Quantitative Economics Problem Set 7

1a $x_{s+t} = \beta_0 + \beta_1 x_{s+t-1} + \varepsilon_{s+t}$

$= \beta_0 + \beta_1(\beta_0 + \beta_1 x_{s+t-2} + \varepsilon_{s+t-1}) + \varepsilon_{s+t}$

$\vdots$

$= \beta_0 + \beta_1\beta_0 + \dots + \beta_1^{t-1}\beta_0$

$+ \beta_1^t x_{s+t-t}$

$+ \varepsilon_{s+t} + \beta_1\varepsilon_{s+t-1} + \dots + \beta_1^{t-1}\varepsilon_{s+t-(t-1)}$

$= \beta_0 \sum_{i=0}^{t-1}\beta_1^i + \beta_1^t x_s + \sum_{i=0}^{t-1}\beta_1^i \varepsilon_{s+t-i}$

For $s=0$, $x_{s+t} = x_t = \beta_0\sum_{i=0}^{t-1}\beta_1^i + \beta_1^t x_0 + \sum_{i=0}^{t-1}\beta_1^i\varepsilon_{s+t-i}$

~~$cov(x_t, \varepsilon_t)$~~

$cov(x_t, \varepsilon_{t-i}) = \beta_1^i \, var(\varepsilon_{t-i})$

$= \beta_1^i \sigma_\varepsilon^2$

Given that x_0 is independent of ε_t for all $t \geq 1$, and that $\{\varepsilon_t\}$ is iid hence each ε_t is independent of every $\varepsilon_{t'}$ where $t \neq t'$, x_t, which is a function of x_0 and $\varepsilon_1, \dots, \varepsilon_t$ is independent of ε_{t+1}, $\varepsilon_{t+2}, \dots$.

b $x_t = \beta_0\sum_{i=0}^{t-1}\beta_1^i + \beta_1^t x_0 + \sum_{i=0}^{t-1}\beta_1^i\varepsilon_{t-i}$

$Ex_t = \beta_0\sum_{i=0}^{t-1}\beta_1^i + \beta_1^t Ex_0 + \sum_{i=0}^{t-1}\beta_1^i E\varepsilon_{t-i}$

$= \beta_0(1-\beta_1^t)/(1-\beta_1) + \beta_1^t\beta_0/(1-\beta_1)$

$= \beta_0/(1-\beta_1)$

$var(x_t) = var(\beta_1^t x_0 + \sum_{i=0}^{t-1}\beta_1^i\varepsilon_{t-i})$

$= \beta_1^{2t} var(x_0) + \sum_{i=0}^{t-1}\beta_1^i var(\varepsilon_{t-i})$

$= \beta_1^{2t}\sigma_\varepsilon^2/(1-\beta_1^2) + \sum_{i=0}^{t-1}\beta_1^i\sigma_\varepsilon^2$

$= \beta_1^{2t}\sigma_\varepsilon^2/(1-\beta_1^2) + (1-\beta_1^{2t})\sigma_\varepsilon^2/(1-\beta_1^2)$

$= \sigma_\varepsilon^2/(1-\beta_1^2)$

~~cov(xs~~

$cov(x_s, x_{s+t}) = cov(x_s, \beta_1^t x_s)$

$= \beta_1^t var(x_s)$

$= \beta_1^t\sigma_\varepsilon^2/(1-\beta_1^2)$

where the first equality follows by the independence of x_s from $\varepsilon_{s+1}, \dots, \varepsilon_{s+t}$.

For all t, $Ex_t = \beta_0/(1-\beta_1)$, $var(x_t) = \sigma_\varepsilon^2/(1-\beta_1^2)$, and for all h, $cov(x_t, x_{t-h}) = \beta_1^h\sigma_\varepsilon^2/(1-\beta_1^2)$. These parameters are time invariant, so x_t is weakly stationary.

c $x_t = t\beta_0 + \sum_{i=0}^{t-1}\varepsilon_{t-i}$

$= t\beta_0 + \sum_{i=1}^{t}\varepsilon_i$

$Ex_t = t\beta_0 + \sum_{i=1}^{t} E\varepsilon_i$

$= t\beta_0$

$var(x_t) = var(\sum_{i=1}^{t}\varepsilon_i)$

$= \sum_{i=1}^{t} var(\varepsilon_i)$

$= t\sigma_\varepsilon^2$

Each of Ex_t and $var(x_t)$ is increasing in t and not independent of t, hence x_t is not weakly stationary.

