \quad j = 1, \dots, p-1. \end{aligned} \quad (35)$$

We will call the parameter α the (*first order*) *persistence* of the process.

¹⁵ One notable exception is a non-invertible MA process. However, such a process would be only rarely encountered in practice, except as an inappropriately first differenced white noise. Long memory fractionally integrated processes (Hosking 1984) may require much higher values of p than will short-memory processes, and may require special consideration. These interesting processes go beyond the scope of this paper.

A necessary condition for stationarity is $\alpha \leq 1$. Although $\alpha < 1$ is not a sufficient condition for stationarity, the case $\alpha = 1$ is the most common form of nonstationarity for the errors in econometric time series regressions. Hence, we will assume that $\alpha = 1$ is the only type of nonstationarity that may be present, and that when such a unit root is present, the θ_j define a stationary process for the first differences. In other words, we assume that the process is either stationary or integrated of order 1. More formally, we assume that the vector $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_p)'$ of AR(p) coefficients is an element of the set Φ for which either all the reciprocal roots of the characteristic polynomial are inside the unit circle, or for which only one root is on the unit circle, that root being 1 and the other reciprocal roots lying inside the unit circle.

If an AR(p) process is covariance stationary, the Yule-Walker equations

$$\begin{aligned}\gamma_0 &= \sum_{j=1}^p \sum_{k=1}^p \varphi_j \varphi_k \gamma_{|j-k|} + \sigma^2, \\ \gamma_j &= \sum_{k=1}^p \varphi_k \gamma_{|j-k|}, \quad j = 1, \dots, p\end{aligned}$$

may be solved for the first $p+1$ autocovariances $\gamma_0, \dots, \gamma_p$ in terms of the AR(p) coefficients φ_j and the innovation variance σ^2 . The remaining autocovariances may then be filled in recursively with

$$\gamma_j = \sum_{k=1}^p \varphi_k \gamma_{j-k}, \quad j = p+1, \dots, n-1.$$

Conversely, the AR(p) coefficients may be found in terms of the first $p+1$ autocovariances by means of

$$\boldsymbol{\varphi} = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix}. \quad (36)$$

The ordinary Asymptotic Method of Moments (AMM) estimator of the AR coefficients directly replaces the asymptotic error autocovariance moments γ_j in (36) with their (weak) empirical counterparts s_j/n as computed from the regression residuals:

$$\hat{\boldsymbol{\varphi}}^{AMM} = \begin{pmatrix} s_0 & \cdots & s_{p-1} \\ \vdots & \ddots & \vdots \\ s_{p-1} & \cdots & s_0 \end{pmatrix}^{-1} \begin{pmatrix} s_1 \\ \vdots \\ s_p \end{pmatrix}. \quad (37)$$

Although this simple estimator is consistent, it takes no account of the distinction between the errors and residuals and so is biased in finite samples, particularly as the process becomes more persistent and approaches a unit root.¹⁶

Generalizing (15), we define the *Moment Ratio Function of Order p* to be the matrix “ratio” of the exact finite sample population moments corresponding to the sample moments used to compute $\hat{\boldsymbol{\varphi}}^{AMM}$:

¹⁶ The MM estimator defined in (37) is virtually equivalent to the results of an OLS regression of the residuals on their own p lags, with the exceptions that it treats the first and last few residuals symmetrically, and uses weak rather than strong autocorrelations.

$$\begin{aligned}
\boldsymbol{\psi}(\boldsymbol{\varphi}; \mathbf{X}) &= \begin{pmatrix} E s_0 & \cdots & E s_{p-1} \\ \vdots & \ddots & \vdots \\ E s_{p-1} & \cdots & E s_0 \end{pmatrix}^{-1} \begin{pmatrix} E s_1 \\ \vdots \\ E s_p \end{pmatrix} \\
&= \begin{pmatrix} \text{tr}_0(\mathbf{MGM}) & \cdots & \text{tr}_{p-1}(\mathbf{MGM}) \\ \vdots & \ddots & \vdots \\ \text{tr}_{p-1}(\mathbf{MGM}) & \cdots & \text{tr}_0(\mathbf{MGM}) \end{pmatrix}^{-1} \begin{pmatrix} \text{tr}_1(\mathbf{MGM}) \\ \vdots \\ \text{tr}_p(\mathbf{MGM}) \end{pmatrix}.
\end{aligned} \tag{38}$$

The scaled autocovariance matrix $\mathbf{G} = \boldsymbol{\Gamma}/\sigma^2$ depends on the true autoregressive coefficients $\boldsymbol{\varphi}$ but not on σ^2 . This Moment Ratio Function depends on the autoregressive coefficients through \mathbf{G} as well as the regressor matrix \mathbf{X} through \mathbf{M} , but fortunately the unknown innovation variance drops out. If the stationarity restriction $\alpha < 1$ were never binding, we could define the Moment Ratio Estimator $\hat{\boldsymbol{\varphi}}^{MR}$ simply by setting $\boldsymbol{\psi}(\hat{\boldsymbol{\varphi}}^{MR}; \mathbf{X}) = \hat{\boldsymbol{\varphi}}^{AMM}$ and solving this system of p nonlinear equations in p unknowns. The Moment Ratio estimator may thus be thought of as the exact finite sample MM estimator, rather than a merely asymptotic MM estimator.

In the unit root case $\alpha = 1$, the Yule-Walker equations readily determine the $n \times n$ normalized covariance matrix \mathbf{H} of the error first differences

$$\begin{aligned}
\xi_t &= \varepsilon_t - \varepsilon_{t-1} \\
\mathbf{H} &= \text{Cov}(\boldsymbol{\xi})/\sigma^2,
\end{aligned} \tag{39}$$

in terms of the θ_j , since we have assumed that at least the first differences are covariance stationary. Conditional on say ε_0 , the errors themselves then have finite covariance matrix,

$$\text{Cov}(\boldsymbol{\varepsilon} | \varepsilon_0) = \sigma^2 \mathbf{NHN}',$$

where \mathbf{N} is the integrator matrix (28). Since the regression is assumed to include an intercept, the residuals sum to 0 regardless of ε_0 , and so have finite conditional *as well as unconditional* covariance, determined by

$$\text{Cov}(\mathbf{e} | \varepsilon_0) = \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{MNHN}'\mathbf{M}.$$

In the case $\alpha = 1$, we therefore define the Moment Ratio Function by

$$\boldsymbol{\psi}(\boldsymbol{\varphi}; \mathbf{X}) = \begin{pmatrix} \text{tr}_0(\mathbf{MNHN}'\mathbf{M}) & \cdots & \text{tr}_{p-1}(\mathbf{MNHN}'\mathbf{M}) \\ \vdots & \ddots & \vdots \\ \text{tr}_{p-1}(\mathbf{MNHN}'\mathbf{M}) & \cdots & \text{tr}_0(\mathbf{MNHN}'\mathbf{M}) \end{pmatrix}^{-1} \begin{pmatrix} \text{tr}_1(\mathbf{MNHN}'\mathbf{M}) \\ \vdots \\ \text{tr}_p(\mathbf{MNHN}'\mathbf{M}) \end{pmatrix}. \tag{40}$$

Generalizing (16) to the AR(p) case, we may now define the Moment Ratio Estimator $\hat{\boldsymbol{\varphi}}^{MR}$ in terms of the point of nearest approach of the Asymptotic Method of Moments Estimator $\hat{\boldsymbol{\varphi}}^{AMM}$ to the Moment Ratio Function, in terms of the Euclidean norm:

$$\hat{\boldsymbol{\phi}}^{MR} = \arg \min_{\boldsymbol{\phi} \in \Phi} \left\| \hat{\boldsymbol{\phi}}^{AMM} - \boldsymbol{\psi}(\boldsymbol{\phi}, \mathbf{X}) \right\|.^{17} \quad (41)$$

This problem was solved numerically for this paper with satisfactory results using a constrained Nelder-Mead simplex algorithm, subject to the explicit constraint $\alpha \leq 1$ and assigning a very high value to the objective function whenever other non-stationary roots were encountered.¹⁸

