Non-nested model comparisons of differencing operators for non-stationary time series

Tucker McElroy[1]
U.S. Census Bureau

University of Florida
Gainesville, FL
September 5, 2024

[1]This presentation is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau (USCB). All time series analyzed in this presentation are from public or external data sources.

Context and Goal

Let $\{X_t\}$ be non-stationary with minimal differencing polynomial $\widetilde{\delta}(z)$.

So $\widetilde{\delta}(L) X_t = Y_t$ is weakly stationary and invertible (where L is the lag operator). All the roots of $\widetilde{\delta}(z)$ have modulus one (i.e., they are unit roots).

Let $\{Y_t\}$ have autocovariance function $\gamma_Y(h)$ and spectral density $f_Y(z) = \sum_{h \in \mathbb{Z}} \gamma_Y(h) z^h$.

Suppose two or more specifications of $\widetilde{\delta}(z)$ are available. Which is better?

Literature

Autoregressive Approach: test whether the roots of an AR operator are unit. Dickey et al. (1986), Schwert (1989), Pantula (1991), De Jong et al. (1992), Perron and Ng (1996), Elliott et al. (1996), Ng and Perron (2001), Breitung (2002), Taylor (2003), Cavaliere and Taylor (2007), Chan (2009), Cavaliere and Xu (2014), and Choi (2015).

Moving Average Approach: over-difference and test whether the spectral density has a zero. Tanaka (1996), Davis and Dunsmuir (1996), Tam and Reinsel (1997), Lacroix (1999), Chen, Davis, and Song (2011), Davis and Song (2011), and Larsson (2014).

Studentization

Non-pivotal Limit: with these approaches the test statistics' limits have a factor involving the long-run variance $f_Y(1)$ (Theorem 9.22 of Tanaka (1996)).

Studentization: one can estimate the long-run variance through a HAC estimator, or use self-normalization (Shao, 2015).

Comparisons of Specifications

In contrast, we compare two specifications of $\widetilde{\delta}(z)$. Elements:

- Nested or non-nested comparisons
- Use sum of squares statistics (motivated by sums of squared forecast errors)
- Use studentization to make pivotal

Definitions

Let $\delta^A(z)$ and $\delta^B(z)$ be two differencing polynomials (the two models).

Their GCD is the polynomial with unit roots common to both, denoted $\delta^{A \cap B}(z)$.

Their LCM is the polynomial with all of their unit roots, denoted $\delta^{A \cup B}(z)$.

We say a model δ is weakly over-specified if its set of unit roots contain all those of $\widetilde{\delta}$; denoted $\delta \supseteq \widetilde{\delta}$.

We say δ is under-specified if it does not include all the unit roots of $\widetilde{\delta}$; denoted $\delta \subset \widetilde{\delta}$.

Framework

- Suppose the models' LCM includes $\widetilde{\delta}(z)$.
- Null hypothesis: models A and B are weakly over-specified.
- Alternative hypothesis: one or both models are mis-specified.
- Apply $\delta^A(L)$ and $\delta^B(L)$, obtaining $\{Y_t^A\}$ and $\{Y_t^B\}$; these need not be stationary.
- Compute average squares of each series, which diverge if the process is non-stationary.

Non-nested Diagram

$$\begin{tikzcd}
 & \delta^{A\cup B} \arrow[dl, dash] \arrow[dr, dash] & \\
\delta^{A} \arrow[dr, dash] & & \delta^{B} \arrow[dl, dash] \\
 & \delta^{A\cap B} &
\end{tikzcd}$$

Nested Case: if $\delta^A \subseteq \delta^B$, then $\delta^{A\cup B} = \delta^B$ and $\delta^{A\cap B} = \delta^A$, and the diagram collapses.

Scenarios

Four Cases: depending on whether $\delta^A(z)$ or $\delta^B(z)$ is (weakly) over- or under-specified.

1. Both are weakly over-specified (Null): both average of squares converge.

2. Model A is weakly over-specified, but not model B: the second average of squares diverges.

3. Model B is weakly over-specified, but not model A: the first average of squares diverges.

4. Neither model is correct: both averages of squares diverge.

United States™ Census Bureau

Test Statistics

This suggests taking the difference of average sums of squares: $T^{-1} \sum_t ({Y_t^A}^2 - {Y_t^B}^2)$.

