Comparing Nested Predictive Regression Models with Persistent Predictors

Tae-Hwy Lee
University of California, Riverside
Jointly with Yan Ge (CUFE) and Mike McCracken (FRBSTL)

August 12, 2015
KAEA/KDI Joint Conference @ KDI

Outline

1. Predictive regression: literature
2. DM and ENC for predictive mean regression
3. DM and ENC for predictive quantile regression
4. Asymptotic distribution of ENC with a stationary predictor
5. Asymptotic distribution of ENC with a persistent predictor
6. Monte Carlo
7. Application: forecasting equity premium
8. When Model 1 is the martingale difference model
9. When the predictor is persistent with a drift
10. Conclusions

1. Predictive regression: literature

Predictive (mean) regression

$$
\begin{aligned}
y_{t+1} &= a + bx_t + e_{t+1} \\
x_{t+1} &= \mu + \phi x_t + v_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\left(\begin{array}{c} e_{t+1} \\ v_{t+1} \end{array} \right) \Bigg| \mathcal{F}_t &\sim N\left[\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma_e^2 & \sigma_{ev} \\ \sigma_{ev} & \sigma_v^2 \end{array} \right) \right] \\
e_{t+1} &= \gamma v_{t+1} + \varepsilon_{t+1} \\
\gamma &= \left(\frac{\sigma_{ev}}{\sigma_{v^2}} \right)
\end{aligned}
$$

Problems when $\gamma \neq 0$

1. **with stationary predictors, finite sample bias:** Stambaugh (1999 *JFE*)
2. **with persistent predictors, nonstandard limit theory and uncorrectable bias:** Elliott and Stock (1994), Campbell and Yogo (2006) and Jansson and Moreira (2006).

See Phillips and JH Lee (2013 *JoE*) for a nice summary of the literature.

Problem #1. Finite sample bias with stationary a predictor.

$$\begin{aligned}
\hat{b} - b &= \frac{\sum_{t=1}^{n}(x_t - \bar{x})\, e_{t+1}}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \\
&= \frac{\sum_{t=1}^{n}(x_t - \bar{x})(\gamma v_{t+1} + \varepsilon_{t+1})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \\
&= \gamma \frac{\sum_{t=1}^{n}(x_t - \bar{x})\, v_{t+1}}{\sum_{t=1}^{n}(x_t - \bar{x})^2} + \frac{\sum_{t=1}^{n}(x_t - \bar{x})\, \varepsilon_{t+1}}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \\
&= \gamma\left(\hat{\phi} - \phi\right) + \frac{\sum_{t=1}^{n}(x_t - \bar{x})\, \varepsilon_{t+1}}{\sum_{t=1}^{n}(x_t - \bar{x})^2}
\end{aligned}$$

Under normality and with a stationary predictor $(|\phi| < 1)$, Stambaugh (1999 *JFE*) shows

$$E\left(\hat{b} - b\right) = -\gamma\left(\frac{1 + 3\phi}{n}\right) + O\left(\frac{1}{n^2}\right),$$

using results of Marriott and Pope (1954) and Kendall (1954)

$$E\left(\hat{\phi} - \phi\right) = -\left(\frac{1 + 3\phi}{n}\right) + O\left(\frac{1}{n^2}\right).$$

Solutions to Problem #1.

- *Bias correction:* Kothari and Shanken (1997 *JFE*), Amihud and Hurvich (2004 *JFQA*), Zhu (2014 *JFEC*)

$$\begin{aligned} \hat{b}_{adj} &= \hat{b} + \hat{\gamma}\left(\frac{1+3\hat{\phi}}{n}\right) \\ \hat{\gamma} &= \left(\frac{\hat{\sigma}_{ev}}{\hat{\sigma}_{v^2}}\right) \end{aligned}$$

- *Biased estimation with monotonicity constraints:* Campbell and Thompson (2008 *RFS*), Lee, Tu, and Ullah (2014 *JoE*, 2015 *JBES*), next slides ...

Forecasting with Monotonicity Constraints: examples

1. Forecasting equity premium
2. Forecasting inflation
3. Forecasting output growth, Forecasting recession probability

Example 1: Forecasting equity premium

Campbell and Thompson (2008 *RFS*), "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?"

$$y_{t+1} = a + bx_t + e_{t+1}, \qquad b > 0$$

$y =$ excess returns on S&P500 over 3-month T-bill interest rate

x	sign(b)	
d/p	$+$	dividend yield
e/p	$+$	earnings yield
se/p	$+$	smoothed earnings yield
b/m	$+$	book-to-market ratio
roe	$+$	smoothed return on equity
t-bill	$-$	3-month Treasury-Bill
lty	$-$	long-term government bond yield
ts	$+$	term spread: long term $-$ short term Treasury yields
ds	$+$	default spread: corporate $-$ Treasury bond yields
inf	$-$	inflation rate
nei	$-$	equity share of new issues

Example 1: Forecasting equity premium (cont.)

$$y_{t+1} = a + bx_t + e_{t+1}$$

1. Unrestricted forecast: $\tilde{a}_n + \tilde{b}_n x_n$ with $\tilde{a}_n, \tilde{b}_n$ unrestricted OLS
2. HM (Historical mean, $b = 0$): HM $= \frac{1}{n} \sum_{t=1}^{n} y_t$
3. Monotonicity ($b > 0$)

- Goyal and Welch (2008): 2 is better than 1.
- Campbell and Thompson (2008): 3 is better than 2.

Example 2: Forecasting inflation

Stock and Watson (2007, *JMCB*), "Why has U.S. inflation become harder to forecast?"

$$y_{t+1} = a + bx_t + e_{t+1}$$

where

- $y = \Delta\pi$, with $\pi =$ change in log price index such as GDP deflator, personal consumption expenditure (PCE) deflator for core items, PCE deflator for all items, or CPI, and
- $x =$ change in activity such as

x	sign(b)	activity
Δu	$-$	unemployment rate (all, 16+)
Δrgdp	$+$	logarithm of real GDP
Δcaputil	$+$	capacity utilization rate
Δpermit	$+$	building permits
cfnai	$+$	Chicago Fed National Activity Index

Example 2: Forecasting inflation (cont.)

$$y_{t+1} = a + bx_t + e_{t+1}$$

1. Unrestricted forecast

$$y_{t+1} = a + bx_t + \gamma(B)y_t + e_{t+1}$$

2. MA(1) $\left(b = 0,\ \gamma(B) = -\theta(1-\theta B)^{-1}\right)$

$$y_{t+1} = \mu + (1-\theta B)\varepsilon_{t+1}$$

3. Monotonicity ($b > 0$)

- SW (2007 *JMCB*): 2 is better than 1.
- Q: 3 is better than 2?

Example 3: Forecasting output growth and recession
Faust and Wright (2009, *JBES*)

$$\begin{aligned} y_{t+1} &= a + bx_t + e_{t+1} \\ \Pr(y_{t+1} < 0) &= \Phi\left(a + bx_t\right) \end{aligned}$$

where $y_{t+1} = \ln(Y_{t+k}/Y_t)$ is the real GDP growth from time t to $t+k$, and x_t is the term spread between 10 year government bond yield and 3 month T-bill rate.

- Estrella and Hardouvelis (1991 *JF*), "The term structure as a predictor of real economic activity"
- Wheelock and Wohar (2009 *FRBSTL Review*), "Can the term spread predict output growth and recession? A survey of the literature"

Example 3: Forecasting output growth and recession (cont.)

$$\begin{aligned} y_{t+1} &= a + bx_t + e_{t+1} \\ \Pr(y_{t+1} < 0) &= \Phi\left(a + bx_t\right) \end{aligned}$$

1. Unrestricted forecast

$$y_{t+1} = a + bx_t + \gamma(B) y_t + e_{t+1}$$

2. AR(4) ($b = 0$)

$$y_{t+1} = a + \gamma(B) y_t + e_{t+1}$$

3. Monotonicity ($b > 0$)

- Faust and Wright (2009 *JBES*): 2 is better than 1.
- Q: 3 is better than 2?

Problem #2. Nonstandard limit theory and uncorrectable bias with persistent regressors.

