MA Advanced Econometrics: Finite-Sample Distributions, Bootstrapping

Karl Whelan

School of Economics, UCD

February 22, 2011

Finite Sample Distributions

- Because it is well known that OLS estimates of time series regression models are consistent when they feature $I(0)$ series while they are inconsistent and generate non-standard distributions when using $I(1)$ series, econometric textbooks tend to stress a strong dichotomy between the stationary and non-stationary series. This gets reflected in a lot of econometric practise.
- The message—that things change drastically when we move from an $I(0)$ series to a unit root series—is somewhat misleading. Practical applications do not use infinite amounts of data and the speeds at which time series estimates converge to their asymptotic distributions is often very slow.
- In truth, for any given sample size, there is no great jump in the behaviour as we go from $\rho < 1$ to $\rho = 1$. Many of the problems that occur with unit root series also apply to high values of ρ.
- Here I'll illustrate these points and then move on to discussing some ways to deal with them.

The Bias of OLS AR(1) Estimates

- Recall that for the AR(1) model, the OLS estimate can be written as

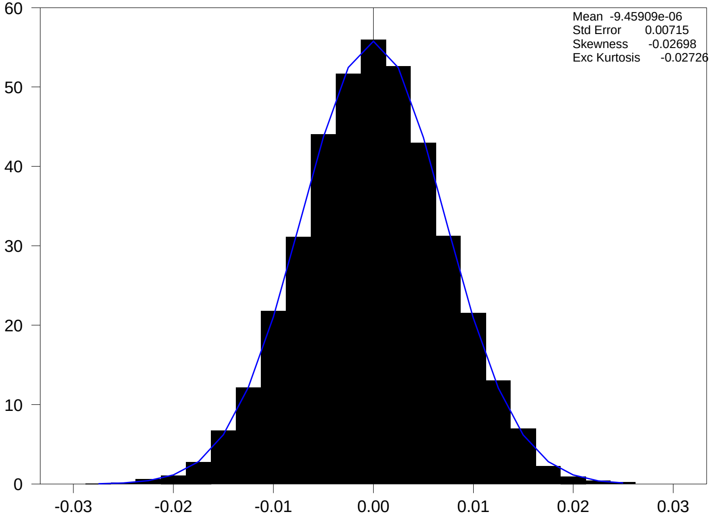
$$\hat{\rho} = \rho + \sum_{t=2}^{T} \left(\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \right) \epsilon_t \tag{1}$$

- ϵ_t is independent of y_{t-1}, so $\mathbb{E}\left(y_{t-1}\epsilon_t\right) = 0$. However, ϵ_t is not independent of the sum $\sum_{t=2}^{T} y_{t-1}^2$.
- If ρ is positive, then a positive shock ϵ_t raises current and future values of y_{t+k}, all of which are in the sum $\sum_{t=2}^{T} y_{t-1}^2$. This means there is a negative correlation between ϵ_t and $\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$, so $\mathbb{E}\,\hat{\rho} < \rho$.
- The size of the bias depends positively on two factors:
  1. The size of ρ: The bigger this is, the stronger the correlation of the shock with future values.
  2. The sample size T: The larger this is, the smaller the fraction of the observations sample that will be highly correlated with the shock.