The last $n-p$ innovations of an $AR(p)$ process may be computed from the errors by

$$\mathbf{u} = \mathbf{D}_{\boldsymbol{\phi}} \boldsymbol{\varepsilon}$$

where the $(n-p) \times n$ innovation-generating operator $\mathbf{D}_{\boldsymbol{\phi}} = (d_{ij})$ is defined by $d_{i,i+p} = 1$, $d_{i,i+p-h} = -\varphi_h$ and $d_{i,j} = 0$ otherwise. The innovations estimated from the residuals are then

$$\hat{\mathbf{u}} = \mathbf{D}_{\boldsymbol{\phi}} \mathbf{e} = \mathbf{D}_{\boldsymbol{\phi}} \mathbf{M} \boldsymbol{\varepsilon} \quad (42)$$

so that in the stationary case $\alpha < 1$,

$$\begin{aligned} E \hat{\mathbf{u}} \hat{\mathbf{u}}' &= \mathbf{D}_{\boldsymbol{\phi}} \mathbf{M} \Gamma \mathbf{M} \mathbf{D}_{\boldsymbol{\phi}}' = \sigma^2 \mathbf{D}_{\boldsymbol{\phi}} \mathbf{M} \mathbf{G} \mathbf{M} \mathbf{D}_{\boldsymbol{\phi}}', \\ E \hat{\mathbf{u}}' \hat{\mathbf{u}} &= \sigma^2 \text{tr}_0(\mathbf{D}_{\boldsymbol{\phi}} \mathbf{M} \mathbf{G} \mathbf{M} \mathbf{D}_{\boldsymbol{\phi}}'), \end{aligned}$$

whence

$$\hat{\sigma}_{MR}^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / \text{tr}_0(\mathbf{D}_{\boldsymbol{\phi}} \mathbf{M} \mathbf{G} \mathbf{M} \mathbf{D}_{\boldsymbol{\phi}}')$$

would be an unbiased and consistent estimator of the innovation variance, if the true value of $\boldsymbol{\phi}$ were known. When $\boldsymbol{\phi}$ is estimated consistently by $\hat{\boldsymbol{\phi}}^{MR}$,

$$\hat{\sigma}_{MR}^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / \text{tr}_0(\mathbf{D}_{\hat{\boldsymbol{\phi}}^{MR}} \mathbf{M} \hat{\mathbf{G}}^{MR} \mathbf{M} \mathbf{D}_{\hat{\boldsymbol{\phi}}^{MR}}')$$

becomes a consistent and at least approximately unbiased estimator.¹⁹ The Moment Ratio estimator of the covariance matrix \mathbf{C} of $\hat{\boldsymbol{\beta}}^{OLS}$ as given in (2) is then

$$\hat{\mathbf{C}}^{MR} = \hat{\sigma}_{MR}^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \hat{\mathbf{G}}^{MR} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1}.$$

When a unit root is present, the estimated innovations instead have finite covariance

$$\text{Cov}(\hat{\mathbf{u}}) = \sigma^2 \mathbf{D}_{\boldsymbol{\phi}} \mathbf{M} \mathbf{N} \mathbf{H} \mathbf{N}' \mathbf{M} \mathbf{D}_{\boldsymbol{\phi}}'$$

whence

$$\hat{\sigma}_{MR}^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / \text{tr}_0(\mathbf{D}_{\boldsymbol{\phi}} \mathbf{M} \mathbf{N} \mathbf{H} \mathbf{N}' \mathbf{M} \mathbf{D}_{\boldsymbol{\phi}}')$$

¹⁷ An earlier version of this paper instead attempted to equate the first p sample autocorrelations directly to individual moment ratio functions. However, this calculation turned out to be very ill-conditioned in the vicinity of a unit root. Transforming the AR coefficients to the ADF persistence form before taking the norm in (41) will give the same answer if the unit root restriction is not binding, but a somewhat different answer when it is binding. This alternative may be worth exploring in future work.

¹⁸ Following *Numerical Recipes* (Press 1992: 402-6), with a linear restriction imposed. In order to retain volume in the simplex, the linear restriction was never allowed to reduce the volume of the simplex in a single step by a factor less than 0.5.

¹⁹ In order to treat the first and last residuals symmetrically, the residuals were in fact computed both forward and backward in time, and the resulting sums of squares averaged, in both the stationary and unit root case in the calculations that follow.

is an unbiased and consistent estimate of the innovation variance when the autoregressive order and parameters are known and have unitary first order persistence. When the autoregressive parameters are consistently estimated and the estimate happens to have a unit root,

$$\hat{\sigma}_{MR}^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / \text{tr}_0 \left(\mathbf{D}_{\hat{\varphi}^{MR}} \mathbf{M} \hat{\mathbf{N}} \hat{\mathbf{H}}^{MR} \mathbf{N}' \mathbf{M} \mathbf{D}_{\hat{\varphi}^{MR}}' \right)$$

is a consistent and approximately unbiased estimator of the innovation variance. The Moment Ratio estimator of the covariance matrix \mathbf{C} of $\hat{\boldsymbol{\beta}}^{OLS}$ as given in (2) and conditional on ε_0 is then

$$\hat{\mathbf{C}}_0^{MR} = \hat{\sigma}_{MR}^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{M} \hat{\mathbf{N}} \hat{\mathbf{H}}^{MR} \mathbf{N}' \mathbf{M} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1}.$$

Conditional on ε_t , similar reasoning yields

$$\hat{\mathbf{C}}_t^{MR} = \hat{\sigma}_{MR}^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{M} \mathbf{N}_t \hat{\mathbf{H}}^{MR} \mathbf{N}_t' \mathbf{M} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1}.$$

The autocovariation function and therefore the coefficient covariance matrix may now be estimated consistently for almost any covariance stationary or I(1) error process by setting the autoregressive order p equal to

$$p = \left\lfloor 4 (n/100)^{2/9} \right\rfloor, \quad [26]$$

the maximum lag considered by the formula (26) commonly used for the Newey-West HAC estimator. As n increases, p as well as $n^{1/2}/p$ increase without bound, so that eventually every autocorrelation is utilized, and estimated consistently. For sample size 100, this formula gives $p = 4$.²⁰

9. Monte Carlo Performance of the MR(p) Estimator

Figures 10 and 11 show the Monte Carlo median and mean of the estimated slope variances using the OLS, HAC ($m=4$), AR($p=4$) estimated by MM, and MR($p=4$) covariance matrices. Throughout the true DGP is AR(1) with the true value of $\varphi = \varphi_1 = \rho_1$ indicated on the horizontal axis, using Gaussian errors and 10,000 replications.

Despite the nuisance parameters that have been added for consistency in case the process is not AR(1), the shapes are qualitatively similar to those in Figures 6 and 7 where the true AR(1) structure is imposed. In terms of median bias, MR(p) clearly dominates AR(p), which in turn clearly dominates HAC. HAC does dominate OLS unless $\varphi < 0.13$. As in the AR(1) case, MR(p) does have a positive mean bias, which peaks at $\varphi = 0.95$ but then reverses above $\varphi = 0.98$.

²⁰ No claim is made that (26) is optimal in any sense, but merely that it is adequate to consistently capture the autocorrelation structure. The reader is of course welcome to substitute any other formula for which both p and $n^{1/2}/p$ increase without bound with n .

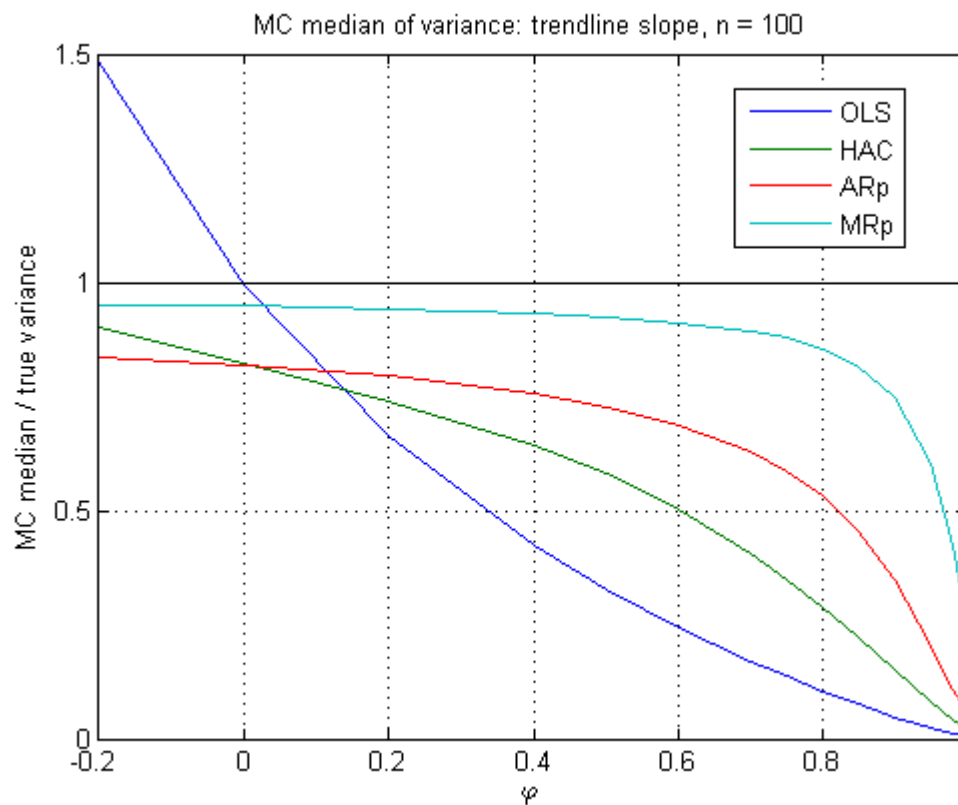


Figure 10

Monte Carlo median of estimated variance of slope coefficient, relative to true variance, versus true AR(1) parameter of error DGP, for a trend line regression with $n = 100$. OLS is shown in blue, NW-HAC in green, conventional AR(p) in red, and MR(p) in cyan.

