Bootstrap Methods for Time Series Data

Robert Taylor
University of Essex

Annual Granger Lecture
University of Nottingham
November 21st, 2023


BOOTSTRAPS

Outline

Moving on to ...

1. Introduction
2. The Basics of Bootstrap Hypothesis Testing
3. Some Popular Bootstrap Resampling Methods
4. Application 1: Unit Root Testing
5. Application 2: Testing for Bubbles
6. Application 3: Testing for Predictability of Returns
7. Conclusions

- Bootstrap methods involve estimating a model many times using simulated data. Quantities computed from the simulated data are then used to make inferences from the actual data. The term *bootstrap* was coined by Efron (1979, AoS). One major reason for their increasing popularity in recent years is the staggering drop in the cost of numerical computation.

- Bootstrap methods involve estimating a model many times using simulated data. Quantities computed from the simulated data are then used to make inferences from the actual data. The term *bootstrap* was coined by Efron (1979, AoS). One major reason for their increasing popularity in recent years is the staggering drop in the cost of numerical computation.
- Although bootstrapping is widely used, it is not always well understood. In practice, bootstrapping is often not as easy to do, and does not work as well, as seems to be widely believed.

- Bootstrap methods involve estimating a model many times using simulated data. Quantities computed from the simulated data are then used to make inferences from the actual data. The term *bootstrap* was coined by Efron (1979, AoS). One major reason for their increasing popularity in recent years is the staggering drop in the cost of numerical computation.
- Although bootstrapping is widely used, it is not always well understood. In practice, bootstrapping is often not as easy to do, and does not work as well, as seems to be widely believed.
- There are many different bootstrap methods. Some are very easy to implement, and some can work extraordinarily well. But bootstrap methods do not always work well, and choosing among alternative ones is often not easy.

- It is well-known that the bootstrap, when correctly implemented, can be an important device to compute critical values or p-values of a statistical test in samples of finite size.

- It is well-known that the bootstrap, when correctly implemented, can be an important device to compute critical values or p-values of a statistical test in samples of finite size.
- The bootstrap may be used either to estimate quantiles of an unknown limiting distribution...

- It is well-known that the bootstrap, when correctly implemented, can be an important device to compute critical values or p-values of a statistical test in samples of finite size.
- The bootstrap may be used either to estimate quantiles of an unknown limiting distribution...
- ... or to deliver finite sample refinements/improved approximations to statistical quantities of interest ...

- It is well-known that the bootstrap, when correctly implemented, can be an important device to compute critical values or p-values of a statistical test in samples of finite size.
- The bootstrap may be used either to estimate quantiles of an unknown limiting distribution...
- ... or to deliver finite sample refinements/improved approximations to statistical quantities of interest ...
- ... or to estimate quantities that might be hard to quantify (eg the standard error of the sample *median*)

- It is well-known that the bootstrap, when correctly implemented, can be an important device to compute critical values or p-values of a statistical test in samples of finite size.
- The bootstrap may be used either to estimate quantiles of an unknown limiting distribution...
- ... or to deliver finite sample refinements/improved approximations to statistical quantities of interest ...
- ... or to estimate quantities that might be hard to quantify (eg the standard error of the sample *median*)
- Widely applied in statistics and econometrics, perhaps less so in time series econometrics.

- Leading applications in time series econometrics include:

- Leading applications in time series econometrics include:
  - Unit root and co-integration testing
  - GARCH volatility modeling
  - Predictive regressions
  - Extreme events/inference without moments
  - Bubble modeling and testing
  - Non causal models
  - Double AR models
  - Improved estimation of VaR models
  - Point process models and extreme returns
  - Realised volatility
  - Fractionally integrated models

- Leading applications in time series econometrics include:
  - Unit root and co-integration testing
  - GARCH volatility modeling
  - Predictive regressions
  - Extreme events/inference without moments
  - Bubble modeling and testing
  - Non causal models
  - Double AR models
  - Improved estimation of VaR models
  - Point process models and extreme returns
  - Realised volatility
  - Fractionally integrated models
- Still, compared to other areas, the bootstrap is arguably under used in time series econometrics and empirical finance. Why?

- Leading applications in time series econometrics include:
    - Unit root and co-integration testing
    - GARCH volatility modeling
    - Predictive regressions
    - Extreme events/inference without moments
    - Bubble modeling and testing
    - Non causal models
    - Double AR models
    - Improved estimation of VaR models
    - Point process models and extreme returns
    - Realised volatility
    - Fractionally integrated models
- Still, compared to other areas, the bootstrap is arguably under used in time series econometrics and empirical finance. Why?
    - computational time?
    - invalidity?
    - difficulty in validly implementing?

- In most of these applications, the “standard” (nonparametric) bootstrap (based on i.i.d. resampling) does not work, for several reasons, possibly including:

- In most of these applications, the "standard" (nonparametric) bootstrap (based on i.i.d. resampling) does not work, for several reasons, possibly including:
  - lack of moments
  - lack of level stationarity
  - non-stationary stochastic volatility
  - the presence of nuisance parameters the bootstrap does not replicate

- In most of these applications, the “standard” (nonparametric) bootstrap (based on i.i.d. resampling) does not work, for several reasons, possibly including:
  - lack of moments
  - lack of level stationarity
  - non-stationary stochastic volatility
  - the presence of nuisance parameters the bootstrap does not replicate
- These features can lead to random limit bootstrap (conditional) measures

- In most of these applications, the “standard” (nonparametric) bootstrap (based on i.i.d. resampling) does not work, for several reasons, possibly including:
  - lack of moments
  - lack of level stationarity
  - non-stationary stochastic volatility
  - the presence of nuisance parameters the bootstrap does not replicate
- These features can lead to random limit bootstrap (conditional) measures
- This does not mean that the bootstrap does not work in general ... rather that (asymptotic) validity requires that the bootstrap is correctly implemented.

Examples of failure of the standard bootstrap in time series:

Examples of failure of the standard bootstrap in time series:

- inference in the presence of unit roots and common stochastic trends/factors

Examples of failure of the standard bootstrap in time series:

- inference in the presence of unit roots and common stochastic trends/factors
- inference in predictive regressions with strongly persistent predictors

Examples of failure of the standard bootstrap in time series:

- inference in the presence of unit roots and common stochastic trends/factors
- inference in predictive regressions with strongly persistent predictors
- regression with persistent stochastic volatility, including IGARCH

Examples of failure of the standard bootstrap in time series:

- inference in the presence of unit roots and common stochastic trends/factors
- inference in predictive regressions with strongly persistent predictors
- regression with persistent stochastic volatility, including IGARCH
- inference in GARCH processes when a parameter is on the boundary of the parameter space

Examples of failure of the standard bootstrap in time series:

- inference in the presence of unit roots and common stochastic trends/factors
- inference in predictive regressions with strongly persistent predictors
- regression with persistent stochastic volatility, including IGARCH
- inference in GARCH processes when a parameter is on the boundary of the parameter space
- inference in causal/non-causal autoregessions with infinite variance innovations

Moving on to ...

1. Introduction
2. The Basics of Bootstrap Hypothesis Testing
3. Some Popular Bootstrap Resampling Methods
4. Application 1: Unit Root Testing
5. Application 2: Testing for Bubbles
6. Application 3: Testing for Predictability of Returns
7. Conclusions

The Basics of Bootstrap Hypothesis Testing

- Suppose that $\hat{\tau}$ is the outcome of a test statistic, τ. If we knew the (exact) cumulative distribution function (CDF) of τ under the null hypothesis, say $F(\tau)$, we would reject the null hypothesis whenever $\hat{\tau}$ is abnormal in some sense. For a test that rejects in the upper tail of the distribution, we might choose to calculate a critical value at level α, say c_α, as defined by the equation, $1 - F(c_\alpha) = \alpha$.

The Basics of Bootstrap Hypothesis Testing

- Suppose that $\hat{\tau}$ is the outcome of a test statistic, τ. If we knew the (exact) cumulative distribution function (CDF) of τ under the null hypothesis, say $F(\tau)$, we would reject the null hypothesis whenever $\hat{\tau}$ is abnormal in some sense. For a test that rejects in the upper tail of the distribution, we might choose to calculate a critical value at level α, say c_α, as defined by the equation, $1 - F(c_\alpha) = \alpha$.
- Then we would reject the null whenever $\hat{\tau} > c_\alpha$. For example, when $F(\tau)$ is the $\chi^2(1)$ distribution and $\alpha = 0.05$, $c_\alpha = 3.84$.

The Basics of Bootstrap Hypothesis Testing

- Suppose that $\hat{\tau}$ is the outcome of a test statistic, τ. If we knew the (exact) cumulative distribution function (CDF) of τ under the null hypothesis, say $F(\tau)$, we would reject the null hypothesis whenever $\hat{\tau}$ is abnormal in some sense. For a test that rejects in the upper tail of the distribution, we might choose to calculate a critical value at level α, say c_α, as defined by the equation, $1 - F(c_\alpha) = \alpha$.
- Then we would reject the null whenever $\hat{\tau} > c_\alpha$. For example, when $F(\tau)$ is the $\chi^2(1)$ distribution and $\alpha = 0.05$, $c_\alpha = 3.84$.
- An alternative approach, is to calculate the *p*-value, or marginal significance level, $p(\hat{\tau}) = 1 - F(\hat{\tau})$ and reject whenever $p(\hat{\tau}) < \alpha$.

The Basics of Bootstrap Hypothesis Testing

- Suppose that $\hat{\tau}$ is the outcome of a test statistic, τ. If we knew the (exact) cumulative distribution function (CDF) of τ under the null hypothesis, say $F(\tau)$, we would reject the null hypothesis whenever $\hat{\tau}$ is abnormal in some sense. For a test that rejects in the upper tail of the distribution, we might choose to calculate a critical value at level α, say c_α, as defined by the equation, $1 - F(c_\alpha) = \alpha$.
- Then we would reject the null whenever $\hat{\tau} > c_\alpha$. For example, when $F(\tau)$ is the $\chi^2(1)$ distribution and $\alpha = 0.05$, $c_\alpha = 3.84$.
- An alternative approach, is to calculate the *p*-value, or marginal significance level, $p(\hat{\tau}) = 1 - F(\hat{\tau})$ and reject whenever $p(\hat{\tau}) < \alpha$.
- In most cases of interest, however, we do not know $F(\tau)$.

- Until recently, the usual approach in such cases has been to replace it by an approximate CDF, say $F^\infty(\tau)$, based on asymptotic theory. This approach works well when $F^\infty(\tau)$ is a good approximation to $F(\tau)$, but that is by no means always true.

- Until recently, the usual approach in such cases has been to replace it by an approximate CDF, say $F^\infty(\tau)$, based on asymptotic theory. This approach works well when $F^\infty(\tau)$ is a good approximation to $F(\tau)$, but that is by no means always true.
- The bootstrap provides another way to approximate $F(\tau)$, which may provide a better approximation. It can be used even when τ is complicated to compute and difficult to analyse theoretically. The asymptotic distribution of τ need not even be known.

- Until recently, the usual approach in such cases has been to replace it by an approximate CDF, say $F^\infty(\tau)$, based on asymptotic theory. This approach works well when $F^\infty(\tau)$ is a good approximation to $F(\tau)$, but that is by no means always true.
- The bootstrap provides another way to approximate $F(\tau)$, which may provide a better approximation. It can be used even when τ is complicated to compute and difficult to analyse theoretically. The asymptotic distribution of τ need not even be known.
- To perform a bootstrap test, we generate B bootstrap samples that satisfy the null. A bootstrap sample is a simulated data set. The procedure for generating the bootstrap samples, which always involves a random number generator, is called a bootstrap data generating process, or bootstrap DGP

- For each of the $j = 1, ..., B$ bootstrap samples, compute a bootstrap statistic, say τ_j^*, usually by the same procedure used to calculate $\hat{\tau}$. The bootstrap *p*-value is then

$$\hat{p}^*(\hat{\tau}) = \frac{1}{B}\sum_{j=1}^{B} I(\tau_j^* > \hat{\tau})$$

where $I(\cdot)$ is the indicator function, equal to 1 (0) when its argument is true (false).