- Large positive values: model A is under-specified (case 3 or 4)
- Large negative values: model B is under-specified (case 2 or 4)
- Moderate values: null is true (case 1)

We use a studentized cusum statistic based on the difference of average sums of squares.

Example 1

Suppose that a quarterly series is suspected of having trend and/or seasonality, so that the initial specification of the differencing operator is $1 - z^4$; this includes trend differencing $1 - z$, as well as seasonal differencing $1 + z$ and $1 + z^2$ for frequencies π and $\pi/2$ respectively.

Suppose we wish to test whether the frequency π factor is needed, versus whether the trend is present. This means that we wish to test $\delta^A(z) = 1 - z + z^2 - z^3$ against $\delta^B(z) = 1 + z + z^2 + z^3$.

Then $\delta^{A \cup B}(z) = (1 - z)(1 + z)(1 + z^2) = 1 - z^4$, and $\delta^{A \cap B}(z) = 1 + z^2$.

Example 2

Again suppose an initial specification of $1 - z^4$, but now are interested in nested testing of whether either of the seasonal components are needed.

1. Test $\delta^A(z) = 1 - z + z^2 - z^3$ against the nested specification $\delta^B(z) = 1 - z^4$, wherein rejections indicate that $1 + z$ (seasonality at frequency π) is needed.

2. Test $\delta^A(z) = 1 - z^2$ against the nested specification $\delta^B(z) = 1 - z^4$, wherein rejections indicate that $1 + z^2$ (seasonality at frequency $\pi/2$) is needed.

3. Test $\delta^A(z) = 1 + z + z^2 + z^3$ against $\delta^B(z) = 1 - z^4$, wherein rejections indicate that $1 - z$ (trend) is needed.

Non-stationary Representation

What is the asymptotics for sum of squares for a non-stationary process?

Let $\{Z_t\}$ be non-stationary with minimal differencing polynomial $\delta(z)$ (of degree d) such that $U_t = \delta(L) Z_t$ and $\{U_t\}$ is stationary invertible; say $U_t = \Psi(L)\epsilon_t$ with $\{\epsilon_t\}$ white noise of variance σ^2.

Let $\xi(z) = \delta(z)^{-1} = \sum_{j \geq 0} \xi_j z^j$; then

$$Z_t = \sum_{h \geq t} \xi_h B_{t-h} + \sum_{j=0}^{t-1} \xi_j U_{t-j},$$

where B_t is a time-varying function depending on initial values $Z_0, Z_{-1}, \ldots, Z_{1-d}$.

Sums of Squares

Let $\{\zeta_n\}_{n=1}^d$ denote the roots of $\delta(z)$; (if the roots are distinct) there exist complex coefficients a_n for $1 \leq n \leq d$ such that $\xi_\ell = \sum_{n=1}^d a_n \zeta_n^{-\ell}$ for $\ell \geq 0$.

Focusing on the second (stochastic) term, and letting $S_j(\zeta) = \sum_{t=1}^j U_t \zeta^t$, the sum of squares can be written

$$\sum_{t=1}^T \left(\sum_{j=0}^{t-1} \xi_j U_{t-j} \right)^2 = \sum_{m,n=1}^d a_m a_n \sum_{j=1}^T S_j(\zeta_m)\, S_j(\zeta_n)\, (\zeta_m \zeta_n)^{-j}.$$

The dominant terms occur for m, n such that $\zeta_m \zeta_n = 1$.

Root Arrangement

Let us arrange the d roots in the following way:

- There are $2c$ complex conjugate roots $\zeta_1, \ldots, \zeta_{2c}$, such that $\zeta_m = \overline{\zeta}_{2c+1-m}$ for $1 \leq m \leq c$.

- There are $d - 2c \leq 2$ real roots $\zeta_{2c+1}, \ldots, \zeta_d$.

- Hence $\zeta_m \zeta_n = 1$ implies that either $m \in \{1, \ldots, 2c\}$ and $n = 2c+1-m$ or $m = n \in \{2c+1, \ldots, d\}$.

