Pitfalls and Possibilities in Predictive Regression

Peter C. B. Phillips

SoFiE Conference June 2013
Overview

Pitfalls in Predictive Regressions

- Bias + Nonstandard Inference
  - spurious predictability
  - difficulties with multiple predictors + strong endogeneity

- Confidence Intervals can have Zero Coverage Probability
  - failure of uniformity (in AR and predictive regression)
  - prediction tests reject with probability one under null

- Quantile Predictive Regression
  - crossing problems
  - invalidity of QR for nonstationary data
  - nonstandard inference for LUR case

- Model Validity under Predictability
  - Meaningful alternatives
  - Equation balancing in time series characteristics
Overview
Possibilities

- **IVX regression**
  - use reconstructed regressor as own instrument (Magdalinos-Phillips, 2009)
  - easy to use in empirical work (Pitarakis & Gonzalo, 2012)
  - has broad asymptotic validity — correct CI

- **Quantile Predictive Regression & IVX versions**
  - resolve quantile crossings & validity
  - provides predictions at all quantiles
  - superior for tail event prediction

- **Nonparametric regression**
  - NP regression addresses balance & validity issues
  - asymptotics are uniform across SM, LM, nonstationary cases
Prediction Model

Predictive regressions: widely used linear predictive regression

- Stock returns vs D/P ratio, Future ER vs forward premium, Consumption growth vs lagged Income, etc
- Designed in simple form

\[
y_{t+1} = \beta x_t + u_{0t+1} \\
x_{t+1} = \rho x_t + u_{xt+1}
\]

- Simple mds innovation structure often used: \( u_t = (u_{0t}, u_{xt})' \)

\[
\mathbb{E}_{\mathcal{F}_{t-1}} u_t = 0, \quad \mathbb{E}_{\mathcal{F}_{t-1}} [u_t u_t'] = \begin{bmatrix} \sigma_{00} & \sigma_{0x} \\ \sigma_{x0} & \sigma_{xx} \end{bmatrix} =: \Sigma
\]

- General WD CHE dependence structure possible/desirable
Prediction Model

Predictive regressions: widely used linear predictive regression

- Stock returns vs D/P ratio, Future ER vs forward premium, Consumption growth vs lagged Income, etc
- Designed in simple form

\[
\begin{align*}
y_{t+1} &= \beta x_t + u_{0t+1} \\
x_{t+1} &= \rho x_t + u_{xt+1}
\end{align*}
\]

- Simple mds innovation structure often used: \( u_t = (u_{0t}, u_{xt})' \)

\[
\mathbb{E}_F u_t = 0, \quad \mathbb{E}_F [u_t u_t'] = \begin{bmatrix} \sigma_{00} & \sigma_{0x} \\ \sigma_{x0} & \sigma_{xx} \end{bmatrix} =: \Sigma
\]

- General WD CHE dependence structure possible/desirable
Pitfalls: (a) Bias
Finite Sample Bias with Stationary Regressors

- Centered OLS decomposition:

\[
(\hat{\beta} - \beta) = \frac{\sum_{t=1}^{n} x_{t-1} u_{0t}}{\sum_{t=1}^{n} x_{t-1}^2} = \frac{\sum_{t=1}^{n} x_{t-1} u_{0xt}}{\sum_{t=1}^{n} x_{t-1}^2} + \left( \frac{\sigma_{0x}}{\sigma_{xx}} \right) \frac{\sum_{t=1}^{n} x_{t-1} u_{xt}}{\sum_{t=1}^{n} x_{t-1}^2}
\]

where \( u_{0xt} = u_{0t} - \frac{\sigma_{0x}}{\sigma_{xx}} u_{xt} \).

- Under normality, stationarity (\(|\rho| < 1|\)), intercept – Stambaugh (1999):

\[
\mathbb{E} (\hat{\beta} - \beta) = - \left( \frac{\sigma_{0x}}{\sigma_{xx}} \right) \left( \frac{1 + 3\rho}{n} \right) + O \left( \frac{1}{n^2} \right)
\]
Centered OLS decomposition:

\[
\hat{\beta} - \beta = \frac{\sum_{t=1}^{n} x_{t-1} u_{0t}}{\sum_{t=1}^{n} x_{t-1}^2} \]

\[
= \frac{\sum_{t=1}^{n} x_{t-1} u_{0,xt}}{\sum_{t=1}^{n} x_{t-1}^2} + \left( \frac{\sigma_{0x}}{\sigma_{xx}} \right) \frac{\sum_{t=1}^{n} x_{t-1} u_{xt}}{\sum_{t=1}^{n} x_{t-1}^2}
\]

\[
= \frac{\sum_{t=1}^{n} x_{t-1} u_{0,xt}}{\sum_{t=1}^{n} x_{t-1}^2} + \left( \frac{\sigma_{0x}}{\sigma_{xx}} \right) (\hat{\rho} - \rho),
\]

where \( u_{0,xt} = u_{0t} - \frac{\sigma_{0x}}{\sigma_{xx}} u_{xt} \).

Under normality, stationarity (|\( \rho \)| < 1), intercept – Stambaugh (1999):

\[
\mathbb{E} (\hat{\beta} - \beta) = - \left( \frac{\sigma_{0x}}{\sigma_{xx}} \right) \left( \frac{1 + 3\rho}{n} \right) + O \left( \frac{1}{n^2} \right),
\]
Consensus that potential explanatory regressors are highly persistent in this model

Local to unity specification $\rho_n = 1 + \frac{c}{n}$ has received much attention: Campbell and Yogo (2006) and Jansson and Moreira (2006)


$$n (\hat{\rho} - \rho) = \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{0t} \quad \Rightarrow \quad \int J_{c}^{x} (r) dB_{0} (r) + \lambda$$

The “finite sample bias” is still present in the limit and not obviously correctable since $c$ is not consistently estimable (but see below)

Not mixed normal and not pivotal
Consensus that potential explanatory regressors are highly persistent in this model

Local to unity specification $\rho_n = 1 + \frac{c}{n}$ has received much attention: Campbell and Yogo (2006) and Jansson and Moreira (2006)


\[ n(\hat{\rho} - \rho) = \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{0t} = \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 \times \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 \times \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 \times \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 \]

The “finite sample bias” is still present in the limit and not obviously correctable since $c$ is not consistently estimable (but see below)

Not mixed normal and not pivotal
Nonnormality comes from endogeneity & LUR limit theory

If \( u_t \sim \text{mds} \left( 0, \Sigma \right) \)

\[
\zeta_{nt} = \begin{bmatrix}
\frac{1}{n} x_{t-1} u_{0t} \\
\frac{1}{\sqrt{n}} u_{xt}
\end{bmatrix}
\]

\[
\sum E_{F_{nt-1}} \zeta_{nt} \zeta'_{nt} = \begin{bmatrix}
\left( \frac{1}{n} \sum x^2_{t-1} \right) \sigma_{00} & \left( \frac{1}{n \sqrt{n}} \sum x_{t-1} \right) \sigma_{0x} \\
\left( \frac{1}{n \sqrt{n}} \sum x_{t-1} \right) \sigma_{x0} & \sigma_{xx}
\end{bmatrix}
\]

not diagonal unless \( \sigma_{0x} \neq 0 \)

Peter C. B. Phillips ( ) Prediction SoFiE Conference June 2013 7 / 90
Pitfalls: (b) Nonstandard Inference
Inference with Persistent Regressors

From the decomposition

\[ n \left( \hat{\beta} - \beta \right) = \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{0,xt} \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 + \frac{\sigma_{x0}}{\sigma_{xx}} \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{xt} \frac{1}{n^2} \sum_{t=1}^{n} (x_{t-1})^2 \]

The second term is a standard LUR distribution

\[ \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{xt} \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 \rightleftharpoons \frac{\int J_c^x(r) dB_x(r)}{\int J_c^x(r)^2 dr} \]

Giving limit theory for \( n \left( \hat{\beta} - \beta \right) \)

\[ MN \left( 0, \sigma_{00,.x} \left( \int J_c^x(r)^2 dr \right)^{-1} \right) + \left( \frac{\sigma_{x0}}{\sigma_{xx}} \right) \frac{\int J_c^x(r) dB_x(r)}{\int J_c^x(r)^2 dr} \]
From the decomposition

\[ n \left( \hat{\beta} - \beta \right) = \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{0,xt} \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 + \frac{\sigma_{x0}}{\sigma_{xx}} \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{xt} \frac{1}{n^2} \sum_{t=1}^{n} (x_{t-1})^2 \]

The second term is a standard LUR distribution

\[ \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{xt} \frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}^2 \Rightarrow \int J_c^x(r) dB_x(r) \int J_c^x(r)^2 dr \]

Giving limit theory for \( n \left( \hat{\beta} - \beta \right) \)

\[
MN \left( 0, \sigma_{00,xx} \left( \int J_c^x(r)^2 dr \right)^{-1} \right) + \left( \frac{\sigma_{x0}}{\sigma_{xx}} \right) \frac{\int J_c^x(r) dB_x(r)}{\int J_c^x(r)^2 dr}
\]
Pitfalls: (b) Nonstandard Inference
Inference with Persistent Regressors

- From the decomposition

\[
 n (\hat{\beta} - \beta) = \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{0.xt} \frac{1}{n^2} \sum_{t=1}^{n} x^2_{t-1} + \frac{\sigma_{x0}}{\sigma_{xx}} \frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{xt} \frac{1}{n^2} \sum_{t=1}^{n} (x_{t-1})^2
\]

- The second term is a standard LUR distribution

\[
\frac{1}{n} \sum_{t=1}^{n} x_{t-1} u_{xt} \frac{1}{n^2} \sum_{t=1}^{n} x^2_{t-1} \Rightarrow \int \frac{J^x_c(r) dB_x(r)}{\int J^x_c(r)^2 dr},
\]

- Giving limit theory for \( n (\hat{\beta} - \beta) \)

\[
MN \left( 0, \sigma_{00.x} \left( \int J^x_c(r)^2 dr \right)^{-1} \right) + \left( \frac{\sigma_{x0}}{\sigma_{xx}} \right) \frac{\int J^x_c(r) dB_x(r)}{\int J^x_c(r)^2 dr}
\]
Nonstandard limit theory from endogeneity and persistent regressor

Endogeneity/bias correction methods such as Phillips and Hansen (1990, FM-OLS) are designed for UR case \((c = 0)\) and give

\[
n \left( \hat{\beta}_{FM} - \beta \right) \Rightarrow MN \left( -c \frac{\sigma_{x0}}{\sigma_{xx}}, \sigma_{00,x} \left( \int J_c^x (r)^2 \, dr \right)^{-1} \right)
\]

Without precise information on \(c\), the bias is not correctable
Pitfalls: (b) Nonstandard Inference

Inference with Persistent Regressors

- t test

\[
\frac{\hat{\beta} - \beta}{\hat{\sigma}_\beta} \implies \left( \frac{\sigma_{00,x}}{\sigma_{00}} \right)^{1/2} Z + \frac{\sigma_{x0}}{(\sigma_{xx}\sigma_{00})^{1/2}} \left( \frac{\int J_c^x(r) dB_x(r)}{(\sigma_{xx} \int J_c^x(r)^2 dr)^{1/2}} \right)
\]

\[
= (1 - \delta^2)^{1/2} Z + \delta \eta_{LUR}(c)
\]

where

\[
\delta = \frac{\sigma_{x0}}{(\sigma_{xx}\sigma_{00})^{1/2}}
\]