- For each of the $j = 1, ..., B$ bootstrap samples, compute a bootstrap statistic, say τ_j^*, usually by the same procedure used to calculate $\hat{\tau}$. The bootstrap *p*-value is then

$$\hat{p}^*(\hat{\tau}) = \frac{1}{B}\sum_{j=1}^{B} I(\tau_j^* > \hat{\tau})$$

  where $I(\cdot)$ is the indicator function, equal to 1 (0) when its argument is true (false).
- This can also be written as $\hat{p}^*(\hat{\tau}) = 1 - \hat{F}^*(\hat{\tau})$ where $\hat{F}^*(\tau)$ is the empirical distribution function [EDF] of the τ_j^*. As $B \to \infty$, $\hat{F}^*(\hat{\tau})$ converges to the true (common) CDF of the τ_j^*, $F^*(\tau)$.
- The bootstrap p-value looks just like the true p-value, but with the EDF of the bootstrap distribution, $\hat{F}^*(\hat{\tau})$, replacing the unknown CDF $F(\hat{\tau})$.
- From this, it is clear that bootstrap tests will generally not be exact. However, most of the problems with bootstrap tests arise not because $\hat{F}^*(\tau)$ is only an estimate of $F^*(\tau)$ but, as alluded to before, because $F^*(\tau)$ may not be a good approximation to $F(\tau)$.

- There is an important special case in which bootstrap tests are exact. For this result to hold, we need two conditions:

- There is an important special case in which bootstrap tests are exact. For this result to hold, we need two conditions:
  1. The test statistic τ is pivotal, which means that its (exact) distribution does not depend on any unknown parameters.

- There is an important special case in which bootstrap tests are exact. For this result to hold, we need two conditions:
  1. The test statistic τ is pivotal, which means that its (exact) distribution does not depend on any unknown parameters.
  2. The number of bootstrap samples B is such that $\alpha(B+1)$ is an integer, where α is the level of the test.

- There is an important special case in which bootstrap tests are exact. For this result to hold, we need two conditions:
  1. The test statistic τ is pivotal, which means that its (exact) distribution does not depend on any unknown parameters.
  2. The number of bootstrap samples B is such that $\alpha(B+1)$ is an integer, where α is the level of the test.
- In such cases a bootstrap test is also called a Monte Carlo test.

- There is an important special case in which bootstrap tests are exact. For this result to hold, we need two conditions:
    1. The test statistic τ is pivotal, which means that its (exact) distribution does not depend on any unknown parameters.
    2. The number of bootstrap samples B is such that $\alpha(B+1)$ is an integer, where α is the level of the test.
- In such cases a bootstrap test is also called a Monte Carlo test.
- The classic one-sample t-test, for example, satisfies the first condition when the data are independent draws from a population which is $N(\mu, \sigma^2)$.

- There is an important special case in which bootstrap tests are exact. For this result to hold, we need two conditions:
  1. The test statistic τ is pivotal, which means that its (exact) distribution does not depend on any unknown parameters.
  2. The number of bootstrap samples B is such that $\alpha(B+1)$ is an integer, where α is the level of the test.
- In such cases a bootstrap test is also called a Monte Carlo test.
- The classic one-sample t-test, for example, satisfies the first condition when the data are independent draws from a population which is $N(\mu, \sigma^2)$.
- The second condition is why you often see choices like $B = 999$.

- Most test statistics we encounter in financial econometrics are not pivotal. Nevertheless, provided they are properly implemented, bootstrap tests will often work better than asymptotic tests. For statistics with pivotal limiting null distributions one can in certain circumstances show that bootstrap methods can deliver a *refinement* to the asymptotic approximation (this is a theoretical device to show that it provides a better approximation to the exact distribution of the statistic).

- Most test statistics we encounter in financial econometrics are not pivotal. Nevertheless, provided they are properly implemented, bootstrap tests will often work better than asymptotic tests. For statistics with pivotal limiting null distributions one can in certain circumstances show that bootstrap methods can deliver a *refinement* to the asymptotic approximation (this is a theoretical device to show that it provides a better approximation to the exact distribution of the statistic).
- For the bootstrap to be asymptotically valid, we need $F^*(\tau)$ and $F(\tau)$ to coincide to first order in large samples. Notice the choice of B has no bearing on this; it is a property that needs to hold for each bootstrap statistic. Where $F(\tau)$ is not asymptotically pivotal the bootstrap can still be asymptotically valid though refinements will not be possible, unless a *double bootstrap* method is used, in which case it is possible under certain circumstances.

Moving on to ...

Some Popular Bootstrap Resampling Methods

- What determines how reliably a bootstrap test performs is how well the bootstrap DGP mimics the features of the true DGP that matter for the distribution of the test statistic. The same thing can also be said for bootstrap confidence intervals and bootstrap standard errors (not discussed in this presentation).

Some Popular Bootstrap Resampling Methods

- What determines how reliably a bootstrap test performs is how well the bootstrap DGP mimics the features of the true DGP that matter for the distribution of the test statistic. The same thing can also be said for bootstrap confidence intervals and bootstrap standard errors (not discussed in this presentation).
- For non-dependent data we will first very briefly review the (standard i.i.d. or) *nonparametric bootstrap*, the *parametric bootstrap*, the *pairs bootstrap*, the *residual bootstrap*, and the *wild bootstrap*.

Some Popular Bootstrap Resampling Methods

- What determines how reliably a bootstrap test performs is how well the bootstrap DGP mimics the features of the true DGP that matter for the distribution of the test statistic. The same thing can also be said for bootstrap confidence intervals and bootstrap standard errors (not discussed in this presentation).
- For non-dependent data we will first very briefly review the (standard i.i.d. or) *nonparametric bootstrap*, the *parametric bootstrap*, the *pairs bootstrap*, the *residual bootstrap*, and the *wild bootstrap*.
- Then for dependent data additionally the *sieve bootstrap* (an extension of the residual bootstrap), *recursive bootstrap*, and the *block bootstrap*. There are lots of other important bootstraps around, but we cannot cover them all!

Some Popular Bootstrap Resampling Methods

- What determines how reliably a bootstrap test performs is how well the bootstrap DGP mimics the features of the true DGP that matter for the distribution of the test statistic. The same thing can also be said for bootstrap confidence intervals and bootstrap standard errors (not discussed in this presentation).
- For non-dependent data we will first very briefly review the (standard i.i.d. or) *nonparametric bootstrap*, the *parametric bootstrap*, the *pairs bootstrap*, the *residual bootstrap*, and the *wild bootstrap*.
- Then for dependent data additionally the *sieve bootstrap* (an extension of the residual bootstrap), *recursive bootstrap*, and the *block bootstrap*. There are lots of other important bootstraps around, but we cannot cover them all!
- Also, we will consider the issue of whether to use restricted or unrestricted parameter estimates in constructing the bootstrap data.

Nonparametric (i.i.d.) and Parametric Bootstraps

- The nonparametric bootstrap samples T points from the original data sample, denoted $\{y_1, ..., y_T\}$, with replacement. The selected data points are chosen as random and independent draws from a given distribution, usually (though not necessarily) assigning equal probability to each data point; ie draws from a uniform distribution over $\{1, ..., T\}$. The statistic of interest can then be calculated from the bootstrap sample. If this is done B times we can obtain the EDF of the bootstrap statistic.

Nonparametric (i.i.d.) and Parametric Bootstraps

- The nonparametric bootstrap samples T points from the original data sample, denoted $\{y_1, ..., y_T\}$, with replacement. The selected data points are chosen as random and independent draws from a given distribution, usually (though not necessarily) assigning equal probability to each data point; ie draws from a uniform distribution over $\{1, ..., T\}$. The statistic of interest can then be calculated from the bootstrap sample. If this is done B times we can obtain the EDF of the bootstrap statistic.
- For hypothesis testing, consider again the one-sample t-test where the data are i.i.d. draws from $N(\mu, \sigma^2)$. Suppose we wish to test the null hypothesis that the population mean is some value μ_0. The t-statistic is given by:

$$t = \frac{\bar{y} - \mu_0}{\hat{\sigma}/\sqrt{T}}$$

where $\bar{y}$ and $\hat{\sigma}^2$ are the sample mean and sample variance of the original sample, respectively.

- There are two ways to i.i.d. bootstrap this hypothesis test.

- There are two ways to i.i.d. bootstrap this hypothesis test.
  1. Calculate the i.i.d. bootstrap sample, as above, denoted $\{y_1^*, ..., y_T^*\}$. Calculate the bootstrap t-statistic $t^* = (\bar{y}^* - \bar{y})/(\hat{\sigma}^*/\sqrt{T})$. Repeat this B times to form the estimated EDF. Notice we centre t^* on $\bar{y}$ because that is the "true value" in the bootstrap universe.

- There are two ways to i.i.d. bootstrap this hypothesis test.
  1. Calculate the i.i.d. bootstrap sample, as above, denoted $\{y_1^*, ..., y_T^*\}$. Calculate the bootstrap t-statistic $t^* = (\bar{y}^* - \bar{y})/(\hat{\sigma}^*/\sqrt{T})$. Repeat this B times to form the estimated EDF. Notice we centre t^* on $\bar{y}$ because that is the "true value" in the bootstrap universe.
  2. Create the i.i.d. bootstrap sample from the data which are centred under the restriction of the null hypothesis: $\{y_1 - \mu_0, ..., y_T - \mu_0\}$. Then calculate the bootstrap t-statistic $t^* = \bar{y}^*/(\hat{\sigma}^*/\sqrt{T})$. Repeat this B times to form the estimated EDF.
- Both are easy enough to calculate, but in more complicated settings it is often preferable to use a restricted approach where we impose the null hypothesis on the bootstrap DGP.

- Notice that the original data points will most likely not appear with equal frequency, taken across the B bootstrap samples. If we want this to be the case and here the *permutation bootstrap* can be used.

- Notice that the original data points will most likely not appear with equal frequency, taken across the B bootstrap samples. If we want this to be the case and here the *permutation bootstrap* can be used.
- The bootstrap also allows us to simulate other quantities. For example we might be interested in the sample median, $\tilde{y}$ say, not the sample mean, and want an estimate of the standard error of the sample median. This quantity depends on the underlying distribution, but can be easily estimated using the i.i.d. bootstrap as (the square root of)

$$B^{-1}\sum_{j=1}^{B}(\tilde{y}_j^* - \tilde{y})^2$$

  where $\tilde{y}^*$ is the sample median of the data generated in bootstrap sample j, with the bootstrap data generated by scheme 1 above.

- The standard t test which compares the t statistic given above to critical values from the t distribution is an exact test of the null hypothesis that the mean of the population is μ_0. This result rests on the assumption the data are (independent) draws from a Gaussian population. If untrue, the t-test won't be correctly sized (it will reject the null hypothesis either more often (giving a oversized test) or less often (undersized test) than the specified significance level, α).

- The standard t test which compares the t statistic given above to critical values from the t distribution is an exact test of the null hypothesis that the mean of the population is μ_0. This result rests on the assumption the data are (independent) draws from a Gaussian population. If untrue, the t-test won't be correctly sized (it will reject the null hypothesis either more often (giving a oversized test) or less often (undersized test) than the specified significance level, α).
- The nonparametric bootstrap, however, does not assume the data are Gaussian and will still deliver an exact test (under the conditions stated earlier). It is therefore considerably more robust.

- The standard t test which compares the t statistic given above to critical values from the t distribution is an exact test of the null hypothesis that the mean of the population is μ_0. This result rests on the assumption the data are (independent) draws from a Gaussian population. If untrue, the t-test won't be correctly sized (it will reject the null hypothesis either more often (giving a oversized test) or less often (undersized test) than the specified significance level, α).
- The nonparametric bootstrap, however, does not assume the data are Gaussian and will still deliver an exact test (under the conditions stated earlier). It is therefore considerably more robust.
- If we were sure the data were Gaussian we could also use the *parametric bootstrap*. Here the bootstrap data, y_t^*, $t = 1, ..., T$, are generated as independent draws from a $N(0, \hat{\sigma}^2)$ distribution and the bootstrap t-statistic $t^* = \bar{y}^*/(\hat{\sigma}^*/\sqrt{T})$ is calculated. Again done B times to form the estimated EDF. Notice, that this is basically a Monte Carlo simulation of the distribution of the original statistic.