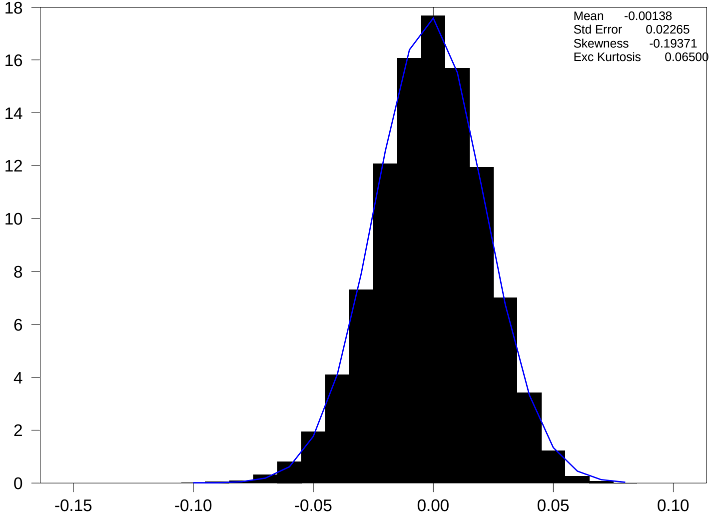
Example: AR(1) Bias When $\rho = 0.7$

- The next few slides illustrate the bias in $\hat{\rho}$ when estimating $AR(1)$ regressions using OLS.
- In each case, we report the distribution of the bias of OLS estimates $\hat{\rho} - \rho$ when the true value of $\rho = 0.7$ but we vary the sample size.
- In the first chart, the sample size is $T = 10,000$ and the asymptotic theory is working very well: There is no bias and the distribution of $\hat{\rho}$ is normal.
- When the sample size if $T = 1,000$, you can just about see the asymptotic theory starting to fail: There is a small average bias of -0.13 and the distribution is a tiny bit skewed.
- As the samples get smaller, the bias gets larger and the distributions become more skewed. By the time we get to $T = 30$, the bias is as large as -0.045 while for $T = 10$ the bias is -0.116.

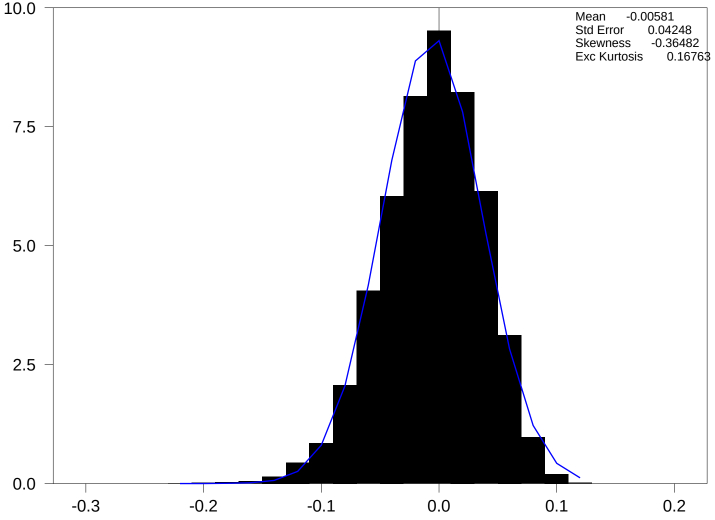
Bias From $AR(1)$ Regression, $\rho = 0.7$, $T = 10000$



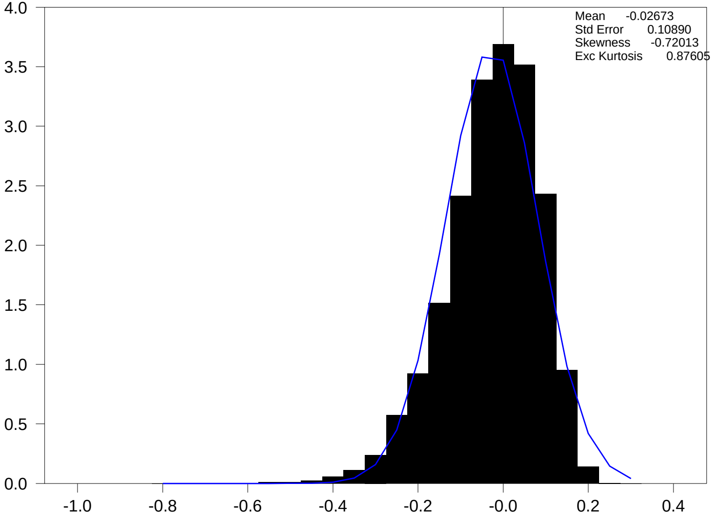
Bias From $AR(1)$ Regression, $\rho = 0.7$, $T = 1000$