Suggested solutions to Problem #2

1. The Bonferroni method: Campbell and Yogo (2006 *JFE*) and Cavanagh et al. (1995 *ET*)
2. A conditional likelihood approach with sufficient statistics: Jansson and Moreira (2006 *Econometrica*)
3. A control function approach: Elliott (2011 *JoE*), Cai and Wang (2014 *JoE*)
4. IVX: Phillips and Magdalinos (2007 *JoE*, 2009), Phillips and JH Lee (2013 *JoE*, 2014), Phillips (2015 *JFEC*)
5. ENC: This paper. ENC $\Rightarrow N(0,1)$ as $\frac{P}{R} \rightarrow \infty$, when the predictor is either stationary or persistent.

Comparing nested predictive regression models

when the predictor is persistent with correlated errors

DGP

$$
\begin{aligned}
y_{t+1} &= a + bx_t + e_{t+1} \\
x_{t+1} &= \mu + \phi x_t + v_{t+1} \\
\phi &= 1 - c/R, c > 0 \\
e_{t+1} &= \gamma v_{t+1} + \varepsilon_{t+1}
\end{aligned}
$$

Models

$$
\begin{aligned}
\text{Model 1} &: \quad y_{t+1} = a_1 + e^{(1)}_{t+1} \\
\text{Model 2} &: \quad y_{t+1} = a_2 + bx_t + e^{(2)}_{t+1}
\end{aligned}
$$

2. DM and ENC for the predictive mean regression

Comparing nested predictive regression models

$$\begin{aligned} \text{Model 1} &: \quad y_{t+1} = a_1 + e_{t+1}^{(1)} \equiv x_{1,t}'\beta_1 + e_{t+1}^{(1)} \\ \text{Model 2} &: \quad y_{t+1} = a_2 + bx_t + e_{t+1}^{(2)} \equiv x_{2,t}'\beta_2 + e_{t+1}^{(2)} \end{aligned}$$

where $x_{2,t}' = \left(x_{1,t}' \ x_t'\right)'$.

$$\begin{aligned} \mathbb{H}_0 &: \quad \mathbb{E}\left[L\left(e_{t+1}^{(1)}\right) - L\left(e_{t+1}^{(2)}\right)\right] = 0 \\ \mathbb{H}_1 &: \quad \mathbb{E}\left[L\left(e_{t+1}^{(1)}\right) - L\left(e_{t+1}^{(2)}\right)\right] > 0 \end{aligned}$$

A method of testing $\mathbb{H}_0$ was discussed in Granger and Newbold (1987, chapter 9) and later extended by Diebold and Mariano (DM 1995).

$$\hat{D}_P = P^{-1} \sum_{t=R}^{T} \left[L\left(\hat{e}_{t+1}^{(1)}\right) - L\left(\hat{e}_{t+1}^{(2)}\right)\right]$$

Issues with DM

1. parameter estimation error (West 1996)
2. comparing multiple models (White 2000, Hansen 2005, Romano and Wolf 2005a, 2005b, 2007)
3. comparing nested models (Clark and McCracken 2001, Clark and West 2006, 2007, Giacomini and White 2005, Chao, Corradi and Swanson 2001).
4. persistent predictors with $\phi = 1 - c/R$ (Stambaugh 1999, Campbell and Yogo 2006, Phillips and Magdalinos 2007, 2009, Phillips and Lee 2013, 2014, Phillips and Chen 2014, Cai and Wang 2014, Breitung and Demetrescu 2015).
5. When $\pi = \infty$, where $\pi := \lim_{T\to\infty} \frac{P}{R}$. (CM 2001, Inoue, Jin, and Rossi 2014, Sun, Hong, and Wang 2015)

This talk deals with Issues #3, #4, #5.

$$\text{Model 1} \quad : \quad y_{t+1} = a_1 + e^{(1)}_{t+1} \equiv x'_{1,t}\beta_1 + e^{(1)}_{t+1}$$
$$\text{Model 2} \quad : \quad y_{t+1} = a_2 + bx_t + e^{(2)}_{t+1} \equiv x'_{2,t}\beta_2 + e^{(2)}_{t+1}$$

$$\begin{aligned}
\hat{D}_P &= P^{-1}\sum_{t=R}^{T}\left[\left(\hat{e}^{(1)}_{t+1}\right)^2 - \left(\hat{e}^{(2)}_{t+1}\right)^2\right] \\
&= P^{-1}\sum_{t=R}^{T}\left(\hat{e}^{(1)}_{t+1} + \hat{e}^{(2)}_{t+1}\right)\left(\hat{e}^{(1)}_{t+1} - \hat{e}^{(2)}_{t+1}\right) \\
&= P^{-1}\sum_{t=R}^{T}\left(2\hat{e}^{(1)}_{t+1} - \hat{e}^{(1)}_{t+1} + \hat{e}^{(2)}_{t+1}\right)\left(\hat{e}^{(1)}_{t+1} - \hat{e}^{(2)}_{t+1}\right) \\
&= P^{-1}\sum_{t=R}^{T}2\hat{e}^{(1)}_{t+1}\left(\hat{e}^{(1)}_{t+1} - \hat{e}^{(2)}_{t+1}\right) - P^{-1}\sum_{t=R}^{T}\left(\hat{e}^{(1)}_{t+1} - \hat{e}^{(2)}_{t+1}\right)^2 \\
&\equiv \hat{B}_P - \hat{A}_P
\end{aligned}$$

Note: $\hat{B}_P$ is the statistic for **forecast encompassing** in the predictive *mean* regression models!

Forecast encompassing in conditional mean

$$\begin{aligned}
\varepsilon_{t+1} &= y_{t+1} - \left[(1-\lambda) f_{t+1}^{(1)} + \lambda f_{t+1}^{(2)}\right] = (1-\lambda)\,\hat{e}_{t+1}^{(1)} + \lambda \hat{e}_{t+1}^{(2)} \\
\hat{e}_{t+1}^{(1)} &= \lambda \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right) + \varepsilon_{t+1} \\
\hat{\lambda}_P &= \arg\min_{\lambda} \frac{1}{P} \sum_{t=R}^{T} \varepsilon_{t+1}^2
\end{aligned}$$

- FOC:

$$\frac{1}{P} \sum_{t=R}^{T} \left[\varepsilon_{t+1} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right)\right] = 0.$$

- If $\lambda = 0$, then $\varepsilon_{t+1} = \hat{e}_{t+1}^{(1)}$. Hence

$$\frac{1}{P} \sum_{t=R}^{T} \left[\hat{e}_{t+1}^{(1)} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right)\right] \xrightarrow{p} 0.$$

Proposition (CW 2007+GL 2014)

1. $\hat{A}_P = P^{-1} \sum_{t=R}^{T} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)} \right)^2 \xrightarrow{p} \mathbb{E}\left(\hat{A}_P\right) \geq 0$
2. $\hat{B}_P = P^{-1} \sum_{t=R}^{T} 2\hat{e}_{t+1}^{(1)} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)} \right) \xrightarrow{p} \mathbb{E}(\hat{B}_P) = 0$ as $P \to \infty$ under $\mathbb{H}_0$.
3. $\hat{D}_P = \hat{B}_P - \hat{A}_P \xrightarrow{p} \mathbb{E}(\hat{D}_P) < 0$ as $P \to \infty$ under $\mathbb{H}_0$.

Three test statistics: We compare DM, ENC, and CCS (Chao, Corradi and Swanson 2001).

1. $DM_P \equiv \hat{S}_P^{-0.5}\sqrt{P}\hat{D}_P$, where

$$\hat{D}_P = \frac{1}{P}\sum_{t=R}^{T}\left[\left(\hat{e}_{t+1}^{(1)}\right)^2 - \left(\hat{e}_{t+1}^{(2)}\right)^2\right]$$

2. $ENC_P \equiv \hat{Q}_P^{-0.5}\sqrt{P}\hat{B}_P$, where

$$\hat{B}_P \equiv \hat{D}_P + \hat{A}_P = \frac{1}{P}\sum_{t=R}^{T}\left[\hat{e}_{t+1}^{(1)}\left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right)\right]$$

3. $CCS_P \equiv \hat{W}_P^{-0.5}\sqrt{P}\hat{M}_P$, where

$$\hat{M}_P = P^{-1}\sum_{t=R}^{T}\hat{e}_{t+1}^{(1)}x_t.$$

$\hat{S}_P, \hat{Q}_P, \hat{W}_P$ are the consistent estimators of $S_P = var\left(\sqrt{P}\hat{D}_P\right), Q_P = var\left(\sqrt{P}\hat{B}_P\right)$, $W_P = var\left(\sqrt{P}\hat{M}_P\right)$.