- Unless \( \delta = 0 \), standard chi-square inference is not valid
Pitfalls: (b) Nonstandard Inference

Bonferroni Bounds Method

- Test has mixture limit theory

\[ t_\beta = \frac{\hat{\beta} - \beta}{\hat{\sigma}_\beta} \implies \delta \eta_{\text{LUR}}(c) + (1 - \delta^2)^{1/2} Z. \]  

- Cavanagh, Elliott and Stock (1995):
  - pretest for \( \delta = 0 \) and t-test is valid
  - if \( \delta \neq 0 \), use a Bonferroni method and most conservative \( c \)
  - construct a 100 \((1 - \alpha_1)\)% CI for \( c \) using Stock (1991) confidence belts
  - For each \( c \) in CI, construct a 100 \((1 - \alpha_2)\)% CI for \( \beta \), denoted as \( \text{CI}_{\beta|c}(\alpha_2) \), using (1).
  - A CI free of \( c \) is obtained as

\[ \text{CI}_\beta(\alpha) = \bigcup_{c \in \text{CI}_c(\alpha_1)} \text{CI}_{\beta|c}(\alpha_2). \]
Pitfalls: constructing a CI for $c$

Stock’s (1991) suggestion

- use LUR limit theory for AR $t_\hat{\rho}$ (Phillips, 1987)

$$t_\hat{\rho} = \frac{\hat{\rho} - 1}{\hat{\sigma}_\rho} \rightarrow \frac{\int J_c dW}{\left(\int J(r)^2 dr\right)^{1/2}} + c \left(\int J_c(r)^2 dr\right)^{1/2} =: \tau_c, \quad (2)$$

- invert statistic $t_\hat{\rho}$ using confidence bands constructed from the limit distribution theory of $\tau_c$

- mechanism: use confidence belts (Kendall, 1946) from intensive simulations of (2)
Confidence belts (levels 2.5%, 50% and 97.5%) with asymptotic functional approximations – Phillips (2013)
Pitfalls: constructing a CI for $c$

Bonferroni Method - specifics

- using $\hat{\rho}$ find $Cl_c(\alpha_1) = [c_l(\alpha_1), c_u(\alpha_1)]$ by inversion - Stock (1991)
- using the cv $d_{t_{\hat{\beta}},c}$ of $t_{\hat{\beta}} \sim \delta \eta_{LUR}(c) + (1 - \delta^2)^{1/2} Z$, find

\[
Cl_{\beta}(\alpha_1, \alpha_2) = \begin{bmatrix}
d^\beta_l(\alpha_1, \alpha_2), d^\beta_u(\alpha_1, \alpha_2)
\end{bmatrix}
= \begin{bmatrix}
\min_{c_l \leq c \leq c_u} d_{t_{\hat{\beta}},c,\frac{1}{2}\alpha_2}, \max_{c_l \leq c \leq c_u} d_{t_{\hat{\beta}},c,1-\frac{1}{2}\alpha_2}
\end{bmatrix}
\]

- Then

\[
Pr \left( t_{\hat{\beta}} \notin \left[ d^\beta_l(\alpha_1, \alpha_2), d^\beta_u(\alpha_1, \alpha_2) \right] \right) \\
\rightarrow Pr \left( \delta \eta_{LUR}(c) + (1 - \delta^2)^{1/2} Z \notin \left[ d^\beta_l(\alpha_1, \alpha_2), d^\beta_u(\alpha_1, \alpha_2) \right] \right) \\
\leq \alpha_1 + \alpha_2. \text{ (By Bonferroni)}
\]
using $\hat{\rho}$ find $Cl_c(\alpha_1) = [c_l(\alpha_1), c_u(\alpha_1)]$ by inversion - Stock (1991)

using the cv $dt_{t_{\hat{\beta}},c}$ of $t_{\hat{\beta}} \sim \delta \eta_{LUR}(c) + (1 - \delta^2)^{1/2} Z$, find

$$Cl_\beta(\alpha_1, \alpha_2) = \left[ d_l^\beta(\alpha_1, \alpha_2), d_u^\beta(\alpha_1, \alpha_2) \right]$$

$$= \left[ \min_{c_l \leq c \leq c_u} d_{t_{\hat{\beta}},c,\frac{1}{2}\alpha_2}, \max_{c_l \leq c \leq c_u} d_{t_{\hat{\beta}},c,1-\frac{1}{2}\alpha_2} \right]$$

Then

$$\Pr\left( t_{\hat{\beta}} \notin \left[ d_l^\beta(\alpha_1, \alpha_2), d_u^\beta(\alpha_1, \alpha_2) \right] \right)$$

$$\rightarrow \Pr\left( \delta \eta_{LUR}(c) + (1 - \delta^2)^{1/2} Z \notin \left[ d_l^\beta(\alpha_1, \alpha_2), d_u^\beta(\alpha_1, \alpha_2) \right] \right)$$

$$\leq \alpha_1 + \alpha_2. \text{ (By Bonferroni)}$$
Pitfalls: constructing a CI for $c$

Bonferroni Method - numerically intensive like grid bootstrap computations

Golly, I just ran 5 billion regressions!!

Construct the $d_{t\hat{\beta},c}$ of $t_{\hat{\beta}} \sim \delta \eta_{LUR}(c) + (1 - \delta^2)^{1/2} Z$ across a grid

$\{20,000 \text{ values of } c\} \times \{5 \text{ values of } \delta\} \times \{50,000 \text{ replications}\}$
Pitfalls: constructing a CI for $c$

Bonferroni Method - numerically intensive like grid bootstrap computations

Golly, I just ran 5 billion regressions!!

Construct the cv $d_{t\hat{\beta}, c}$ of $t_{\hat{\beta}} \sim \delta \eta_{LUR} (c) + (1 - \delta^2)^{1/2} Z$ across a grid

\[ \{20,000 \text{ values of } c\} \times \{5 \text{ values of } \delta\} \times \{50,000 \text{ replications}\} \]
Pitfalls: constructing a CI for $c$

Bonferroni Method - Campbell & Yogo (2006, CY)

- Campbell and Yogo (2006): the same idea with DF-GLS
- Often employed in applied work and in other methodological work

Undesirable Properties:
- Conservative by construction: numerical size often much less than nominal size
- Results in biased test, since there are local alternatives for rejecting less often than nominal size.
- No extensions to multivariate systems/multiple regressors
- .... plus ...

A Major New Problem!!
Pitfalls: constructing a CI for $c$

Bonferroni Method - Campbell & Yogo (2006, CY)

- Campbell and Yogo (2006): the same idea with DF-GLS
- Often employed in applied work and in other methodological work
- Undesirable Properties:
  - Conservative by construction: numerical size often much less than nominal size
  - Results in biased test, since there are local alternatives for rejecting less often than nominal size.
  - No extensions to multivariate systems/multiple regressors
  - .... plus ...

A Major New Problem!!
Pitfalls: constructing a CI for \( c \)

Bonferroni Method - Campbell & Yogo (2006, CY)

- Campbell and Yogo (2006): the same idea with DF-GLS
- Often employed in applied work and in other methodological work
- Undesirable Properties:
  - Conservative by construction: numerical size often much less than nominal size
  - Results in biased test, since there are local alternatives for rejecting less often than nominal size.
  - No extensions to multivariate systems/multiple regressors
  - .... plus ...

- A Major New Problem!!
Unnoticed Problem:

- Stock (1991) confidence intervals have ZERO coverage probability when $|\rho| < 1$ (and more generally when $n^{\frac{1}{3}} (1 - \rho) \to \infty$)
- UR test statistic $\tau_c$ fails tightness
- $C_l c(\alpha_1) = [c_l(\alpha_1), c_u(\alpha_1)]$ diverges
- probability mass escapes as $c \to -\infty$ and stationary region $|\rho| < 1$ is approached

Consequences - serious failure of robustness to $\rho$

- CY confidence intervals have zero coverage probability in stationary case
- CY Q test indicates predictability with probability unity under the null when $|\rho| < 1$
Unnoticed Problem:

- Stock (1991) confidence intervals have ZERO coverage probability when $|\rho| < 1$ (and more generally when $n^{1/3} (1 - \rho) \to \infty$)
- UR test statistic $\tau_c$ fails tightness
- $CI_c(\alpha_1) = [c_l(\alpha_1), c_u(\alpha_1)]$ diverges
- probability mass escapes as $c \to -\infty$ and stationary region $|\rho| < 1$ is approached

Consequences - serious failure of robustness to $\rho$

- CY confidence intervals have zero coverage probability in stationary case
- CY $Q$ test indicates predictability with probability unity under the null when $|\rho| < 1$
Asymptotic analysis - local to unity case
Form of the confidence belts

**Theorem**

As \( c \to -\infty \) the asymptotic form of the distribution of \( \tau_c \) is

\[
\tau_c \sim N \left( -\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2} |c|^{1/2}}, \frac{1}{4} \right) + O_p \left( \frac{1}{|c|^{1/2}} \right),
\]

inducing the (2.5%, 97.5%) confidence belts

\[
\left\{ -\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2} |c|^{1/2}} \pm \frac{1.96}{2} \right\}
\]
Asymptotic analysis - local to unity case

Expansion of the limit distribution

As \( c \to -\infty \), we expand \( \tau_c \) as

\[
\tau_c = \frac{\int J_c dW}{(\int J(r)^2 dr)^{1/2}} + c \left( \frac{\int J_c(r)^2 dr}{(\int J(r)^2 dr)^{1/2}} \right)^{1/2}
\]

\[
= \frac{(-2c)^{1/2} \int J_c dW}{\{(-2c) \int J(r)^2 dr\}^{1/2}} - \frac{|c|^{1/2}}{2^{1/2}} \left\{ (-2c)^{1/2} \int J_c(r)^2 dr \right\}^{1/2}
\]

\[
\sim \frac{\zeta}{\left\{ 1 + \frac{2\zeta}{(-2c)^{1/2}} \right\}^{1/2}} - \frac{|c|^{1/2}}{2^{1/2}} \left\{ 1 + \frac{2\zeta}{(-2c)^{1/2}} \right\}^{1/2}
\]

\[
= -\frac{|c|^{1/2}}{2^{1/2}} + \frac{1}{2} \zeta + O_p \left( \frac{1}{|c|^{1/2}} \right), \quad \zeta \equiv N(0,1)
\]
Asymptotic analysis - local to unity case

Form of the confidence belts

Asymptotic confidence belts: \[
\left\{ -\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2} |c|^{1/2}} \pm \frac{1.96}{2} \right\}
\]

Confidence belts with asymptotic functional approximations
Asymptotic analysis - local to unity case

Form of the confidence belts

Family

\[
\begin{align*}
\tau_c &= N \left( -\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2} |c|^{1/2}}, \frac{1}{4} \right)
\end{align*}
\]

is not tight as \( c \to -\infty \).

Probability mass escapes as \( c \to -\infty \), i.e as stationary region \( |\rho| < 1 \) is approached.
Family

\[ \tau_c = N \left( -\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2} |c|^{1/2}}, \frac{1}{4} \right) \]

is not tight as \( c \to -\infty \).