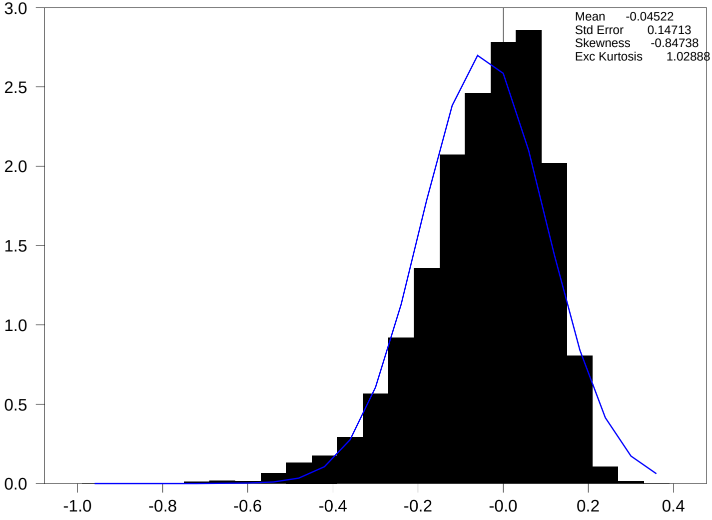
Bias From $AR(1)$ Regression, $\rho = 0.7$, $T = 300$



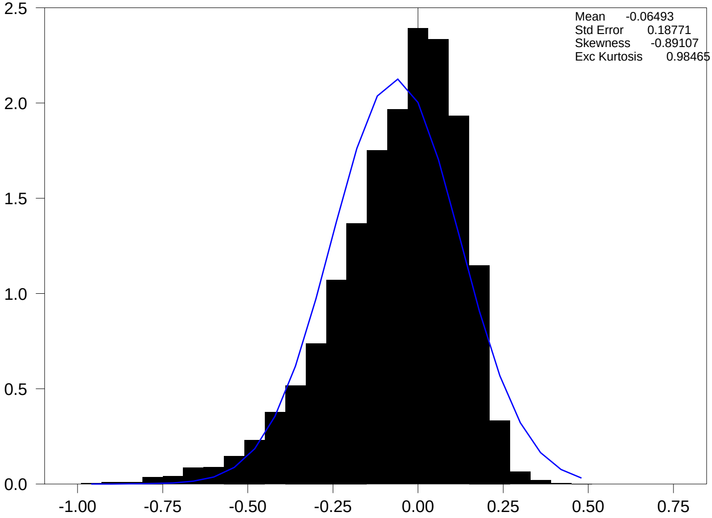
Bias From $AR(1)$ Regression, $\rho = 0.7$, $T = 50$



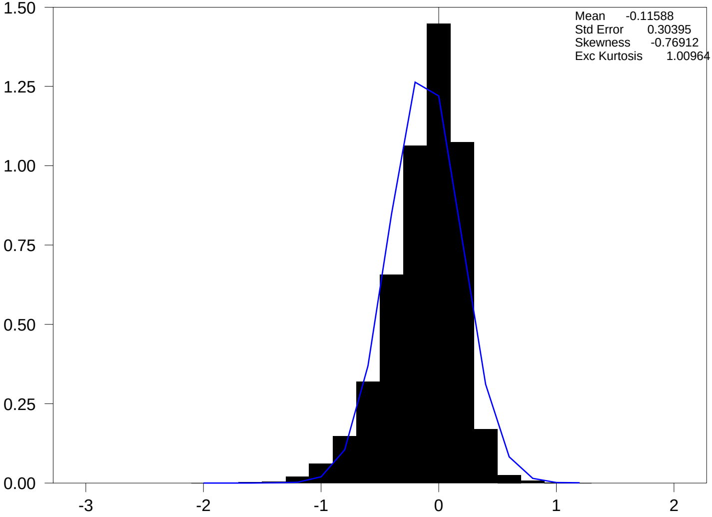
Bias From $AR(1)$ Regression, $\rho = 0.7$, $T = 30$



Bias From $AR(1)$ Regression, $\rho = 0.7$, $T = 20$



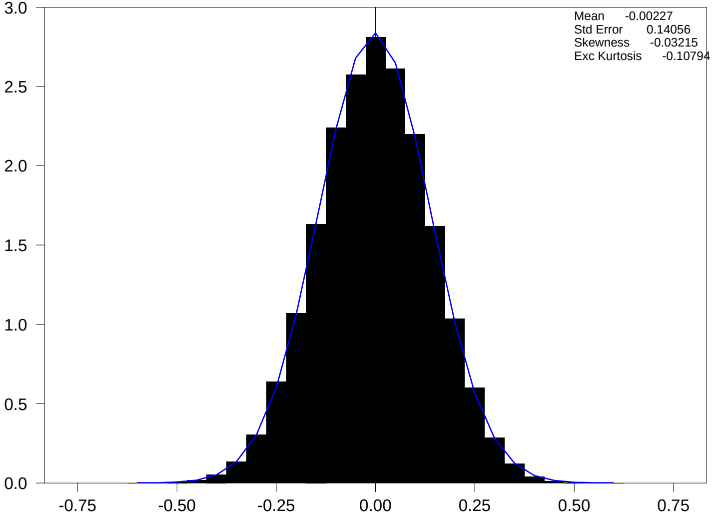
Bias From $AR(1)$ Regression, $\rho = 0.7$, $T = 10$