d $x_{s+h} = \beta_0\sum_{i=0}^{h-1}\beta_1^i + \beta_1^h x_s + \sum_{i=0}^{h-1}\beta_1^i\varepsilon_{s+h-i}$

$x_{s+h|s} := E[x_{s+h} \mid x_s, x_{s-1}, \dots]$

$= E[\beta_0\sum_{i=0}^{h-1}\beta_1 + \beta_1^h x_s + \sum_{i=0}^{h-1}\beta_1^i\varepsilon_{s+h-i}]$

$= \beta_0\sum_{i=0}^{h-1}\beta_1 + \beta_1^h x_s + \cancel{\sum_{i=0}^{h-1}\beta_1^i\varepsilon_{s+h-i}}$

where the third equality follows from the independence of $\varepsilon_{s+1}, \dots, \varepsilon_{s+h}$ from $x_s, x_{s-1}, \dots$

~~$MSFE(x_{s+h|s}) := E(x_{s+h} - x_{s+h|s})^2$~~

~~$= (\sum_{i=0}^{h-1}\beta_1^i\varepsilon_{s+h-i})^2$~~

~~$= E(\sum_{i=0}^{h-1}\beta_1^i\varepsilon_{s+h-i})^2$~~

~~$= E(\sum_{i=0}^{h-1}\sum_{j=0}^{h-1}\beta_1^i\beta_1^j\varepsilon_{s+h-i}\varepsilon_{s+h-j})$~~

~~$= \sum_{i=0}^{h-1}\sum_{j=0}^{h-1}\beta_1^{i+j}\sigma_\varepsilon^2$~~

~~$= \sigma_\varepsilon^2(\sum_{i=0}^{h-1}\beta_1^i)^2$~~

~~when $|\beta_1| < 1$,~~

~~$MSFE(x_{s+h|s}) = [(1-\beta_1^h)/(1-\beta_1)]^2\sigma_\varepsilon^2$~~

~~As h becomes large, β_1^h converges to 0, $MSFE(x_{s+h|s})$ converges to $1/(1-\beta_1)^2\sigma_\varepsilon^2$~~

~~when $\beta_1 = 1$~~

~~$MSFE(x_{s+h|s}) = h^2\sigma_\varepsilon^2$~~

~~As h becomes large, $MSFE(x_{s+h|s})$ increases exponentially with h, and approaches ∞ as h becomes large~~

$MSFE(x_{s+h|s}) := E(x_{s+h} - x_{s+h|s})^2$

$= E(\sum_{i=0}^{h-1}\beta_1^i\varepsilon_{s+h-i})^2$

$= var(\sum_{i=0}^{h-1}\beta_1^i\varepsilon_{s+h-i})$

$= \sum_{i=0}^{h-1}\beta_1^{2i} var(\varepsilon_{s+h-i})$

$= \sigma_\varepsilon^2\sum_{i=0}^{h-1}\beta_1^{2i}$

~~$= \sigma_\varepsilon^2(1-\beta_1^{2h})/(1-\beta_1)$~~

when $|\beta_1| < 1$ $MSFE = \sigma_\varepsilon^2\sum_{i=0}^{h-1} = \sigma_\varepsilon^2(1-\beta_1^{2h})/(1-\beta_1)$

$MSFE(x_{s+h|s})$ converges to $\sigma_\varepsilon^2/(1-\beta_1) = var(x_t)$ as h becomes large because $x_{s+h|s}$ converges to Ex_t.

when $\beta_1 = 1$

$MSFE(x_{s+h|s}) = \sigma_\varepsilon^2\sum_{i=0}^{h-1} 1$

$= h\sigma_\varepsilon^2$

$MSFE(x_{s+h|s})$ increases in direct proportion to h. The series is a random walk with drift, so the forecast error is equal to the random walk component. This component is mean zero, so the MSFE is equal to its variance, which is $h\sigma_\varepsilon^2$ given that the random shocks are iid.