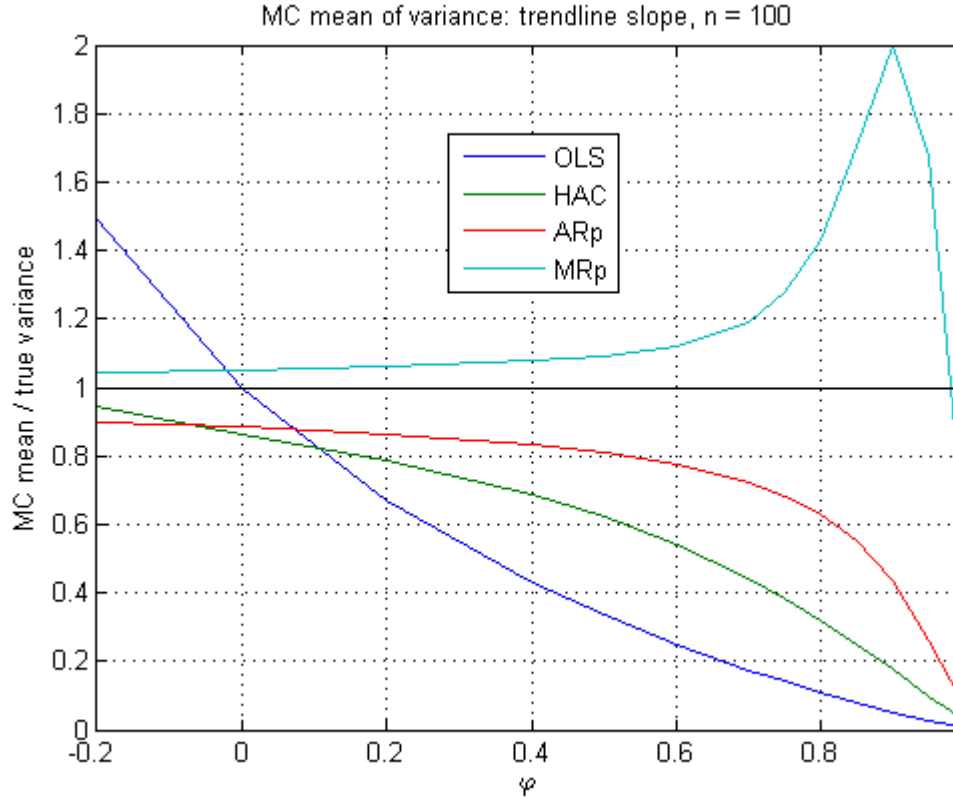


Figure 11

Monte Carlo mean of estimated variance of slope coefficient, relative to true variance, versus true AR(1) parameter of error DGP, for a trend line regression with $n = 100$. OLS is shown in blue, NW-HAC in green, conventional AR(p) in red, and MR(p) in cyan.

Figures 12 and 13 give the 95% confidence interval coverages using the MR(p) estimator of the slope standard error, versus OLS, HAC and AR(p) as in Figures 10 and 11. Despite the pronounced positive mean bias of the MR(p) estimator, its coverage is under 95%. Nevertheless, MR(p) dominates all the competitors in terms of coverage bias, except of course for OLS when φ is very close to 0.

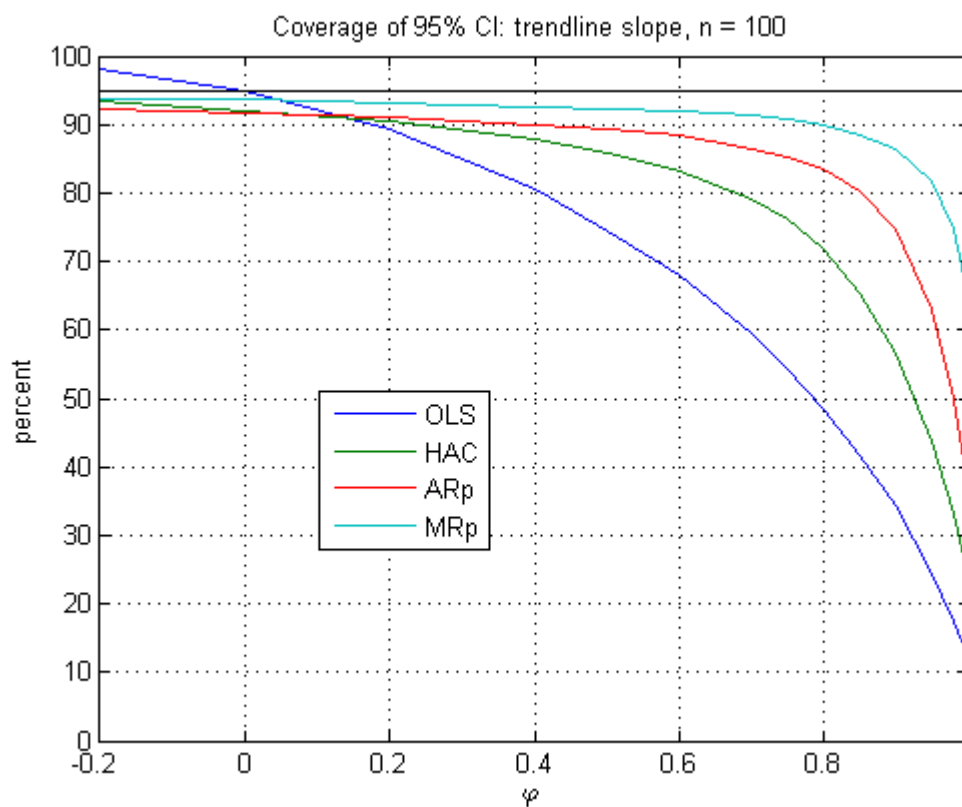


Figure 12

Monte Carlo coverage of 95% confidence interval for slope coefficient in a trend line regression with $n = 100$. OLS is shown in blue, NW-HAC in green, conventional $AR(p)$ in red, and $MR(p)$ in cyan.

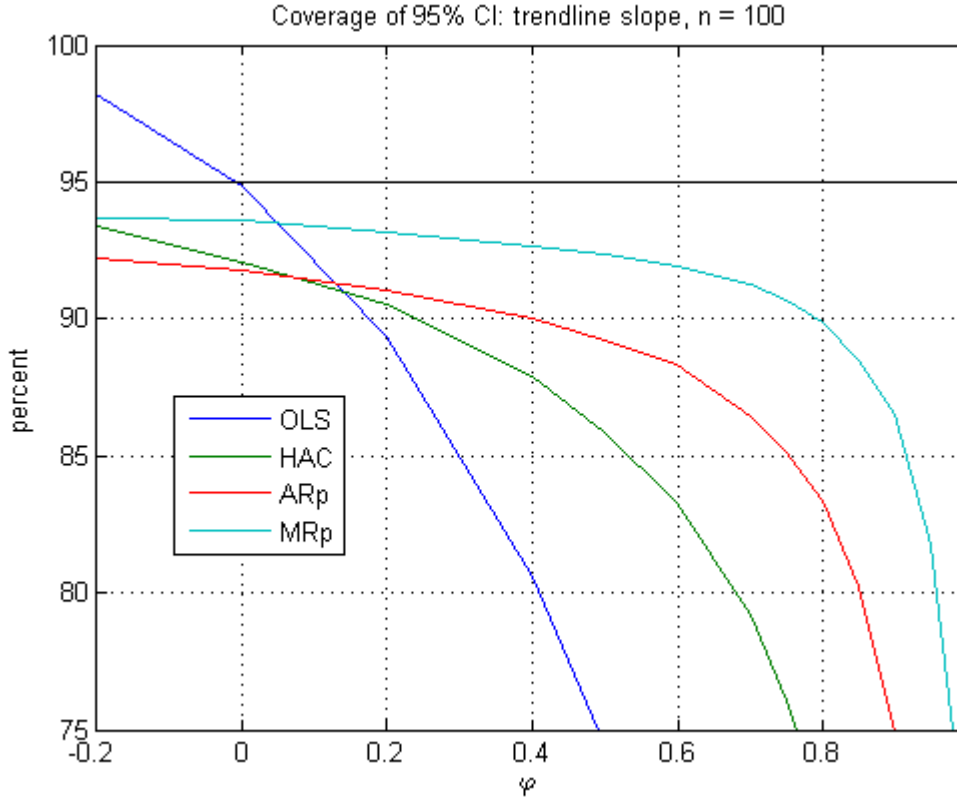


Figure 13
Detail of Figure 12.