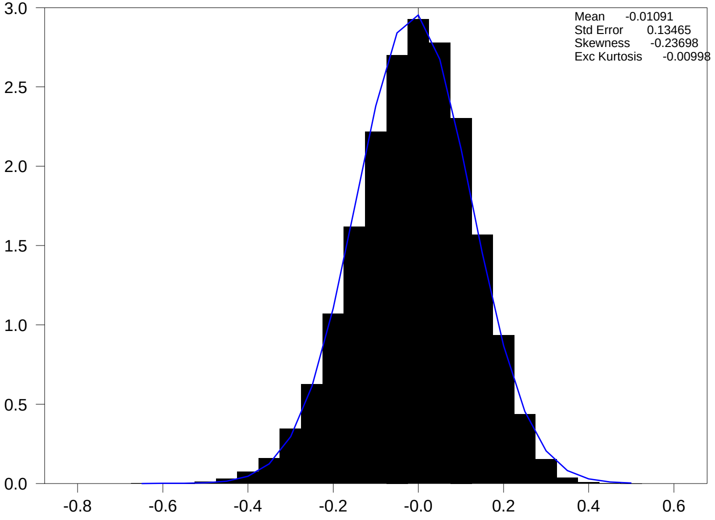
Example: AR(1) Bias for $T = 50$ as ρ Increases

- The next few slides repeat the process of showing distributions of the bias of OLS estimates $\hat{\rho} - \rho$ but in this case, we vary the value of ρ instead of the sample size, which is kept fixed at $T = 50$.
- Our first chart shows the bias when $\rho = 0.05$, so the series is almost white noise, meaning the observations are close to being i.i.d. The logic of the Lindberg-Levy Central Limit Theorem for i.i.d. observations works well here and the estimator has a Normal distribution.
- For $\rho = 0.3$ and $\rho = 0.5$, there is some bias and the distribution becomes a bit more skewed.
- By $\rho = 0.8$, the distribution is highly skewed and the bias is -0.03.
- The skewness in the distribution increases all the way up to $\rho = 1$. But note that there is no great jump in the size of the bias or the shape of the distribution as ρ goes from 0.99 to 1. The asymptotic theory for $\rho = 0.99$ may be completely different from the theory for $\rho = 1$ but in finite samples there is no great difference.

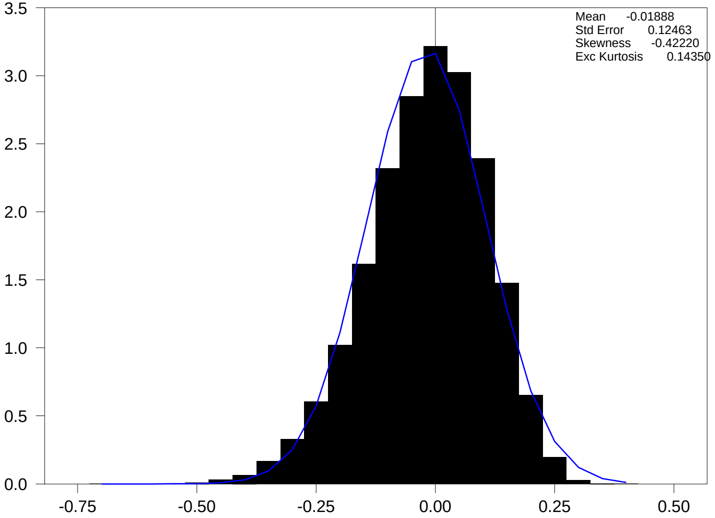
Bias From $AR(1)$ Regression, $\rho = 0.05$, $T = 50$