Regression-based Bootstraps

- Consider the usual Classic Linear Regression Model (CLRM),

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + u_t, \qquad E(u_t|\boldsymbol{X}_t) = 0, \qquad E(u_s, u_t) = 0, \forall s \neq t \quad (1)$$

where $\boldsymbol{X}_t$ is a k-vector of (exogenous) regressors and $\boldsymbol{\beta}$ is a k-vector.

Regression-based Bootstraps

- Consider the usual Classic Linear Regression Model (CLRM),

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + u_t, \qquad E(u_t|\boldsymbol{X}_t) = 0, \qquad E(u_s, u_t) = 0, \forall s \neq t \quad (1)$$

  where $\boldsymbol{X}_t$ is a k-vector of (exogenous) regressors and $\boldsymbol{\beta}$ is a k-vector.
- Assume, for the present, that the u_t are IID with variance σ^2.

Regression-based Bootstraps

- Consider the usual Classic Linear Regression Model (CLRM),

$$y_t = \boldsymbol{X}_t \boldsymbol{\beta} + u_t, \qquad E(u_t|\boldsymbol{X}_t) = 0, \qquad E(u_s, u_t) = 0, \forall s \neq t \quad (1)$$

  where $\boldsymbol{X}_t$ is a k-vector of (exogenous) regressors and $\boldsymbol{\beta}$ is a k-vector.
- Assume, for the present, that the u_t are IID with variance σ^2.
- The *(fixed regressor) residual bootstrap* resamples from the residuals (usually Ordinary Least Squares, OLS) from estimating (1). The bootstrap DGP is $y_t^* = \boldsymbol{X}_t \hat{\boldsymbol{\beta}} + u_t^*$, where u_t^* are i.i.d. resampled from the (often rescaled and centred) OLS residuals, $\hat{u}_t$.

Regression-based Bootstraps

- Consider the usual Classic Linear Regression Model (CLRM),

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + u_t, \qquad E(u_t|\boldsymbol{X}_t) = 0, \qquad E(u_s, u_t) = 0, \forall s \neq t \quad (1)$$

  where $\boldsymbol{X}_t$ is a k-vector of (exogenous) regressors and $\boldsymbol{\beta}$ is a k-vector.
- Assume, for the present, that the u_t are IID with variance σ^2.
- The *(fixed regressor) residual bootstrap* resamples from the residuals (usually Ordinary Least Squares, OLS) from estimating (1). The bootstrap DGP is $y_t^* = \boldsymbol{X}_t\hat{\boldsymbol{\beta}} + u_t^*$, where u_t^* are i.i.d. resampled from the (often rescaled and centred) OLS residuals, $\hat{u}_t$.
- A *(fixed regressor) parametric residual bootstrap* draws the u_t^* as eg $NIID(0, s^2)$, s^2 the OLS variance estimate from (1).

Regression-based Bootstraps

- Consider the usual Classic Linear Regression Model (CLRM),

$$y_t = \boldsymbol{X}_t \boldsymbol{\beta} + u_t, \qquad E(u_t|\boldsymbol{X}_t) = 0, \qquad E(u_s, u_t) = 0, \forall s \neq t \quad (1)$$

  where $\boldsymbol{X}_t$ is a k-vector of (exogenous) regressors and $\boldsymbol{\beta}$ is a k-vector.
- Assume, for the present, that the u_t are IID with variance σ^2.
- The *(fixed regressor) residual bootstrap* resamples from the residuals (usually Ordinary Least Squares, OLS) from estimating (1). The bootstrap DGP is $y_t^* = \boldsymbol{X}_t \hat{\boldsymbol{\beta}} + u_t^*$, where u_t^* are i.i.d. resampled from the (often rescaled and centred) OLS residuals, $\hat{u}_t$.
- A *(fixed regressor) parametric residual bootstrap* draws the u_t^* as eg $NIID(0, s^2)$, s^2 the OLS variance estimate from (1).
- To perform hypothesis tests on the elements of $\boldsymbol{\beta}$ it is simplest to use a restricted estimate of $\boldsymbol{\beta}$ that imposes the restriction(s) imposed by the null hypothesis, $\tilde{\boldsymbol{\beta}}$ say, in the bootstrap DGP.

- The residual bootstrap imposes independence of the bootstrap errors, u_t^*, from the regressors. This of course implies (but is much stronger than) the conditional mean restriction on the CLRM holds.

- The residual bootstrap imposes independence of the bootstrap errors, u_t^*, from the regressors. This of course implies (but is much stronger than) the conditional mean restriction on the CLRM holds.
- The residual bootstrap is invalid if u_t is not i.i.d. If the u_t are independent but possibly heteroskedastic then the (fixed regressor) wild bootstrap of Wu (1986, AoS) can be validly used. Here the bootstrap DGP is $y_t^* = \boldsymbol{X}_t\hat{\boldsymbol{\beta}} + \hat{u}_t^*$ where $\hat{u}_t^* = \hat{u}_t \times w_t$, where the w_t's are a sequence of independent random variables with mean zero and variance 1.
- Examples used for w_t include $NIID(0, 1)$, and independent draws from the Rademacher distribution, which takes either the value 1 or -1, each with probability 0.5.

- The residual bootstrap imposes independence of the bootstrap errors, u_t^*, from the regressors. This of course implies (but is much stronger than) the conditional mean restriction on the CLRM holds.
- The residual bootstrap is invalid if u_t is not i.i.d. If the u_t are independent but possibly heteroskedastic then the (fixed regressor) wild bootstrap of Wu (1986, AoS) can be validly used. Here the bootstrap DGP is $y_t^* = \boldsymbol{X}_t\hat{\boldsymbol{\beta}} + \hat{u}_t^*$ where $\hat{u}_t^* = \hat{u}_t \times w_t$, where the w_t's are a sequence of independent random variables with mean zero and variance 1.
- Examples used for w_t include $NIID(0,1)$, and independent draws from the Rademacher distribution, which takes either the value 1 or -1, each with probability 0.5.
- The choice of distribution for w_t can be important for the finite sample accuracy of the bootstrap. It is less relevant in large samples though in some cases further restrictions, such as symmetry, need to be imposed for validity.

- Like the residual bootstrap, the wild bootstrap generates bootstrap errors, $\hat{u}_t^*$, which are conditionally (on the regressor matrix $\boldsymbol{X}$) mean zero, and so the bootstrap pairs $(y_t^*, \boldsymbol{X}_t)$ satisfy a linear regression with the "true" coefficient $\hat{\boldsymbol{\beta}}$.

- Like the residual bootstrap, the wild bootstrap generates bootstrap errors, $\hat{u}_t^*$, which are conditionally (on the regressor matrix $\boldsymbol{X}$) mean zero, and so the bootstrap pairs $(y_t^*, \boldsymbol{X}_t)$ satisfy a linear regression with the "true" coefficient $\hat{\boldsymbol{\beta}}$.
- But unlike the residual bootstrap, the conditional variance of $\hat{u}_t^*$ equals $\hat{u}_t^2$; ie the wild bootstrap errors will, on average, have about the same variance as the u_t - i.e. the wild bootstrap does not impose independence between $\hat{u}_t^*$ and the regressors.

- Like the residual bootstrap, the wild bootstrap generates bootstrap errors, $\hat{u}_t^*$, which are conditionally (on the regressor matrix $\boldsymbol{X}$) mean zero, and so the bootstrap pairs $(y_t^*, \boldsymbol{X}_t)$ satisfy a linear regression with the "true" coefficient $\hat{\boldsymbol{\beta}}$.
- But unlike the residual bootstrap, the conditional variance of $\hat{u}_t^*$ equals $\hat{u}_t^2$; ie the wild bootstrap errors will, on average, have about the same variance as the u_t - i.e. the wild bootstrap does not impose independence between $\hat{u}_t^*$ and the regressors.
- An interesting property of the wild bootstrap is that it annihilates any (weak) correlations present in the data set(s) it is applied to, because of the independence of the w_t's. This can be either a blessing or a curse as we will see.

- Like the residual bootstrap, the wild bootstrap generates bootstrap errors, $\hat{u}_t^*$, which are conditionally (on the regressor matrix $\boldsymbol{X}$) mean zero, and so the bootstrap pairs $(y_t^*, \boldsymbol{X}_t)$ satisfy a linear regression with the "true" coefficient $\hat{\boldsymbol{\beta}}$.
- But unlike the residual bootstrap, the conditional variance of $\hat{u}_t^*$ equals $\hat{u}_t^2$; ie the wild bootstrap errors will, on average, have about the same variance as the u_t - i.e. the wild bootstrap does not impose independence between $\hat{u}_t^*$ and the regressors.
- An interesting property of the wild bootstrap is that it annihilates any (weak) correlations present in the data set(s) it is applied to, because of the independence of the w_t's. This can be either a blessing or a curse as we will see.
- A dependent wild bootstrap has recently been proposed by Shao (2010, JASA).

- The wild bootstrap also imposes the conditional mean restriction of the CLRM. An alternative, which allows for some forms of heteroskedasticity, is the *pairs bootstrap* of Freedman (1981, AoS). Here we re-sample the data and not the residuals. Using the nonparametric bootstrap we re-sample the pairs $(y_t^*, \boldsymbol{X}_t^*)$ from $\{(y_t, \boldsymbol{X}_t)\}_{t=1}^{T}$. Generally inaccurate as does not condition on $\boldsymbol{X}$ and so it does not impose the conditional mean assumption, which obviously holds on the original data.

Bootstrap Methods For Dependent Data

- All of the bootstrap DGPs that have been discussed so far treat the error terms (or the data, in the case of the pairs bootstrap) as independent. When that is not the case, these methods are not appropriate. In particular, resampling (whether of residuals or data) breaks up whatever dependence there may be and is therefore unsuitable for use when there is dependence.

Bootstrap Methods For Dependent Data

- All of the bootstrap DGPs that have been discussed so far treat the error terms (or the data, in the case of the pairs bootstrap) as independent. When that is not the case, these methods are not appropriate. In particular, resampling (whether of residuals or data) breaks up whatever dependence there may be and is therefore unsuitable for use when there is dependence.
- Numerous bootstrap DGPs for dependent data have been proposed. The two most popular approaches are the *sieve bootstrap* and the *block bootstrap*. The former attempts to model the dependence using a parametric model. The latter resamples blocks of consecutive observations instead of individual observations. They can be appropriately combined with the methods discussed before such as the wild bootstrap.

The Sieve Bootstrap

- Suppose that the error terms u_t in (1) follow a weakly stationary process with (conditionally) homoskedastic innovations. The sieve bootstrap attempts to approximate this process, generally by using an $AR(p)$ process with p chosen either by some sort of model selection criterion (eg BIC) or by sequential testing. Technically, the sieve bootstrap imposes a rate condition on p so that it increases with the sample size, T. If p is fixed (but not necessarily known), the sieve bootstrap is sometimes called a *recoloured bootstrap*.

The Sieve Bootstrap

- Suppose that the error terms u_t in (1) follow a weakly stationary process with (conditionally) homoskedastic innovations. The sieve bootstrap attempts to approximate this process, generally by using an $AR(p)$ process with p chosen either by some sort of model selection criterion (eg BIC) or by sequential testing. Technically, the sieve bootstrap imposes a rate condition on p so that it increases with the sample size, T. If p is fixed (but not necessarily known), the sieve bootstrap is sometimes called a *recoloured bootstrap*.
- The first step is to estimate the model (1), preferably imposing the null hypothesis if one is to be tested, to obtain residuals $\hat{u}_t$. The next step is to estimate the $AR(p)$ model

$$\hat{u}_t = \sum_{i=1}^{p} \phi_i \hat{u}_{t-i} + e_t \tag{2}$$

make some choice of p, then estimate the AR by either OLS or Yule-Walker (the latter ensures the fitted model satisfies stationarity conditions).