The greater generality of the $MR(p)$ and $AR(p)$ estimators over $MR(1)$ and $AR(1)$ comes at the cost of somewhat greater bias and reduced coverage, because they are not restricted to the true (but in practice unknown) $AR(1)$ DGP. However, graphical comparisons (omitted here in the interest of space) show that this cost is small in comparison to the superiority of $MR(p)$ over $AR(p)$ and of $AR(p)$ over HAC.

Although the $MR(p)$ estimator greatly improves the coverage of 95% confidence intervals based on the standard Student t critical values with $n-k$ degrees of freedom, it still has a systematic downward size distortion because estimation of the $AR(p)$ parameters introduces an element of uncertainty that is not taken into account by this distribution, so that the true distribution of the t -ratio is in fact even more leptokurtic. Bartlett (1935) and Quenouille (1952) have suggested instead using a critical value based on the t distribution with a reduced “effective degrees of freedom” implied by the increase in variance caused by the serial correlation adjustment. Using the MRE for $\hat{\beta}_j$, this would be $EDOF_j = (n - k)\hat{\epsilon}_{j,j}^{OLS} / \hat{\epsilon}_{j,j}^{MR}$, and similarly for linear combinations of coefficients and for the denominator DOF in tests of joint hypotheses. This modification would be worth trying in future research but is not implemented in the present paper.

10. An $MR(p)$ Unit Root Test

The MR(1) unit root test discussed in Section 7 above may easily be extended to the MR(p) case as follows: 1. Estimate $\hat{\phi}^{AMM}$ via (37). Designate the sum of these coefficients by $\hat{\alpha}^{AMM}$, which will be our test statistic. 2. Then, compute Moment Ratio estimates of the autoregressive coefficients using (41), but with the null hypothesis restriction that their sum $\hat{\alpha}^{UR}$ be unity. 3. Convert the resulting restricted standard coefficients $\hat{\phi}^{UR}$ to ADF form $(\hat{\alpha}^{UR}, \hat{\theta}^{UR})$ using (35). 4. Generate the desired number of random samples with covariance determined by $\hat{\theta}^{UR}$. 5. Sum these to obtain unit root regression errors, and re-estimate the equation with the constructed unit root errors. 6. With synthetic data set i , re-estimate $\hat{\phi}_i^{AMM}$ and thence $\hat{\alpha}_i^{AMM}$ via (35). The probability of obtaining a persistence as low as $\hat{\alpha}^{AMM}$ under the null of a unit root (and with $\theta = \hat{\theta}^{UR}$) is then estimated as the proportion p^{UR} of the $\hat{\alpha}_i^{AMM}$ that are less than $\hat{\alpha}^{AMM}$. Since the simulations do not require solving (41) on each replication, they run quite fast: 10,000 replications take less than 1.2 seconds in MATLAB on a laptop with a 2.50 GHz Intel® Core™ i5-2520M CPU for the case illustrated.

In the MR(1) case of Section 7, the unit root test size is correct by construction, to within Monte Carlo sampling error. In the MR(p) case, however, the test depends on using $\hat{\theta}^{UR}$ as a proxy for θ , and therefore may have some size distortion. This was investigated by constructing 10,000 trend line regressions with random walk errors and conducting the unit root test described above with 10,000 replications each.

Figure 14 plots the actual size versus the nominal size. The size distortion is generally moderate, though there is some positive distortion for nominal size in the range 0 – 0.18. Table 1 tabulates selected values. For example, if the nominal size is 5%, the test in fact falsely rejects a unit root 6.0% of the time. In order to obtain a true type I error of 5%, a nominal size of 4.1% would have to be used. At a 5% test size, the Monte Carlo sampling standard error is approximately 0.2% on both p^{UR} and the estimated actual size.

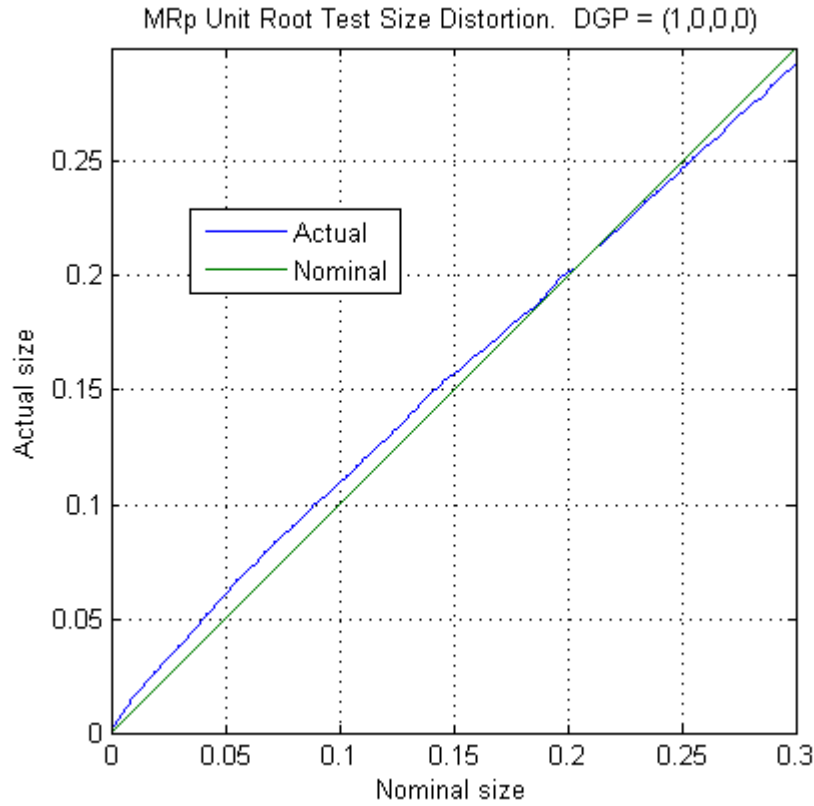


Figure 14
Monte Carlo size distortion of MR(p) unit root test.
Trendline regression with random walk errors, $n = 100$.

Table 1
Nominal versus Actual MR(p) Unit Root Test Size.
Trendline regression with random walk errors, $n = 100$.

Nominal	Actual	s.e.
.200	.201	.004
.100	.104	.003
.050	.060	.002
.010	.015	.001
.198	.200	.004
.089	.100	.003
.041	.050	.002
.006	.010	.001

11. A Blended Covariance Matrix

Figure 15 shows the coverage of a 95% confidence interval for the slope coefficient in our illustrative trend line regression, at the top (in blue) when a unit root is

imposed and at the bottom (in red) when the $MR(p)$ standard errors are always used, as a function of the true $AR(1)$ parameter φ of the error DGP, along with a horizontal black line indicating the 95% ideal value. Always imposing a unit root leads to a “95% CI” that includes the true value in 10,000 cases out of 10,000, unless φ is 0.95 or higher, and therefore is ordinarily far too conservative a strategy. On the other hand, the pure $MR(p)$ estimator, which only uses the unit root formula when the point estimate of the persistence is actually unity, can yield far too small a confidence interval.