Probability mass escapes as \( c \to -\infty \), i.e. as stationary region \( |\rho| < 1 \) is approached.
Asymptotic analysis - stationary case

Stock CIs have zero coverage probability in stationary/near stationary case

**Theorem**

100(1 – \(\alpha\)) % CI for |\(\rho\)| < 1 is

\[
[\rho_L, \rho_U] = \left[ 1 - 2A_\rho + 2 \frac{2\zeta \pm z_{\alpha/2}}{\sqrt{n}} A_\rho^{1/2} \right], \quad A_\rho = \frac{1 - \rho}{1 + \rho}
\]

(4)

where \(\zeta = N(0,1)\). Coverage probability is

\[
P\{\rho \in [\rho_L, \rho_U]\} = P_\zeta \left\{ \zeta \in \left[ \frac{\sqrt{n}}{4} (1 - \rho) A_\rho^{1/2} \pm \frac{z_{\alpha/2}}{2} \right] \right\}
\]
Asymptotic analysis - stationary case

Stock CIs have zero coverage probability in stationary/near stationary case

$$[\rho_L, \rho_U] \xrightarrow{p} \frac{3\rho - 1}{1 + \rho} = \bar{\rho}$$
Asymptotic analysis - stationary case
Stock CIs have zero coverage probability in stationary/near stationary case

Coverage probability

\[ P \{ \rho \in [\rho_L, \rho_U] \} = P_{\zeta} \left\{ \zeta \in \left[ \frac{\sqrt{n}}{4} (1 - \rho) A_{\rho}^{1/2} \pm \frac{z_{\alpha/2}}{2} \right] \rightarrow \infty \right\} \rightarrow 0 \]

i.e. if \( \sqrt{n} (1 - \rho) A_{\rho}^{1/2} \rightarrow \infty \).
Asymptotic analysis - stationary case

Stock CIs have zero coverage probability in stationary/near stationary case

Zero coverage probability whenever \( \sqrt{n} (1 - \rho)^{3/2} \to \infty \)

That is: for all fixed \(|\rho| < 1\) and for all mildly integrated \(\rho\) with

\[ \rho = 1 + \frac{c}{n^{\delta}}, \quad \delta < \frac{1}{3}, \quad c < 0 \]
Implications for Campbell Yogo Predictive Tests

Complete failure of robustness to AR coefficient

Coverage probabilities of Campbell-Yogo and stationary confidence intervals for the predictive regression coefficient $\beta$.
Implications for Campbell Yogo Predictive Tests

Prediction Failure

- CY Q test employs t ratio $t_{\hat{\beta}(\rho)}$ based on regression coefficient $\hat{\beta}(\rho)$ conditional on $\rho$:

$$\hat{\beta}(\rho) = \frac{\sum_{t=1}^{n} x_{t-1} \left[ y_{t} - \frac{\hat{\sigma}_{x_0}}{\hat{\sigma}_{xx}} (x_{t} - \rho x_{t-1}) \right]}{\sum_{t=1}^{n} x_{t-1}^2}$$

- Confidence interval from Stock inversion giving $[\rho_L, \rho_U]$ and

$$[\hat{\beta}_L (\rho_U), \hat{\beta}_U (\rho_L)] = \left[ \hat{\beta} (\rho_U) - z_{\alpha/2} / \sigma_{\hat{\beta}(\rho)}, \hat{\beta} (\rho_L) + z_{\alpha/2} / \sigma_{\hat{\beta}(\rho)} \right]$$
Implications for Campbell Yogo Predictive Tests

Prediction Failure

$\text{CY confidence interval is inconsistent for all } |\rho| < 1$

$$[\beta_L(\rho_U), \beta_U(\rho_L)] \rightarrow \beta + \frac{\sigma_{x0}}{\sigma_{xx}} \frac{(\rho - 1)^2}{\rho + 1} \neq \beta$$

for all $\sigma_{x0} \neq 0$
Implications for Campbell Yogo Predictive Tests

Prediction Failure

Under $H_0: \beta = 0$, Size $\to$ Unity for $|\rho| < 1$

$$\hat{\beta}(\rho) \to p \frac{\sigma_{x0}}{\sigma_{xx}} \frac{(\rho - 1)^2}{\rho + 1} \neq 0$$

- So rejection probability $\to 1$
- Always find evidence of predictability with this test!!
Possibilities: (a) & (b) Correct centering

Correct centering of AR test statistic - then use Bonferroni construction

- Idea: do the inversion based on \( t_\rho = \frac{\hat{\rho} - \rho}{\sigma_{\hat{\rho}}} \)

\[
t_\rho \implies \begin{cases} 
\frac{\int J_c dW}{(\int J(r)^2 dr)^{1/2}} \\
N(0, 1)
\end{cases} 
\Rightarrow \lim_{c \to -\infty} N(0, 1) \quad (Phillips, 1987)
\]

for \( n(1 - \rho) \to \infty \) (Giraitis & Phillips, 2006)

- CI construction for \( \rho \) is then uniform in \( \rho \in (0, 1] \) (Mikusheva, 2007).

- But: numerically intensive, applies only for AR(1), no extension to multivariate regressors

Idea: do the inversion based on $t_\hat{\rho} = \frac{\hat{\rho} - \rho}{\sigma_{\hat{\rho}}}$

\[
\begin{align*}
t_\hat{\rho} & \quad \rightarrow \quad \frac{\int J_c dW}{\left(\int J(r)^2 dr\right)^{1/2}} \\
\rightarrow & \quad N(0, 1) \quad (\text{Phillips, 1987}) \\
\text{for } n(1 - \rho) & \quad \rightarrow \quad \infty \quad (\text{Giraitis & Phillips, 2006})
\end{align*}
\]

CI construction for $\rho$ is then uniform in $\rho \in (0, 1]$ (Mikusheva, 2007).

But: numerically intensive, applies only for AR(1), no extension to multivariate regressors.
Possibilities: (a) & (b) IVX Inference

IVX approach – Magdalinos and Phillips (MP, 2009)

- Idea: self generation - create less persistent IV using own regressor $x_t$

$$
\tilde{z}_t = \sum_{j=1}^{t} \rho_{nz}^{t-j} \triangle x_j
$$

$$
\rho_{nz} = 1 + \frac{c_z}{n^\varphi}, \quad \varphi \in (0, 1), \quad c_z < 0,
$$

- IVX $\Rightarrow$ filter the regressor $x_t$ to produce a valid instrument for $x_t$
  - reduce persistence when $x_t$ is LUR
  - applicable when $x_t$ is UR, LUR, MI, ME
Possibilities: (a) & (b) IVX Inference

IVX approach – Magdalinos and Phillips (MP, 2009)

- Idea: self generation - create less persistent IV using own regressor $x_t$

$$\tilde{z}_t = \sum_{j=1}^{t} \rho_{nz}^{t-j} \Delta x_j$$

$$\rho_{nz} = 1 + \frac{c_z}{n\varphi}, \varphi \in (0, 1), c_z < 0,$$

- IVX ⇒ filter the regressor $x_t$ to produce a valid instrument for $x_t$
  - reduce persistence when $x_t$ is LUR
  - applicable when $x_t$ is UR, LUR, MI, ME
Possibilities: (a) & (b) IVX Inference

IVX approach - mechanics

- IVX is based on the regressor $x_t$

\[ \tilde{z}_t = \sum_{j=1}^{t} \rho_{n_z}^{t-j} \triangle x_j, \quad \rho_{n_z} = 1 + \frac{c_z}{n^\varphi}, \quad \varphi \in (0, 1), \quad c_z < 0 \]

- Since $\triangle x_j = \frac{c}{n} x_{j-1} + u_{xt}$, we have the decomposition

\[ \tilde{z}_t = \sum_{j=1}^{t} \rho_{n_z}^{t-j} \left( \frac{c}{n} x_{j-1} + u_{xt} \right) = \sum_{j=1}^{t} \rho_{n_z}^{t-j} u_{xt} + \frac{c}{n} \sum_{j=1}^{t} \rho_{n_z}^{t-j} x_{j-1} \]

\[ = z_t + \frac{c}{n} \psi_{nt} \]

- $z_t = \rho_{n_z} z_{t-1} + u_{xt}$ plays the role of a mildly integrated instrument.
Possibilities: (a) & (b) IVX Inference
IVX approach - mechanics

- IVX is based on the regressor $x_t$

$$\tilde{z}_t = \sum_{j=1}^{t} \rho_{nz}^{t-j} \Delta x_j, \quad \rho_{nz} = 1 + \frac{c_z}{n^{\phi}}, \quad \phi \in (0, 1), \quad c_z < 0$$

- Since $\Delta x_j = \frac{c}{n} x_{j-1} + u_{xt}$, we have the decomposition

$$\tilde{z}_t = \sum_{j=1}^{t} \rho_{nz}^{t-j} \left( \frac{c}{n} x_{j-1} + u_{xt} \right) = \sum_{j=1}^{t} \rho_{nz}^{t-j} u_{xt} + \frac{c}{n} \sum_{j=1}^{t} \rho_{nz}^{t-j} x_{j-1}$$

$$= z_t + \frac{c}{n} \psi_{nt}$$

- $z_t = \rho_{nz} z_{t-1} + u_{xt}$ plays the role of a mildly integrated instrument.
Possibilities: (a) & (b) IVX Inference

IVX approach - mechanics

- IVX is based on the regressor $x_t$

$$\tilde{z}_t = \sum_{j=1}^t \rho_{nz}^{t-j} \triangle x_j, \quad \rho_{nz} = 1 + \frac{c_z}{n^\varphi}, \quad \varphi \in (0, 1), \quad c_z < 0$$

- Since $\triangle x_j = \frac{c}{n} x_{j-1} + u_{xt}$, we have the decomposition

$$\tilde{z}_t = \sum_{j=1}^t \rho_{nz}^{t-j} \left( \frac{c}{n} x_{j-1} + u_{xt} \right) = \sum_{j=1}^t \rho_{nz}^{t-j} u_{xt} + \frac{c}{n} \sum_{j=1}^t \rho_{nz}^{t-j} x_{j-1}$$

$$= z_t + \frac{c}{n} \psi_{nt}$$

- $z_t = \rho_{nz} z_{t-1} + u_{xt}$ plays the role of a mildly integrated instrument.
Possibilities: (a) & (b) IVX Inference

IVX estimation & limit theory

- IVX estimator:
  \[
  \hat{\beta}_{IVX} = \frac{\sum_{t=1}^{n} \tilde{z}_{t-1}y_t}{\sum_{t=1}^{n} \tilde{z}_{t-1}x_{t-1}} = \beta + \frac{\sum_{t=1}^{n} \tilde{z}_{t-1}u_{0t}}{\sum_{t=1}^{n} \tilde{z}_{t-1}x_{t-1}}
  \]

- Nice limit theory under the rate restriction \( \varphi \in (0, 1) \)
  \[
  n^{\frac{1+\varphi}{2}} (\hat{\beta}_{IVX} - \beta) \longrightarrow \Psi
  \]
  where \( \Psi \) is a correctly centered mixed normal random variable

- The t-ratio:
  \[
  t_{\hat{\beta}_{IVX}} = \frac{\hat{\beta}_{IVX} - \beta}{\hat{\sigma}_{IVX}} \longrightarrow Z
  \]
  which is a standard normal
asymptotic independence between MG parts of numerator/denominator.