3. DM and ENC for the predictive quantile regression

Ge and Lee (2014)

Comparing nested predictive quantile regression models

$$\text{Model 1} \quad : \quad y_{t+1} = a_{1,\alpha} + e^{(1)}_{t+1,\alpha}$$
$$\text{Model 2} \quad : \quad y_{t+1} = a_{2,\alpha} + b_\alpha x_t + e^{(2)}_{t+1,\alpha}.$$

- Loss function:

$$\rho_\alpha\left(e^{(i)}_{t+1,\alpha}\right) \equiv g_\alpha\left(e^{(i)}_{t+1,\alpha}\right) e^{(i)}_{t+1,\alpha} \qquad i = 1, 2,$$

where $g_\alpha(z) = \alpha - 1(z < 0)$.

- DM (with α omitted for simplicity)

$$\mathbb{H}_0 \quad : \quad \mathbb{E}\left[\rho\left(e^{(1)}_{t+1}\right) - \rho\left(e^{(2)}_{t+1}\right)\right] = 0$$
$$\mathbb{H}_1 \quad : \quad \mathbb{E}\left[\rho\left(e^{(1)}_{t+1}\right) - \rho\left(e^{(2)}_{t+1}\right)\right] > 0.$$

$$\hat{D}_P = \frac{1}{P}\sum_{t=R}^{T}\left[g\left(\hat{e}^{(1)}_{t+1}\right)\hat{e}^{(1)}_{t+1} - g\left(\hat{e}^{(2)}_{t+1}\right)\hat{e}^{(2)}_{t+1}\right]$$

Proposition (GL 2014)

1. $\hat{A}_P \equiv P^{-1} \sum_{t=R}^{T} \left[g\left(\hat{e}_{t+1}^{(2)}\right) - g\left(\hat{e}_{t+1}^{(1)}\right) \right] \hat{e}_{t+1}^{(2)} \geq 0.$

2. $\hat{B}_P \equiv P^{-1} \sum_{t=R}^{T} \left[g\left(\hat{e}_{t+1}^{(1)}\right) \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right) \right]$
   $\xrightarrow{p} \mathbb{E}(\hat{B}_P) = 0$ as $P \to \infty$, under $\mathbb{H}_0$.

3. $\hat{D}_P = \hat{B}_P - \hat{A}_P \xrightarrow{p} \mathbb{E}(\hat{D}_P) \leq 0$ as $P \to \infty$, under $\mathbb{H}_0$.

Note: $\hat{B}_P$ is the statistic for **forecast encompassing** in the predictive *quantile* regression models!

Forecast encompassing in conditional quantile

$$\begin{aligned}
\varepsilon_{t+1} &= y_{t+1} - \left[(1-\lambda) f_{t+1}^{(1)} + \lambda f_{t+1}^{(2)} \right] = (1-\lambda)\,\hat{e}_{t+1}^{(1)} + \lambda \hat{e}_{t+1}^{(2)} \\
\hat{e}_{t+1}^{(1)} &= \lambda \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)} \right) + \varepsilon_{t+1} \\
\hat{\lambda}_P &= \arg\min_{\lambda} \frac{1}{P} \sum_{t=R}^{T} \rho_\alpha \left(\varepsilon_{t+1} \right)
\end{aligned}$$

- FOC:
$$\frac{1}{P} \sum_{t=R}^{T} \left[g_\alpha(\varepsilon_{t+1}) \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)} \right) \right] = 0.$$
- If $\lambda = 0$, then $\varepsilon_{t+1} = \hat{e}_{t+1}^{(1)}$. Hence
$$\hat{B}_P = \frac{1}{P} \sum_{t=R}^{T} \left[g_\alpha \left(\hat{e}_{t+1}^{(1)} \right) \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)} \right) \right] \xrightarrow{p} 0$$

Three statistics

1. $DM_P \equiv \hat{S}_P^{-1/2}\sqrt{P}\hat{D}_P$, where

$$\hat{D}_P = \frac{1}{P}\sum_{t=R}^{T}\left[g\left(\hat{e}_{t+1}^{(1)}\right)\hat{e}_{t+1}^{(1)} - g\left(\hat{e}_{t+1}^{(2)}\right)\hat{e}_{t+1}^{(2)}\right]$$

2. $ENC_P \equiv \hat{Q}_P^{-1/2}\sqrt{P}\hat{B}_P$, where

$$\hat{B}_P \equiv \hat{D}_P + \hat{A}_P = \frac{1}{P}\sum_{t=R}^{T}\left[g\left(\hat{e}_{t+1}^{(1)}\right)\left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right)\right]$$

3. $CCS_P \equiv \hat{W}_P^{-1/2}\sqrt{P}\hat{M}_P$, where

$$\hat{M}_P \quad = \quad P^{-1}\sum_{t=R}^{T} g\left(\hat{e}_{t+1}^{(1)}\right)x_t$$

$\hat{S}_P, \hat{Q}_P, \hat{W}_P$ are the consistent estimators of $S_P = var\left(\sqrt{P}\hat{D}_P\right), Q_P = var\left(\sqrt{P}\hat{B}_P\right)$, $W_P = var\left(\sqrt{P}\hat{M}_P\right)$.

4. Asymptotic distribution of ENC with a stationary predictor

Assumptions

Let $x_t = \phi x_{t-1} + v_t$.
Assumption 1a. $\{x_t\}$ is weakly stationary with $|\phi| < 1$.
Assumption 1b. $\{x_t\}$ has the AR root local to unity, $\phi = 1 - c/T, c \geq 0$.

Let $\pi = \lim_{P,R\to\infty} P/R$ and $\xi = R/T = R/\left(P+R\right)$.
Assumption 2a. $0 < \pi < \infty$.
Assumption 2b. $\pi = 0$ (or $\xi \to 1$).
Assumption 2c. $\pi = \infty$ (or $\xi \to 0$).

Proposition 1 (CM 2001). Under Assumption 1a and Assumption 2a ($0 < \pi < \infty$),

$$ENC_P \Rightarrow \frac{\int_\xi^1 \xi^{-1} \left[W(s) - W(s-\xi)\right] \mathrm{d}W(s)}{\sqrt{\int_\xi^1 \xi^{-2} \left[W(s) - W(s-\xi)\right]^2 \mathrm{d}s}},$$

under $\mathbb{H}_0$, where $W(s)$ is a Wiener process and $s \in [0,1]$. Under Assumption 2a, RHS is *not* standard normal.

Proposition 2 (CM 2001). Under Assumption 1a and Assumption 2b ($\pi = 0, \xi \to 1$),

$$ENC_P \Rightarrow N\left(0,1\right),$$

under $\mathbb{H}_0$.

Proposition 3 (GLM 2015). Under Assumption 1a and Assumption 2c $(\pi = \infty, \xi \to 0)$,

$$ENC_P \Rightarrow N\left(0,1\right),$$

under $\mathbb{H}_0$.