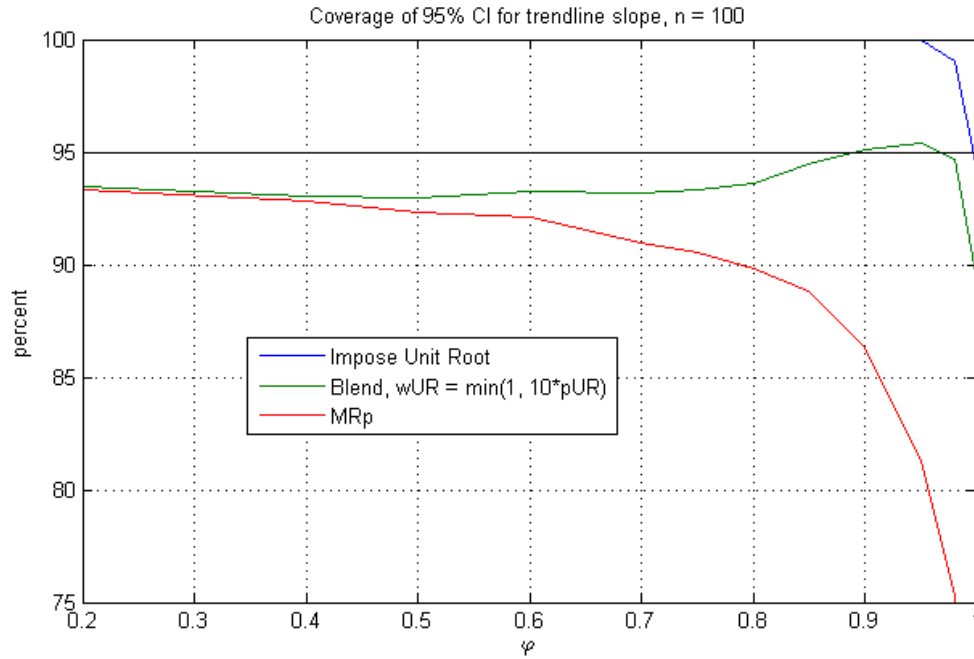


Figure 15

Coverage of 95% confidence interval for trend line slope coefficient when a unit root is imposed always (blue line), or never (red line). The intermediate green line uses a blended covariance matrix based on a linear combination the $MR(p)$ and unit root precision matrices, assigning weight $w^{UR} = \min(1, 10 p^{UR})$ to the unconditional unit root covariance matrix.

Ever since the seminal work of Nelson and Plosser (1982), econometric protocol has typically been to test for a unit root and then to either ignore the possibility of a unit root if it can be rejected at some convention test size, say 5%, or else to impose a unit root (typically by re-estimating the equation in first differences) if it cannot be rejected, even when the point estimate falls short of a unit root. However, this practice unfortunately produces results that change discontinuously and substantially with the data.²¹

²¹ Peron and Yabu (2009) discontinuously switch to a first differenced estimator whenever the estimated persistence is within an $n^{-1/2}$ neighborhood of unity, where n is the sample size. This has a very similar effect to basing the switch on the size of a unit root test.

This discontinuous behavior can easily be corrected, however, simply by blending the unit root and non-unit root covariance matrices using a continuous function of the unit root test statistic p^{UR} .²² It would be inappropriate to blend the covariance matrices themselves directly, since the unconditional unit root covariance matrix indicates infinite variance for the intercept. However, the unconditional unit root *precision* matrix (inverse covariance matrix) merely has a zero in this position, which will not dominate the blend unless the unit root case is given a 100% weight. Therefore our proposed blended covariance matrix takes the form:

$$\hat{\mathbf{C}}^{Blend} = \left(w^{UR} \hat{\mathbf{C}}^{UR^{-1}} + (1 - w^{UR}) \hat{\mathbf{C}}^{MR^{-1}} \right)^{-1},$$

where $\hat{\mathbf{C}}^{UR}$ is the unconditional unit root covariance matrix, constructed from $\hat{\mathbf{C}}_0^{MR}$ as discussed at the end of Section 5.²³

The central green line in Figure 15 shows the coverage using $\hat{\mathbf{C}}^{Blend}$, with $w^{UR} = \min(1, 10p^{UR})$.

This formula effectively embraces the unit root estimate if a unit root cannot be rejected at the 10% level, but does not completely ignore it unless it can be absolutely rejected with $p^{UR} = 0$. It may be seen that this choice of w^{UR} provides a very good correction of the coverage for all values of φ , and even slightly over-corrects for φ between 0.9 and 0.96. Replacing the 10 in the formula with 20 (corresponding to embracing the unit root if it cannot be rejected at the 5% level) was found to over-correct excessively, while replacing it with 5 (corresponding to a 20% test size) was found to under-correct consistently.

12. A Heteroskedasticity-Consistent MR Covariance Matrix

Although the primary emphasis of the present paper has been the case of regressor-conditional homoskedasticity, a Heteroskedasticity and Autocorrelation Consistent MR estimator (MR-HAC) may be constructed as a generalization of a slightly improved version of the White (1980) HCC estimator: The original HCC estimator for the simple case of OLS regression is

$$\hat{\mathbf{C}}^{HCC} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{F}^{HCC} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

where

$$\mathbf{F}^{HCC} = \left(f_{ij}^{HCC} \right), \text{ with } f_{ii}^{HCC} = e_i^2; f_{ij}^{HCC} = 0, i \neq j.$$

This estimator has a downward finite sample bias because the regression residuals generally have smaller variance than the errors themselves. Furthermore, when non-trivial regressors are present, the residuals will be heteroskedastic even when the errors are homoskedastic. In the benchmark case of regressor-conditional homoskedasticity and in the absence of serial correlation, (6) becomes

²² Similarly, Bunzel and Vogelsang (2005) smoothly transition from I(0) to I(1) covariance estimators using a continuous function of a unit root test, while Harvey et al. (2007) interpolate I(0) and I(1) t-statistics using a continuous function of both an ADF unit root test and a KPSS stationarity test.

²³ MATLAB will appropriately invert $\hat{\mathbf{C}}^{UR}$ despite the ∞ entry in its (1,1) position, but will generate a non-fatal error message. This distracting message can be avoided by replacing the ∞ with 1, inverting the resulting proper matrix, and then replacing its (1,1) entry with 0.

$$E\mathbf{e}\mathbf{e}' = E(\mathbf{M}\mathbf{\varepsilon}\mathbf{\varepsilon}'\mathbf{M}) = \sigma^2\mathbf{M} = \sigma^2(m_{ij})$$

rather than $\sigma^2\mathbf{I}$. However, both the bias and the induced heteroskedasticity may easily be removed by defining the *Unbiased HCC estimator* (HCC-U) as

$$\hat{\mathbf{C}}^{HCC-U} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}^{HCC-U}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (43)$$

where

$$\mathbf{F}^{HCC-U} = (f_{ij}^{HCC-U}), \text{ with } f_{ii}^{HCC-U} = e_i^2 / m_{ii}; f_{ij}^{HCC-U} = 0, i \neq j.$$

When serial correlation is present, our strategy will be to first approximate the serial correlation with a homoskedastic AR(p) process and to construct its MR estimator as above, and then to construct an unbiased and heteroskedasticity-consistent estimator in terms of the estimated *innovations* to the approximating process rather than the *residuals* themselves.

When the approximating homoskedastic AR(p) process is stationary, let \mathbf{Q} be the lower-triangular Cholesky decomposition of its scaled error autocovariance matrix \mathbf{G} :

$$\mathbf{G} = \mathbf{\Gamma} / \sigma^2 = \mathbf{Q}\mathbf{Q}'.$$

Then the $n \times 1$ extended innovation vector²⁴

$$\mathbf{u} = \mathbf{Q}^{-1}\mathbf{\varepsilon}$$

has covariance matrix $\sigma^2\mathbf{I}_n$ and

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\varepsilon}\mathbf{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \\ &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{u}\mathbf{u}'\mathbf{Q}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \end{aligned}$$

A natural estimator of \mathbf{u} is

$$\hat{\mathbf{u}} = \mathbf{Q}^{-1}\mathbf{e} = \mathbf{Q}^{-1}\mathbf{M}\mathbf{\varepsilon},$$

which has covariance matrix

$$\begin{aligned} E(\mathbf{u}\mathbf{u}') &= \mathbf{Q}^{-1}\mathbf{M}\mathbf{\Gamma}\mathbf{M}'\mathbf{Q}'^{-1} \\ &= \sigma^2\mathbf{Q}^{-1}\mathbf{M}\mathbf{G}\mathbf{M}'\mathbf{Q}'^{-1} \\ &= \sigma^2\mathbf{S} = \sigma^2(s_{ij}) \end{aligned}$$

Define

$$\mathbf{V} = (v_{ij}) \text{ with } v_{ii} = \hat{u}_i^2 / s_{ii} \text{ and } v_{ij} = 0, i \neq j.$$

Since $E\hat{u}_i^2 = \sigma^2 s_{ii}$, $E\mathbf{V} = \sigma^2\mathbf{I}$. Replacing \mathbf{G} , \mathbf{Q} , \mathbf{S} and \mathbf{V} by their MR estimates, we define

$$\hat{\mathbf{C}}^{MR-HAC} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}^{MR}\mathbf{V}^{MR}\mathbf{Q}^{MR'}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

In the absence of serial correlation, so that $\mathbf{G} = \mathbf{Q} = \mathbf{I}$, this MR-HAC estimator reduces to the HCC-U estimator (43).