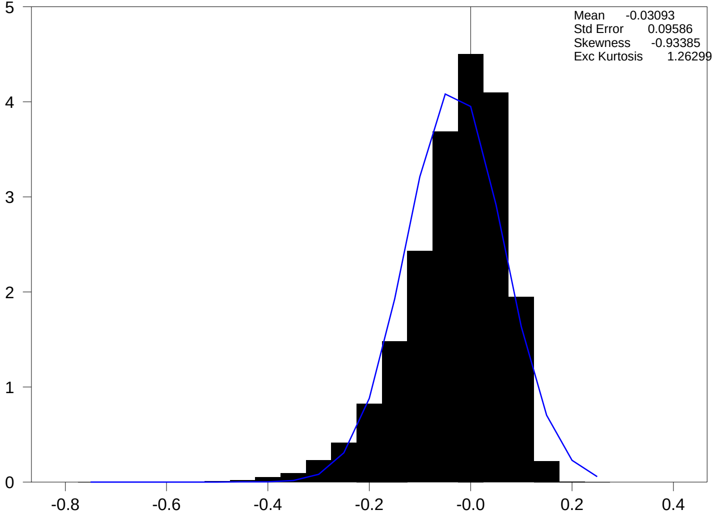
Bias From $AR(1)$ Regression, $\rho = 0.30$, $T = 50$



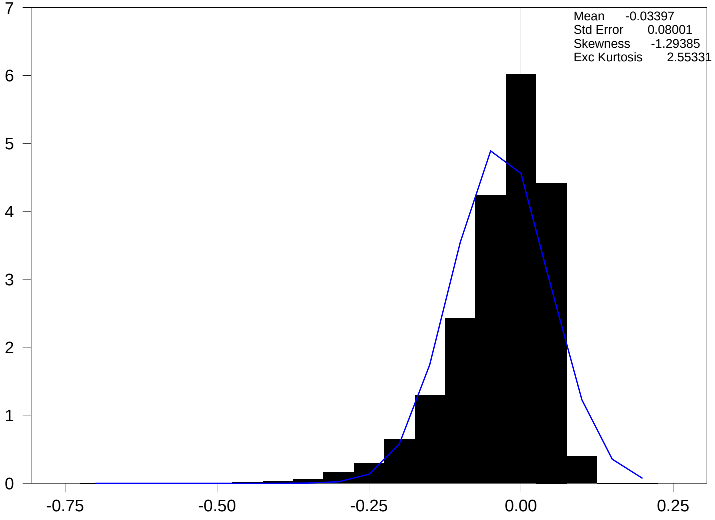
Bias From $AR(1)$ Regression, $\rho = 0.50$, $T = 50$



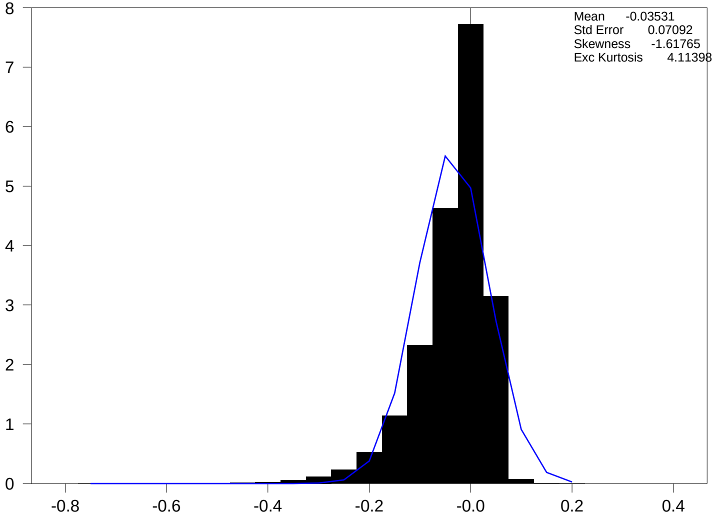
Bias From $AR(1)$ Regression, $\rho = 0.80$, $T = 50$