2a $Ex_t = E(\varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q})$

$= E\varepsilon_t + \theta_1 E\varepsilon_{t-1} + \dots + \theta_q E\varepsilon_{t-q}$

$= 0$

$var(x_t) = var(\varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q})$

$= var\,\varepsilon_t + \theta_1^2 var\,\varepsilon_{t-1} + \dots + \theta_q^2 var\,\varepsilon_{t-q}$

$= (1 + \theta_1^2 + \dots + \theta_q^2)\, var(\varepsilon_{t-q})$

$= (\sum_{i=0}^{q}\theta_i^2)\sigma_\varepsilon^2$

where $\theta_0 = 1$

b $\text{cov}(x_t, x_{t-h})$
$= \text{cov}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}, \varepsilon_{t-h} + \theta_1\varepsilon_{t-h-1} + \dots + \theta_q\varepsilon_{t-h-q})$
$= 0$

for $h > q$, given that $\{\varepsilon_t\}_t$ is iid hence $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$ for all $t \neq s$, by linearity of covariance.

c $\text{cov}(x_t, x_t) = \text{var}(x_t) = \sigma_\varepsilon^2 \sum_{i=0}^{q} \theta_i^2$
$\text{cov}(x_t, x_{t-1})$
$= \cancel{\text{cov}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}, \varepsilon_{t-1} + \theta_1\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q-1})}$
$= \cancel{\text{cov}(\theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{t-q}, \varepsilon_{t-1} + \theta_1\varepsilon_{t-2}}$
$= \text{cov}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \varepsilon_{t-1} + \theta_1\varepsilon_{t-2} + \theta_2\varepsilon_{t-3})$
$= \text{cov}(\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \varepsilon_{t-1} + \theta_1\varepsilon_{t-2})$
$= \theta_1 \text{var}(\varepsilon_{t-1}) + \theta_1\theta_2 \text{var}(\varepsilon_{t-2})$
$= \theta_1(1+\theta_2)\sigma_\varepsilon^2$
$\text{cov}(x_t, x_{t-2})$
$= \text{cov}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \varepsilon_{t-2} + \theta_1\varepsilon_{t-3} + \theta_2\varepsilon_{t-4})$
$= \theta_2 \text{var}(\varepsilon_{t-2})$
$= \theta_2\sigma_\varepsilon^2$

$$\text{cov}(x_t, x_{t-h}) = \begin{cases} \sigma_\varepsilon^2 \sum_{i=0}^{2} \theta_i^2 = (1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2 & \text{if } h=0 \\ \theta_1(1+\theta_2)\sigma_\varepsilon^2 & \text{if } h=1 \\ \theta_2\sigma_\varepsilon^2 & \text{if } h=2 \\ 0 & \text{if } h>2 \end{cases}$$

By inspection, each of Ex_t, $\text{var}(x_t)$ and $\text{cov}(x_t, x_{t-h})$ is independent of t, so x_t is weakly stationary.

di $x_{t-h} = \varepsilon_{t-h} + \theta_1\varepsilon_{t-h-1}$
For $h \geq 1$, $\varepsilon_{t-h}, \varepsilon_{t-h-1} \neq \varepsilon_t$, then, given that $\{\varepsilon_t\}$ is iid, $\varepsilon_{t-h}, \varepsilon_{t-h-1} \perp\!\!\!\perp \varepsilon_t$, then x_{t-h}, which is ~~of~~ entirely a function of $\varepsilon_{t-h}, \varepsilon_{t-h-1}$, is independent of ε_t.

ii $\varepsilon_t = x_t - \theta_1\varepsilon_{t-1}$
$= x_t - \theta_1(x_{t-1} - \theta_1\varepsilon_{t-2})$
$= x_t - \theta_1 x_{t-1} + (-\theta_1)^2\varepsilon_{t-2}$
$= x_t - \theta_1 x_{t-1} + (-\theta_1)^2 x_{t-2} + (-\theta_1)^3 \varepsilon_{t-3}$
$\vdots$
$= \sum_{i=0}^{t-1} (-\theta_1)^i x_{t-i} + (-\theta_1)^t \varepsilon_0$
$= \sum_{i=0}^{t-1} (-\theta_1)^i x_{t-i}$