- The bootstrap errors are then generated recursively by the equation

$$u_t^* = \sum_{i=1}^{p} \hat{\phi}_i u_{t-i}^* + e_t^*, \quad (3)$$

where the $\hat{\phi}_i$ are the estimated parameters from (2), and the e_t^* are resampled from the (possibly rescaled) associated residuals, say $\hat{e}_t$. This could be eg i.i.d. or wild resampling.

- The bootstrap errors are then generated recursively by the equation

$$u_t^* = \sum_{i=1}^{p} \hat{\phi}_i u_{t-i}^* + e_t^*, \tag{3}$$

  where the $\hat{\phi}_i$ are the estimated parameters from (2), and the e_t^* are resampled from the (possibly rescaled) associated residuals, say $\hat{e}_t$. This could be eg i.i.d. or wild resampling.
- Usually *initialised* at zero. With i.i.d. resampling the recursion can be started (possibly well) before $t = 1$ to allow the DGP to “warm in”.

- The bootstrap errors are then generated recursively by the equation

$$u_t^* = \sum_{i=1}^{p} \hat{\phi}_i u_{t-i}^* + e_t^*, \tag{3}$$

where the $\hat{\phi}_i$ are the estimated parameters from (2), and the e_t^* are resampled from the (possibly rescaled) associated residuals, say $\hat{e}_t$. This could be eg i.i.d. or wild resampling.
- Usually *initialised* at zero. With i.i.d. resampling the recursion can be started (possibly well) before $t = 1$ to allow the DGP to "warm in".
- The final step is to generate the bootstrap data by the equation

$$y_t^* = \boldsymbol{X}_t \hat{\boldsymbol{\beta}} + u_t^*$$

The regression parameters $\boldsymbol{\beta}$ could be estimated by OLS or a more efficient method such as GLS, as long as the method is consistent under the null hypothesis.

- Goncalves and Killian (2004, Jnl Econometrics) look at the large sample behaviour of bootstrap tests and confidence intervals based on the AR parameters when p is known and finite. G&K (2007, Econometric Reviews) extend this to the case where p is a function of T, which allows u_t to follow a very general linear process, including stationary and invertible $ARMA(p, q)$ processes.

- Goncalves and Killian (2004, Jnl Econometrics) look at the large sample behaviour of bootstrap tests and confidence intervals based on the AR parameters when p is known and finite. G&K (2007, Econometric Reviews) extend this to the case where p is a function of T, which allows u_t to follow a very general linear process, including stationary and invertible $ARMA(p, q)$ processes.
- G & K demonstrate that if the e_t^* in (3) are obtained by i.i.d. resampling, then the sieve bootstrap is invalid when u_t is conditionally heteroskedastic (eg GARCH), because even the large sample distributions of estimators of the AR coefficients depend on nuisance parameters arising from the conditional heteroskedasticity, which the bootstrap does not replicate. BIG problem for eg finance applications then!

- Same is true when bootstrapping fractional integration tests - see eg Cavaliere, Nielsen and Taylor (2017, Jnl Econometrics).

- Same is true when bootstrapping fractional integration tests - see eg Cavaliere, Nielsen and Taylor (2017, Jnl Econometrics).
- G & K demostrate the (asymptotic) validity of the *recursive-design wild bootstrap* (as outlined above with wild bootstrap resampling), as well as for a related *fixed-design wild bootstrap*, and for the *pairs bootstrap*. Cavaliere, Nielsen and Taylor also use a recursive-design wild bootstrap to solve the problem in the fractional case.

- Same is true when bootstrapping fractional integration tests - see eg Cavaliere, Nielsen and Taylor (2017, Jnl Econometrics).
- G & K demostrate the (asymptotic) validity of the *recursive-design wild bootstrap* (as outlined above with wild bootstrap resampling), as well as for a related *fixed-design wild bootstrap*, and for the *pairs bootstrap*. Cavaliere, Nielsen and Taylor also use a recursive-design wild bootstrap to solve the problem in the fractional case.
- These methods will then, not surprisingly, be very useful for testing applications in macroeconometrics and financial time series econometrics, as we will shortly see with some leading examples. Each is designed to highlight particular problems with obtaining a valid bootstrap bootstrap implementation and how these are solved.

- In the methods just given, if we use the wild bootstrap to generate the e_t^*, then this is the recursive-design wild bootstrap. For the fixed-design wild bootstrap the errors are instead generated by the equation (same e_t^*'s are used)

$$u_t^* = \sum_{i=1}^{p} \hat{\phi}_i u_{t-i} + e_t^*$$

- In the methods just given, if we use the wild bootstrap to generate the e_t^*, then this is the recursive-design wild bootstrap. For the fixed-design wild bootstrap the errors are instead generated by the equation (same e_t^*'s are used)

$$u_t^* = \sum_{i=1}^{p} \hat{\phi}_i u_{t-i} + e_t^*$$

- In the pairs bootstrap, at each point in time we sample the tuples $(y_t, y_{t-1}, ...y_{t-p})$ to give $(y_t^*, y_{t-1}^*, ...y_{t-p}^*)$, and then stack the T such draws together.

- In the methods just given, if we use the wild bootstrap to generate the e_t^*, then this is the recursive-design wild bootstrap. For the fixed-design wild bootstrap the errors are instead generated by the equation (same e_t^*'s are used)

$$u_t^* = \sum_{i=1}^{p} \hat{\phi}_i u_{t-i} + e_t^*$$

- In the pairs bootstrap, at each point in time we sample the tuples $(y_t, y_{t-1}, ...y_{t-p})$ to give $(y_t^*, y_{t-1}^*, ...y_{t-p}^*)$, and then stack the T such draws together.
- G& K show that the recursive-design method requires slightly stronger regularity conditions for validity than the other two methods, but displays the best finite sample accuracy of the three. Because of this, it is much more widely used than the other two.

- The bootstrap can also be used to estimate (and, hence, correct for) the finite sample bias in estimating the AR slope coefficients. Can be done with either the bootstrap-after-bootstrap or the double bootstrap. Put simply, one uses a second bootstrap to estimate the bias in estimation in the bootstrap DGP where the AR parameters are 'known'. Then apply this estimated discrepancy as a bias correction to the original estimates.

- The bootstrap can also be used to estimate (and, hence, correct for) the finite sample bias in estimating the AR slope coefficients. Can be done with either the bootstrap-after-bootstrap or the double bootstrap. Put simply, one uses a second bootstrap to estimate the bias in estimation in the bootstrap DGP where the AR parameters are 'known'. Then apply this estimated discrepancy as a bias correction to the original estimates.
- The sieve bootstrap can also be validly implemented in cases where u_t is a stationary and invertible fractionally integrated process - see Poskitt (2008, Journal of Time Series Analysis). Can also be implemented for non-stationary fractional models - see Kapetanios, Papailias and Taylor (2019, JTSA).

The Block Bootstrap

- Block bootstrap methods, originally proposed by Künsch (1989, AoS), divide the quantities that are being resampled, which might be either rescaled residuals or $[\boldsymbol{y}, \boldsymbol{X}]$ pairs, into blocks of m consecutive observations. The blocks, which may be either overlapping or nonoverlapping and may be either fixed or variable in length, are then resampled. It appears that the best approach is to use overlapping blocks of fixed length; see Lahiri (1999, AoS). This is called the moving-block bootstrap.

The Block Bootstrap

- Block bootstrap methods, originally proposed by Künsch (1989, AoS), divide the quantities that are being resampled, which might be either rescaled residuals or $[\boldsymbol{y}, \boldsymbol{X}]$ pairs, into blocks of m consecutive observations. The blocks, which may be either overlapping or nonoverlapping and may be either fixed or variable in length, are then resampled. It appears that the best approach is to use overlapping blocks of fixed length; see Lahiri (1999, AoS). This is called the moving-block bootstrap.
- In theory block bootstrap methods can handle weak dependence and conditional heteroskedasticity in the model but their finite-sample performance is often not very good. Finite sample performance also very strongly dependent on the choice of block length. Moreover, since they do not impose the null hypothesis, any test statistic must be adjusted so that it is testing a hypothesis that is true for the bootstrap DGP.

Moving on to ...

1. Introduction
2. The Basics of Bootstrap Hypothesis Testing
3. Some Popular Bootstrap Resampling Methods
4. Application 1: Unit Root Testing
5. Application 2: Testing for Bubbles
6. Application 3: Testing for Predictability of Returns
7. Conclusions

Bootstrap Unit Root Tests

- The sieve approach outlined above has a long tradition in econometrics, and no more so than in unit root and co-integration testing. Sieve approximations are valid for any stationary and invertible linear $MA(\infty)$ (linear process). They avoid the need for complicated non-linear estimation of models with MA components.

Bootstrap Unit Root Tests

- The sieve approach outlined above has a long tradition in econometrics, and no more so than in unit root and co-integration testing. Sieve approximations are valid for any stationary and invertible linear $MA(\infty)$ (linear process). They avoid the need for complicated non-linear estimation of models with MA components.
- Sieve methods were originally used in the context of the augmented Dickey Fuller [ADF] tests by Saïd and Dickey (1984, Biometrika).

- Consider the usual ADF model (d_t is some deterministic component)

$$\Delta y_t = d_t + \rho y_{t-1} + \sum_{j=1}^{k} \phi_j \Delta y_{t-j} + u_t. \qquad (4)$$

- S&D test the unit root null hypothesis, $H_0 : \rho = 0$, against the (trend) stationary alternative, $H_0 : \rho < 0$.
- S&D show that if $1/k + k^3/T \to 0$ as $T \to \infty$ then the ADF regression t-statistic

$$t_{DF} = \hat{\rho}/\text{s.e.}(\hat{\rho})$$

has the usual ADF limiting null distribution (originally derived for u_t i.i.d.) if u_t is a stationary and invertible $ARMA(p, q)$ driven by an i.i.d. innovation.

- Chang and Park (2002, Econometric Reviews) generalise Saïd and Dickey (1984) to allow u_t to follow a general linear process driven by conditionally heteroskedastic innovations. Interestingly, and unlike the stationary AR case in G & K, the limiting distribution of the ADF statistic does not depend on any nuisance parameters arising from the conditional heteroskedasticity.
- The ADF limiting null distribution can however be a very poor approximation to the finite sample null distribution of the sieve-based ADF statistic. Call the bootstrap!

- Basawa *et al.* (1991, AoS) illustrates the potential for danger of not imposing the null when computing a bootstrap DGP. Using a recursive bootstrap, and looking at the t-statistic whose numerator is the difference between the OLS estimator $\hat{\rho}$ from a DF ($k = 0$) regression and from a bootstrap $AR(1)$ generated recursively as $y_t^* = \hat{\rho} y_{t-1}^* + u_t^*$ they show that the bootstrap statistic does not converge to the DF null distribution and, as a result, conclude that bootstrap unit root testing is infeasible. WRONG!

- Basawa *et al.* (1991, AoS) illustrates the potential for danger of not imposing the null when computing a bootstrap DGP. Using a recursive bootstrap, and looking at the t-statistic whose numerator is the difference between the OLS estimator $\hat{\rho}$ from a DF ($k = 0$) regression and from a bootstrap $AR(1)$ generated recursively as $y_t^* = \hat{\rho} y_{t-1}^* + u_t^*$ they show that the bootstrap statistic does not converge to the DF null distribution and, as a result, conclude that bootstrap unit root testing is infeasible. WRONG!
- Ferretti and Romo (1996, Biometrika) for the $AR(1)$ case, Park (2003, Econometrica) for the $AR(p)$ case, and Chang and Park (2003, JTSA) for the $AR(\infty)$ (sieve) case show how to develop valid bootstrap implementations of the ADF tests. Each use an i.i.d. residual bootstrap approach.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4) satisfying S&D's rate condition on k.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4) satisfying S&D's rate condition on k.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k} \phi_j e_{t-j} + u_{k,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k$, and the residuals, $\tilde{u}_t$.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4) satisfying S&D's rate condition on k.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k} \phi_j e_{t-j} + u_{k,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \bar{\tilde{u}}$, to get bootstrap residuals, u_t^*.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4) satisfying S&D's rate condition on k.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k} \phi_j e_{t-j} + u_{k,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \overline{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4) satisfying S&D's rate condition on k.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k} \phi_j e_{t-j} + u_{k,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \overline{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the e_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} e_j^*$, with y_0^* set to eg zero.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4) satisfying S&D's rate condition on k.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k} \phi_j e_{t-j} + u_{k,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \bar{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the e_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} e_j^*$, with y_0^* set to eg zero.
- Step 6: Calculate the bootstrap analogue of t_{DF} in (4) applied to y_t^*.