Compared to earlier discussion

\[
\begin{align*}
\zeta_{nt} & = \begin{bmatrix} \frac{1}{n^{1+\phi}} z_{t-1}u_0t \\ \frac{1}{\sqrt{n}} u_{xt} \end{bmatrix}, \\
\sum \mathbb{E}f_{nt-1} \zeta_{nt} \zeta'_{nt} & = \begin{bmatrix} \frac{1}{n^{1+\phi}} \sum z_{t-1}^2 \sigma_{00} & \left( \frac{1}{n^{1+\phi/2}} \sum z_{t-1} \right) \sigma_{0x} \\ \left( \frac{1}{n^{1+\phi/2}} \sum z_{t-1} \right) \sigma_{0x} & \sigma_{xx} \end{bmatrix} \\
\implies & \text{ diagonal}
\end{align*}
\]

achieved by reducing order of magnitude of the instrument \( z_t \)
Ongoing work:

- allowing multivariate $x_t$ with mixed orders of persistence
- allowing intercepts, deterministic trends, and structural breaks

Desirable properties:

- pivotal mixed normal limit theory under general persistent regressors
- accommodates multivariate systems + many regressors
- applicable when regressors have differing levels of persistence
- good size & power properties of test
Possibilities: (a) & (b) IVX Inference

IVX approach - ongoing research

**Ongoing work:**
- allowing multivariate $x_t$ with mixed orders of persistence
- allowing intercepts, deterministic trends, and structural breaks

**Desirable properties:**
- pivotal mixed normal limit theory under general persistent regressors
- accommodates multivariate systems + many regressors
- applicable when regressors have differing levels of persistence
- good size & power properties of test
QR predictive model - conditional quantile formulation:

\[ Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \beta(\tau)x_{t-1} + \sigma_{00}^{1/2}\Phi^{-1}_{u_0}(\tau) \]  

\[ x_t = x_{t-1} + u_{xt} \]

\[ \Phi = cdfN(0, 1), \quad Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \text{quantile function}, \quad u_t = (u_{0t}, u_{xt})' \sim \text{mds} \]

- Quantile crossings occur for \( \rho > \tau \) when natural order is reversed:

\[ \{\beta(\rho) - \beta(\tau)\} x_{t-1} + \sigma_{00}^{1/2} \{\Phi^{-1}_{u_0}(\rho) - \Phi^{-1}_{u_0}(\tau)\} < 0 \]

- that is whenever

\[ x_{t-1} < \frac{\sigma_{00}^{1/2} \{\Phi^{-1}_{u_0}(\tau) - \Phi^{-1}_{u_0}(\rho)\}}{\beta(\rho) - \beta(\tau)} \quad \text{if} \quad \beta(\rho) - \beta(\tau) > 0 \]
Pitfalls: (c) Invalidity of Quantile Predictive Regression

QR Model Invalidity under the Alternative

QR predictive model - conditional quantile formulation:

\[ Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \beta(\tau) x_{t-1} + \sigma_{00}^{1/2} \Phi_u^{-1}(\tau) \]

\[ x_t = x_{t-1} + u_{xt} \]

\[ \Phi = cdfN(0,1),\ Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \text{quantile function},\ u_t = (u_{0t}, u_{xt})' \sim \text{mds} \]

Quantile crossings occur for \( \rho > \tau \) when natural order is reversed:

\[ \{\beta(\rho) - \beta(\tau)\} x_{t-1} + \sigma_{00}^{1/2} \{\Phi_u^{-1}(\rho) - \Phi_u^{-1}(\tau)\} < 0 \]

that is whenever

\[ x_{t-1} < \frac{\sigma_{00}^{1/2} \{\Phi_u^{-1}(\tau) - \Phi_u^{-1}(\rho)\}}{\beta(\rho) - \beta(\tau)}, \quad \text{if } \beta(\rho) - \beta(\tau) > 0 \]
Theorem

Quantile crossing frequency given by

\[
n^{-1} \sum_{t=1}^{n} 1 \left\{ \frac{x_{t-1}}{\sqrt{n}} < \frac{1}{\sqrt{n}} \sigma_{00}^{1/2} \left\{ \Phi_{u_0}^{-1}(\tau) - \Phi_{u_0}^{-1}(\rho) \right\} \right\} \Rightarrow \int_{0}^{1} 1 \{ B_x(r) < 0 \} dx
\]

- distribution is the arc sine law with density \( \frac{1}{\pi x^{1/2}(1-x)^{1/2}} \) over \( x \in [0, 1] \)
- Most likely many or few crossings for a given trajectory \( \{x_t\}_{t=1}^{n} \)
- Quantile prediction requires constant slope coefficient for validity!

- But ...
Pitfalls: (c) Invalidity of Quantile Predictive Regression

QR Model Invalidity under the Alternative

Theorem

Quantile crossing frequency given by

\[
\frac{1}{n} \sum_{t=1}^{n} 1 \left\{ \frac{x_{t-1}}{\sqrt{n}} < \frac{1}{\sqrt{n}} \frac{\sigma_{00}^{1/2} \{ \Phi_{u0}^{-1}(\tau) - \Phi_{u0}^{-1}(\rho) \}}{\beta(\rho) - \beta(\tau)} \right\} \Rightarrow \int_{0}^{1} 1 \{ B_x(r) < 0 \} \, dr
\]

- distribution is the arc sine law with density \( \frac{1}{\pi x^{1/2} (1-x)^{1/2}} \) over \( x \in [0, 1] \)
- Most likely many or few crossings for a given trajectory \( \{ x_t \}_{t=1}^{n} \).
- Quantile prediction requires constant slope coefficient for validity!

But ...
Pitfalls: (c) Invalidity of Quantile Predictive Regression
QR Model Invalidity under the Alternative

Theorem

Quantile crossing frequency given by

\[ n^{-1} \sum_{t=1}^{n} 1 \left\{ \frac{x_{t-1}}{\sqrt{n}} < \frac{1}{\sqrt{n}} \frac{\sigma_{00}^{1/2}}{\beta(\rho) - \beta(\tau)} \right\} \Rightarrow \int_0^1 1 \left\{ B_x(r) < 0 \right\} dr \]

- distribution is the arc sine law with density \( \frac{1}{\pi x^{1/2}(1-x)^{1/2}} \) over \( x \in [0, 1] \)
- Most likely many or few crossings for a given trajectory \( \{x_t\}_{t=1}^n \).
- Quantile prediction requires constant slope coefficient for validity!

But ...
Local to constant slope $\beta(\tau) = \beta + \frac{b(\tau)}{k_n}$ with $k_n \to \infty$

With localizing coefficient function $b(\tau) \in [b_L, b_U]$

Local QR model

$$Q_{yt}(\tau | \mathcal{F}_{t-1}) = \left( \beta + \frac{b(\tau)}{k_n} \right) x_{t-1} + \sigma_{00}^{1/2} \Phi_{u_0}^{-1}(\tau)$$
Possibilities: (c) Quantile Predictive Regression
QR Model Validity by Marginalizing the Alternative Condition for no quantile crossing

\[
\sqrt{n} \left\{ \frac{b(\rho) - b(\tau)}{k_n} \right\} \frac{x_{t-1}}{\sqrt{n}} + \sigma_{00}^{1/2} \left\{ \Phi_{u_0}^{-1}(\rho) - \Phi_{u_0}^{-1}(\tau) \right\} > 0
\]

holds if \( \frac{\sqrt{n}}{k_n} \rightarrow 0 \) as \( n \rightarrow \infty \).

Probability of quantile crossing tends to zero as \( n \rightarrow \infty \).

\( k_n \rightarrow \infty \) fast enough so quantiles do not cross \( \left( \frac{\sqrt{n}}{k_n} \rightarrow 0 \right) \)

\( k_n \rightarrow \infty \) not too fast \( \left( \frac{k_n}{n} \rightarrow 0 \right) \) so marginal deviations can be consistently estimated
Condition for no quantile crossing

\[
\sqrt{n} \left\{ \frac{b(\rho) - b(\tau)}{k_n} \right\} \frac{x_{t-1}}{\sqrt{n}} + \sigma_{00}^{1/2} \left\{ \Phi^{-1}_u(\rho) - \Phi^{-1}_u(\tau) \right\} > 0
\]

holds if \( \frac{\sqrt{n}}{k_n} \to 0 \) as \( n \to \infty \).

Probability of quantile crossing tends to zero as \( n \to \infty \).

\( k_n \to \infty \) fast enough so quantiles do not cross \( \left( \frac{\sqrt{n}}{k_n} \to 0 \right) \)

\( k_n \to \infty \) not too fast \( \left( \frac{k_n}{n} \to 0 \right) \) so marginal deviations can be consistently estimated
FM-QR limit theory for $x_t \equiv UR$

$$n \left( \hat{\beta}^+ (\tau) - \beta (\tau) \right) \Rightarrow MN (0, V), \quad V = \frac{\sigma^2_{\psi, x}}{f (F^{-1} (\tau))^2} \left( \int_0^1 B_x^2 \right)^{-1}$$

Allowance for marginal alternatives $\beta (\tau) = \beta + \frac{b(\tau)}{k_n}$ with $k_n \to \infty$

$$\frac{n}{k_n} (\hat{b}^+ (\tau) - b (\tau)) \Rightarrow MN (0, V)$$

Extensions to IVX-QR for $x_t \equiv LUR, UR, MI, ME$ (Lee, 2013)
• Predictive model formulation:

\[ y_{t+1} = \beta x_t + u_{0t+1}; \quad x_{t+1} = \rho x_t + u_{xt+1} \]

• Equation unbalanced under \( \beta_A \neq 0 \) as \( \beta_A x_{t-1} = O_p(\sqrt{n}) \) conflicts with \( y_t = I(0) \)

• But if we reject the null and \( \beta_A \neq 0 \) and \( y_t \) is \( I(0) \) we have

\[
\hat{\beta} = \frac{1}{n} \frac{1}{n} \sum_{t=1}^{n} x_{t-1} y_t
= O_p \left( \frac{1}{n} \right) \rightarrow_p 0
\]

so \( \hat{\beta} \rightarrow_p 0 \) conflicts with the alternative hypothesis \( \beta = \beta_A \)
Pitfalls: (d) Misbalancing in General

Model Invalidity under the Alternative

- Predictive model formulation:

\[ y_{t+1} = \beta x_t + u_{0t+1}; \quad x_{t+1} = \rho x_t + u_{xt+1} \]

- Equation unbalanced under \( \beta_A \neq 0 \) as \( \beta_A x_{t-1} = O_p (\sqrt{n}) \) conflicts with \( y_t = I(0) \)

- But if we reject the null and \( \beta_A \neq 0 \) and \( y_t \) is \( I(0) \) we have

\[ \hat{\beta} = \frac{1}{n} \frac{1}{n} \sum_{t=1}^{n} x_{t-1} y_t = O_p \left( \frac{1}{n} \right) \rightarrow^p 0 \]

so \( \hat{\beta} \rightarrow^p 0 \) conflicts with the alternative hypothesis \( \beta = \beta_A \)
Possibility: (d) Achieving Balance under Alternative
Resolving Balance by localization of coefficient

- Local to zero predictive regression:

\[ y_{t+1} = \beta_n x_t + u_{0t+1} \]
\[ x_{t+1} = \rho_n x_t + u_{xt+1} \]

- Suppose \( \beta_n = \frac{b}{n^\gamma} \) with \( b \neq 0 \), marginal deviation from null