When the approximating homoskedastic AR(p) process has a unit root, let \mathbf{Z} be the lower Cholesky decomposition of the scaled covariance matrix \mathbf{H} of the error first differences as defined in (39):

²⁴ This extended innovation vector contains p more terms up front than the truncated innovation vector estimated by (42).

$$\mathbf{H} = \mathbf{E}(\xi\xi')/\sigma^2 = \mathbf{Z}\mathbf{Z}'.$$

Then the $n \times 1$ extended innovation vector is given by

$$\mathbf{u} = \mathbf{Z}^{-1}\xi$$

and may be estimated by

$$\hat{\mathbf{u}} = \mathbf{Z}^{-1}\hat{\xi} = \mathbf{Z}^{-1}\mathbf{N}^{-1}\mathbf{e} = \mathbf{Z}^{-1}\mathbf{N}^{-1}\mathbf{M}\boldsymbol{\varepsilon} = \mathbf{Z}^{-1}\mathbf{N}^{-1}\mathbf{M}\mathbf{N}\xi.$$

This has covariance matrix

$$\begin{aligned} \mathbf{E}(\hat{\mathbf{u}}\hat{\mathbf{u}}') &= \sigma^2 \mathbf{Z}^{-1}\mathbf{N}^{-1}\mathbf{M}\mathbf{N}\mathbf{H}\mathbf{N}'\mathbf{M}\mathbf{N}'^{-1}\mathbf{Z}'^{-1} \\ &= \sigma^2 \mathbf{S}^{UR} = \sigma^2 (s_{ij}^{UR}). \end{aligned}$$

Conditional on say ε_0 ,

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}} | \varepsilon_0) &= \mathbf{E}((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \varepsilon_0) \\ &= \mathbf{E}((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{N}\mathbf{Z}\mathbf{u}\mathbf{u}'\mathbf{Z}'\mathbf{N}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \varepsilon_0) \end{aligned}$$

Setting

$$\mathbf{V}^{UR} = (\mathbf{v}_{ij}^{UR}), \text{ with } v_{ii}^{UR} = \hat{u}_i^2 / s_{ii}^{UR} \text{ and } v_{ij}^{UR} = 0, i \neq j,$$

and replacing \mathbf{H} , \mathbf{Z} , \mathbf{S}^{UR} and \mathbf{V}^{UR} by their MR estimates, we define the unit root MR-HAC covariance matrix, conditional on say ε_0 , to be:

$$\hat{\mathbf{C}}_0^{MR-HAC} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{N}\mathbf{Z}^{MR} \mathbf{V}^{UR-MR} \mathbf{Z}^{MR'} \mathbf{N}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Figures 16 and 17 illustrate the Monte Carlo median bias and 95% CI coverage for both the MR(p) and MR-HAC estimators. Surprisingly, there is essentially no difference in performance, despite the here-unnecessary greater generality of MR-HAC.

The comparisons made above in Figures 10 - 13 above between the MR(p) and NW-HAC estimators were not entirely fair, since NW-HAC attempts to be more general with its RCH correction. This greater generality naturally costs it precision when the additional generality is not really necessary. The comparisons between MR-HAC and NW-HAC in Figures 16 and 17 overcome this objection.

An alternative way to make a fair comparison between the two approaches is to eliminate the RCH adjustment in NW-HAC by defining the “NW-AC” (for Autocorrelation Corrected) estimator, simply by replacing the $e_i e_j$ in (24) with their average value $s_{|i-j|} / (n - |i - j|)$. In other words,

$$\hat{\mathbf{C}}^{NW-AC} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\hat{\mathbf{H}}^{NW-AC} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (44)$$

where

$$\hat{\mathbf{H}}^{NW-AC} = (\mathbf{K}(|i - j|) s_{|i-j|} / (n - |i - j|)).$$

(Compare equation (27).) Coverages based on this covariance matrix are identified as “NW-AC” in Figures 16 and 17.

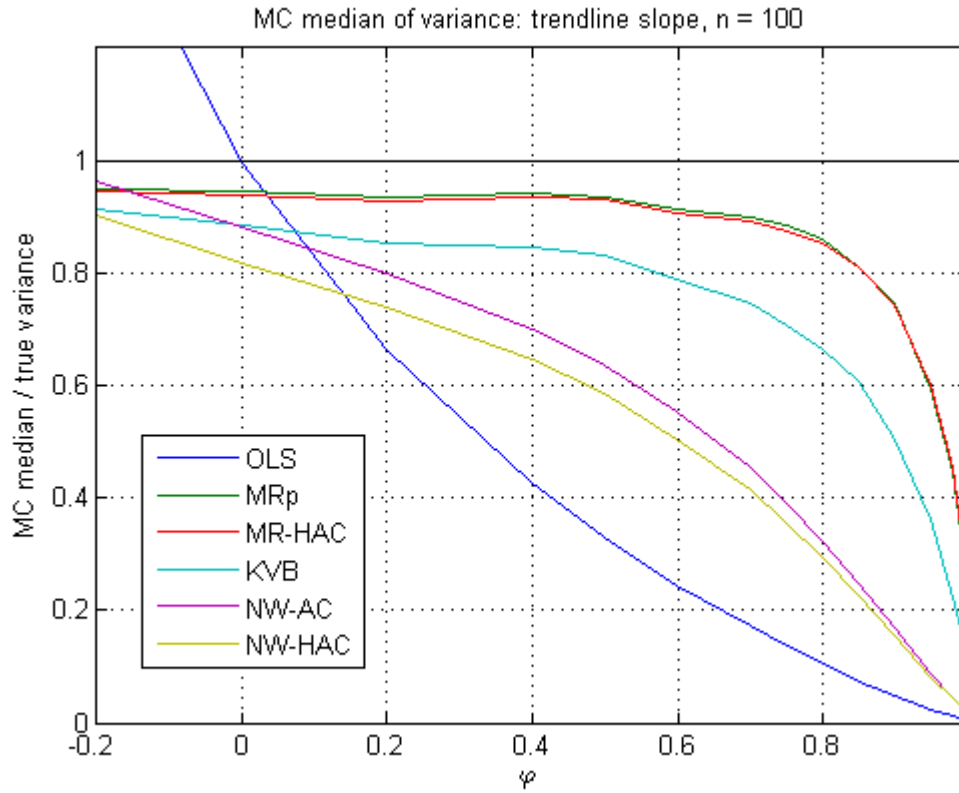


Figure 16

Monte Carlo median of estimated variance of slope coefficient, relative to true variance, versus true AR(1) parameter of error DGP, for a trend line regression with $n = 100$. OLS is shown in blue, MR(p) in green, MR-HAC in red, Kiefer-Vogelsang-Bunzel in cyan, NW-AC in magenta, and NW-HAC in dark yellow.

As expected, NW-HAC, with its here-unnecessary heteroskedasticity adjustment, is distinctly outperformed by NW-AC. Nevertheless, MR(p) clearly outperforms even NW-AC except for some negative values of φ .