5. Asymptotic distribution of ENC with a persistent predictor

$$\begin{aligned} y_{t+1} &= bx_t + e_{t+1} \\ x_{t+1} &= \phi x_t + v_{t+1} \end{aligned}$$

Stationary process: $|\phi| < 1$

$$T^{-1} \sum_{t=1}^{T} x_t^2 \xrightarrow{p} \frac{\sigma_v^2}{1-\phi^2}$$

$$T^{-0.5} \sum_{t=1}^{T} x_t v_{t+1} \Rightarrow N\left(0, \frac{\sigma_v^2}{1-\phi^2}\right)$$

Local to unit root process[1]: $\phi = 1 - c/T$

- Let $t = [Tr]$, $r \in [0\ 1]$. Then

$$x_{[Tr]}/\sqrt{T} \Rightarrow J_x^c(r) = \int_0^r e^{(r-s)c} dB_x(s)$$

where $J_x^c(r)$ is an Onstein-Uhlenbeck process and $B_x(s)$ is a Brownian motion,

$$T^{-2} \sum_{t=1}^{T} x_t^2 \Rightarrow \int_0^1 J_x^c(r)^2 \, dr,$$

$$T^{-1} \sum_{t=1}^{T} x_t v_{t+1} \Rightarrow \int_0^1 J_x^c(r) \, dB_x(r).$$

[1] Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987), Stock (1991), Choi (1993), Elliott and Stock (1994), Cavanagh, Elliott, and Stock (1995), Toda and Yamamoto (1995), Dolado and Lutkepohl (1996), Park (2003), Amihud and Hurvich (2004), Torus, Valkanov, Yan (2004), Giraitis and Phillips (2006), Campbell and Yogo (2006), Jansson and Moreira (2006), Mikusheva (2007, 2014), Phillips and Magdalinos (2007), Phillips and Lee (2013), Phillips and Chen (2014), Cai and Wang (2014), Breitung and Demetrescu (2015), among others.

- Let $t \equiv [Ts]$ and $\xi \equiv R/T$.
  Then we have $t/T \to s$ and $(t - R + 1)/T \to (s - \xi)$.

$$T^{-2} \sum_{j=t-R+1}^{t} x_j^2 \;\Rightarrow\; \int_{s-\xi}^{s} J_x^c(r)^2 \, dr,$$

$$T^{-1} \sum_{j=t-R+1}^{t} x_j v_{j+1} \;\Rightarrow\; \int_{s-\xi}^{s} J_x^c(r) \, dB_x(r), \qquad t = R, \ldots, T,$$

Now we state the key result for $\sqrt{P}\hat{B}_P$.

Proposition 4. Under Assumption 1b and Assumption 2c $(\pi = \infty, \xi \to 0)$, we have

$$P^{-1}\sum_{t=R}^{T}\hat{e}_{t+1}^{(1)}\left(\hat{e}_{t+1}^{(1)}-\hat{e}_{t+1}^{(2)}\right)=P^{-1}\sum_{t=R}^{T}-e_{t+1}\left(e_{t+1}-\hat{e}_{t+1}^{(1)}\right)+o_p\left(1\right)$$

under $\mathbb{H}_0$.

Proof of Proposition 4: Under $\mathbb{H}_0$, $e_{t+1}^{(1)} = e_{t+1}^{(2)} =: e_{t+1}$. Note that $\hat{e}_{t+1}^{(i)} = e_{t+1} - x_{i,t}' \left(\hat{\beta}_{i,t} - \beta_i\right)$. Then

$$\begin{aligned}
\sum_{t=R}^{T} \hat{e}_{t+1}^{(1)} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right) &= \sum_{t=R}^{T} e_{t+1} \left[-x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_1\right)\right] \\
&\quad + \sum_{t=R}^{T} e_{t+1} \left[x_{2,t}' \left(\hat{\beta}_{2,t} - \beta_2\right)\right] \\
&\quad + \sum_{t=R}^{T} \left(\hat{\beta}_{1,t} - \beta_1\right) x_{1,t} x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_1\right) \\
&\quad - \sum_{t=R}^{T} \left(\hat{\beta}_{1,t} - \beta_1\right) x_{1,t} x_{2,t}' \left(\hat{\beta}_{2,t} - \beta_2\right) \\
&\equiv A_1 + A_2 + A_3 + A_4.
\end{aligned}$$

Proposition 4 follows from Lemmas 1, 2, 3 below which show that

$$A_1 + A_2 + (A_3 + A_4) = O_p\left(\frac{T}{R}\right) + O_p\left(\frac{P}{T}\right) + O_p\left(T^{-0.5}\right).$$

Lemma 1: Under Assumption 2c ($\pi = \infty, \xi \to 0$), under $\mathbb{H}_0$,
$A_1 = \sum_{t=R}^{T} e_{t+1} \left[-x'_{1,t} \left(\hat{\beta}_{1,t} - \beta_1 \right) \right] = O_p \left(\frac{T}{R} \right)$.

Proof of Lemma 1:

$$\begin{aligned}
A_1 &= \sum_{t=R}^{T} e_{t+1} \left[-x'_{1,t} \left(\hat{\beta}_{1,t} - \beta_1 \right) \right] \\
&= -\sum_{t=R}^{T} e_{t+1} \left(e_{t+1} - \hat{e}_{t+1}^{(1)} \right) \\
&= -\sum_{t=R}^{T} e_{t+1} \left(R^{-1} \sum_{j=t-R+1}^{t} e_j \right) \\
&= -\sum_{t=R}^{T} e_{t+1} \left(T^{-1} \sum_{j=t-R+1}^{t} e_j \right) / \xi \\
&= -\sum_{t=R}^{T} \left[\frac{1}{\sqrt{T}} e_{t+1} \left(T^{-1/2} \sum_{j=1}^{t} e_j - T^{-1/2} \sum_{j=1}^{t-R} e_j \right) \right] / \xi \\
&\Rightarrow -\frac{\sigma_e^2}{\xi} \int_{\xi}^{1} \left[W(s) - W(s - \xi) \right] \mathrm{d}V_e(s) = O_p \left(\frac{1}{\xi} \right) = O_p \left(\frac{T}{R} \right)
\end{aligned}$$

Lemma 2. Under Assumption 2c ($\pi = \infty$), under $\mathbb{H}_0$,
$A_2 = \sum_{t=R}^{T} e_{t+1} \left[x'_{2,t} \left(\hat{\beta}_{2,t} - \beta_{2,t} \right) \right] = O\left(\frac{P}{T}\right)$.

Proof of Lemma 2: Let $G_T = \text{diag}\left(T^{0.5}, T\right)$.

$$
\begin{aligned}
A_2 &= \sum_{t=R}^{T} e_{t+1} x'_{2,t} \left(\hat{\beta}_{2,t} - \beta_2 \right) \\
&= \sum_{t=R}^{T} e_{t+1} x'_{2,t} \left(\sum_{j=t-R}^{t-1} x_{2,j} x'_{2,j} \right)^{-1} \left(\sum_{j=t-R}^{t-1} x_{2,j} e_{j+1} \right) \\
&= \sum_{t=R}^{T} e_{t+1} x'_{2,t} G_T^{-1} \left[G_T^{-1} \sum_{j=t-R+1}^{t} x_{2,j} x'_{2,j} G_T^{-1} / \xi \right]^{-1} \\
&\quad \times \left[G_T^{-1} \sum_{j=t-R+1}^{t} x_{2,j} e_{j+1} / \xi \right],
\end{aligned}
$$

where

$$G_T^{-1}\left(\sum_{j=t-R+1}^{t} x_{2,j}x_{2,j}'\right)G_T^{-1}/\xi$$
$$\Rightarrow \left(\begin{array}{cc} \xi & \int_{s-\xi}^{s} J_x^c(r)\,\mathrm{d}r \\ \int_{s-\xi}^{s} J_x^c(r)\,\mathrm{d}r & \int_{s-\xi}^{s} (J_x^c(r))^2\,\mathrm{d}r \end{array}\right)/\xi = O_p(1),$$

$$G_T^{-1}\sum_{j=t-R+1}^{t} x_{2,j}e_{j+1}/\xi \Rightarrow \left(\begin{array}{c} \int_{s-\xi}^{s} 1\mathrm{d}V_e(r) \\ \int_{s-\xi}^{s} J_x^c(r)\,\mathrm{d}V_e(r) \end{array}\right)/\xi = O_p(1)$$

where $J_x^c(r)$ is an OU process, and $V_e(r)$ is a Wiener process.