Then

\[ y_t \sim \begin{cases} 
  u_{0t} + b n^{\frac{1}{2} - \gamma} J_x \left( \frac{t}{n} \right) \neq I(0) & \gamma < 0.5 \\
  u_{0t} + b J_x \left( \frac{t}{n} \right) = O_p(1) & \gamma = 0.5 \\
  u_{0t} = O_p(1) & \gamma > 0.5 
\end{cases} \]

- Test is consistent for \( \gamma \in [0.5, 1) \) and inconsistent for \( \gamma \geq 1 \) since

\[ n\hat{\beta}_n = n \left( \hat{\beta}_n - \beta_n \right) + n\beta_n = O_p(1) + O_p \left( n^{1-\gamma} \right) \]
Local to zero predictive regression:

\[ y_{t+1} = \beta_n x_t + u_{0t+1} \]
\[ x_{t+1} = \rho_n x_t + u_{xt+1} \]

Suppose \( \beta_n = \frac{b}{n^{\gamma}} \) with \( b \neq 0 \), marginal deviation from null

Then

\[
\begin{align*}
    y_t & \sim \begin{cases} 
        u_{0t} + bn^{\frac{1}{2}-\gamma} J_x \left( \frac{t}{n} \right) \neq I(0) & \gamma < 0.5 \\
        u_{0t} + b J_x \left( \frac{t}{n} \right) = O_p(1) & \gamma = 0.5 \\
        u_{0t} = O_p(1) & \gamma > 0.5 
    \end{cases}
\end{align*}
\]

Test is consistent for \( \gamma \in [0.5, 1) \) and inconsistent for \( \gamma \geq 1 \) since

\[
n \hat{\beta}_n = n (\hat{\beta}_n - \beta_n) + n \beta_n = O_p(1) + O_p(n^{1-\gamma})
\]
Local to zero predictive regression:

\[ y_{t+1} = \beta_n x_t + u_{0t+1} \]
\[ x_{t+1} = \rho_n x_t + u_{xt+1} \]

Suppose \( \beta_n = \frac{b}{n^{\gamma}} \) with \( b \neq 0 \), marginal deviation from null

Then

\[
\begin{cases}
  u_{0t} + b n^{\frac{1}{2} - \gamma} J_x \left( \frac{t}{n} \right) \neq I(0) & \gamma < 0.5 \\
  u_{0t} + b J_x \left( \frac{t}{n} \right) = O_p(1) & \gamma = 0.5 \\
  u_{0t} = O_p(1) & \gamma > 0.5 
\end{cases}
\]

Test is consistent for \( \gamma \in [0.5, 1) \) and inconsistent for \( \gamma \geq 1 \) since

\[
n\hat{\beta}_n = n (\hat{\beta}_n - \beta_n) + n\beta_n = O_p(1) + O_p\left(n^{1-\gamma}\right)
\]
Possibilities: (d) Achieving Balance under Alternative
Resolving Balance by localization of coefficient

Marginal deviations from null are detectable!!
Nonlinearity & local linearity

- Linear specifications can be satisfactory but require coefficient localization or error localization for balance
  - localization is a form of nonlinearity where coefficients change with $n$
- Direct nonlinear specifications also offer reasonable solutions
- And may also benefit from coefficient localization
Motivation for nonlinear specifications
Resolving Balance by Nonlinearity

- Model

\[ y_{t+1} = g(x_t) + u_{0_{t+1}}, \]
\[ x_{t+1} = \left(1 + \frac{c}{n}\right)x_t + u_{x_{t+1}}, \]

- Park and Phillips (1998): \( y_t \) can be \( I(0) \) for some \( g \) such as \( g \in L_1 \)
  - attenuates signal
  - transforms \( I(1) \) process to a process with memory \( d \sim \frac{1}{4} \)
Motivation for nonlinear specifications
Resolving Balance by Nonlinearity

Integrable Transform \( g \in L_1 \) of a random walk

\[ \text{RW vs } g(\text{RW}) \text{ (red)} \]

\[ g(\text{RW}) + \text{white noise (green)} \]
Motivation for nonlinear specifications
Resolving Balance by Nonlinearity

- Model

\[ y_{t+1} = g(x_t) + u_{0t+1}, \]
\[ x_{t+1} = \left(1 + \frac{c}{n}\right)x_t + u_{xt+1}, \]

- Wang and Phillips (2009a,b; WP): kernel estimation gives

\[ (\sqrt{n}h)^{1/2} \hat{g} = (\sqrt{n}h)^{1/2} [\hat{g} - g] + (\sqrt{n}h)^{1/2} g = O_p \left( (\sqrt{n}h)^{1/2} \right) \]

divergent, thereby achieving (2)

- Limit theory is pivotal and mixed normal – so (1) and (2) are covered; and effects are small (3) **without** coefficient localization for \( g \in L_1 \).
Motivation for nonlinear specifications
Resolving Balance by Nonlinearity with Localization

- For more general $g$ (e.g. asymptotically homogeneous) local to zero predictive power may be achieved with

$$y_{t+1} = \beta_n g(x_t) + u_{0t+1},$$
$$x_{t+1} = x_t + u_{xt+1}$$

with some function $g(x_t) = g(x_t, \pi)$ known up to $\pi$

- Consistent estimation possible for $\beta_n$ local to zero, e.g.

$$|\beta_n| \lesssim \frac{b}{n^{1/4}}, \text{ or } \frac{b}{n^{1/2} \kappa(n^{1/2}, \pi_o)}$$

- Asymptotic theory in Shi and Phillips (2012) for weakly identified $\pi$

- Solutions to (a), (b) and (c) can be achieved in this framework.
For more general $g$ (e.g. asymptotically homogeneous) local to zero predictive power may be achieved with

$$y_{t+1} = \beta_n g(x_t) + u_{0t+1},$$
$$x_{t+1} = x_t + u_{xt+1}$$

with some function $g(x_t) = g(x_t, \pi)$ known up to $\pi$

Consistent estimation possible for $\beta_n$ local to zero, e.g.

$$|\beta_n| \lesssim \frac{b}{n^{1/4}}, \text{ or } \frac{b}{n^{1/2} \kappa (n^{1/2}, \pi_0)}$$

Asymptotic theory in Shi and Phillips (2012) for weakly identified $\pi$

Solutions to (a), (b) and (c) can be achieved in this framework.
Pitfalls + Possibilities in Predictive Regression

sum up

Pitfalls

- bias in estimation, testing, CIs & nonstandard inference
- uncertain degrees of persistence - can be fatal as in Stock & CY
- need to cope with multiple regressors and marginal predictability
- omitted predictors misspecification
- issues of model validity under alternative

Possibilities

- IVX & Nonparametrics
- Application to long horizon prediction (Phillips & Lee, 2013)
- Quantile Regression balancing & IVX (Phillips, 2013; Lee, 2013)
Pitfalls + Possibilities in Predictive Regression

### Pitfalls
- Bias in estimation, testing, CIs & nonstandard inference
- Uncertain degrees of persistence - can be fatal as in Stock & CY
- Need to cope with multiple regressors and marginal predictability
- Omitted predictors misspecification
- Issues of model validity under alternative

### Possibilities
- IVX & Nonparametrics
- Application to long horizon prediction (Phillips & Lee, 2013)
- Quantile Regression balancing & IVX (Phillips, 2013; Lee, 2013)
Pitfalls + Possibilities in Predictive Regression

Pitfalls

- bias in estimation, testing, CIs & nonstandard inference
- uncertain degrees of persistence - can be fatal as in Stock & CY
- need to cope with multiple regressors and marginal predictability
- omitted predictors misspecification
- issues of model validity under alternative

Possibilities

- IVX & Nonparametrics
- Application to long horizon prediction (Phillips & Lee, 2013)
- Quantile Regression balancing & IVX (Phillips, 2013; Lee, 2013)
Thank You!
Robustness Issues

- Triangular system with UR or LUR regressors:

\[ y_t = Ax_t + u_{0t} \]

- \( x_t \) UR, or Local to Unity (LUR)

- Invalid asymptotic inference when \( x_t \) LUR
  - Estimators for \( A \) are not asymptotically \( MN(0, \cdot) \)
  - Wald test is not asymptotically \( \chi^2 \)
  - FM and other correction methods do not work

- Inference relies upon knowledge of type of regressor persistence
Triangular system with unknown degree of persistence:

\[
\begin{align*}
    y_t &= Ax_t + u_{0t} \\
    x_t &= R_n x_{t-1} + u_{xt} \\
    R_n &= I_K + \frac{C}{n^\alpha}, \quad C \leq 0, \quad \alpha > 0
\end{align*}
\]

- \(x_t\) UR if \(C = 0\) or \(\alpha > 1\)
- \(x_t\) LUR if \(C < 0\) and \(\alpha = 1\)
- \(x_t\) MI if \(C < 0\) and \(\alpha \in (0, 1)\)
Choose

\[ R_{nz} = I_K + C_z/n^\phi \quad \text{for } \phi \in (2/3, 1), \ C_z < 0 \]

Instrument Construction: \[ \tilde{z}_t = \sum_{j=1}^{t} R_{nz}^{t-j} \Delta x_j \]

Since \[ \Delta x_j = u_{xj} + C / n^\alpha x_{j-1} \]

\[ \tilde{z}_t = \sum_{j=1}^{t} R_{nz}^{t-j} u_{xj} + \frac{C}{n^\alpha} \sum_{j=1}^{t} R_{nz}^{t-j} x_{j-1} = z_t + \text{error} \]

\[ z_t \ R_{nz}-\text{mildly integrated} \]

**Two Cases:** (i) \( \alpha < \phi \), (ii) \( \phi \leq \alpha \leq 1 \)
Feasible IVX Inference

IVX estimation

\[ \tilde{A}_n = (Y' \tilde{Z} - n\hat{\Delta}_{0x}) (X' \tilde{Z})^{-1} \]

**Theorem**

**Case (i):** When \(2/3 < \varphi < \min(\alpha, 1)\),

\[ n^{\frac{1+\varphi}{2}} \text{vec} (\tilde{A}_n - A) \Rightarrow MN \left( 0, (\tilde{\Psi}_{xx}^{-1})' C_z V_{zz} C_z \tilde{\Psi}_{xx}^{-1} \otimes \Omega_{00} \right) \]

as \(n \to \infty\), where

\[ \tilde{\Psi}_{xx} = \begin{cases} 
\Omega_{xx} + \int_0^1 B_x dB'_x & x_t \text{ UR} \\
\Omega_{xx} + \int_0^1 J_C dJ'_C & x_t \text{ LUR} \\
\Omega_{xx} + V_{xx} C & x_t \text{ MI} 
\end{cases} \]
Feasible IVX Inference

- Let $V_{xz} = \int_0^\infty e^{rC} V_{xx} e^{rC_z} \, dr$

**Theorem**

Case (ii): For $\alpha \in (1/3, \varphi]$ and $\varphi \in (2/3, 1)$:

$$n^{1+\alpha/2} \text{vec} (\tilde{A}_n - A) \rightarrow \begin{cases} N \left( 0, V_{xx}^{-1} \otimes \Omega_{00} \right) & \text{if } \alpha < \varphi \\ N \left( 0, V_{xz}^{-1} C^{-1} V_{xx} C^{-1} (V_{xz}')^{-1} \otimes \Omega_{00} \right) & \text{if } \alpha = \varphi \end{cases}$$