iii $x_{T+1|T}$
$:= E[x_{T+1} | x_T, x_{T-1}, \dots]$
$= E[\varepsilon_{T+1} + \theta_1\varepsilon_T | x_T, x_{T-1}, \dots]$
$= E[\varepsilon_{T+1} | x_T, x_{T-1}, \dots] + \theta_1 E[\varepsilon_T | x_T, x_{T-1}, \dots]$
$= E[\varepsilon_{T+1}] + \theta_1 E[\sum_{i=0}^{t-1}(-\theta_1)^i x_{t-i} | x_T, x_{T-1}, \dots]$
$= \theta_1 \sum_{i=0}^{t-1} (-\theta_1)^i x_{t-i}$
where the fourth equality follows by the fact that each of $x_T, x_{T-1}, \dots$ is entirely a function of $\varepsilon_T, \varepsilon_{T-1}, \dots$, and the fact that ε_{T+1} is independent from $\varepsilon_T, \varepsilon_{T-1}, \dots$, hence also from $x_T, x_{T-1}, \dots$.

3a Monthly excess returns. The time series has low persistence because it does not have a smooth trajectory, takes only short excursions from the mean, and its autocorrelation decays to zero rapidly with increasing lags.

The early 1930s appear to constitute a different epoch from the subsequent range of periods. In the former, the time series appears to ~~exhibit much larger~~ have larger variance. It does not appear to ~~be~~ be non stationary otherwise.

No further transformation appears necessary for stationarity.

b Dividend-price ratio. The series appears highly persistent. The plot of the series is ~~smo~~ relatively smooth and takes long excursions from the mean. Autocorrelation decays to zero ~~slo~~ slowly with increasing lags.

The 1930s appear to constitute a different epoch from the subsequent range of periods. In the former range, dividend price ratio appears to have a higher mean and variance.

~~The~~ ~~The series~~ Plausibly, differencing the series will yield a series that ~~is~~ is more approximately stationary hence more ~~appropriately estimated to~~ appropriate for estimation by an autoregressive model

c USD-AUD exchange rate. The series appears to be highly persistent. The plot is relatively smooth and takes lengthy excursions from the mean. ~~The~~ Autocorrelation decays to zero slowly with increasing lags.

There appear to be at least three different epochs. First from 1980 to the early 2000s, then from the early 2000s to the early 2010s, then from the early 2010s ~~to~~ onward. Another source of non stationarity within each epoch is an apparent deterministic / stochastic trend.

Plausibly, differencing the series will yield a series that is more approximately stationary, hence more appropriate for approximation by an autoregressive model.

d US real PCE quarterly growth rate. The series appears to have low persistence. The plot is not smooth and the series takes only short excursions from the mean. Autocorrelation decays to zero rapidly with increasing lags.

The range ~~of~~ of periods ~~of to~~ ending around the 1950s appears to constitute an epoch of high variance.

No transformation appears necessary to restore stationarity.

4a No. ~~The monthly percentage change in industrial production measured in percentage points per annum is $1200\times(IP_t/IP_{t-1}-1)$.~~

Suppose that the compound annual growth rate is x. Then $IP_t = IP_{t-1}(1+x)^{1/12}$, $\ln(IP_t) = \ln(IP_{t-1}) + 1/12\ln(1+x)$, $\ln(IP_t/IP_{t-1}) = 1/12\ln(1+x)$, $1200\ln(IP_t/IP_{t-1}) = 100\ln(1+x) \simeq 100x$ (for small x). So Y_t is approximately equal to the monthly percentage change in percentage points per annum.

~~13 M7 M8 M9 M10 M11 M12~~
~~IP 99.0 99.6 100.2 100.4 101.0 101.4~~
~~Y~~

b

13	M7	M8	M9	M10	M11	M12
IP	99.0	99.6	100.2	100.4	101.0	101.4
Y		7.2508	7.2072	2.3928	7.1500	4.7431

14	M1	M2	~~M3~~	~~M4~~
$\hat{Y}$	4.0892	3.9230		

c The forecaster's intention is to accommodate potential seasonality in the model.