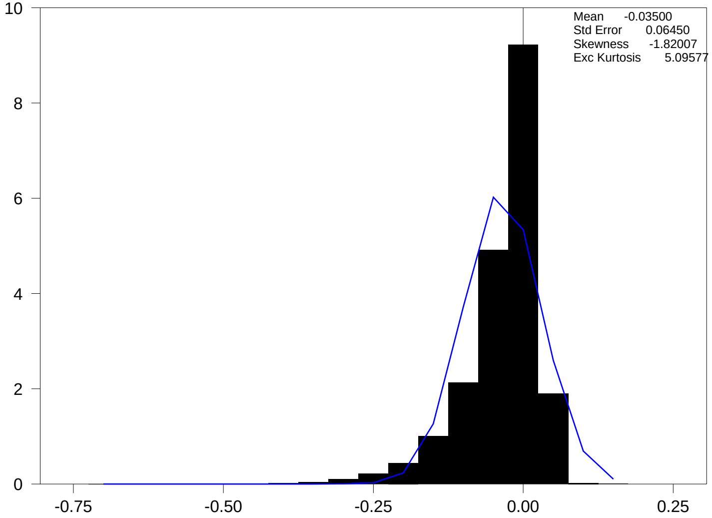
Bias From $AR(1)$ Regression, $\rho = 0.90$, $T = 50$



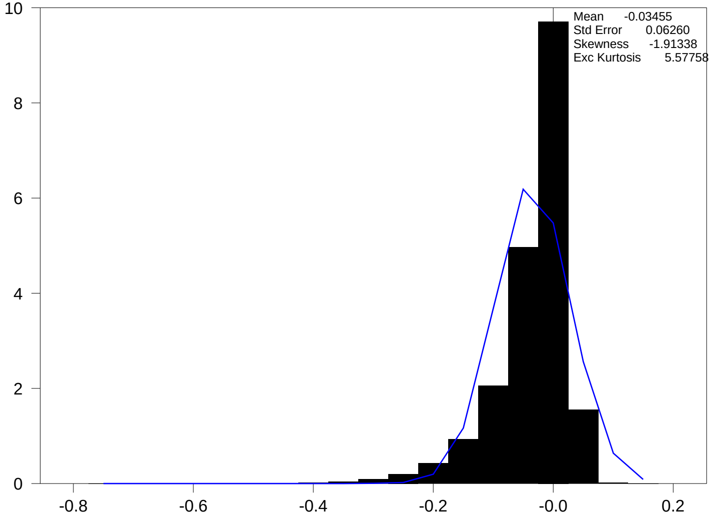
Bias From $AR(1)$ Regression, $\rho = 0.95$, $T = 50$



Bias From $AR(1)$ Regression, $\rho = 0.99$, $T = 50$



Bias From $AR(1)$ Regression, $\rho = 1$, $T = 50$



Spurious Regressions Without Nonstationarity

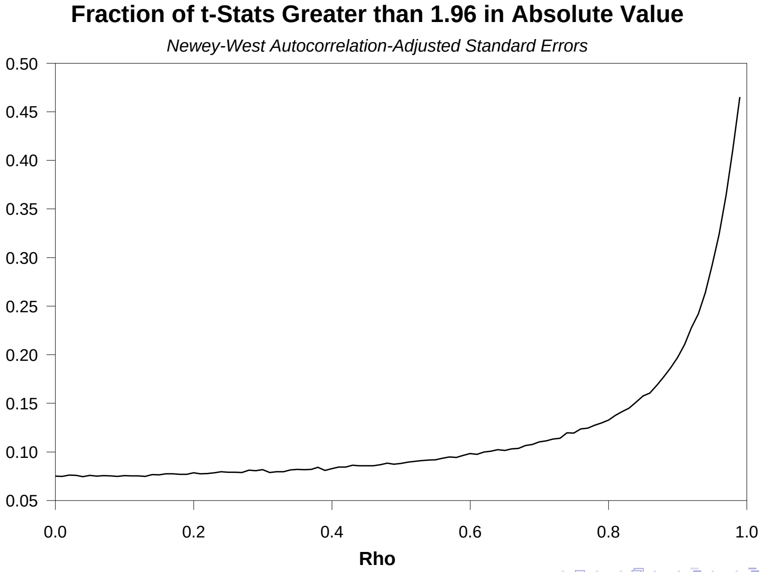
- We know that when we regress one $I(1)$ series on another, we can get spuriously significant coefficients. What is less well known is that the problem of spuriously significant results can also occur with stationary series.
- The next page illustrates results from simulations in which we take two stationary series

$$y_t = \rho y_{t-1} + \epsilon_t^y \quad (2)$$

$$x_t = \rho x_{t-1} + \epsilon_t^x \quad (3)$$

  and regress y_t and x_t for various values of ρ for a sample of $T = 200$.
- According to the asymptotic distribution, t statistics greater than 1.96 in absolute value should only be observed 5% of the time.
- However, the figure shows that even when we adjust for autocorrelation (using the Newey-West heteroskedastic and autocorrelation consistent covariance matrix) the fraction of t statistics greater than 1.96 rises well above five percent as ρ increases.