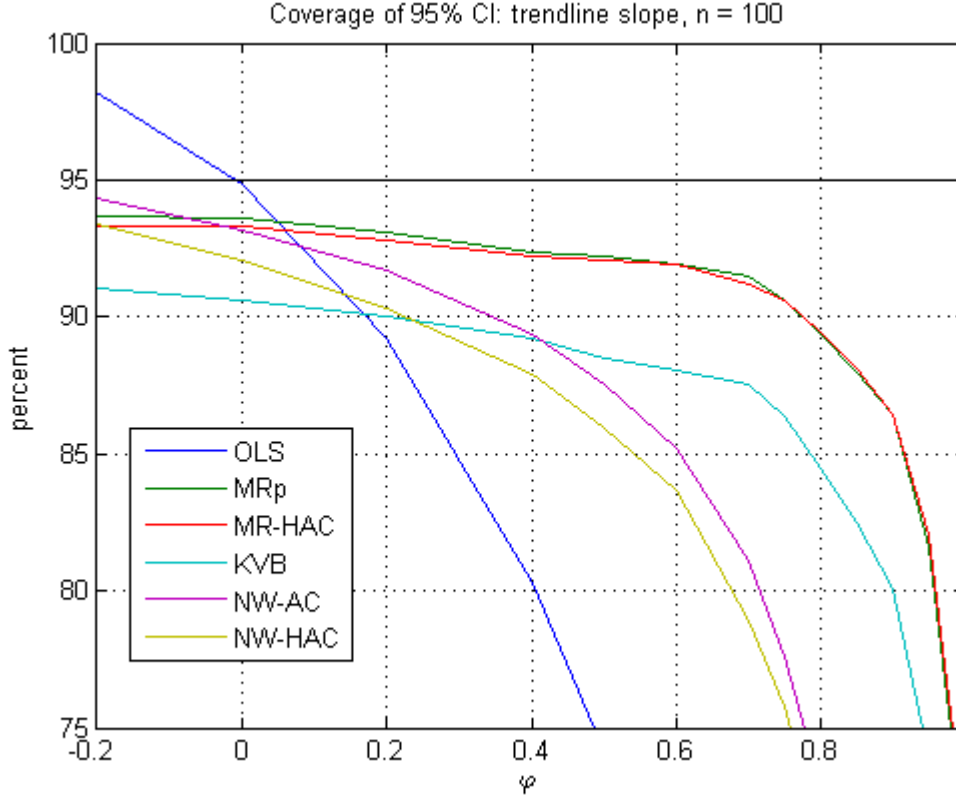


Figure 17

Coverage of 95% confidence interval for the trend slope coefficient with $n = 100$. OLS is shown in blue, MR(p) in green, MR-HAC in red, Kiefer-Vogelsang-Bunzel in cyan, NW-AC in magenta, and NW-HAC in dark yellow.

The additional lines “KVB” in Figures 16 and 17 are based on the covariance estimator of Kiefer, Vogelsang and Bunzel (KVB 2000), as modified by Kiefer and Vogelsang (2002):

$$\hat{\mathbf{C}}^{KVB} = c(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{B}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where

$$\mathbf{B} = (e_i e_j (n - |i - j|) / n)$$

and c is an asymptotic correction for bias. The analytical computations of Abadir and Paruolo (1996, 2002) demonstrate that in terms of the above 2002 version of the statistic (which differs by a factor of 2 from that in the 2000 paper), $c = 5.588756592\dots$ While KVB has the advantage that it makes no parametric assumptions about the form of the serial correlation, even as an approximation that loses its restrictiveness in large samples, it is inconsistent in that the limiting covariance estimate is merely an unbiased random variable, rather than a limit in probability.

It may be seen from Figures 16 and 17 that the KVB estimator is superior to NW-HAC in terms of median bias for all positive values of φ , and in terms of 95% CI

coverage for all values of φ greater than approximately 0.25. However, KVB in turn is dominated by MR-HAC by either measure for all values of φ illustrated.

13. Real GDP Growth

Figure 19 shows the logarithm of real US GDP for 1947Q1-2014Q2 ($n = 270$ observations), along with a quadratic trendline fit by OLS:

$$\log y_t = a + b(t/4) + c(t/4)^2 + \varepsilon_t$$

All variables were normalized to 0 in the last observation. As a consequence, $-a$ is a measure of the “output gap” during 2014Q2, and b is a measure of the trend rate of growth during the same quarter. The integer time index was divided by 4 so that $100b$ has units of percent per annum. The annual rate of change of the growth rate is $2c$.

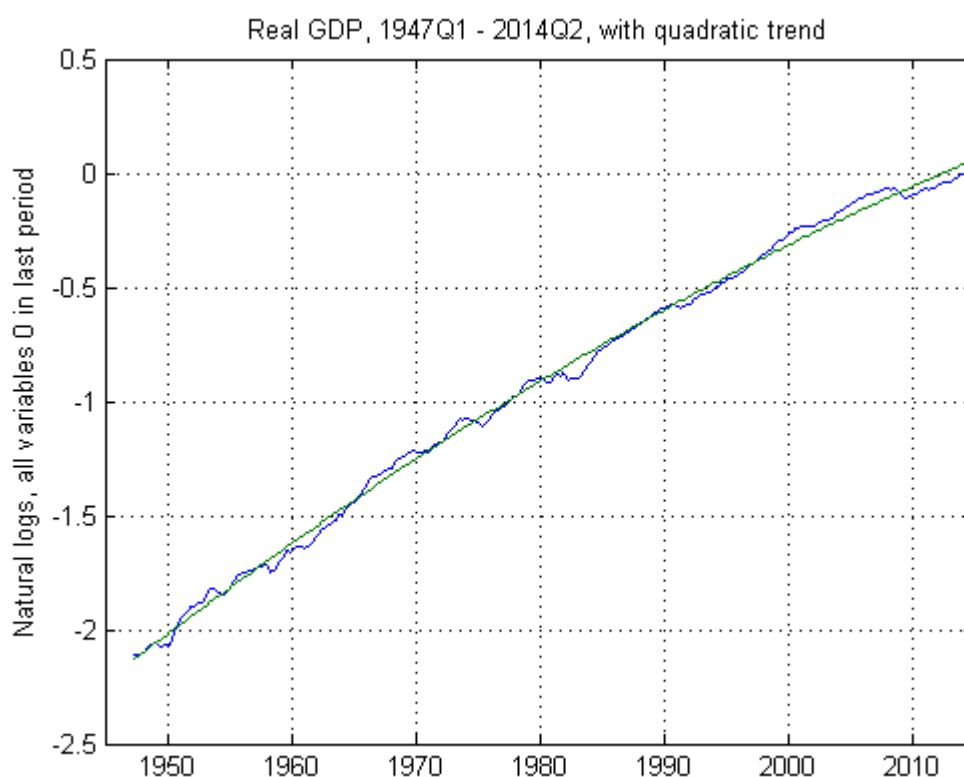


Figure 18
Log Real GDP (blue) and quadratic trend (black).
All variables normalized to 0 in last quarter.

Table 2 gives the OLS estimates of the quadratic parameters, along with several measures of their standard deviations. The estimate of the intercept a indicates that the “output gap” in 2014Q2 was -4.80% relative to the quadratic trend. The estimate of the linear term b indicates that in the same quarter, the expected growth rate of GDP was

down to 2.30% per annum. The estimate of the quadratic term c indicates that GDP growth is falling at a rate of 2.82 (2×1.41) percent per annum per century.

Table 2

$$\log y_t = a + b(t/4) + c(t/4)^2 + \varepsilon_t$$

1947Q1 – 2014Q2 (n = 270 obs.)
 $p^{\text{UR}} = 0.104$

coefficient:	100a	100b	c×10⁴
OLS estimate:	4.796	2.293	-1.408
s.e.:			
OLS	0.613**	0.042**	0.064**
HAC(4)	1.576**	0.096**	0.127**
AR(4)	2.263*	0.155**	0.223**
MR(4)	3.294	0.225**	0.323**
Blend	∞	0.871**	1.144

†, *, ** indicate t-statistic exceeds nominal 10%, 5%, 1% critical value, resp.

The OLS standard errors, taken at face value, indicate that all three coefficients are strongly significantly different from zero. The popular Newey-West HAC standard errors shown in the next row of Table 2 are 98 to 157% greater than the OLS standard errors, when a bandwidth of $m = 4$ is employed, as determined by (26), but still indicate that all three coefficients are strongly significantly different from zero. However, although these standard errors are consistent, in finite samples they are triply biased downwards, as noted in Section 4 above, and hence still give pronounced size distortion, as demonstrated in the simulations above.

The standard errors identified as AR(4) in Table 2 were obtained by fitting an AR(4) model to the first 4 residual autocorrelations, as shown in Table 3, by means of the Yule-Walker equations. These standard errors are another 44 to 76% larger than the HAC standard errors, depending on the coefficient. With these standard errors, the output gap falls from “strongly significant” (test size .01 or lower) to merely “significant” (test size .05 or lower), but the other two coefficients remain strongly significant. However, these standard errors still understate the uncertainty because residual autocorrelations generally show less persistence than the true autoregressive process.

Table 3
Residual autocorrelations and Method of Moments versus
Moment Ratio autoregressive coefficient estimates for
regression of Table 2.

lag j	r_j	$\hat{\varphi}_j^{MM}$	r_j^{MR}	$\hat{\varphi}_j^{MR}$	$\hat{\varphi}_j^{UR}$
1	0.956	1.251	0.971	1.311	1.434
2	0.885	-0.212	0.921	-0.255	-0.362
3	0.800	-0.112	0.860	-0.130	-0.145
4	0.715	0.009	0.799	0.033	0.073
α		0.936		0.958	1.000

The MR(4) coefficients in Table 3 show even more persistence than the AR(4) coefficients. Although the difference in persistence is not great ($\alpha = 0.958$ versus 0.936), the MR(4) standard errors in Table 2 are another 45 to 46% higher than the unadjusted AR(4) standard errors. The “output gap” is now insignificant at even the very weak 10% level, but the other two coefficients remain highly significant.