FCLT for DFT

Utilize a functional CLT for the DFT: for $r \in (0, 1]$

$$T^{-1/2} S_{[rT]}(\zeta_m) \stackrel{\mathcal{L}}{\Longrightarrow} \sigma \Psi(\zeta_m) \, B_r(\zeta_m)$$

as $T \to \infty$ jointly in $1 \leq m \leq d$, where $\{B_r(\zeta_m)\}_{r \in [0,1]}$ are standard Brownian Motion processes. (For conjugate roots the processes are linked, but otherwise are independent.)

Non-stationary Case Asymptotics

Then for $s \in (0,1]$,

$$T^{-2} \sum_{t=1}^{[sT]} \left(\sum_{j=0}^{t-1} \xi_j U_{t-j} \right)^2 \stackrel{\mathcal{L}}{\Longrightarrow} Q_s(\delta)$$

$$Q_s(\delta) = \sum_{m=1}^{2c} a_m a_{2c+1-m} f_U(\zeta_m) \int_0^s B_r(\zeta_m) B_r(\zeta_{2c+1-m})\, dr$$

$$+ \sum_{m=2c+1}^{d} a_m^2 f_U(\zeta_m) \int_0^s (B_r(\zeta_m))^2\, dr.$$

Stationary Case Asymptotics

When $\{Z_t\}$ is stationary (and mean zero), $T^{-1}\sum_{t=1}^{T} Z_t^2 \xrightarrow{P} \text{Var}[Z]$ and

$$T^{-1/2}\sum_{t=1}^{[sT]}(Z_t^2 - \text{Var}[Z]) \stackrel{\mathcal{L}}{\Longrightarrow} \tau B_s,$$

where $\tau^2 = f_{Z^2}(1)$ is the long-run variance (f_{Z^2} is the autocovariance generating function of $\{Z_t^2\}$).

Difference of Average of Squares

Apply both $\delta^A(L)$ and $\delta^B(L)$, so that $Y_t^A = \delta^A(L) X_t$ and $Y_t^B = \delta^B(L) X_t$. The asymptotic distribution of

$$\widehat{\theta}_t = \frac{1}{t} \sum_{s=1}^{t} {Y_s^A}^2 - \frac{1}{t} \sum_{s=1}^{t} {Y_s^B}^2$$

depends on whether δ^A or δ^B (or neither) are weakly over-specified.

Cases for Model A

Consider two cases for model A:

- If $\widetilde{\delta} \subseteq \delta^A$, then $\{Y_t^A\}$ is stationary, and $T^{-1} \sum_{t=1}^{T} {Y_t^A}^2 \xrightarrow{P} \text{Var}[Y^A]$.
- If $\widetilde{\delta} \not\subseteq \delta^A$, then $T^{-2} \sum_{t=1}^{T} {Y_t^A}^2 \xRightarrow{\mathcal{L}} Q_1(\widetilde{\delta} \setminus \delta^A)$.

All Four Cases

1. $\widetilde{\delta} \subseteq \delta^A$ and $\widetilde{\delta} \subseteq \delta^B$. Then $\widehat{\theta}_T \xrightarrow{P} \mathsf{Var}[Y^A] - \mathsf{Var}[Y^B]$.

2. $\widetilde{\delta} \subseteq \delta^A$ but $\widetilde{\delta} \not\subseteq \delta^B$. Then $T^{-1}\widehat{\theta}_T \xRightarrow{\mathcal{L}} -Q_1(\widetilde{\delta} \setminus \delta^B)$.

3. $\widetilde{\delta} \not\subseteq \delta^A$ but $\widetilde{\delta} \subseteq \delta^B$. Then $T^{-1}\widehat{\theta}_T \xRightarrow{\mathcal{L}} Q_1(\widetilde{\delta} \setminus \delta^A)$.

4. $\widetilde{\delta} \not\subseteq \delta^A$ but $\widetilde{\delta} \not\subseteq \delta^B$. Then $T^{-1}\widehat{\theta}_T \xRightarrow{\mathcal{L}} Q_1(\widetilde{\delta} \setminus \delta^A) - Q_1(\widetilde{\delta} \setminus \delta^B)$.