Regressing Two $AR(1)$ Series With Common Value of ρ On Each Other, $T = 200$



Median Unbiased Estimates and Confidence Intervals

- Given that we know OLS estimates of $AR(1)$ models are biased, is there a way to get better estimates?
- Andrews (1993) provided calculations of the distributions of OLS estimators for various values of ρ and for various sample sizes under the assumption of Normally distributed errors. These kinds of calculations can be used to provide new estimates of ρ and confidence intervals.
- Use Monte Carlo simulation methods to simulate the distribution of OLS estimators for each value of ρ for a sample size of T. Label the 5th percentile of the resulting OLS estimators $q_5(\rho)$, the median $q_{50}(\rho)$ and the 95th percentile value $q_{95}(\rho)$. Define the inverse function q_α^{-1} such that $q_\alpha^{-1}(q_\alpha(\rho)) = \rho$.
- If one obtains a value of $\hat{\rho}$ from a sample of size T, then the **median-unbiased estimator** of ρ is the value such $q_{50}(\rho) = \hat{\rho}$. In other words, it's the value of ρ such that when this is the true value, you are as likely to get an OLS estimate above $\hat{\rho}$ as you are to get one below.
- A $2\alpha\%$ confidence interval can be constructed as $\left(q_{1-\alpha}^{-1}(\hat{\rho}), q_\alpha^{-1}(\hat{\rho})\right)$. The probability of observing $\hat{\rho}$ equals α percent for the values at both ends of this interval.

Bootstrap Confidence Intervals for AR Models

- In many cases, it is not accurate to assume that the error terms in AR models are Normally distributed.
- An alternative is to use **bootstrap** methods. For example, consider the case of the $AR(1)$ model, $y_t = \alpha + \rho y_{t-1} + \epsilon_{t-1}$. We can use simulation methods that mimic the distribution of the in-sample residuals, whether or not these residuals appear to be normally distributed.
- Bruce Hansen (1999) describes a **grid bootstrap** method that works roughly as follows:
  1. Estimate the model via OLS to obtain residuals $\hat{\epsilon}_t$.
  2. For a wide range of values of ρ, construct new simulated series by making an assumption about the initial value y_0^* and setting $y_k^* = \alpha + \rho y_{k-1}^* + \epsilon_k^*$ by picking the ϵ_k^* from randomly choosing values from from $\hat{\epsilon}_t$.
  3. For each value of ρ generate a distribution of OLS estimates from the simulated series and save the quantiles $q_p(\rho)$.
  4. As with the Andrews method, median unbiased estimates can be defined as $q_{50}^{-1}(\hat{\rho})$ and confidence intervals constructed as $\left(q_{1-\alpha}^{-1}(\hat{\rho}), q_{\alpha}^{-1}(\hat{\rho})\right)$

Bootstrapping Standard Errors for VARs

- After estimating a VAR model $Z_t = AZ_{t-1} + \epsilon_t$ it is common to present the impulse response functions. In this reduced-form VAR, these IRFs are $I, A, A^2, ...$ What is the sampling distribution of these estimates?
- If the VAR is estimated via OLS, then the standard asymptotic results apply, and the coefficients in A have a limiting normal distribution. The IRFs are nonlinear functions of these coefficients so we can use the Delta method to get approximations to the asymptotic distributions of the IRF estimates. Unfortunately, these estimates are not very accurate in finite samples.
- Most VAR practitioners now use bootstrap methods.
  1. Estimate the VAR via OLS and save the errors $\hat{\epsilon}_t$.
  2. Randomly sample from these errors to create, for example, 10,000 simulated data series $Z_t^* = \hat{A}Z_{t-1}^* + \epsilon_t^*$.
  3. Estimate a VAR model on the simulated data and save the 10,000 IRFs associated with these estimate.
  4. Calculate quantiles of the simulated IRFs, e.g. of the 10,000 estimates of the effect in period 2 on variable i of shock j.
  5. Use the 5th and 95th quantiles of the simulated IRFs as confidence intervals.