Although the MR(4) estimate of the persistence α falls short of unity, the unit root test statistic $p^{UR} = 0.105$ so that a unit root cannot be even weakly rejected. This exceeds 10%, so that the preferred “blended” covariance matrix simply equals the unconditional unit root covariance matrix. These standard errors are shown in the last line of Table 2. Although the terminal growth rate of 2.30% per annum remains highly significant, the rate of change of the growth rate is now not even weakly significant.

In sum, although a quadratic trend line fits US log GDP quite well visually, there is no evidence that the estimated “output gap” of -4.80% for 2012Q2 is anything more than a statistical illusion. Even by the pure MR(4) standard error of 3.29%, this value is not significantly different from zero. The preferred blended standard errors are actually infinite. The blended standard errors indicate that the apparent declining trend in the growth rate is not statistically significant. However, the terminal growth rate estimate of 2.30% per annum remains strongly significant, even using the blended covariance matrix, which here is equal to the unconditional unit root covariance matrix and therefore assumes no cointegrating relationship.²⁵

14. Conclusions

The proposed Moment Ratio Estimator for the autoregressive parameters of the errors in an OLS regression is computed from the conventional residual autocorrelation coefficients, but gives far less bias, and provides coefficient confidence intervals with far less size distortion, than the available alternatives. The estimator is in the spirit of the Median Unbiased estimator of Andrews (1993) and McCulloch (2008), but does

²⁵ FGLS based on the MR(4) autocovariance matrix would of course be more efficient than OLS and could give different results. The present paper makes no attempt to implement this.

not require their Monte Carlo simulation, and provides smaller standard error bias in the illustrated case.

The presence of a unit root in the errors, and therefore the absence of a cointegrating relationship, does require reposing the problem, but does not by itself indicate that an OLS correlation between the variables is spurious.

It has been shown that despite their consistency, the popular HAC standard errors of Newey and West (1987) can greatly overstate the precision of OLS coefficient estimates with sample sizes and serial correlation commonly found in economic studies. Although the author does not believe that the regressor-conditional heteroskedasticity targeted by the Newey-West HAC estimator is typically as serious a problem in econometric data as serial correlation, a “Moment Ratio-HAC” procedure is proposed that attempts to incorporate this consideration. Surprisingly, there is relatively little cost in terms of size distortion to employing it when errors are in fact independent of the regressors, at least in the simple model considered.

Although the factorizations that motivate the MR estimator rely on the very strong assumption that the errors are actually independent of the regressors, the author is optimistic that the MR estimator developed here will have favorable properties when the regressors are merely weakly exogenous. Feasible GLS estimates would provide more efficient estimates of the regression coefficients than the OLS estimates considered here, but any such GLS estimates should be based on preliminary MR estimates of the covariance structure.

The author thanks participants at the 2008 Econometric Society North American Summer Meetings, the 14th, 16th, and 17th annual Conference on Computing in Economics and Finance, and the Ohio State University and University of Washington Econometrics Seminars, several anonymous NSF and journal referees, and in particular Robert DeJong, Charles Nelson, Masao Ogaki, Frank Schorfheide, and James Stock for helpful comments and suggestions on previous versions of this paper. All remaining errors and omissions remain the responsibility of the author.

REFERENCES

- Abadir, Karim M., and Paolo Paruolo. 1996. Two Mixed Normal Densities from Cointegration Analysis. *Econometrica* 65:671-80.
- _____. 2002. Simple Robust Testing of Regression Hypotheses: A Comment. *Econometrica* 70:2097-9.
- Andrews, Donald W.K.. 1993. Exactly Median-Unbiased Estimation of First Order Autoregressive / Unit Root Models. *Econometrica* 61:139-65.

- Andrews, Donald W.K., and J. Christopher Monahan. 1992. An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica* 60:953-66.
- Andrews and Chen. 1994. Approximately Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models. *Journal of Business and Economic Statistics* 61:139-65.
- Bartlett, M.S. 1935. Some aspects of the time-correlation problem in regard to tests of significance. *Journal of the Royal Statistical Society* 98: 536-43.
- Bollerslev, Tim. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *J. Econometrics* 31: 1-50.
- Bunzel, Helle, and Timothy L. Vogelsang. 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* 23: 381-94.
- Canjels, Eugene, and Mark W. Watson. 1997. Estimating Deterministic Trends in the Presence of Serially Correlated Errors. *Review of Economics and Statistics* 79: 184-200.
- Choi, Chi-Young, Ling Hu, and Masao Ogaki. 2008. Robust Estimation for Structural Spurious Regressions and a Hausman-type Cointegration Test. *Journal of Econometrics* 142: 327-351.
- Den Haan, Wouter J., and Andrew Levin. 1997. A Practitioner's Guide to Robust Covariance Matrix Estimation. In G.S. Maddala and C.R. Rao, eds., *Robust Inference (Handbook of Statistics 15)*. Elsevier.
- Greene, William H. 2003. *Econometric Analysis*, 5th ed. Prentice Hall.
- Harvey, David I., Stephen J. Leybourne and A.M. Robert Taylor. A Simple, Robust and Powerful Test of the Trend Hypothesis. *Journal of Econometrics* 141: 1302-30.
- Hayashi, Fumio. 2000. *Econometrics*. Princeton Univ. Press.
- Hosking, J.M.R. 1984. Modeling Persistence in Hydrological Time Series Using Fractional Differencing. *Water Resources Research* 20: 1898-1908.
- Kiefer, Nicholas M., and Timothy J. Vogelsang. 2002. Heteroskedasticity-and-Autocorrelation Robust Standard Errors using the Bartlett Kernel without Truncation. *Econometrica* 70: 2093-5.
- Kiefer, Nicholas M., Timothy J. Vogelsang, and Helle Bunzel. 2000. Simple Robust Testing of Regression Hypotheses. *Econometrica* 68: 695-714.

McCulloch, J. Huston. 1985. Interest Risk-Sensitive Deposit Insurance Premia: Stable ACH Estimates. *Journal of Banking and Finance* 9: 137-56.

_____. 2008. Median-Unbiased Estimation of Higher Order Autoregressive/Unit Root Processes and Autocorrelation Consistent Covariance Estimation in a Money Demand Model. Presented at Econometric Society 2008 North American Summer Meetings, Pittsburgh, June 19-22, and at 14th International Conference on Computing in Economics and Finance, Paris, June 26-28, Online at <<http://www.econ.ohio-state.edu/jhm/papers/MUARM1S.pdf>>

Nelson, Charles R., and Heejoon Kang. 1981. Spurious Periodicity in Inappropriately Detrended Time Series. *Econometrica* 49: 741-51.

Nelson, Charles R., and Charles Plosser. 1982. Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. *Journal of Monetary Economics* 10: 139-62.

Newey, Whitney, and Kenneth West. 1987. A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703-8.

_____. 1994. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economics Studies* 61: 631-53.

Paolella, Marc S. 2003. Computing Moments of Ratios of Quadratic Forms in Normal Variables. *Computational Statistics & Data Analysis* 42: 313-331.

Percival, Donald B. 1993. Three Curious Properties of the Sample Variance and Autocovariance for Stationary Processes with Unknown Mean. *The American Statistician* 47: 274-6.

Peron, Pierre, and Tomoyoshi Yabu. 2009. Estimating Deterministic Trends with an Integrated or Stationary Noise Component. *Journal of Econometrics* 151: 56-69.

Politis, Dimitris N. 2011. Higher-Order Accurate, Positive Semidefinite Estimation of Large-Sample Covariance and Spectral Density Matrices. *Econometric Theory* 27: 703-44.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. 1992. *Fortran Numerical Recipes*, Second Edition. Cambridge University Press.

Quenouille, M.H. 1952. *Associated Measurements*. London: Butterworths Scientific Publications.

Stock, James H., and Mark W. Watson. 2007. *Introduction to Econometrics*, 2nd ed. Pearson/Addison-Wesley.

Sun, Yixiao, Peter C.B. Phillips, and Sainan Jin. 2008. Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing. *Econometrica* **76**: 175-94.

White, Halbert. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48:817-38.