Therefore

$$
\begin{aligned}
A_2 &= \sum_{t=R}^{T} e_{t+1} x'_{2,t} G_T^{-1} \left[G_T^{-1} \sum_{j=t-R+1}^{t} x_{2,j} x'_{2,j} G_T^{-1} / \xi \right]^{-1} \\
&\quad \times \left[G_T^{-1} \sum_{j=t-R+1}^{t} x_{2,j} e_{j+1} / \xi \right] \\
&\Rightarrow \int_{\xi}^{1} \begin{pmatrix} 1 & J_x^c(s) \end{pmatrix} \begin{pmatrix} \xi & \int_{s-\xi}^{s} J_x^c(r)\, \mathrm{d}r \\ \int_{s-\xi}^{s} J_x^c(r)\, \mathrm{d}r & \int_{s-\xi}^{s} J_x^c(r)^2\, \mathrm{d}r \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} \int_{s-\xi}^{s} 1 \mathrm{d}V_e(r) \\ \int_{s-\xi}^{s} J_x^c(r)\, \mathrm{d}V_e(r) \end{pmatrix} \mathrm{d}V_e(r) \\
&= \int_{\xi}^{1} \begin{pmatrix} 1 & J_x^c(s) \end{pmatrix} \begin{pmatrix} O_p(\xi) & O_p(\xi) \\ O_p(\xi) & O_p(\xi) \end{pmatrix}^{-1} \begin{pmatrix} O_p(\xi) \\ O_p(\xi) \end{pmatrix} \mathrm{d}V_e(r) \\
&= O_p(1)
\end{aligned}
$$

Lemma 3. Under Assumption 2c ($\pi = \infty$),

$$A_3 + A_4 = \sum_{t=R}^{T} \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) x_{1,t} x'_{1,t} \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) - \sum_{t=R}^{T} \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) x_{1,t} x'_{2,t} \left(\hat{\beta}_{2,t} - \beta_{2,t}\right)$$

is $o_p(1)$ under $\mathbb{H}_0$.

Proof of Lemma 3: Let $B_i(t) = \left(R^{-1}\sum_{j=t-R+1}^{t} x_{i,j}x_{i,j}^{'}\right)^{-1}$, $H_i(t) = R^{-1}\sum_{j=t-R+1}^{t} x_{i,j}e_{j+1}^{(i)}$, and $\hat{\beta}_{i,t} - \beta_i = B_i(t)\, H_i(t)$. Let $B_i = \left(\mathbb{E}x_{i,t}x_{i,t}^{'}\right)^{-1}$. Then

$$
\begin{aligned}
A_3 + A_4 &= \sum_{t=R}^{T}\left(\hat{\beta}_{1,t} - \beta_1\right) x_{1,t}x_{1,t}'\left(\hat{\beta}_{1,t} - \beta_1\right) \\
&\quad - \sum_{t=R}^{T}\left(\hat{\beta}_{1,t} - \beta_1\right) x_{1,t}x_{2,t}'\left(\hat{\beta}_{2,t} - \beta_2\right) \\
&\equiv \sum_{t=R}^{T} H_1'(t)\, B_1(t) q_{1,t} B_1(t) H_1(t) \\
&\quad - \sum_{t=R}^{T} H_1'(t)\, B_1(t) x_{1,t}x_{2,t}' B_2(t) H_2(t),
\end{aligned}
$$

Let

$$\begin{aligned} E_T &= \text{diag}\left(T^0, T^{0.5}\right) \\ F_T &= \text{diag}\left(T^1, T^{1.5}\right) \\ G_T &= \text{diag}\left(T^{0.5}, T^1\right) \end{aligned}$$

Let

$$\begin{aligned} \ddot{x}'_{2,t} &\equiv x'_{2,t} E_T^{-1} \\ \ddot{B}_2(t) &\equiv E_T \times R^{-1} B_2(t) \times F_T \times \xi \\ \ddot{H}_2(t) &\equiv F_T^{-1} \times R H_2(t) / \xi \end{aligned}$$

Then

$$\begin{aligned} & \sum_{t=R}^{T} H_1'(t) B_1(t) x_{1,t} x'_{2,t} B_2(t) H_2(t) \\ = & \sum_{t=R}^{T} H_1'(t) B_1(t) x_{1,t} \ddot{x}'_{2,t} \ddot{B}_2(t) \ddot{H}_2(t) \end{aligned}$$

$$\ddot{x}'_{2,t} \equiv x'_{2,t} E_T^{-1} \Rightarrow \begin{pmatrix} 1 & J_x^c(r) \end{pmatrix} = O_p(1)$$

If $E_T F_T = G_T G_T$ then $E_T \times K \times F_T = G_T \times K \times G_T$ for any 2×2 matrix K.

$$\begin{aligned}
\ddot{B}_2(t) &\equiv E_T \times R^{-1} B_2(t) \times F_T \times \xi \\
&= G_T \left[R^{-1} B_2(t) \right] G_T \times \xi \\
&= \left[G_T^{-1} \left[R^{-1} B_2(t) \right]^{-1} G_T^{-1} \right]^{-1} \times \xi \\
&= \left[G_T^{-1} \sum_{j=t-R+1}^{t} x_{2,j} x'_{2,j} G_T^{-1} \right]^{-1} \times \xi \\
&\Rightarrow \begin{pmatrix} \xi & \int_{s-\xi}^{s} J_x^c(r)\, dr \\ \int_{s-\xi}^{s} J_x^c(r)\, dr & \int_{s-\xi}^{s} \left(J_x^c(r)\right)^2 dr \end{pmatrix}^{-1} \times \xi = O_p(1)
\end{aligned}$$

$$
\begin{aligned}
\ddot{H}_2(t) &\equiv F_T^{-1} \times RH_2(t)/\xi \\
&= F_T^{-1} \sum_{j=t-R+1}^{t} x_{2,j} e_{j+1}/\xi \\
&= \begin{pmatrix} T^{-1}/\xi \times \sum_{j=t-R+1}^{t} e_{j+1} \\ T^{-1.5}/\xi \times \sum_{j=t-R+1}^{t} x_j e_{j+1} \end{pmatrix} \\
&= \begin{pmatrix} T^{-0.5}/\xi \times T^{-0.5} \sum_{j=t-R+1}^{t} e_{j+1} \\ T^{-0.5}/\xi \times T^{-1} \sum_{j=t-R+1}^{t} x_j e_{j+1} \end{pmatrix} \\
&\Rightarrow \begin{pmatrix} T^{-0.5}/\xi \times \int_{s-\xi}^{s} 1 \mathsf{d} V_e(r) \\ T^{-0.5}/\xi \times \int_{s-\xi}^{s} J_x^c(r) \mathsf{d} V_e(r) \end{pmatrix} \\
&= \begin{pmatrix} O_p\left(T^{-0.5}/\xi\right) \times O_p(\xi) \\ O_p\left(T^{-0.5}/\xi\right) \times O_p(\xi) \end{pmatrix} = O_p\left(T^{-0.5}\right).
\end{aligned}
$$

Hence $\ddot{x}'_{2,t}, \ddot{B}_2(t), \ddot{H}_2\left(t\right)$ have the same orders of magnitude as $x'_{2,t}, B_2(t), H_2\left(t\right)$ in stationary case. Therefore $A_3 + A_4$ is $o_p\left(1\right)$

Proposition 5. Under Assumption 1b and Assumption 2c ($\pi = \infty, \xi \to 0$), $\lim_{\xi \to 0} ENC_P \Rightarrow N(0,1)$ under $\mathbb{H}_0$.