- Asymptotic mixed normality applies for all $1/3 < \alpha \leq \varphi \in (2/3, 1)$
Let \( V_{xz} = \int_0^\infty e^{rC} V_{xx} e^{rCz} dr \)

**Theorem**

*Case (ii): For \( \alpha \in (1/3, \varphi] \) and \( \varphi \in (2/3, 1) \):

\[
\frac{1+\alpha}{2} \, \text{vec} \left( \tilde{A}_n - A \right) \Rightarrow \begin{cases} 
N \left( 0, V_{xx}^{-1} \otimes \Omega_{00} \right) & \text{if } \alpha < \varphi \\
N \left( 0, V_{xz}^{-1} C^{-1} V_{xx} C^{-1} (V'_{xz})^{-1} \otimes \Omega_{00} \right) & \text{if } \alpha = \varphi
\end{cases}
\]

Asymptotic mixed normality applies for all

\[
1/3 < \alpha \leq \varphi \in (2/3, 1)
\]
Feasible IV Inference (testing)

- Testing for linear restrictions:

\[ H_0 : H vec (A) = h, \hspace{1em} rank (H) = r \]

- Wald test based on IV procedure:

\[
W_n = (H vec \tilde{A}_n - h)' \left[ H \left\{ (X' \tilde{P} \tilde{Z} X)^{-1} \otimes \hat{\Omega}_{00} \right\} H' \right]^{-1} (H vec \tilde{A}_n - h)
\]

**Theorem**

*If \( \alpha > 1/3 \) and \( \varphi \) is chosen in \((2/3, 1)\)*

\[
W_n \Rightarrow_{H_0} \chi^2 (r)
\]

irrespective of whether \( x_t \) is UR, or LUR or MI
Feasible IVX Inference – Remarks

- Standard \( \chi^2 \) inference applies without \textit{a priori} knowledge of the order of persistence of \( x_t \): a full solution to Elliott (1998)
- Price paid for robustness: reduction in the RofC of \( \tilde{A}_n \) to \( O \left( n^{\frac{1+\phi}{2}} \right) \) from \( O(n) \).
- Exception - no cost when: \( x_t \) mildly integrated with \( \alpha \leq \varphi \)
- IVX method is feasible: instruments constructed from the regressors without extra information.
- Hence the requirement \( \alpha > 1/3 \): untractable simultaneity
- Mildly integrated processes provide good instruments because

\[ z_t \text{ stationary: } O_p(1) \ll z_t \text{ mild: } O_p \left( n^{\varphi/2} \right) \ll z_t \text{ LUR: } O_p \left( n^{1/2} \right) \]

- reduced RoC \( \implies \) eliminates endogeneity
- increased RoC \( \implies \) tractable asymptotic bias
Nonparametric predictive regression

Nonparametric estimation

• Model with single lagged regressor

\[ y_t = g(x_{t-\ell}) + u_{0t}, \]
\[ x_t = (1 + \frac{c}{n}) x_{t-1} + u_{xt} \]

unknown \( g \), mds \( u_{0t} \), and SM, LM, AP \( u_{xt} \)

• Nonparametric NW estimation

\[ \hat{g}(x) = \frac{\sum_{t=\ell+1}^{n} K_h (x_{t-\ell} - x) y_t}{\sum_{t=\ell+1}^{n} K_h (x_{t-\ell} - x)} \]

• Limit theory (Wang & Phillips, 2009)

\[ \left( \sum_{t=1+\ell}^{n} K \left( \frac{x_{t-\ell} - x}{h_n} \right) \right)^{1/2} \left( \hat{g}(x) - g(x) \right) \xrightarrow{d} N \left( 0, \sigma_0^2 \int_{-\infty}^{\infty} K(x)^2 dx \right) \]

• Includes SM, LM, AP \( u_{xt} \) with \(|d| < \frac{1}{2}\) (Kasparis, Andreou & Phillips, 2013)
Model with single lagged regressor

\[
y_t = g(x_{t-\ell}) + u_{0t},
\]

\[
x_t = \left(1 + \frac{c}{n}\right) x_{t-1} + u_{xt}
\]

unknown \( g \), mds \( u_{0t} \), and SM, LM, AP \( u_{xt} \)

Nonparametric NW estimation

\[
\hat{g}(x) = \frac{\sum_{t=\ell+1}^{n} K_h (x_{t-\ell} - x) y_t}{\sum_{t=\ell+1}^{n} K_h (x_{t-\ell} - x)}
\]

Limit theory (Wang & Phillips, 2009)

\[
\left( \sum_{t=1+\ell}^{n} K \left( \frac{x_{t-\ell} - x}{h_n} \right) \right)^{1/2} (\hat{g}(x) - g(x)) \xrightarrow{d} N \left( 0, \sigma_0^2 \int_{-\infty}^{\infty} K(x)^2 dx \right)
\]

Includes SM, LM, AP \( u_{xt} \) with \(|d| < \frac{1}{2}\) (Kasparis, Andreou & Phillips, 2013)
Test constant regression function $H_0 : g(x) = \mu$ using a t-ratio

$$\hat{t}(x, \mu) := \left( \frac{\sum_{t=1+\ell}^{n} K \left( \frac{x_{t-\ell} - x}{h_n} \right)}{\hat{\sigma}_0^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda} \right)^{1/2}$$

$$\hat{\sigma}_0^2 = n^{-1} \sum_{t=1+\ell}^{n} (y_t - \hat{\mu})^2$$

with $\hat{\mu} = \sum_{t=1+\ell}^{n} y_t / n$ and $\hat{\mu} = n^{-1} \sum_{t=1+\ell}^{n} (y_t - \hat{\mu})^2$ under the null

Test functionals over isolated points $X_s = \{ \bar{x}_1, \ldots, \bar{x}_s \}$ for some fixed $s$

$$\hat{F}_{\text{sum}} := \sum_{x \in X_s} [\hat{t}(x, \hat{\mu})]^2$$

and

$$\hat{F}_{\text{max}} := \max_{x \in X_s} [\hat{t}(x, \hat{\mu})]^2.$$
Nonparametric predictive regression
Nonlinear model testing

- Test constant regression function $\mathbb{H}_0 : g(x) = \mu$ using a t-ratio
  \[
  \hat{t}(x, \mu) := \left( \frac{\sum_{t=1+\ell}^n K \left( \frac{x_t - \ell - x}{h_n} \right)}{\hat{\sigma}_0^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda} \right)^{1/2} (\hat{g}(x) - \hat{\mu}) \xrightarrow{d} N(0, 1)
  \]
  with $\hat{\mu} = \sum_{t=1+\ell}^n y_t / n$ and $\hat{\sigma}_0^2 = n^{-1} \sum_{t=1+\ell}^n (y_t - \hat{\mu})^2$ under the null.

- Test functionals over isolated points $X_s = \{\bar{x}_1, ..., \bar{x}_s\}$ for some fixed $s$
  \[
  \hat{F}_{\text{sum}} := \sum_{x \in X_s} [\hat{t}(x, \hat{\mu})]^2 \quad \text{and} \quad \hat{F}_{\text{max}} := \max_{x \in X_s} [\hat{t}(x, \hat{\mu})]^2.
  \]

- Limit theory under $H_0$ as $n \to \infty$
  \[
  \hat{F}_{\text{sum}} \xrightarrow{d} \chi^2_s \quad \text{and} \quad \hat{F}_{\text{max}} \xrightarrow{d} Y \sim_d \max_{i \leq s} (\chi^2_{1i}).
  \]
Nonparametric predictive regression
Nonlinear model testing

- Test constant regression function \( H_0 : g(x) = \mu \) using a t-ratio

\[
\hat{t}(x, \mu) := \left( \frac{\sum_{t=1+\ell}^{n} K \left( \frac{x_{t-\ell}-x}{h_n} \right)}{\hat{\sigma}_0^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda} \right)^{1/2} (\hat{g}(x) - \hat{\mu}) \xrightarrow{d} N(0, 1)
\]

with \( \hat{\mu} = \sum_{t=1+\ell}^{n} y_t / n \) and \( \hat{\sigma}_0^2 = n^{-1} \sum_{t=1+\ell}^{n} (y_t - \hat{\mu})^2 \) under the null

- Test functionals over isolated points \( X_s = \{ \bar{x}_1, \ldots, \bar{x}_s \} \) for some fixed \( s \)

\[
\hat{F}_{\text{sum}} := \sum_{x \in X_s} [\hat{t}(x, \hat{\mu})]^2 \quad \text{and} \quad \hat{F}_{\text{max}} := \max_{x \in X_s} [\hat{t}(x, \hat{\mu})]^2.
\]

- Limit theory under \( H_0 \) as \( n \to \infty \)

\[
\hat{F}_{\text{sum}} \xrightarrow{d} \chi_s^2 \quad \text{and} \quad \hat{F}_{\text{max}} \xrightarrow{d} Y \sim_d \max_{i \leq s} (\chi_{1i}^2).
\]
Nonparametric predictive regression

Simulations - Model design

- Model

\[ y_t = g(x_{t-1}) + u_{0t}, \quad x_t = \left(1 + \frac{c}{n}\right)x_{t-1} + u_{xt}, \quad x_0 = 0 \]

- Error mechanisms

\[ u_{xt} = \rho_x u_{xt-1} + \eta_t, \quad \rho_x = \{0, 0.3\} \quad \text{(SM)} \]

\[ (I - L)^d u_{xt} = \eta_t, \quad d = \{-0.25, 0.25\} \quad \text{(AP & LM)} \]

\[
\begin{bmatrix}
  u_{0t} \\
  \eta_t
\end{bmatrix}
\sim iidN \left(0, \begin{bmatrix}
  1 & r \\
  r & 1
\end{bmatrix} \right), \quad |r| < 1.
\]
Functional forms

\[ g_0(x) = 0 \text{ (null hypothesis)} \]
\[ g_1(x) = 0.015x, \text{ (linear)} \]
\[ g_2(x) = \frac{1}{4} \text{sign}(x) |x|^{1/4} \text{ (polynomial)} \]
\[ g_3(x) = \frac{1}{5} \ln(|x| + 0.1) \text{ (logarithmic)} \]
\[ g_4(x) = \left(1 + e^{-x}\right)^{-1} \text{ (logistic)} \]
\[ g_5(x) = \left(1 + |x|^{0.9}\right)^{-1} \text{ (recip’al)} \]
\[ g_6(x) = e^{-5x^2} \text{ (integrable)} \]
Nonparametric predictive regression

Simulation Results - Size (Nominal 5%), Tests - NP, FM, JM

\( n = 500, \ h = n^{-b}, \ x_t \sim \text{local to unity process (c)} \)

Fig. 1(a) \( c = 0 \)

Fig. 1(b) \( c = -50 \)

- NPP \( b = 0.1 \)
- NPP \( b = 0.2 \)
- FMLS
- J&M
Nonparametric predictive regression

Simulation Results - Size (Nominal 5%), Tests - NP, FM, JM

\[ n = 500, \ h = n^{-b}, \ \Delta x_t \sim ARFIMA(d) \]

Fig. 2(a) \( d = -0.25 \)

Fig. 2(b) \( d = 0.25 \)
Nonparametric predictive regression

Simulation Results - Power, Tests - NP, FM, JM

\[ n = 1000, \ h = n^{-b}, \ g(x) = 0.015x: \ \rho_x = 0.3 \]