Chang and Park (2003) Sieve Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4) satisfying S&D's rate condition on k.
- Step 2: Imposing H_0, define $e_t = \Delta y_t$. Then estimate (OLS or YW) the sieve regression, $e_t = d_t + \sum_{j=1}^{k} \phi_j e_{t-j} + u_{k,t}$, to obtain the restricted estimates $\tilde{\phi}_j$, $j = 1, ..., k$, and the residuals, $\tilde{u}_t$.
- Step 3: i.i.d. resample from the (centred) residuals, $\tilde{u}_t - \bar{\tilde{u}}$, to get bootstrap residuals, u_t^*.
- Step 4: Recursively generate $e_t^* = \sum_{j=1}^{k} \tilde{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the e_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} e_j^*$, with y_0^* set to eg zero.
- Step 6: Calculate the bootstrap analogue of t_{DF} in (4) applied to y_t^*.
- Step 7: Perform Steps 2-6 B times to form the estimated bootstrap EDF. Obtain bootstrap p-value.

- C&P Demonstrate the asymptotic validity of their sieve bootstrap unit root test. However, they impose that the shocks, u_t are i.i.d. Their bootstrap is still valid with conditionally heteroskedastic shocks, but won't mimic such effects in the bootstrap data.

- C&P Demonstrate the asymptotic validity of their sieve bootstrap unit root test. However, they impose that the shocks, u_t are i.i.d. Their bootstrap is still valid with conditionally heteroskedastic shocks, but won't mimic such effects in the bootstrap data.
- Their bootstrap is not valid if there is unconditional heteroskedasticity (*non-stationary volatility*).

- C&P Demonstrate the asymptotic validity of their sieve bootstrap unit root test. However, they impose that the shocks, u_t are i.i.d. Their bootstrap is still valid with conditionally heteroskedastic shocks, but won't mimic such effects in the bootstrap data.
- Their bootstrap is not valid if there is unconditional heteroskedasticity (*non-stationary volatility*).
- Although C&P argue that their sieve bootstrap loses no power relative to the test based on asymptotic critical values, their own simulations show large power losses relative to the standard ADF test under H_1. This occurs because Step 2 imposes H_0 on the sieve stage. Under H_1, e_t is non-invertible, violating the usual conditions for sieve validity. In practice in fitting the sieve we will fit many, many lags! Imposing H_0 on estimation is not always the best thing to do.

- C&P Demonstrate the asymptotic validity of their sieve bootstrap unit root test. However, they impose that the shocks, u_t are i.i.d. Their bootstrap is still valid with conditionally heteroskedastic shocks, but won't mimic such effects in the bootstrap data.
- Their bootstrap is not valid if there is unconditional heteroskedasticity (*non-stationary volatility*).
- Although C&P argue that their sieve bootstrap loses no power relative to the test based on asymptotic critical values, their own simulations show large power losses relative to the standard ADF test under H_1. This occurs because Step 2 imposes H_0 on the sieve stage. Under H_1, e_t is non-invertible, violating the usual conditions for sieve validity. In practice in fitting the sieve we will fit many, many lags! Imposing H_0 on estimation is not always the best thing to do.
- Cavaliere and Taylor (2008, Econometric Theory) address these problems proposing wild bootstrap ADF tests.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4).

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4).
- Step 2: (Optional): Estimate (4) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4).
- Step 2: (Optional): Estimate (4) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4).
- Step 2: (Optional): Estimate (4) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4).
- Step 2: (Optional): Estimate (4) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the u_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} u_j^*$, with eg y_0^* set to zero [replace u_j^* by e_j^* if Step 4 is performed].

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4).
- Step 2: (Optional): Estimate (4) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the u_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} u_j^*$, with eg y_0^* set to zero [replace u_j^* by e_j^* if Step 4 is performed].
- Step 6: Calculate the wild bootstrap analogue of t_{DF} in (4) applied to y_t^*.

Cavaliere and Taylor (2008) Wild Bootstrap ADF Test

- Step 1: Calculate the ADF statistic, t_{DF}, from (4).
- Step 2: (Optional): Estimate (4) to obtain the estimates $\hat{\phi}_j$, $j = 1, ..., k$.
- Step 3: Wild bootstrap resample from the first differences, $e_t = \Delta y_t$, to get bootstrap residuals, $u_t^* = e_t \times w_t$.
- Step 4: (Optional) Recursively generate $e_t^* = \sum_{j=1}^{k} \hat{\phi}_j e_{t-j}^* + u_t^*$, setting pre-sample values to eg zero.
- Step 5: Impose H_0 on the bootstrap DGP by cumulating the u_t^*'s; ie $y_t^* = y_0^* + \sum_{j=1}^{t} u_j^*$, with eg y_0^* set to zero [replace u_j^* by e_j^* if Step 4 is performed].
- Step 6: Calculate the wild bootstrap analogue of t_{DF} in (4) applied to y_t^*.
- Step 7: Perform Steps 2-6 B times to form the estimated bootstrap EDF. Obtain bootstrap p-value.

- Because the wild bootstrap annihilates weak correlations, there's no need to perform the sieve element for asymptotic validity, unlike with C&P's bootstrap. But including a sieve stage can improve finite sample size. Indeed k can be set to zero in the bootstrap version of (4) in Step 6 if the sieve stage is omitted.

- Because the wild bootstrap annihilates weak correlations, there's no need to perform the sieve element for asymptotic validity, unlike with C&P's bootstrap. But including a sieve stage can improve finite sample size. Indeed k can be set to zero in the bootstrap version of (4) in Step 6 if the sieve stage is omitted.
- C&T's algorithm does not actually impose H_0 on the (optional) sieve (again because the wild bootsrap kills serial correlation) and, as a result, C&T's wild bootstrap ADF tests avoid the power losses seen with C&P's tests.

- Because the wild bootstrap annihilates weak correlations, there's no need to perform the sieve element for asymptotic validity, unlike with C&P's bootstrap. But including a sieve stage can improve finite sample size. Indeed k can be set to zero in the bootstrap version of (4) in Step 6 if the sieve stage is omitted.
- C&T's algorithm does not actually impose H_0 on the (optional) sieve (again because the wild bootsrap kills serial correlation) and, as a result, C&T's wild bootstrap ADF tests avoid the power losses seen with C&P's tests.
- C&T show that the wild bootstrap t^*_{DF} statistic has the same first order limiting null distribution as the limiting null distribution of t_{DF} under null, local and fixed alternatives. Hence, behaves like an infeasible size-corrected ADF test.

- Because the wild bootstrap annihilates weak correlations, there's no need to perform the sieve element for asymptotic validity, unlike with C&P's bootstrap. But including a sieve stage can improve finite sample size. Indeed k can be set to zero in the bootstrap version of (4) in Step 6 if the sieve stage is omitted.
- C&T's algorithm does not actually impose H_0 on the (optional) sieve (again because the wild bootsrap kills serial correlation) and, as a result, C&T's wild bootstrap ADF tests avoid the power losses seen with C&P's tests.
- C&T show that the wild bootstrap t^*_{DF} statistic has the same first order limiting null distribution as the limiting null distribution of t_{DF} under null, local and fixed alternatives. Hence, behaves like an infeasible size-corrected ADF test.
- C&T in various papers (eg 2019 Econometric Theory, Econometric Reviews, 2009) show that the bootstrap ADF tests perform very well in the presence of both conditional heteroskedasticity and unconditional heteroskedasticity of many forms (eg volatility breaks, trending volatility, IGARCH, AR-SV, various GARCH -type models).

- Cavaliere and Taylor (2008, Jnl Econometrics) propose an alternative parametric bootstrap approach based on a consistent estimate of the volatility profile of the series. Does not perform as well as the wild bootstrap in finite samples.

- Cavaliere and Taylor (2008, Jnl Econometrics) propose an alternative parametric bootstrap approach based on a consistent estimate of the volatility profile of the series. Does not perform as well as the wild bootstrap in finite samples.
- Cavaliere, Rahbek and Taylor (2012, Econometrica) develop bootstrap implementations of Johansen's sequential procedure for determining the co-integration rank of a system of variables. This work again shows the importance of imposing the co-integration rank postulated by the null hypothesis on the bootstrap DGP at each stage of the sequential procedure. They show that not doing so, as in the bootstrap procedure developed in Swensen (2009, Econometrica), leads to an inconsistent bootstrap procedure.

Moving on to ...

Testing for Bubbles

- Phillips *et al.* (2011, IER) (PWY), focus on testing the null hypothesis of a fixed unit root across the whole sample against the alternative of explosive autoregressive behaviour in some subset of the sample using the supremum of a set of forward recursive (ie sequences of sub-samples) right-tailed (A)DF tests applied to the price and dividend series. If the test finds explosive autoregressive behaviour for the prices but not for the dividends, this indicates that an explosive rational bubble exists.

Testing for Bubbles

- Phillips *et al.* (2011, IER) (PWY), focus on testing the null hypothesis of a fixed unit root across the whole sample against the alternative of explosive autoregressive behaviour in some subset of the sample using the supremum of a set of forward recursive (ie sequences of sub-samples) right-tailed (A)DF tests applied to the price and dividend series. If the test finds explosive autoregressive behaviour for the prices but not for the dividends, this indicates that an explosive rational bubble exists.
- PWY implement their test based on finite sample Monte Carlo critical values assuming Gaussian IID innovations. Harvey, Leybourne, Sollis and Taylor (2016, Jnl Empirical Finance) [HLST] propose wild bootstrap implementations of the PWY test which allow for nonstationary volatility in the innovations.

Bubble DGP

- In its simplest form, the bubble DGP of PWY is of the form:

$$y_t = \mu + u_t \tag{5}$$

$$u_t = \begin{cases} u_{t-1} + \varepsilon_t, & t = 2, \ldots, \lfloor \tau_{1,0} T \rfloor, \\ (1 + \delta_{1,T}) u_{t-1} + \varepsilon_t, & t = \lfloor \tau_{1,0} T \rfloor + 1, \ldots, \lfloor \tau_{2,0} T \rfloor, \\ u_{t-1} + \varepsilon_t, & t = \lfloor \tau_{2,0} T \rfloor + 1, \ldots, T \end{cases}$$

  where $\delta_{1,T} \geq 0$.
- When $\delta_{1,T} > 0$, y_t follows a unit root up to time $\lfloor \tau_{1,0} T \rfloor$, after which it displays explosive AR behaviour over $t = \lfloor \tau_{1,0} T \rfloor + 1, \ldots, \lfloor \tau_{2,0} T \rfloor$. When applied to asset prices, and assuming unit root behaviour in the corresponding dividend series, this can be interpreted as a bubble regime.
- At the end of the bubble period, y_t reverts to unit root dynamics. The DGP admits a bubble regime continuing to the end of the sample period if $\tau_{2,0} = 1$.

- The null hypothesis, H_0, is that no bubble is present in the series and y_t follows a unit root process throughout the sample period, i.e. $H_0 : \delta_{1,T} = 0$. The alternative hypothesis is given by $H_1 : \delta_{1,T} > 0$, and corresponds to the case where a bubble is present in the series, which either runs to the end of the sample (if $\tau_{2,0} = 1$), or terminates in-sample.

- The null hypothesis, H_0, is that no bubble is present in the series and y_t follows a unit root process throughout the sample period, i.e. $H_0 : \delta_{1,T} = 0$. The alternative hypothesis is given by $H_1 : \delta_{1,T} > 0$, and corresponds to the case where a bubble is present in the series, which either runs to the end of the sample (if $\tau_{2,0} = 1$), or terminates in-sample.
- To test H_0 against H_1, PWY propose a test based on the supremum of recursive right-tailed (A)DF tests.