Proof of Proposition 5: From Proposition 2 and Lemma 1, $A_1 = O_p\left(\xi^{-1}\right)$ is the dominant term in $\hat{B}_P$ and hence ENC_P is

$$
\begin{aligned}
ENC_P &= A_1/\sqrt{\mathsf{var}\left(A_1\right)} + o_p\left(1\right) \\
&= \frac{-\sum_{t=R}^{T} e_{t+1}\left(e_{t+1} - \hat{e}_{t+1}^{(1)}\right)}{\sqrt{\sum_{t=R}^{T}\left[-e_{t+1}\left(e_{t+1} - \hat{e}_{t+1}^{(1)}\right) - \hat{c}_P\right]^2}} + o_p\left(1\right) \\
&\Rightarrow \lim_{\xi \to 0} \frac{-\sigma_e^2 \xi^{-1} \int_\xi^1 \left[W(s) - W(s-\xi)\right] \mathsf{d}V_e(s)}{\sqrt{\sigma_e^4 \times \xi^{-2} \int_\xi^1 \left[W(s) - W(s-\xi)\right]^2 \mathsf{d}s}} \sim N\left(0,1\right),
\end{aligned}
$$

where $A_1 \Rightarrow -\sigma_e^2 \xi^{-1} \int_\xi^1 [V_e(s) - V_e(s-\xi)] \mathsf{d}V_e(s)$, $c_{t+1} = -e_{t+1}\left(e_{t+1} - \hat{e}_{t+1}^{(1)}\right)$ and $\hat{c}_P = P^{-1}\sum_{t=R}^{T} c_{t+1} = P^{-1}A_1$. The denominator follows from Lemma 4.

We divide $[0,1]$ to n equal segments, let $t=[sn]$, $s\in[0,1]$, $\xi=1/n$.
Let $\{u_i\}_{i=1}^{n}$ be a mixing sequence with $E\left(u\right)=0$ and $\mathrm{var}(u)=1$.

$$n^{-1}\sum_{t=1}^{n}u_{t-1}u_t\Rightarrow\int_{\xi}^{1}\left[W(s)-W(s-\xi)\right]\mathrm{d}W(s).$$

$$n^{-2}\sum_{t=1}^{n}u_{t-1}^{2}\Rightarrow\int_{\xi}^{1}\left[W(s)-W(s-\xi)\right]^{2}\mathrm{d}s.$$

By using a CLT,

$$\frac{n^{-1}\sum_{t=1}^{n}u_{t-1}u_t}{\sqrt{n^{-2}\sum_{t=1}^{n}u_{t-1}^{2}}}\Rightarrow N\left(0,1\right).$$

Therefore, $\lim_{\xi\to 0}\frac{-\sigma_e^2\xi^{-1}\int_{\xi}^{1}[W(s)-W(s-\xi)]\mathrm{d}V_e(s)}{\sqrt{\sigma_e^4\times\xi^{-2}\int_{\xi}^{1}[W(s)-W(s-\xi)]^2\mathrm{d}s}}\sim N\left(0,1\right)$ under $\mathbb{H}_0$.

Lemma 4. $\sum_{t=R}^{T}\left[-e_{t+1}\left(e_{t+1}-\hat{e}_{t+1}^{(1)}\right)-\hat{c}_P\right]^2 \Rightarrow$
$\sigma_e^4 \times \xi^{-2}\int_{\xi}^{1}[W(s)-W(s-\xi)]^2 \mathsf{d}s.$

Proof of Lemma 4:

$$
\begin{aligned}
& \sum_{t=R}^{T}\left[-e_{t+1}\left(e_{t+1}-\hat{e}_{t+1}^{(1)}\right)-\hat{c}_P\right]^2 \\
= & \sum_{t=R}^{T}\left[e_{t+1}\left(e_{t+1}-\hat{e}_{t+1}^{(1)}\right)\right]^2-P\hat{c}_P^2 \\
= & \sum_{t=R}^{T}\left[e_{t+1}\left(R^{-1}\sum_{j=t-R+1}^{t}e_j\right)\right]^2+O_p\left(\frac{1}{P\xi^2}\right) \\
= & \frac{T^2}{R^2}\sum_{t=R}^{T}\left[\left(e_{t+1}\right)^2\left(\frac{1}{\sqrt{T}}\sum_{j=1}^{t}e_j-\frac{1}{\sqrt{T}}\sum_{j=1}^{t-R}e_j\right)^2\right]\frac{1}{T}+O_p\left(\frac{1}{P\xi^2}\right) \\
\Rightarrow & \ \xi^{-2}\int_{\xi}^{1}\sigma_e^2\left[\sigma_e W(s)-\sigma_e W(s-\xi)\right]^2 \mathrm{d}s,
\end{aligned}
$$

where line 3 follows from Lemma 1 for $P\hat{c}_P = A_1 = O_p\left(\xi^{-1}\right)$, where $A_1 \Rightarrow -\sigma_e^2\xi^{-1}\int_{\xi}^{1}\left[V_e(s)-V_e(s-\xi)\right]\mathrm{d}V_e(s)$.

6. Monte Carlo

Monte-Carlo Simulation

$$\text{DGP} : y_{t+1} = a + bx_t + e_{t+1}$$

$a = 1$
$b \in \left\{0, \frac{1}{10}\right\}$
$e_{t+1} \sim IID\ N\left(0, \sigma_e^2\right), \sigma_e \in \left\{\frac{1}{10}, 1\right\}$
$x_{t+1} = \phi x_t + v_{t+1},\ \phi \in \{0, 0.5, 0.9, 0.95, 0.99, 1\}, v_{t+1} \sim IID$
$N(0, 1)$
$R \in \{60, 120, 240\},\ P \in \{48, 240, 1200\},\ R + P = T + 1$
2000 replications

$$\text{Model 1} \quad : \quad y_{t+1} = a_1 + e_{t+1}^{(1)}$$
$$\text{Model 2} \quad : \quad y_{t+1} = a_2 + bx_t + e_{t+1}^{(2)}$$

Table: Rejection frequency under 5% level, $b = 0$ (With intercept on small model)

		$P = 48$			$P = 1200$		
		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
$\phi = 0$	$R = 60$	0.005	0.027	0.044	0.000	0.030	0.045
$\sigma_e = 1$	$R = 120$	0.016	0.037	0.050	0.000	0.038	0.047
	$R = 240$	0.029	0.046	0.064	0.000	0.031	0.057
$\phi = 0.5$	$R = 60$	0.010	0.038	0.060	0.000	0.044	0.042
$\sigma_e = 1$	$R = 120$	0.019	0.036	0.051	0.000	0.039	0.058
	$R = 240$	0.023	0.039	0.056	0.000	0.035	0.049
$\phi = 0.9$	$R = 60$	0.007	0.035	0.050	0.000	0.048	0.055
$\sigma_e = 1$	$R = 120$	0.016	0.035	0.057	0.000	0.037	0.056
	$R = 240$	0.023	0.040	0.051	0.000	0.036	0.058
$\phi = 0.95$	$R = 60$	0.005	0.028	0.041	0.000	0.040	0.030
$\sigma_e = 1$	$R = 120$	0.013	0.033	0.058	0.000	0.047	0.046
	$R = 240$	0.017	0.031	0.059	0.000	0.034	0.048
$\phi = 0.99$	$R = 60$	0.003	0.033	0.043	0.000	0.045	0.004
$\sigma_e = 1$	$R = 120$	0.010	0.032	0.052	0.000	0.040	0.018
	$R = 240$	0.010	0.032	0.052	0.000	0.040	0.018
$\phi = 1$	$R = 60$	0.003	0.030	0.039	0.000	0.037	0.001
$\sigma_e = 1$	$R = 120$	0.012	0.029	0.054	0.000	0.037	0.002
	$R = 240$	0.028	0.048	0.063	0.000	0.033	0.006

Table: Rejection frequency under 5% level, $b = 0.1$ (With intercept on small model)

		$P = 48$			$P = 1200$		
		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
$\phi = 0$	$R = 60$	0.023	0.107	0.121	0.001	0.601	0.935
$\sigma_e = 1$	$R = 120$	0.049	0.124	0.106	0.021	0.728	0.918
	$R = 240$	0.070	0.142	0.111	0.129	0.823	0.927
$\phi = 0.5$	$R = 60$	0.032	0.122	0.123	0.002	0.751	0.970
$\sigma_e = 1$	$R = 120$	0.078	0.311	0.260	0.511	1.000	1.000
	$R = 240$	0.081	0.164	0.120	0.238	0.923	0.976
$\phi = 0.9$	$R = 60$	0.078	0.311	0.260	0.511	1.000	1.000
$\sigma_e = 1$	$R = 120$	0.133	0.390	0.293	0.936	1.000	1.000
	$R = 240$	0.170	0.445	0.322	0.993	1.000	1.000
$\phi = 0.95$	$R = 60$	0.110	0.441	0.328	0.926	1.000	1.000
$\sigma_e = 1$	$R = 120$	0.220	0.571	0.440	1.000	1.000	1.000
	$R = 240$	0.255	0.586	0.455	1.000	1.000	1.000
$\phi = 0.99$	$R = 60$	0.252	0.623	0.400	1.000	1.000	0.999
$\sigma_e = 1$	$R = 120$	0.434	0.770	0.533	1.000	1.000	1.000
	$R = 240$	0.519	0.824	0.615	1.000	1.000	1.000
$\phi = 1$	$R = 60$	0.286	0.679	0.589	1.000	1.000	0.988
$\sigma_e = 1$	$R = 120$	0.529	0.817	0.693	1.000	1.000	0.999
	$R = 240$	0.643	0.881	0.726	1.000	1.000	0.996