![Graph](image.png)

**Fig. 3(a)**  \( c = 0 \)

**Fig. 3(b)**  \( c = -50 \)
Nonparametric predictive regression

Simulation Results - Power, Tests - NP, FM, JM

\[ n = 1000, \quad h = n^{-b}, \quad c = 0, \quad \rho_x = 0.3 \]

Fig. 4(a) \( g(x) = (1 + |x|^{0.9})^{-1} \)

Fig. 4(b) \( g(x) = e^{-x^2} \)

---

Peter C. B. Phillips ()
Prediction
SoFiE Conference June 2013 65 / 90
Predictability Test Results for SP500 Returns 1926:1 - 2010:12 ($n = 1009$)

Two common predictors, Various grids and bandwidths $h_n = \hat{\sigma}_x n^{-b}$

<table>
<thead>
<tr>
<th>Tests</th>
<th>Grid pts</th>
<th>Lag</th>
<th>Log(D/P) $b$</th>
<th>Log(E/P) $b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum</td>
<td>10</td>
<td>1</td>
<td>-</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.3,0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Sum</td>
<td>35</td>
<td>1</td>
<td>-</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.2,0.3</td>
<td>0.1,0.2</td>
</tr>
<tr>
<td>Sum</td>
<td>50</td>
<td>1</td>
<td>0.2,0.3,0.4</td>
<td>0.1,0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>0.1,0.2,0.3,0.4</td>
</tr>
</tbody>
</table>
Pitfalls

- bias & nonstandard inference
- uncertain degrees of persistence - can be fatal
- zero coverage probability in CIs for stationarity/MI regressors
- need to cope with multiple regressors and marginal predictability
- omitted predictors misspecification - need to allow for temporal dependence
- model validity under alternative
Pitfalls + Possibilities in Predictive Regression

Pitfalls

- bias & nonstandard inference
- uncertain degrees of persistence - can be fatal
- zero coverage probability in CIs for stationarity/MI regressors
- need to cope with multiple regressors and marginal predictability
- omitted predictors misspecification - need to allow for temporal dependence
- model validity under alternative
Possibilities - IVX

- easy to use, no simulations/integration, standard chi-square inference
- holds for local to unity, mildly integrated, mildly explosive, and stationary regressors
- handles multiple regressors - many fundamentals, varying degrees of persistence
- handles multiple dependent variables, e.g., size sorted, value sorted, momentum sorted returns can all be regressed simultaneously
Possibilities - Nonparametrics

- easy to use
- copes with predictors of general functional form
- assures model validity under alternative
- particularly wide applicability regarding regressor persistence
- good size and power properties against parametric methods
- but challenged by multiple regressors
Motivation for dependent innovations
Scalar regressor misspecification with mds innovations

- Many papers impose mds innovations in predictive regression

\[ y_{t+1} = \beta_n x_t + u_{0t+1}, \quad \mathbb{E}_{\mathcal{F}_t}(u_{0t+1}) = 0 \]  \hspace{1cm} (7)

- Bivariate regression with mds innovations is unconvincing .... when
  - Two or more fundamentals \( x_{1t} \) and \( x_{2t} \) are known to have predictive power
  - As many authors have found (e.g. Campbell & Yogo, 2006)

- Omitted variable misspecification produces contradiction to (7)

\[ y_{t+1} = \beta_n x_{1t} + u_{0t+1}, \quad \mathbb{E}_{\mathcal{F}_t}(u_{0t+1}) \neq 0 \text{ since } x_{2t} \in \mathcal{F}_t. \]
The Model

- Multivariate system of predictive regressions:
  \[ y_{t+1} = A x_t + u_{0t+1} \]
  \[ x_{t+1} = R_n x_t + u_{xt+1} \]
  \[ R_n = I_p + \frac{C}{n^\alpha}, \text{ for some } \alpha \in (0, 1] \]

- \( A : m \times p \) coefficient matrix and \( C = \text{diag} (c_1, c_2, ..., c_p) \) are unknown.

- Each \( x_{it} \) can be in the general vicinity of unity
  - mildly integrated \((C < 0, \alpha \in (0, 1))\),
  - nearly integrated \((C < 0, \alpha = 1)\),
  - integrated \((C = 0)\),
  - locally explosive \((C > 0, \alpha = 1)\),
  - mildly explosive \((C > 0, \alpha \in (0, 1))\).
Multivariate system of predictive regressions:

\[ y_{t+1} = Ax_t + u_{0t+1} \]
\[ x_{t+1} = R_n x_t + u_{xt+1} \]

\[ R_n = I_p + \frac{C}{n^\alpha}, \text{ for some } \alpha \in (0, 1] \]

- \( A : m \times p \) coefficient matrix and \( C = diag (c_1, c_2, ..., c_p) \) are unknown.
- Each \( x_{it} \) can be in the general vicinity of unity
  - mildly integrated \( (C < 0, \alpha \in (0, 1)) \),
  - nearly integrated \( (C < 0, \alpha = 1) \),
  - integrated \( (C = 0) \),
  - locally explosive \( (C > 0, \alpha = 1) \),
  - mildly explosive \( (C > 0, \alpha \in (0, 1)) \).
**Weakly dependent errors:**

\[
u_t = \begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix} = \sum_{j=0}^{\infty} F_j \varepsilon_{t-j}, \quad \varepsilon_t \sim iid \ (0, \Sigma),
\]

\[\Sigma > 0, \quad \mathbb{E} \| \varepsilon_1 \|^4 < \infty, \quad F_0 = I_{m+p}, \quad \sum_{j=0}^{\infty} j \| F_j \| < \infty,\]

\[F(z) = \sum_{j=0}^{\infty} F_j z^j \text{ and } F(1) = \sum_{j=0}^{\infty} F_j > 0.\]

\[\Omega = \sum_{h=-\infty}^{\infty} \mathbb{E} (u_t u_{t-h}^\prime) = F(1) \Sigma F(1)', \quad \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix}.\]
Predictive Regression - IVX Limit Theory
Preliminary tools

- **FCLT:**
  \[
  \frac{1}{\sqrt{n}} \sum_{j=1}^{ns} u_j = \frac{1}{\sqrt{n}} \left[ \sum_{j=1}^{ns} u_{0j} \right] = \begin{bmatrix} B_{0n}(s) \\ B_{xn}(s) \end{bmatrix} = BM \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix}.
  \]

- **Local to unity asymptotics:**
  \[
  \frac{x_t}{\sqrt{n}} \rightarrow J_c^x(r) \text{ with } \left\lfloor nr \right\rfloor = t,
  \]
  where
  \[
  J_c^x(r) = \int_0^r e^{(r-s)c} dB_x(s),
  \]
  encompassing the unit root case
  \[
  J_c^x(r) = B_x(r) \text{ if } C = 0.
  \]
Predictive Regression - IVX Limit Theory

Preliminary tools

- BN decomposition (Phillips and Solo, 1992):

\[
\begin{align*}
    u_t &= F(1)\varepsilon_t - \Delta \tilde{\varepsilon}_t = \begin{bmatrix} F_0(1) \\ F_x(1) \end{bmatrix} \varepsilon_t - \Delta \tilde{\varepsilon}_t \\
    \tilde{\varepsilon}_t &= \sum_{j=0}^{\infty} \tilde{F}_j \varepsilon_{t-j}, \quad \tilde{F}_j = \sum_{s=j+1}^{\infty} F_s
\end{align*}
\]

- in partitioned form

\[
\begin{align*}
    u_{0t} &= F_0(1)\varepsilon_t - \Delta \tilde{\varepsilon}_{0t} \\
    u_{xt} &= F_x(1)\varepsilon_t - \Delta \tilde{\varepsilon}_{xt}
\end{align*}
\]
Consider first $\alpha = 1$ (local to unity $x_t$). We impose, for technical reasons, the rate condition:

$$\frac{n^{1/2}}{n^\phi} + \frac{n^\phi}{k} + \frac{k}{n} \rightarrow 0.$$  

Reasonable since, e.g., $k = n^{0.6}$ corresponds to 4-5 years among 60-80 years for monthly data.

**IVX Estimator:**

$$\hat{B}^{IVX} - B = \left( \sum_{t=1}^{n-k} u_{0t+k} \left( \tilde{z}_t^k \right)' \right) \left( \sum_{t=1}^{n-k} x_t^k \left( \tilde{z}_t^k \right)' \right)^{-1}.$$
Numerator decomposition

\[
\frac{1}{k^{\frac{1}{2}} n^{\frac{1}{2}} + \varphi} \sum_{t=1}^{n-k} u_{0t+k} (z_t^k)' + \frac{1}{k^{\frac{1}{2}} n^{\frac{3}{2}} + \varphi} \sum_{t=1}^{n-k} u_{0t+k} (\psi_{nt}^k)' \]

\[= \frac{1}{k^{\frac{1}{2}} n^{\frac{1}{2}} + \varphi} \sum_{t=1}^{n-k} u_{0t+k} (z_t^k)' + o_p(1)
\]
Component limits

Lemma

1. \( \frac{C_z}{k^{\frac{1}{2}} n^\varphi} z_t^k \implies V_x(t) \equiv N(0, \Omega_{xx}) \) for any \( t \)

2. \( \frac{1}{n} \sum_{t=1}^{n-k} \left( \frac{C_z}{k^{\frac{1}{2}} n^\varphi} z_t^k \right) \left( \frac{C_z}{k^{\frac{1}{2}} n^\varphi} z_t^k \right)' \xrightarrow{p} \Omega_{xx} = F_x(1) \Sigma F_x(1)' \)

3. \( \frac{x_t^k}{kn^{\frac{1}{2}}} \implies J^x_c(r) \)
Limit Theory

Numerator

- Use BN decomposition and
- Apply MGCLT to the dominant component of the numerator

Lemma

For \( \frac{n\phi}{k} + \frac{k}{n} \rightarrow 0 \),

\[
\frac{1}{k^{\frac{1}{2}} n^{\frac{1}{2}} + \rho} \sum_{t=1}^{n-k} u_{0t+k} \left( z_t^k \right)' \quad \implies \quad N \left( 0, C_z^{-1} \Omega_{xx} C_z^{-1} \otimes \Omega_{00} \right)
\]

Since

\[
\frac{1}{k^{\frac{1}{2}} n^{\frac{1}{2}} + \rho} \sum_{t=1}^{n-k} u_{0t+k} \left( z_t^k \right)' \sim \frac{1}{\sqrt{n}} F_0 \left( 1 \right) \sum_{t=1}^{n-k} \varepsilon_{t+k} \left( \frac{z_t^k}{k^{\frac{1}{2}} n^{\frac{\phi}{2}}} \right)'
\]
Denominator decomposition

\[
\frac{1}{k^2 n^{1+\phi}} \sum_{t=1}^{n-k} x_t^k (\tilde{z}_t^k)'
\]

\[= \frac{1}{n} \sum_{t=1}^{n-k} \frac{x_t^k}{kn^{1/2}} \left( \frac{z_t^k}{kn^\phi} \right)' + \frac{1}{n} \sum_{t=1}^{n-k} \frac{x_t^k}{kn^{1/2}} \left( \frac{C_z}{kn^{1/2+\phi} \psi_{nt}^k} \right) C_z^{-1} C
\]
Component limits