~~In consider th~~ The associated t statistic for the hypothesis test of the null that the true value of the coefficient on Y_{t-12} in the population regression is zero against the alternative that it is non-zero is $-0.063/0.045 = -1.4$. Fail to ~~Reject~~ the null at the 10% level of significance which has critical value 1.645. ~~Then, this rejection should be interpreted as the coeff the~~ Y_{t-12} should not be included in the model because the relationship between Y_t and Y_{t-12} is not statistically significant.

Plausibly, ~~a one sided test on~~ given that Y_{t-12} is included to accommodate potential seasonality, the ~~expected~~ coefficient on Y_{t-12} is expected to be positive, a priori, and a one-sided test should be conducted instead. Then, ~~the critical values~~ fail to reject the null at all levels of significance. Again, Y_{t-12} should not be included as a regressor.

d $AIC = \ln SSR_m/T + m\,2/T$

$BIC = \ln SSR_m/T + m\ln T/T$

$T = 12\times(2013-1986+1) = 336$

p	0	1	2	3	4	5	6
m	1	2	3	4	5	6	7
SSRm	19533	18643	17377	16285	15842	15824	15824
AIC	4.0687	4.0280	3.9636	3.9047	3.8831	3.8819	3.8938
BIC	4.0801	4.0507	3.9977	3.9501	3.9399	3.9560	3.9734

Both AIC and BIC select the model with $p=4$ lags.

e One variable Granger causes another iff the mean-squared error of the estimate of the latter variable on the basis of both variables is less than the mean-squared error of the estimate of the latter variable on the basis of the latter variable alone. In symbols, ~~this is iff~~ X_t Granger causes Y_t iff

$E[Y_t - E[Y_t \mid Y_{t-1}, X_{t-1}]]^2 < E[Y_t - E[Y_t \mid Y_{t-1}]]^2$.

Equivalently, X_t does not Granger cause Y_t iff $E[Y_t - E[Y_t \mid Y_{t-1}, X_{t-1}]]^2 = E[Y_t - E[Y_t \mid Y_{t-1}]]^2$ which is iff $E[Y_t \mid Y_{t-1}, X_{t-1}]^2 = E[Y_t \mid Y_t]$.

The test for Granger causality given the ADL(4,4) model

$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_4 Y_{t-4} + \delta_1 X_{t-1} + \cdots + \delta_4 X_{t-4} + u$

is a test of

$H_0: \delta_1 = \cdots = \delta_4 = 0$ against

$H_1: \delta_1 \neq 0, \ldots,$ or $\delta_4 \neq 0$

The test statistic is the ~~adjusted~~ F statistic

$F = \frac{n-k-1}{q}\, SSR_{rs} - SSR_{un}/SSR_{un}$

$= \frac{T-8-1}{4}\, 15842 - 13147/13147$

$= 16.758$

Under the null, $F \xrightarrow{d} F_{4,\infty}$

The critical value at the 1% level of significance from the $F_{4,\infty}$ distribution is 3.32. At the 1% level of significance, reject the null, and conclude that ΔR_t Granger causes Y_t.

f Y_t is more plausibly stationary than IP_t because IP_t appears to have a deterministic and/or stochastic trend. The plot of IP_t is smooth, this suggests that IP_t is highly persistent hence non stationary. Likewise for R_t and ΔR_t.

OLS estimators are not asymptotically normal for non-stationary series, so the standard methods for conducting hypothesis tests (and constructing confidence intervals) do not apply. Then, ~~it is not~~ the above test of Granger causality cannot be conducted.

5a $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{q} \delta_i X_{t-i} + u_t$.

$\hat\beta_0 = \bar Y - \sum_{i=1}^{p} \hat\beta_i \bar Y_{t-i} - \sum_{i=1}^{q} \hat\delta_i \bar X_{t-i}$

$\hat\beta_i = \widehat{cov}(Y_t, Y_{t-i}) / \widehat{var}(Y_{t-i})$ for $i \in \{1, \dots, p\}$

$\hat\delta_i = \widehat{cov}(Y_t, X_{t-i}) / \widehat{var}(X_{t-i})$ for $i \in \{1, \dots, q\}$

$Y_t^\dagger = \beta_0^\dagger + \sum_{i=1}^{p} \beta_i^\dagger Y_{t-i}^\dagger - \sum_{i=1}^{q} \delta_i^\dagger X_{t-i}^\dagger u_t^\dagger$