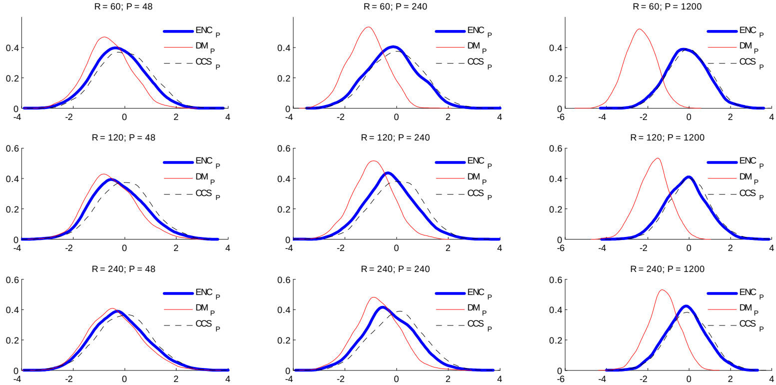

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0$, $b = 0$, $\sigma_e = 1$. 2000 Repeats.

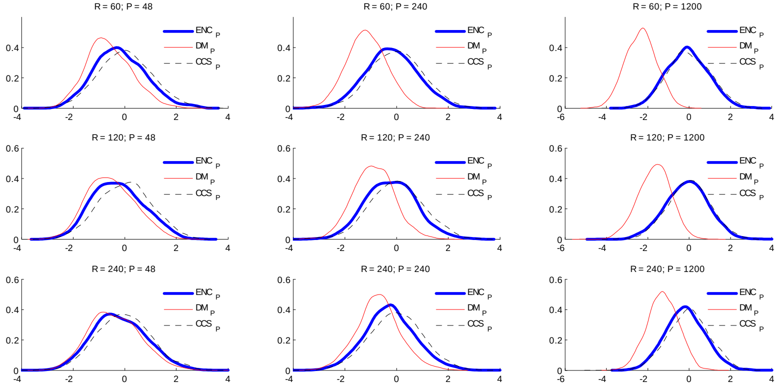

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0$, $b = 0$, $\sigma_e = 0.1$. 2000 Repeats.

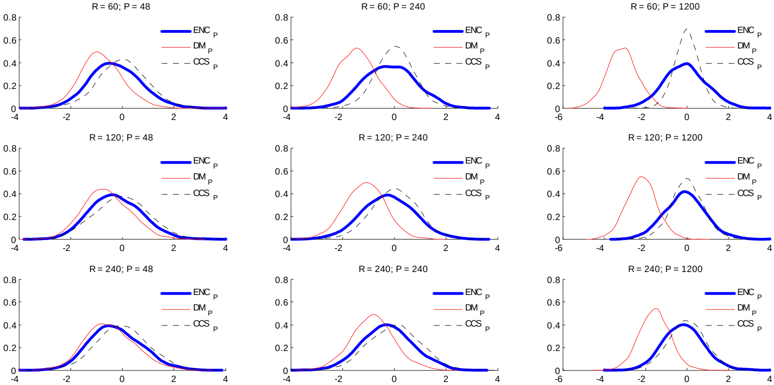

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0.99$, $b = 0$, $\sigma_e = 1$. 2000 Repeats.

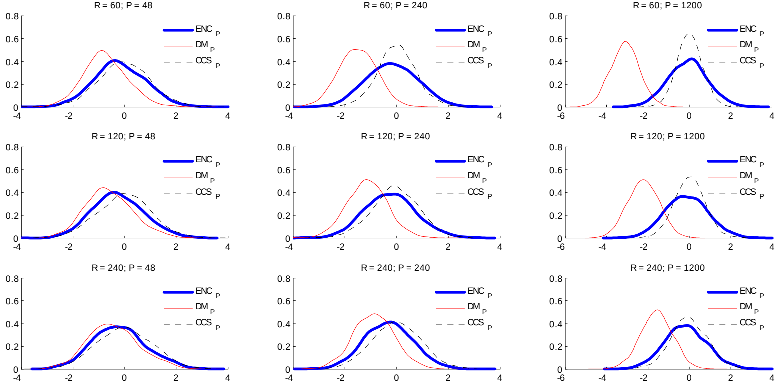

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0.99$, $b = 0$, $\sigma_e = 0.1$. 2000 Repeats.

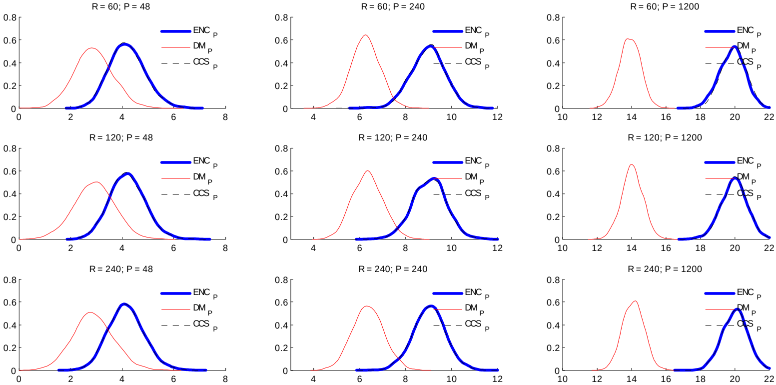

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 0.1$, $\sigma_e = 0.1$. 2000 Repeats.

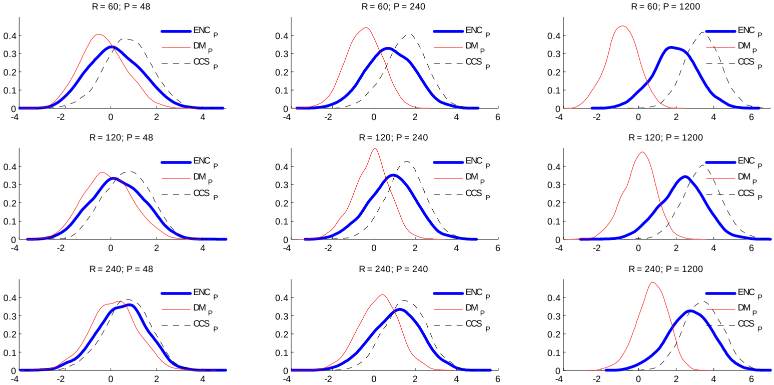

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 0.1$, $\sigma_e = 1$. 2000 Repeats.

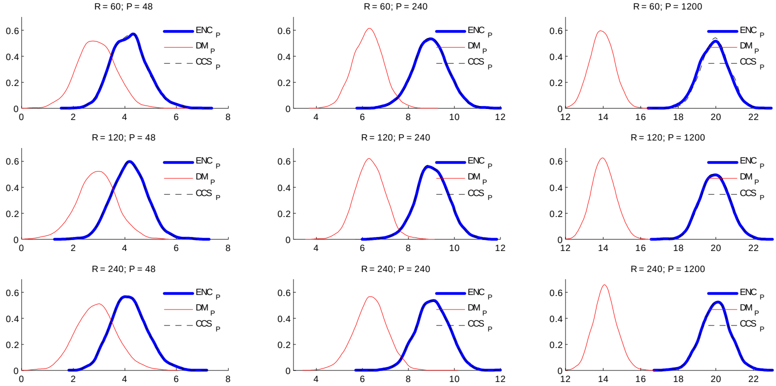

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 1$, $\sigma_e = 1$. 2000 Repeats.