Cusum form of Statistics

First case is null hypothesis; to obtain a non-degenerate limit, we should center. We propose the cusum form

$$S_T = \widehat{\theta}_T - \binom{T+1}{2}^{-1} \sum_{t=1}^{T} t \widehat{\theta}_t.$$

Cusum Statistics' Asymptotics

For the four cases:

$$T^{1/2} S_T \overset{\mathcal{L}}{\Longrightarrow} \tau(B_1 - 2\int_0^1 B_s \, ds)$$

$$T^{-1} S_T \overset{\mathcal{L}}{\Longrightarrow} -\left(Q_1(\widetilde{\delta} \setminus \delta^B) - 2\int_0^1 Q_s(\widetilde{\delta} \setminus \delta^B) \, ds\right)$$

$$T^{-1} S_T \overset{\mathcal{L}}{\Longrightarrow} \left(Q_1(\widetilde{\delta} \setminus \delta^A) - 2\int_0^1 Q_s(\widetilde{\delta} \setminus \delta^A) \, ds\right)$$

$$T^{-1} S_T \overset{\mathcal{L}}{\Longrightarrow} \left(Q_1(\widetilde{\delta} \setminus \delta^A) - 2\int_0^1 Q_s(\widetilde{\delta} \setminus \delta^A) \, ds\right) - \left(Q_1(\widetilde{\delta} \setminus \delta^B) - 2\int_0^1 Q_s(\widetilde{\delta} \setminus \delta^B) \, ds\right)$$

Cusum Studentization

Studentize to remove τ from the null limit:

$$W_T = T^{-2} \sum_{t=1}^{T} t^2 \left(\widehat{\theta}_t - \widehat{\theta}_T \right)^2.$$

Under the null hypothesis (case 1), this converges weakly to $\tau^2 \int_0^1 (B_s - sB_1)^2 \, ds$, so τ will cancel out asymptotically in $S_T / \sqrt{W_T}$. This is $O_P(T^{-1/2})$.

Problem with Power

Under the alternative hypothesis (cases 2, 3, or 4), $T^{-3} W_T$ converges weakly to a functional of the Q Brownian Motion processes, and we find that again $S_T / \sqrt{W_T} = O_P(T^{-1/2})$.

So the studentized statistic has same rate of convergence under both null and alternative scenarios; there is no power!

Rate Estimation

We only want to studentize when the null is true. We can estimate whether this is the case with a consistent rate estimator.

Let β be defined such that $T^{-\beta} W_T = O_P(1)$, i.e., $\beta = 0$ under the null and $\beta = 3$ under the alternative. We can estimate β with

$$\widehat{\beta} = \frac{\log W_T}{\log T},$$

which is consistent at rate $\log T$: $\log T(\widehat{\beta} - \beta) = \log(T^{-\beta} W_T) = O_P(1)$.

Cusum Studentization

Consider a studentization that smoothly depends upon $\widehat{\beta}$:

$$P_T = \cos(\pi\widehat{\beta}/6)^2 \sqrt{W_T} + \sin(\pi\widehat{\beta}/6)^2 \widehat{\beta}/3.$$

This "weights" $\sqrt{W_T}$ and 1, ensuring that the only the former is asymptotically present when $\widehat{\beta} \xrightarrow{P} 0$, while also enforcing that the normalization tends to 1 as $\widehat{\beta} \xrightarrow{P} 3$.

Cusum Studentization

Test the null hypothesis with $\sqrt{T} S_T / P_T$. Under the null,

$$T^{1/2} \frac{S_T}{P_T} \xrightarrow{\mathcal{L}} \frac{\left(B_1 - 2 \int_0^1 B_s \, ds\right)}{\left\{\int_0^1 (B_s - sB_1)^2 \, ds\right\}^{1/2}}$$

Under the alternative,

$$T^{1/2} \frac{S_T}{P_T} = O_P((\log T)^2),$$

indicating the test is consistent.

Top-down Identification

Begin with a specification, and test whether to remove factors.

Multiple Routes for Roots: depending on the order in which we test factors, we could end with different (non-nested) specifications.

Top-down Example

$$\begin{tikzcd}[column sep=small]
 & \boxed{1 - z^4} \arrow[dl, dash] \arrow[dr, dash] & \\
\boxed{1 - z + z^2 - z^3} \arrow[dr, dash] & & \boxed{1 + z + z^2 + z^3} \arrow[dl, dash] \\
 & \boxed{1 + z^2} &
\end{tikzcd}$$

1. Test whether $1 + z$ is needed.
2. Test whether $1 - z$ is needed.

Ordering

Given a collection of m viable models, we wish to make $\binom{m}{2}$ comparisons, and order them.