Lemma

\[ \frac{1}{k^2 n^{2+\phi}} \sum_{t=1}^{n-k} x^k_t (\psi^k_{nt})' C \longrightarrow \left( \int_0^1 J^x_c (r) J^x_c (r)' dr \right) C^{-1}_z \]

\[ \frac{1}{k^2 n^{1+\phi}} \sum_{t=1}^{n-k} x^k_t (z^k_t)' C_z \rightarrow \rho \frac{1}{2} \Omega_{xx} \]

Denominator limit

\[ \frac{1}{k^2 n^{1+\phi}} \sum_{t=1}^{n-k} x^k_t (\tilde{z}_t^k)' \longrightarrow \frac{1}{2} \Omega_{xx} C^{-1}_z + \left( \int_0^1 J^x_c (r) J^x_c (r)' dr \right) C^{-1}_z \]
Combining gives the limit theory for

\[ \hat{B}^{IVX} - B = \left( \sum_{t=1}^{n-k} u_{0t+k} \left( \hat{z}_t^k \right)' \right) \left( \sum_{t=1}^{n-k} x_t^k \left( \hat{z}_t^k \right)' \right)^{-1}. \]

**Theorem**

\[ n^{\frac{1}{2}} k^{\frac{3}{2}} \left( \hat{B}^{IVX} - B \right) \xrightarrow{\text{d}} MN \left( 0, \Sigma_B \right), \]

where

\[ \Sigma_B = \left( \Psi^{-1}_{cxz} \right)' C_z^{-1} \Omega_{xx} C_z^{-1} \left( \Psi^{-1}_{cxz} \right) \otimes \Omega_{00}, \]

\[ \Psi_{cxz} = \frac{1}{2} \Omega_{xx} C_z^{-1} + \left( \int_0^1 J_c^x (r) J_c^x (r)' \, dr \right) C_z^{-1} C. \]
Mixed normality holds b/c

\[ \zeta_{nt+k} := \begin{bmatrix} \frac{1}{k^{1/2}n^{1/2+\phi}} z_t^k \otimes \varepsilon_{t+k} \\ \frac{1}{\sqrt{n}} \varepsilon_{t+k} \end{bmatrix} \]

has asymptotically orthogonal components:

\[
\sum \mathbb{E} \mathcal{F}_{nt+k-1} \zeta_{nt+k} \zeta'_{nt+k} = \begin{bmatrix}
\frac{1}{n} \sum_{t=1}^{n-k} \left( \frac{1}{k^{1/2}n^{\phi}} z_t^k \right) \left( \frac{1}{k^{1/2}n^{\phi}} z_t^k \right)' \otimes \Sigma \\
\frac{1}{n} \sum_{t=1}^{n-k} \left( \frac{1}{k^{1/2}n^{\phi}} z_t^k \right) \otimes \Sigma
\end{bmatrix}
\rightarrow_p \begin{bmatrix}
C_z^{-1} \Omega_{xx} C_z^{-1} \otimes \Sigma \\
0 \\
0 \otimes \Sigma
\end{bmatrix}.
\]
Limit Theory of Tests
Self-normalized IVX Estimator for use in Testing

Lemma

\[ \frac{1}{kn^{1+2\phi}} \sum_{t=1}^{n-k} (\tilde{z}_t^k) (\tilde{z}_t^k)' = \frac{1}{kn^{1+2\phi}} \sum_{t=1}^{n-k} (z_t^k) (z_t^k)' + o_p(1). \]

\[ nk^3 (X'P_{\tilde{Z}}X)^{-1} \otimes \hat{\Omega}_0 = (\Psi_{cxz}^{-1}) C_z^{-1} (\Omega_{xx}) C_z^{-1} (\Psi_{cxz}^{-1})' \otimes \Omega_0 \]

where

\[ (X'P_{\tilde{Z}}X)^{-1} = \left\{ \left( \sum_{t=1}^{n-k} x_t^k (\tilde{z}_t^k)' \right) \left( \sum_{t=1}^{n-k} (\tilde{z}_t^k) (\tilde{z}_t^k)' \right)^{-1} \left( \sum_{t=1}^{n-k} x_t^k (\tilde{z}_t^k)' \right)' \right\}^{-1}. \]
Limit Theory
Self-normalized IVX Estimator for use in Testing

Theorem

\[
\begin{align*}
\text{Vec} \left\{ \left( \hat{\mathcal{B}}^{IVX} - \mathcal{B} \right) \right\}' \left[ (X' \tilde{Z} X)^{-1} \otimes \hat{\Omega}_{00} \right]^{-1} \text{Vec} \left\{ \left( \hat{\mathcal{B}}^{IVX} - \mathcal{B} \right) \right\} \\
\implies \chi^2 (mp),
\end{align*}
\]

using a consistent estimator \( \hat{\Omega}_{00} \) for \( \Omega_{00} \).

- Standard Wald statistic \( W_{IVX} \) - easy to compute
- Standard chi-squared test
Model Validity & Consistency under Alternative
Balance by localization of coefficient

- Local to zero predictive regression:

\[ y_{t+k} = B_n x_t^k + u_{0t+k}, \quad x_t^k = \sum_{j=1}^k x_{t+j-1} \]

- Suppose \( B_n = \frac{B}{n^{\gamma_k}} \) with \( B \neq 0 \). Now \( x_{t=\lfloor nr \rfloor}^{\frac{1}{kn^{\gamma}}} \Rightarrow J_c^x(r) \)

- Then

\[ y_{t+k} \sim \begin{cases} 
  u_{0t+k} + Bn^{1-\gamma} J_c^x \left( \frac{t}{n} \right) \neq l(0) & \gamma < 0.5 \\
  u_{0t+k} + BJ_c^x \left( \frac{t}{n} \right) = O_p(1) & \gamma = 0.5 \\
  u_{0t+k} = O_p(1) & \gamma > 0.5 
\end{cases} \]

- Balance for \( \gamma \geq 0.5 \)
Local to zero predictive regression:

\[ y_{t+k} = B_n x_t^k + u_{0t+k}, \quad x_t^k = \sum_{j=1}^{k} x_{t+j-1} \]

Suppose \( B_n = \frac{B}{n^{\gamma_k}} \) with \( B \neq 0 \). Now \( \frac{x_t^k}{k^{1/2}} \) \( \Rightarrow \) \( J_c^x (r) \)

Then

\[ y_{t+k} \sim \begin{cases} 
  u_{0t+k} + B n^{1/2 - \gamma} J_c^x (\frac{t}{n}) & \neq I(0) \quad \gamma < 0.5 \\
  u_{0t+k} + B J_c^x (\frac{t}{n}) & = O_p(1) \quad \gamma = 0.5 \\
  u_{0t+k} = O_p(1) & \gamma > 0.5 
\end{cases} \]

Balance for \( \gamma \geq 0.5 \)
Model Validity & Consistency under Alternative Balance by localization of coefficient

- Test consistency

\[ n^{1/2} k^{3/2} \hat{B}_{IVX} = n^{1/2} k^{3/2} \left( \hat{B}_{IVX} - B_n \right) + n^{1/2} k^{3/2} B_n = O_p(1) + O_p \left( \frac{k^{1/2}}{n^{\gamma-1/2}} \right) \]

- Test is consistent whenever

\[ \frac{n^{1/2} k^{3/2}}{n^{\gamma} k} = \frac{k^{1/2}}{n^{\gamma-1/2}} \to \infty \]

- Convergence rate of \( \hat{B}_{IVX} \) is fast enough to distinguish \( B_n \neq 0 \)
Consider first $\alpha < 1$, $C < 0$ (mildly integrated $x_t$). We use the rate condition:

$$\frac{n^{1/2}}{n^\phi} + \frac{n^\phi}{k} + \frac{k}{n^\alpha} + \frac{n^\alpha}{n} \to 0.$$ 

Set horizon $k << n^\alpha << n$

IVX limit theory:

**Theorem**

$$n^{1/2} k^{3/2} \left( \hat{B}^{IVX} - B \right) \implies N \left( 0, 4\Omega_{xx}^{-1} \otimes \Omega_{00} \right)$$

$$W_{IVX} \implies \chi^2 (mp)$$

- Same convergence rate as for $\alpha = 1$
Consider first $\alpha < 1$, $C > 0$ (mildly explosive $x_t$ with $R_n = I + n^{-\alpha} C$). We use the same rate condition:

$$\frac{n^{1/2}}{n^q} + \frac{n^q}{k} + \frac{k}{n^\alpha} + \frac{n^\alpha}{n} \rightarrow 0.$$ 

IVX limit theory

**Theorem**

$$kn^\alpha \left( \hat{B}^{IVX} - B \right) R_n^n \quad \Rightarrow \quad MN \left( 0, \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp \otimes \Omega_{00} \right)$$

$$W_{IVX} \quad \Rightarrow \quad \chi^2 (mp)$$

Faster convergence rate than $\alpha = 1$
Localizing Coefficient Estimation
Univariate Model - illustration

- Localizing coefficient $C$ and index $\alpha$ in

$$x_{t+1} = R_n x_t + u_{xt+1}$$

$$R_n = 1 + \frac{c}{n^\alpha}, \text{ for some } \alpha \in (0, 1]$$

- $\alpha$ is consistently estimable

  - Estimators: $\hat{\alpha} = -\frac{\log|R_n - 1|}{\log n} \rightarrow_p \alpha$; $\tilde{\alpha} = \frac{\log(\frac{2}{n} \sum_{t=1}^n x_t^2)}{\log n} \rightarrow_p \alpha$

  - Limit distribution

$$n^{\frac{1-\alpha}{2}} \log n \left\{ \hat{\alpha} - \alpha + \frac{1}{\log n} \log \left| \frac{c \sigma_x^2}{\omega_x^2} \right| \right\} \Rightarrow \frac{\omega_x^2}{c \sigma_x^2} \xi_c \quad \alpha \in (0, 1)$$

$$n^{\frac{1}{2}} \log n \left\{ \hat{\alpha} - 1 \right\} \Rightarrow -\log |c + \xi_c| \quad \alpha = 1$$
Localizing Coefficient Estimation
Univariate Model - illustration

- Localizing coefficient $C$ and index $\alpha$ in

$$x_{t+1} = R_n x_t + u_{xt+1}$$
$$R_n = 1 + \frac{c}{n^\alpha}, \text{ for some } \alpha \in (0, 1]$$

- $\alpha$ is consistently estimable
- Estimators: $\hat{\alpha} = -\frac{\log|\hat{R}_n - 1|}{\log n} \to_p \alpha$; $\tilde{\alpha} = \frac{\log\left(\frac{2}{n} \sum_{t=1}^{n} x_t^2\right)}{\log n} \to_p \alpha$
- Limit distribution

$$n^{\frac{1-\alpha}{2}} \log n \left\{ \hat{\alpha} - \alpha + \frac{1}{\log n} \log \left| \frac{c \sigma_x^2}{\omega_x^2} \right| \right\} \Rightarrow \frac{\omega_x^2}{c \sigma_x^2} \bar{\zeta}, \quad \alpha \in (0, 1)$$

$$(\log n) \left\{ \hat{\alpha} - 1 \right\} \Rightarrow -\log \left| c + \bar{\zeta}_c \right| \quad \alpha = 1$$
Pitfalls: constructing a CI for $c$

Stock’s confidence belts for $c$

Confidence belts (levels 2.5%, 50% and 97.5%) with asymptotic functional approximations – Phillips (2013)