- For serially uncorrelated ε_t, the PWY statistic is

$$PWY = \sup_{\tau \in [\tau_0, 1]} DF_\tau$$

where DF_τ is the standard DF statistic, ie the t-ratio for $\hat{\phi}_\tau = 0$ in the fitted OLS regression

$$\Delta y_t = \hat{\alpha} + \hat{\phi}_\tau y_{t-1} + \hat{\varepsilon}_t \tag{6}$$

calculated over the sub-sample $t = 1, ..., \lfloor \tau T \rfloor$, i.e.

$$DF_\tau = \frac{\hat{\phi}_\tau}{\sqrt{\hat{\sigma}_\tau^2 / \sum_{t=2}^{\lfloor \tau T \rfloor} (y_{t-1} - \bar{y}_\tau)^2}}$$

where $\bar{y}_\tau = (\lfloor \tau T \rfloor - 1)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} y_{t-1}$ and $\hat{\sigma}_\tau^2 = (\lfloor \tau T \rfloor - 3)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{\varepsilon}_t^2$.

- The *PWY* statistic is therefore the supremum of a sequence of forward recursive DF statistics with minimum sample length $\lfloor \tau_0 T \rfloor$.

- The *PWY* statistic is therefore the supremum of a sequence of forward recursive DF statistics with minimum sample length $\lfloor \tau_0 T \rfloor$.
- PWY assume that ε_t in (5) is i.i.d. or an $AR(p)$ driven by i.i.d. innovations.

- The *PWY* statistic is therefore the supremum of a sequence of forward recursive DF statistics with minimum sample length $\lfloor \tau_0 T \rfloor$.
- PWY assume that ε_t in (5) is i.i.d. or an $AR(p)$ driven by i.i.d. innovations.
- Many studies find evidence of structural breaks in the unconditional variance of asset returns, often with the breaks linked to major financial and macroeconomic crises such as the 1970s oil price shocks, the East Asian currency crisis in the late-1990s, the dot-com crash in 2001 and the global financial crisis in 2008-2009.

- The *PWY* statistic is therefore the supremum of a sequence of forward recursive DF statistics with minimum sample length $\lfloor \tau_0 T \rfloor$.
- PWY assume that ε_t in (5) is i.i.d. or an $AR(p)$ driven by i.i.d. innovations.
- Many studies find evidence of structural breaks in the unconditional variance of asset returns, often with the breaks linked to major financial and macroeconomic crises such as the 1970s oil price shocks, the East Asian currency crisis in the late-1990s, the dot-com crash in 2001 and the global financial crisis in 2008-2009.
- Apparent volatility changes in asset returns could be induced by the presence of a speculative bubble, and the converse could also be true. It is therefore critically important to have available a reliable method for detecting an explosive period in a series that is robust to the potential presence of nonstationary volatility, particularly if the evidence is to be used to inform future monetary policy.

- HLST show that nonstationary volatility leads to a non-pivotal null limiting distribution for the *PWY* test. Simulations for various common patterns of nonstationary volatility show that the *PWY* test can be badly over-sized.

- HLST show that nonstationary volatility leads to a non-pivotal null limiting distribution for the *PWY* test. Simulations for various common patterns of nonstationary volatility show that the *PWY* test can be badly over-sized.
- HLST propose a wild bootstrap, applied to the first differences of the data to replicate in the bootstrap DGP the pattern of nonstationary volatility present in the original innovations.

Wild Bootstrap PWY Algorithm

- *Step 1. Generate T bootstrap innovations ε_t^*, as follows: $\varepsilon_1^* = 0$, $\varepsilon_t^* = w_t \Delta y_t$, $t = 2, ..., T$, where $\{w_t\}_{t=2}^T$ is a $NIID(0, 1)$ sequence.*
- *Step 2. Construct the bootstrap sample as the partial sum*

$$y_t^* = \sum_{j=1}^{t} \varepsilon_j^*, \quad t = 1, ..., T.$$

- *Step 3. Compute the bootstrap test statistic*

$$\mathit{PWY}^* = \sup_{\tau \in [\tau_0, 1]} \mathit{DF}_\tau^*$$

where DF_τ^* is the t-ratio on $\hat{\phi}_\tau^*$ in the fitted OLS regression

$$\Delta y_t^* = \hat{\alpha}^* + \hat{\phi}_\tau^* y_{t-1}^* + \hat{\varepsilon}_t^*$$

calculated over the sub-sample period $t = 1, ..., \lfloor \tau T \rfloor$, i.e.

Wild Bootstrap PWY Algorithm

$$DF^*_\tau = \frac{\hat{\phi}^*_\tau}{\sqrt{\hat{\sigma}^{*2}_\tau / \sum_{t=2}^{\lfloor \tau T \rfloor} \left(y^*_{t-1} - \bar{y}^*_\tau \right)^2}}$$

where $\bar{y}^*_\tau = (\lfloor \tau T \rfloor - 1)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} y^*_{t-1}$ *and* $\hat{\sigma}^{*2}_\tau = (\lfloor \tau T \rfloor - 3)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{\varepsilon}^{*2}_t$.

- *Step 4. Bootstrap p-values can then be computed in the usual way by repeating Steps 1-3 B times.*

- HLST show that the wild bootstrap *PWY* statistic shares the same first-order (non-pivotal) limiting null distribution as the original *PWY* statistic within a broad class of nonstationary volatility processes (essentially the same as C&T consider in the context of conventional unit root testing).

- HLST show that the wild bootstrap *PWY* statistic shares the same first-order (non-pivotal) limiting null distribution as the original *PWY* statistic within a broad class of nonstationary volatility processes (essentially the same as C&T consider in the context of conventional unit root testing).
- HLST's bootstrap PWY tests achieve the asymptotic power function of (infeasibly) size-corrected variant of the original PWY statistic, under locally explosive alternatives.

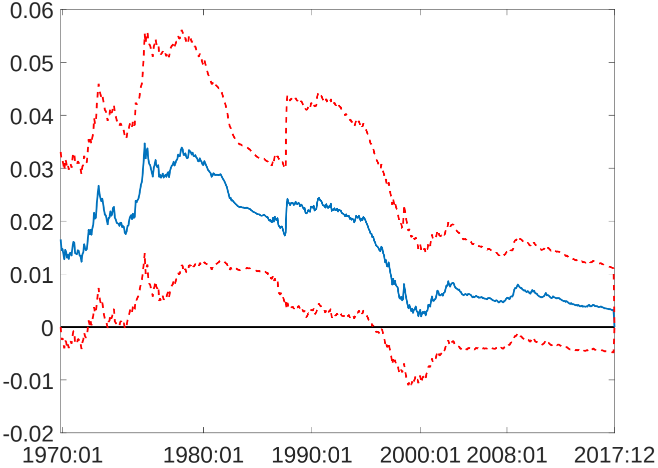
- HLST show that the wild bootstrap *PWY* statistic shares the same first-order (non-pivotal) limiting null distribution as the original *PWY* statistic within a broad class of nonstationary volatility processes (essentially the same as C&T consider in the context of conventional unit root testing).
- HLST's bootstrap PWY tests achieve the asymptotic power function of (infeasibly) size-corrected variant of the original PWY statistic, under locally explosive alternatives.
- Under fixed magnitude explosive alternatives the booststrap PWY statistic diverges with the sample size, T, but at a slower rate that does the original PWY statistic and hence the bootstrap PWY is still consistent.

- HLST show that the wild bootstrap *PWY* statistic shares the same first-order (non-pivotal) limiting null distribution as the original *PWY* statistic within a broad class of nonstationary volatility processes (essentially the same as C&T consider in the context of conventional unit root testing).
- HLST's bootstrap PWY tests achieve the asymptotic power function of (infeasibly) size-corrected variant of the original PWY statistic, under locally explosive alternatives.
- Under fixed magnitude explosive alternatives the booststrap PWY statistic diverges with the sample size, T, but at a slower rate that does the original PWY statistic and hence the bootstrap PWY is still consistent.
- It is often believed that a bootstrap statistic must replicate the limiting null distribution of the statistic under both the null and alternative to be valid and consistent, but this is not the case!

- HLST show that the wild bootstrap *PWY* statistic shares the same first-order (non-pivotal) limiting null distribution as the original *PWY* statistic within a broad class of nonstationary volatility processes (essentially the same as C&T consider in the context of conventional unit root testing).
- HLST's bootstrap PWY tests achieve the asymptotic power function of (infeasibly) size-corrected variant of the original PWY statistic, under locally explosive alternatives.
- Under fixed magnitude explosive alternatives the booststrap PWY statistic diverges with the sample size, T, but at a slower rate that does the original PWY statistic and hence the bootstrap PWY is still consistent.
- It is often believed that a bootstrap statistic must replicate the limiting null distribution of the statistic under both the null and alternative to be valid and consistent, but this is not the case!
- The same phenomenon occurs with the wild bootstrap KPSS tests proposed in Cavaliere and Taylor (2005, Econometric Theory).

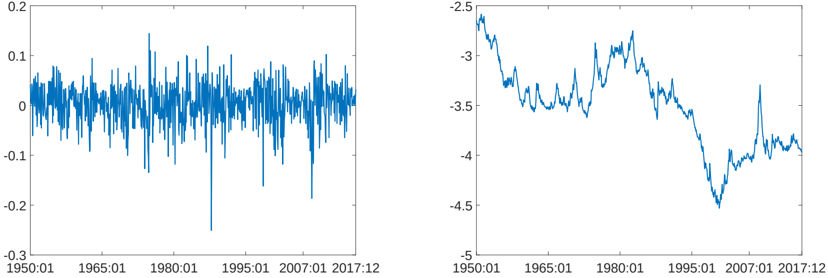
Moving on to ...

Is there any Predictability in the Equity Premium?



Dividend yield: Forward Recursive IV regression estimates and pointwise CIs, 1950-2017 (Goyal-Welch 2008 updated monthly data).

... what about the persistence of the predictor?



The equity premium looks very mean reverting etc (almost noise), but the dividend yield looks strongly persistent (usual ADF test has p-value of 0.41).

The Basic Predictive Regression Set-up

Consider the predictive regression

$$y_t \quad = \quad \alpha + \beta \, x_{t-1} + u_t \tag{7}$$

where

$$(x_t - \mu_x) \quad = \quad \rho(x_{t-1} - \mu_x) + v_t, \tag{8}$$

with $(u_t, v_t)' \sim iid(0, \mathbf{\Sigma})$ where

$$\mathbf{\Sigma} = \mathrm{E}\left(\begin{pmatrix} u_t \\ v_t \end{pmatrix} \begin{pmatrix} u_t & v_t \end{pmatrix} \right) = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}.$$

Null hypothesis: x_{t-1} does not predict y_t, i.e.

$$H_0 : \quad \beta = 0 \, .$$

Yet, even in this simplest setup...

Endogeneity and (high) Persistence

Should

- the shocks u_t and v_t correlate (so that $\phi = \sigma_{uv}/\sigma_u\sigma_v \neq 0$; for the EP-DY data above this correlation is estimated to be $\hat{\phi} = -0.98$), and
- the regressor x_t be autocorrelated,

one speaks of endogeneity (albeit a bit of a misnomer).

Endogeneity and (high) Persistence

Should

- the shocks u_t and v_t correlate (so that $\phi = \sigma_{uv}/\sigma_u\sigma_v \neq 0$; for the EP-DY data above this correlation is estimated to be $\hat{\phi} = -0.98$), and
- the regressor x_t be autocorrelated,

one speaks of endogeneity (albeit a bit of a misnomer).