Weak Favoring: a significant upper 1-sided test indicates B is significantly favored over A. Say B is "weakly favored" over A, written $B \geq A$, if test statistic is non-negative.

Compatibility of Weak Favoring

$B \geq A$ iff $S_T \geq 0$ iff $\widehat{\theta}_T \geq \left(\binom{T+1}{2}\right)^{-1} \sum_{t=1}^{T} t\widehat{\theta}_t$ iff

$$\sum_{s=1}^{T} (T+1-s)\, {Y_s^A}^2 \geq \sum_{s=1}^{T} (T+1-s)\, {Y_s^B}^2.$$

Hence the testing framework is transitive: $B \geq A$ and $C \geq B$ implies $C \geq A$. So we avoid M.C. Escher contradictions (e.g., $B \geq A$ and $C \geq B$, but $C \leq A$).

Non-transitive (M.C. Escher)

Procedure

This discussion suggests the following procedure:

1. Obtain m models via nested unit-root testing in a top-down approach.

2. Perform $\binom{m}{2}$ nested/non-nested comparisons of these models.

3. For each pair, determine which model is weakly favored: the left model is weakly favored over the right model if and only if the upper 1-sided test does not reject, and the right model is weakly favored over the left model if and only if the lower 1-sided test does not reject.

4. Rank the models by weak favoring.

Test Statistic's Null Distribution

The null distribution has been generated by 10^6 simulations, utilizing a discretization of the Brownian Motion at mesh size 10^3.

It has an "M" shape, with very light tails; the majority of the mass falling between -2 and 2.

Simulation in Example 1

Null process has $\widetilde{\delta}(z) = 1 + z^2$, and $\delta^A(z) = 1 - z + z^2 - z^3$ and $\delta^B(z) = 1 + z + z^2 + z^3$. Let $\{U_t\}$ to be i.i.d. standard normal.

Settings: samples size $T = 50, 100, 200, 400, 800$, and $\alpha = .01, .05, .10$.

Summary: small size distortions (first column), 1-sided tests (second and third columns) have power exceeding 70% even for small α and T. Improvement in T is slow. Slightly higher power for 2-sided tests (fourth column).

T	α	$\widetilde{\delta} = \delta^{A\cap B}$	$\widetilde{\delta} = \delta^{A}$	$\widetilde{\delta} = \delta^{B}$	$\widetilde{\delta} = \delta^{A\cup B}$
$T = 50$	.01	.0257	.7105	.7079	.8112
	.05	.0594	.7157	.7133	.8212
	.10	.1011	.7212	.7183	.8297
$T = 100$	.01	.0234	.7475	.7477	.8629
	.05	.0624	.7513	.7517	.8698
	.10	.1078	.7548	.7553	.8752
$T = 200$	.01	.0222	.7713	.7725	.8940
	.05	.0635	.7742	.7756	.8989
	.10	.1105	.7769	.7783	.9023
$T = 400$	.01	.0203	.7850	.7846	.9152
	.05	.0636	.7872	.7868	.9188
	.10	.1120	.7891	.7890	.9219
$T = 800$	.01	.0184	.7942	.7965	.9304
	.05	.0614	.7960	.7983	.9335
	.10	.1126	.7978	.7999	.9359

Other Simulations

Size Distortion: consider testing a stationary null versus alternative $\widetilde{\delta}(z) = 1 - z$ (as alternative), where the process is an AR(1) with parameter ϕ; size degrades as ϕ approaches unity.

Power Distortion: consider testing a stationary null versus alternative $\widetilde{\delta}(z) = 1 - z$ (as alternative), where the process is an ARIMA(0,1,1) with parameter θ; power degrades as θ approaches -1.