$\hat\beta_0^\dagger = \bar Y^\dagger - \sum_{i=1}^{p} \hat\beta_i^\dagger Y_{t-i}^\dagger - \sum_{i=1}^{q} \hat\delta_i^\dagger X_{t-i}^\dagger$

$\hat\beta_i^\dagger = \widehat{cov}(Y_t^\dagger, Y_{t-i}^\dagger) / \widehat{var}(Y_{t-i}^\dagger)$

$= \widehat{cov}(aY_t, aY_{t-i}) / \widehat{var}(aY_{t-i})$

$= \widehat{cov}(Y_t, Y_{t-i}) / \widehat{var}(Y_{t-i})$

$= \hat\beta_i$ for $i \in \{1, \dots, q\}$

$\hat\delta_i^\dagger = \widehat{cov}(Y_t^\dagger, X_{t-i}^\dagger) / \widehat{var}(X_{t-i}^\dagger)$

$= \widehat{cov}(aY_t, bX_t) / \widehat{var}(bX_t)$

$= a/b \, \hat\delta_i$ for $i \in \{1, \dots, q\}$

$\hat\beta_0^\dagger = a\bar Y_t - a\sum_{i=1}^{p} \hat\beta_i Y_{t-i} - a\sum_{i=1}^{q} \hat\delta_i X_{t-i}$

Then $\hat u_t^\dagger = a\hat u_t$,

$SSR_m^\dagger = \sum_{t=1}^{T} \hat u_t^{\dagger 2} = \sum_{t=1}^{T} a^2 \hat u_t^2 = a^2 SSR_m$

For $a > 1$, $a^2 > 1$, then $SSR_m^\dagger > SSR_m$, and $SSR_m^\dagger$ decreases more rapidly with increasing m, hence $\ln SSR_m^\dagger / T$ decreases more rapidly with increasing m, so the m such that $IC^\dagger$ is minimised is larger than the m such that IC is minimised. IC selects a larger model.

For $a < 1$, $a^2 < 1$, by an analogous argument, IC selects a smaller model.

b Denote the quantities for the smaller model with $p-1$ lags by $'$.

$F = \frac{n-k-1}{q} \frac{SSR' - SSR}{SSR}$

$= \frac{T-p-1}{1} \frac{SSR' - SSR}{SSR}$

$= T-p-1 \, (SSR'/SSR - 1)$

$IC' = \ln SSR'/T - (m-1)\eta_T/T$

$IC = \ln SSR/T - m\eta_T/T$

$IC' - IC = \ln(SSR'/SSR) - \eta_T/T$

$= \ln(1 + F/(T-p-1)) - \eta_T/T$

$= \ln(1 + t^2/(T-p-1)) - \eta_T/T$

c Given large T, $t^2/T-p-1$ is small, then

$IC' - IC \approx t^2/(T-p-1) - \eta_T/T$

$IC' - IC \gtreqless 0 \iff t^2/(T-p-1) \gtreqless \eta_T/T$

$\iff t^2 \gtreqless (T-p-1)/T \, \eta_T \iff t \gtreqless ((T-p-1)/T)^{1/2} \eta_T^{1/2}$

$\iff t \gtreqless \eta_T^{1/2}$ given large T.

d Under Ho, the probability that the short model is chosen is the probability that $IC' < IC$, $\iff IC' - IC < 0 \iff t < \eta_T^{1/2}$ is equal to $\cancel{\Phi(\eta_T^{1/2})}$ $P(|z| \le \eta_T^{1/2})$ $\cancel{2\Phi(-\eta_T^{1/2})}$. If $\eta_T = \eta$ for all T, this probability is independent of T. If $\eta_T \to \infty$ as $T \to \infty$, this probability converges to 1 as $\eta \to \infty$.

e For AIC, $\eta_T = 2$, for BIC, $\eta_T = \ln T$, which approaches ∞ as T becomes large. Under the null, that $\beta_p = 0$, as T becomes large, the probability that BIC selects the AR(p-1) model over the AR(p) model converges to 1 whereas the probability that AIC does so remains ~~positive~~ less than 1. ~~AIC is more conservative (erring on the side of overfitting) than BIC~~. AIC selects a larger model than necessary with positive probability, even when T is large.