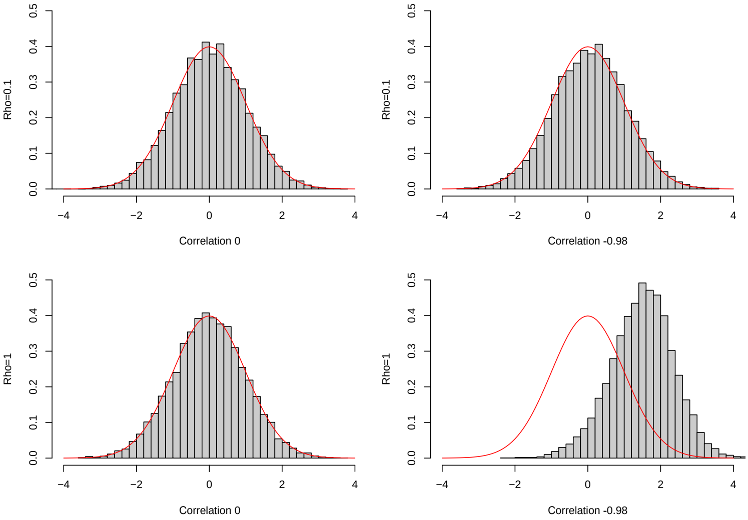
Under endogeneity and high persistence (near integration, $\rho = 1 - c/T$),

- the OLS estimator is 2nd order biased and
- the *t*-statistic has a non-normal limiting distribution.

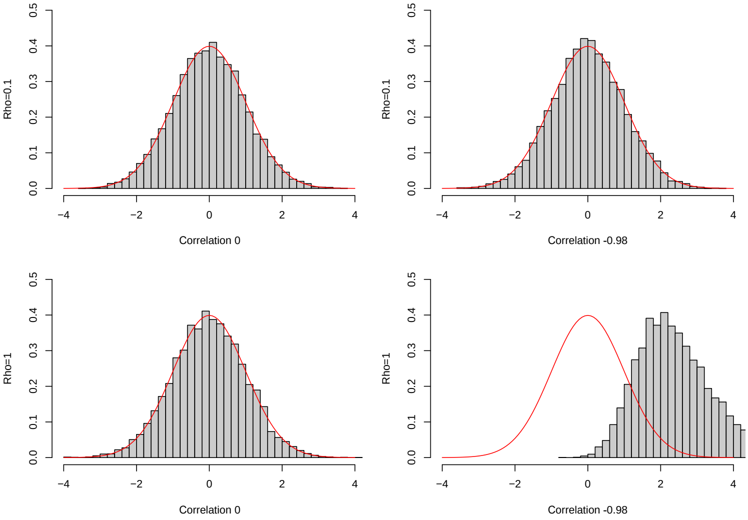
See Cavanagh, Elliott and Stock (1995, Econometric Theory), Stambaugh (1999, Jnl Fin Economics), Campbell and Yogo (2006, Jnl Fin Economics) etc.

No problem when regressors are weakly persistent.

OLS t-statistics, $T = 305$



More trouble with variance breaks - volatility of both shocks 3 times higher in the first 20% of the sample



- In the near-unit root case, the limiting null distributions of the standard t-statistic for testing $\beta = 0$ in (7) depends on both ϕ and c, whenever neither is zero.

- In the near-unit root case, the limiting null distributions of the standard t-statistic for testing $\beta = 0$ in (7) depends on both ϕ and c, whenever neither is zero.
- If ρ were known, one could employ GLS estimation. For unknown ρ, there are a number of 'solutions' proposed in the literature:
    - Bonferroni - Cavanagh *et al.* (1995), Campbell and Yogo (2006)
    - Restricted log-likelihood - Jansson and Moreira (2006, Econometrica)
    - Almost optimal tests - Elliott *et al.* (2015, Econometrica)
    - Generic IV estimation - Breitung and Demetrescu (2015, Jnl Econometrics)
    - Extended Instrumental Variables [IVX] method of Kostakis *et al.* (2015, Review of Financial Studies) [KMS]

- In the near-unit root case, the limiting null distributions of the standard t-statistic for testing $\beta = 0$ in (7) depends on both ϕ and c, whenever neither is zero.
- If ρ were known, one could employ GLS estimation. For unknown ρ, there are a number of 'solutions' proposed in the literature:
  - Bonferroni - Cavanagh *et al.* (1995), Campbell and Yogo (2006)
  - Restricted log-likelihood - Jansson and Moreira (2006, Econometrica)
  - Almost optimal tests - Elliott *et al.* (2015, Econometrica)
  - Generic IV estimation - Breitung and Demetrescu (2015, Jnl Econometrics)
  - Extended Instrumental Variables [IVX] method of Kostakis *et al.* (2015, Review of Financial Studies) [KMS]
- The IVX method has become very popular. It delivers tests with standard pivotal limiting null distributions. Unlike many of the other methods, this holds regardless of whether the predictor is weakly or strongly persistent. Also implementable with multiple predictors.

- In the near-unit root case, the limiting null distributions of the standard t-statistic for testing $\beta = 0$ in (7) depends on both ϕ and c, whenever neither is zero.
- If ρ were known, one could employ GLS estimation. For unknown ρ, there are a number of 'solutions' proposed in the literature:
  - Bonferroni - Cavanagh *et al.* (1995), Campbell and Yogo (2006)
  - Restricted log-likelihood - Jansson and Moreira (2006, Econometrica)
  - Almost optimal tests - Elliott *et al.* (2015, Econometrica)
  - Generic IV estimation - Breitung and Demetrescu (2015, Jnl Econometrics)
  - Extended Instrumental Variables [IVX] method of Kostakis *et al.* (2015, Review of Financial Studies) [KMS]
- The IVX method has become very popular. It delivers tests with standard pivotal limiting null distributions. Unlike many of the other methods, this holds regardless of whether the predictor is weakly or strongly persistent. Also implementable with multiple predictors.
- However, the asymptotic approximation can be very poor. So bootstrap implementations seem worth exploring.

The IVX Approach

- KMS develop asymptotically valid methods of estimation and inference in the context of (7)-(8) based on the use of the mildly integrated IVX instrument

$$z_{I,t} = \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j} = (1 - \varrho L)_+^{-1} \Delta x_t$$

where $\varrho = 1 - aT^{-\gamma}$ with $\gamma \in (0, 1)$ and $a \geq 0$.

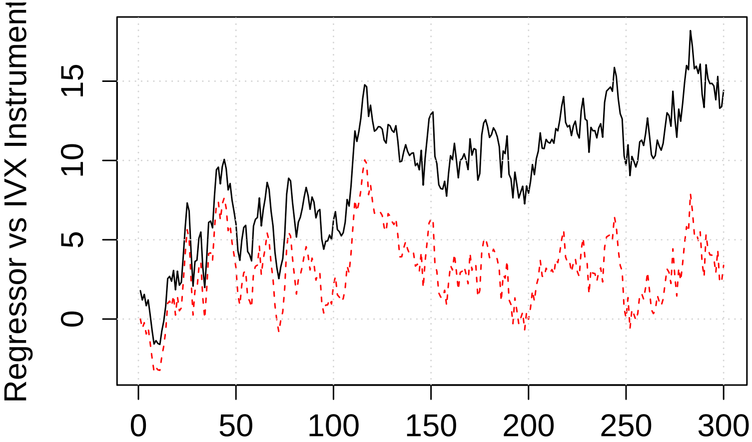
The IVX Approach

- KMS develop asymptotically valid methods of estimation and inference in the context of (7)-(8) based on the use of the mildly integrated IVX instrument

$$z_{I,t} = \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j} = (1 - \varrho L)_+^{-1} \Delta x_t$$

  where $\varrho = 1 - aT^{-\gamma}$ with $\gamma \in (0,1)$ and $a \geq 0$.
- KMS develop IV-based tests using this instrument for x_t. They allow u_t to follow a serially uncorrelated $GARCH(p,q)$ process and v_t to be a linear process driven by general (conditionally heteroskedastic) martingale difference [MD] innovations.

The IVX trick applied to a random walk - using $a = 1$, $\gamma = 0.95$ as in KMS



The IVX Test of KMS

The IVX-based t-ratio of KMS for testing $H_0 : \beta = 0$ in (7) instruments the endogenous predictor x_{t-1} with the IVX instrument $z_{I,t-1}$, and is given by

$$t_{zx} = \frac{\hat{\beta}_{zx}}{s.e.(\hat{\beta}_{zx})} \tag{9}$$

where $\hat{\beta}_{zx}$ is the IVX estimator of β,

$$\hat{\beta}_{zx} = \frac{\sum_{t=1}^{T} z_{t-1} \left(y_t - \bar{y} \right)}{\sum_{t=1}^{T} z_{t-1} \left(x_{t-1} - \bar{x}_{-1} \right)} \tag{10}$$

with $\bar{y} = T^{-1} \sum_{t=1}^{T} y_t$ and $\bar{x}_{-1} = T^{-1} \sum_{t=1}^{T} x_{t-1}$, and $s.e.(\hat{\beta}_{zx})$ is its standard error (White standard error if one allows for conditional heteroskedasticity) formed from the OLS residuals from estimating (7).

Bootstrap IVX Tests

- Demetrescu, Georgiev, Rodrigues and Taylor (2022, Jnl Econometrics) [DGRT] explore two bootstrap resampling schemes. The first, a residual wild bootstrap [RWB]. The second is a fixed regressor wild bootstrap [FRWB]. DGRT show that both are first-order asymptotically valid.

Bootstrap IVX Tests

- Demetrescu, Georgiev, Rodrigues and Taylor (2022, Jnl Econometrics) [DGRT] explore two bootstrap resampling schemes. The first, a residual wild bootstrap [RWB]. The second is a fixed regressor wild bootstrap [FRWB]. DGRT show that both are first-order asymptotically valid.
- DGRT show that these allow one to replace the $GARCH(1,1)$ assumption on u_t with a much more general bivariate MD assumption on the two innovations. Moreover, non-stationary volatility in each innovation can be allowed, and the endogeneity correlation, ϕ, can be allowed to vary over time.

Bootstrap IVX Tests

- Demetrescu, Georgiev, Rodrigues and Taylor (2022, Jnl Econometrics) [DGRT] explore two bootstrap resampling schemes. The first, a residual wild bootstrap [RWB]. The second is a fixed regressor wild bootstrap [FRWB]. DGRT show that both are first-order asymptotically valid.
- DGRT show that these allow one to replace the $GARCH(1,1)$ assumption on u_t with a much more general bivariate MD assumption on the two innovations. Moreover, non-stationary volatility in each innovation can be allowed, and the endogeneity correlation, ϕ, can be allowed to vary over time.
- DGRT show that these bootstraps also allow for valid subsample implementation of the IVX tests of KMS (pockets of predictability). These have non-pivotal limiting null distributions which depend in a complex way on nuisance parameters arising from both the serial correlation and heteroskedastic aspects of the DGP, and constructing an asymptotic test is not even feasible.

A Residual Wild Bootstrap

1. Fit the predictive regression (7) to the sample data $(y_t, x_{t-1})'$ to obtain the residuals $\hat{u}_t$, $t = 1, ..., T$.
2. Fit by OLS an autoregression of order $p + 1$ to x_t; viz,

$$x_t = \hat{m} + \sum_{j=1}^{p+1} \hat{a}_j x_{t-j} + \hat{v}_t$$

   and compute the OLS residuals $\hat{v}_t$, $t = p + 1, \ldots, T$. Set $\hat{v}_t = 0$ for $t = 1, \ldots, p$.
3. Generate bootstrap innovations $(u_t^*, v_t^*)' = (w_t \hat{u}_t,\, w_t \hat{v}_t)'$, $t = 1 \ldots, T$, where w_t, $t = 1, ..., T$, is a scalar $i.i.d.(0, 1)$ sequence with $\mathrm{E}(w_t^4) < \infty$, which is independent of the sample data.

4. Define the bootstrap data $(y_t^*, x_{t-1}^*)'$ where $y_t^* = u_t^*$ (so that the null hypothesis is imposed on the bootstrap y_t^*) and where x_t^* is generated according to the recursion

$$x_t^* = \sum_{j=1}^{p+1} \hat{a}_j x_{t-j}^* + v_t^*, \quad t = 1, ..., T$$

with initial conditions $x_0^* = \ldots = x_{-p}^* = 0$. Create the associated bootstrap IVX instrument, z_t^*, as:

$$z_0^* = 0 \quad \text{and} \quad z_t^* = \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j}^*, \ t = 1, \ldots, T,$$

where ϱ is the same value as used in constructing the original IVX instrument, z_t.