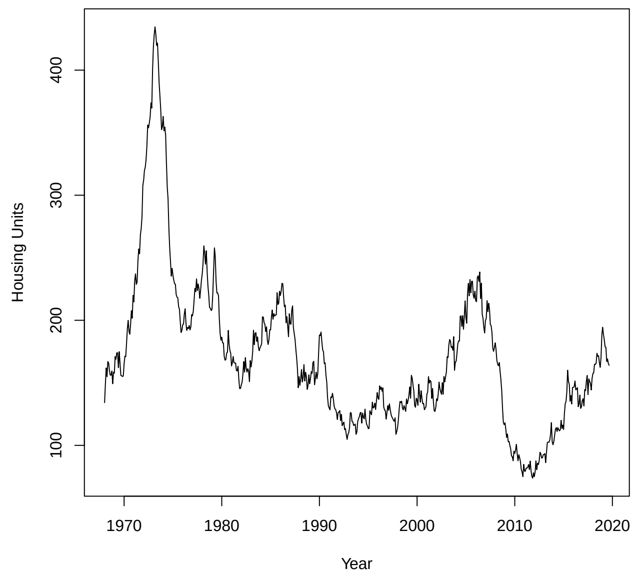

Figure 1: Housing series: "Housing units authorized but not started," January 1968 through September 2019.

Unstarts Data

Consider "Housing units authorized but not started," which is a monthly time series covering January 1968 through September 2019 (downloaded on November 5, 2019). The data, which we refer to as Unstarts for short, has seasonal, trend, and cyclical effects: see Figure 1.

Spectral Density: use an AR(15) spectral estimator (the order is determined by AIC, and the fitting is OLS), having marked with red vertical lines in Figure 2 the locations of the five seasonal frequencies $2\pi j/12$ for $j = 1, 2, 3, 4, 5$.

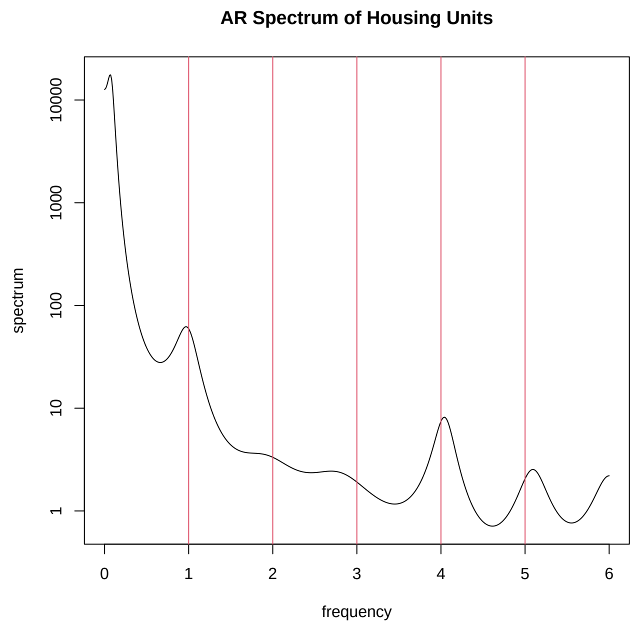

Figure 2: AR(15) spectral density plot of Unstarts series. Red vertical lines correspond to seasonal frequencies.

Omit Second and Third Seasonal Frequencies?

Note that $1 - z^{12} = (1-z)\prod_{j=1}^{5}(1 - 2\cos(2\pi j/12)z + z^2)\,(1+z)$.

$$1 - z^{12}$$

$$(1-z)\prod_{j=1,3,4,5}(1 - 2\cos(2\pi j/12)z + z^2)\,(1+z)$$

$$(1-z)\prod_{j=1,2,4,5}(1 - 2\cos(2\pi j/12)z + z^2)\,(1+z)$$

$$(1-z)\prod_{j=1,4,5}(1 - 2\cos(2\pi j/12)z + z^2)\,(1+z)$$

Unstarts Results

- A is favored against $A \cup B$ with test statistic of 6.66.
- A is favored against $A \cap B$ with test statistic of -85.59.
- $A \cup B$ is favored against B with test statistic of -83.34.
- B against $A \cap B$ is rejected with test statistic of 5.19.
- $A \cup B$ is favored over $A \cap B$ with test statistic of -69.47.
- A is favored over B with a test statistic of -102.87.
- So $A \geq A \cup B \geq A \cap B \geq B$.

Summary

1. Testing framework for non-nested (and nested) comparisons, to find the best differencing operator $\delta(z)$ for non-stationary time series.

2. Employ a studentized cusum statistic based on a difference of sums of squares, for time series differenced according to both specifications of $\delta(z)$.

3. Studentization is adjusted with a rate estimation.

Contact: tucker.s.mcelroy@census.gov