5. Using the bootstrap sample data, $\left(y_t^*, x_{t-1}^*, z_{t-1}^*\right)'$, in place of the original sample data, $(y_t, x_{t-1}, z_{t-1})'$, construct the bootstrap analogues of the IVX statistics.

A Fixed-Regressor Wild Bootstrap

1. Construct the wild bootstrap innovations $y_t^* = \hat{y}_t w_t$, where $\hat{y}_t = y_t - \frac{1}{T}\sum_{t=1}^{T} y_t$ are the demeaned sample observations on y_t.
2. Using the bootstrap sample data $\left(y_t^*, x_{t-1}, \boldsymbol{z}_{t-1}'\right)'$, in place of the original sample data $\left(y_t, x_{t-1}, \boldsymbol{z}_{t-1}'\right)'$, construct the bootstrap analogues of the IVX statistics.

Key Differences?

- A key difference between the RWB and FRWB surrounds the generation of the bootstrap analogue data for x_t and z_t. While the RWB rebuilds into the bootstrap data (an estimate of) the correlation between the innovations u_t and v_t (it is crucial in doing so that the same R_t is used to multiply both $\hat{u}_t$ and $\hat{v}_t$), the FRWB does not. This is an important distinction because the finite sample behaviour of the IVX statistics is heavily dependent on the correlation between u_t and v_t when x_t is strongly persistent.

Key Differences?

- A key difference between the RWB and FRWB surrounds the generation of the bootstrap analogue data for x_t and z_t. While the RWB rebuilds into the bootstrap data (an estimate of) the correlation between the innovations u_t and v_t (it is crucial in doing so that the same R_t is used to multiply both $\hat{u}_t$ and $\hat{v}_t$), the FRWB does not. This is an important distinction because the finite sample behaviour of the IVX statistics is heavily dependent on the correlation between u_t and v_t when x_t is strongly persistent.
- A further difference is that because the RWB uses the bootstrap data x_t^* and z_t^*, one is implicitly using an estimate of ρ. Under strong persistence c, cannot be consistently estimated and so x_t^* will not be generated with the same local-to-unity parameter as x_t. However, the IVX statistics instrument x_{t-1} by z_{t-1}, and their bootstrap analogues instrument x_{t-1}^* by z_{t-1}^*. But both z_t and z_t^* are, by construction, mildly integrated processes, regardless of the value of c. There is therefore no necessity for the estimate of c to be consistent.

Monte Carlo Results from DGRT I

Case 1: Empirical Size: Scalar Predictor, IID errors

- DGP (7)-(8) with $\beta = 0$. Set $\alpha = \mu_x = 0$, w.n.l.o.g.
- $\rho := 1 - c/T$ with $c \in \{-0.5, -0.25, 0, 2.5, 5, 10, 25, ..., 250\}$
- $(u_t, v_t)'$ is zero-mean IID bivariate Gaussian with covariance matrix $\boldsymbol{\Sigma} := \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}$ and $\phi = -0.95$
- IVX with $a = 1$, $\gamma = 0.95$, and KMS's finite-sample correction
- Report: $t_{zx}^{*,RWB}$ and $t_{zx}^{*,FRWB}$ (RWB and FRWB implementations of t_{zx}); t_{zx}^{EW} (asymptotic IVX test with conventional ses) and t_{zx} (asymptotic IVX test with White ses)
- $T = 250$, 10000 MC replications, 999 bootstrap replications. Nominal 5% level. In Step 2 of RWB p chosen by BIC over the search set $p \in \{0, ..., \lfloor 4(T/100)^{0.25} \rfloor\}$.

Table 1: Size of Left-sided Tests
Gaussian IID innovations

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
-5	0.046	0.004	0.004	0.003
-2.5	0.045	0.000	0.000	0.001
0	0.041	0.001	0.001	0.001
2.5	0.062	0.005	0.005	0.005
5	0.068	0.010	0.011	0.010
10	0.064	0.019	0.019	0.018
25	0.057	0.029	0.030	0.028
50	0.056	0.034	0.036	0.035
75	0.056	0.037	0.038	0.037
100	0.054	0.038	0.040	0.038
125	0.054	0.039	0.042	0.041
150	0.055	0.043	0.046	0.042
200	0.054	0.046	0.048	0.045
250	0.054	0.048	0.051	0.048

Table 2: Size of Right-sided Tests
Gaussian IID innovations

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
-5	0.046	0.074	0.080	0.073
-2.5	0.041	0.094	0.097	0.093
0	0.053	0.105	0.114	0.110
2.5	0.064	0.112	0.116	0.115
5	0.062	0.107	0.116	0.112
10	0.062	0.097	0.102	0.099
25	0.057	0.078	0.084	0.080
50	0.052	0.067	0.072	0.067
75	0.053	0.064	0.068	0.065
100	0.053	0.061	0.065	0.062
125	0.052	0.060	0.063	0.060
150	0.053	0.056	0.060	0.059
200	0.050	0.054	0.056	0.053
250	0.051	0.051	0.055	0.053

Table 3: Size of Two-sided Tests
Gaussian IID innovations

c	$t_{zx}^{*,RWB}$	$t_{zx}^{*,FRWB}$	t_{zx}^{EW}	t_{zx}
-5	0.048	0.038	0.044	0.039
-2.5	0.038	0.040	0.048	0.044
0	0.047	0.051	0.057	0.053
2.5	0.053	0.058	0.062	0.060
5	0.054	0.058	0.063	0.060
10	0.055	0.060	0.066	0.060
25	0.056	0.056	0.060	0.058
50	0.051	0.051	0.054	0.052
75	0.049	0.047	0.052	0.049
100	0.049	0.048	0.052	0.050
125	0.050	0.049	0.053	0.051
150	0.051	0.049	0.054	0.052
200	0.050	0.048	0.054	0.050
250	0.049	0.048	0.053	0.050

Monte Carlo Results from DGRT II

Case 2: Empirical Size: Multiple Predictors

- Multiple predictor simulation DGP:

$$\begin{aligned} y_t &= \alpha + \mathbf{x}'_{t-1}\boldsymbol{\beta} + u_t, \qquad t = 1, \ldots, T, \\ \mathbf{x}_t &= \boldsymbol{\rho}\mathbf{x}_{t-1} + \mathbf{v}_t, \qquad t = 0, \ldots, T, \end{aligned}$$

  where $\mathbf{x}_t := (x_{1,t}, ..., x_{K,t})'$ is a $K \times 1$ vector of predictor variables, $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters, $\alpha = 0.25$, $\boldsymbol{\rho}$ is a $K \times K$ diagonal matrix with common diagonal element ρ, i.e., $\boldsymbol{\rho} := \text{diag}(\rho, ..., \rho)$.
- The AR parameter ρ is again set equal to $1 - c/T$ with $c \in \{-5, -2.5, 0, 2.5, 5, 10, 25, ..., 250\}$

- The innovations are generated as $(u_t, \mathbf{v}_t')' \sim NIID(\mathbf{0}, \mathbf{\Sigma})$ where

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{u,v_1} & 0 & \cdots & 0 \\ \sigma_{u,v_1} & \sigma_{v_1}^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_{v_2}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{v_K}^2 \end{pmatrix} \tag{11}$$

  with $\sigma_u^2 = 0.037$, $\sigma_{u,v_1} = -0.035$, $\sigma_{v_1}^2 = ... = \sigma_{v_K}^2 = 0.045$.
- Notice, therefore, that the first predictor, $x_{1,t}$ is endogenous (with an endogeneity correlation parameter $\phi_1 = -0.83$), while the remaining predictors $x_{2,t}, ..., x_{K,t}$ are exogenous.
- Empirical sizes of the Wald tests for the joint significance of the K predictors. NB RWB uses obvious VAR generalisation of Step 2.

Table 4: Size of joint Wald Tests. $K = 3$ predictors.

c	$W_{zx}^{*,\text{RWB}}$	$W_{zx}^{*,\text{FRWB}}$	W_{zx}^{EW}	W_{zx}
-5	0.085	0.352	0.385	0.366
-2.5	0.097	0.176	0.193	0.177
0	0.075	0.105	0.117	0.104
2.5	0.067	0.086	0.103	0.090
5	0.059	0.077	0.095	0.083
10	0.054	0.066	0.083	0.071
25	0.052	0.061	0.075	0.066
50	0.053	0.057	0.070	0.061
75	0.053	0.053	0.069	0.058
100	0.051	0.053	0.069	0.057
125	0.052	0.054	0.070	0.058
150	0.052	0.054	0.069	0.058
200	0.052	0.055	0.071	0.059
250	0.053	0.055	0.071	0.060

Table 5: Size of joint Wald Tests. $K = 5$ predictors.

c	$W_{zx}^{*,\text{RWB}}$	$W_{zx}^{*,\text{FRWB}}$	W_{zx}^{EW}	W_{zx}
-5	0.074	0.402	0.466	0.421
-2.5	0.091	0.239	0.281	0.241
0	0.082	0.157	0.186	0.156
2.5	0.069	0.120	0.156	0.129
5	0.063	0.105	0.138	0.116
10	0.062	0.086	0.120	0.098
25	0.053	0.067	0.100	0.080
50	0.052	0.059	0.089	0.069
75	0.051	0.055	0.085	0.063
100	0.049	0.053	0.082	0.062
125	0.049	0.053	0.080	0.062
150	0.046	0.052	0.078	0.061
200	0.047	0.051	0.079	0.060
250	0.044	0.049	0.077	0.058

Table 6: Size of joint Wald Tests. $K = 10$ predictors.

c	$W_{zx}^{*,\text{RWB}}$	$W_{zx}^{*,\text{FRWB}}$	W_{zx}^{EW}	W_{zx}
-5	0.058	0.513	0.635	0.559
-2.5	0.072	0.398	0.505	0.425
0	0.087	0.306	0.406	0.324
2.5	0.075	0.238	0.342	0.262
5	0.067	0.191	0.301	0.225
10	0.060	0.141	0.244	0.175
25	0.050	0.089	0.174	0.118
50	0.048	0.067	0.142	0.091
75	0.046	0.060	0.129	0.081
100	0.046	0.056	0.120	0.077
125	0.043	0.053	0.117	0.074
150	0.042	0.052	0.116	0.071
200	0.039	0.049	0.116	0.070
250	0.036	0.050	0.116	0.072

Moving on to ...

Conclusions

- Bootstrap methods provide a very powerful, albeit under-used, toolkit for improved inference in time series econometrics.

Conclusions

- Bootstrap methods provide a very powerful, albeit under-used, toolkit for improved inference in time series econometrics.
- When appropriately implemented, bootstrap methods can approximate the finite sample null distributions of statistics, even in highly complex settings where the null distributions of the statistics depend in very complex ways on nuisance parameters, even in large samples.

Conclusions

- Bootstrap methods provide a very powerful, albeit under-used, toolkit for improved inference in time series econometrics.
- When appropriately implemented, bootstrap methods can approximate the finite sample null distributions of statistics, even in highly complex settings where the null distributions of the statistics depend in very complex ways on nuisance parameters, even in large samples.
- The crucial property needed is that the bootstrap DGP, and hence the generated bootstrap data, mimics the features of the original data which impact on the (exact or limiting) distribution of the statistic of interest. If the bootstrap DGP misses an important feature of the data then it cannot be expected to perform well.

Conclusions

- Bootstrap methods provide a very powerful, albeit under-used, toolkit for improved inference in time series econometrics.
- When appropriately implemented, bootstrap methods can approximate the finite sample null distributions of statistics, even in highly complex settings where the null distributions of the statistics depend in very complex ways on nuisance parameters, even in large samples.
- The crucial property needed is that the bootstrap DGP, and hence the generated bootstrap data, mimics the features of the original data which impact on the (exact or limiting) distribution of the statistic of interest. If the bootstrap DGP misses an important feature of the data then it cannot be expected to perform well.
- As a result, if the user does not establish the limiting null distribution of the statistic of interest then it is best to ensure the bootstrap DGP mimics all of the data features in the original DGP.

Thank you for listening!