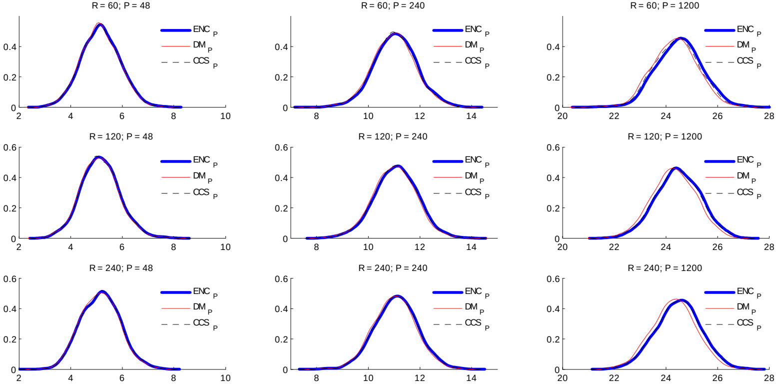

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 1$, $\sigma_e = 0.1$. 2000 Repeats.

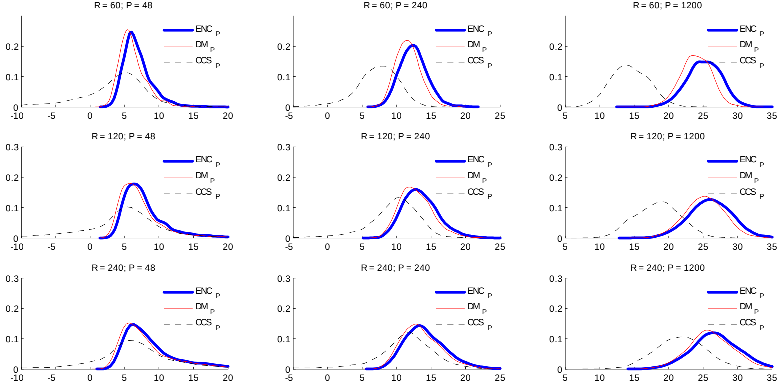

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 0.1$, $\sigma_e = 0.1$. 2000 Repeats.

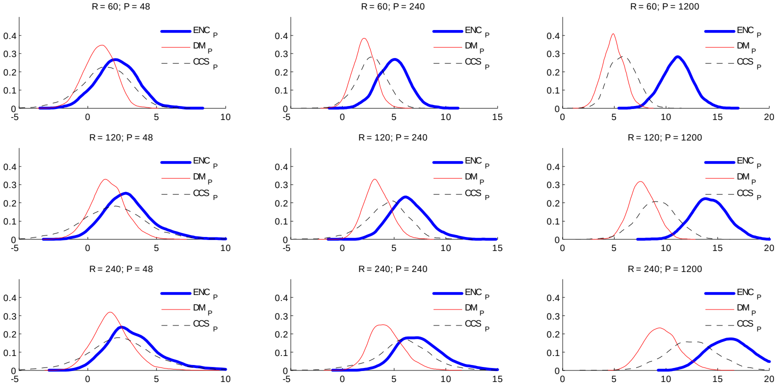

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 0.1$, $\sigma_e = 1$. 2000 Repeats.

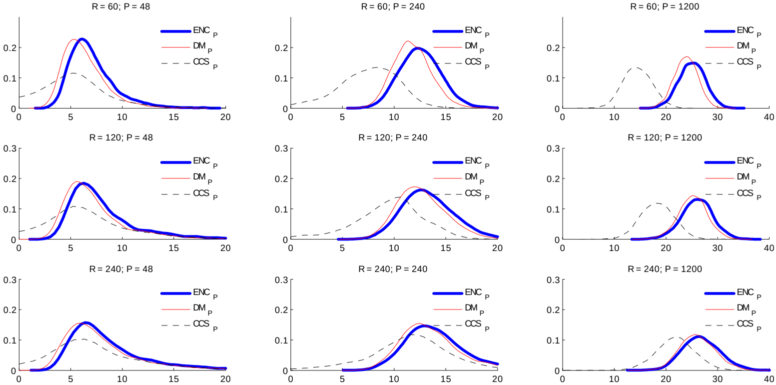

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 1$, $\sigma_e = 1$. 2000 Repeats.

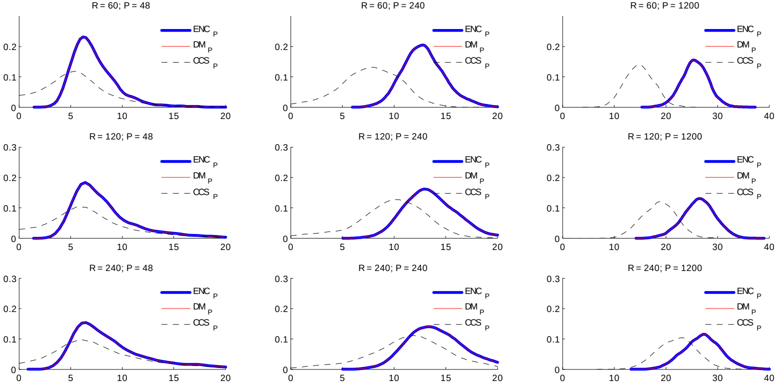

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 1$, $\sigma_e = 0.1$. 2000 Repeats.

7. Application

Predictive mean regression for equity premium

$$\text{Model 1} \quad : \quad y_{t+1} = a_1 + e_{t+1}^{(1)}$$
$$\text{Model 2} \quad : \quad y_{t+1} = a_2 + bx_t + e_{t+1}^{(2)}$$

- Data: Goyal and Welch (2008), Monthly, 1926:01 to 2011:12, $T = 1032$.
- $y_{t+1} =$ equity premium
- $x_t =$ persistent predictors such as dividend-yield ratio (DY), dividend-price ratio (DP), long term rate of yield (LTY), and inflation (INFL).
- Rolling estimation windows of size R.
- (R, P) ranges from $(258, 774), (259, 773), (260, 772), \ldots, (773, 259), (774, 258)$.
- $\frac{P}{R}$ ranges from $\frac{774}{258} = 3$ to $\frac{258}{774} = \frac{1}{3}$.

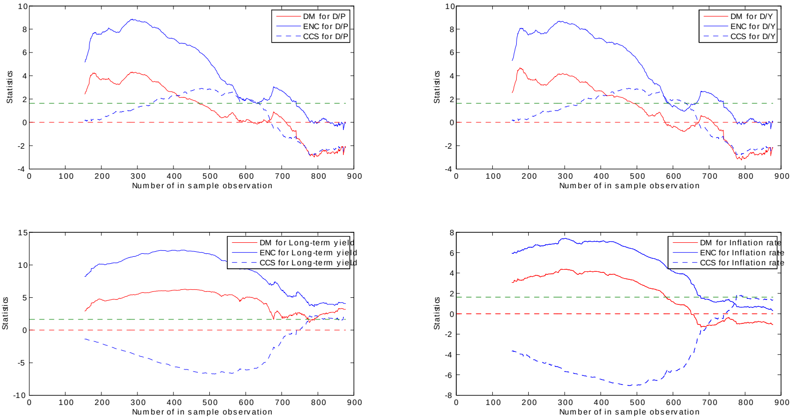

Figure: Testing for predictive ability of persistent predictors for equity premium

8. When Model 1 is the martingale difference model

Clark and West (2006) when $a_1 = 0$

$$\begin{aligned} \text{Model 1} &: \quad y_{t+1} = 0 + e_{t+1}^{(1)} \\ \text{Model 2} &: \quad y_{t+1} = a_2 + bx_t + e_{t+1}^{(2)} \end{aligned}$$

$$\begin{aligned} x_t &= \phi x_{t-1} + v_t \\ \phi &= 1 - c/R, c > 0 \end{aligned}$$

Proposition 6. Under Assumption 1b and Assumption 2c, CCS_P is asymptotically standard normal under $\mathbb{H}_0$.

Proof: Under Model 1, $\hat{e}_{t+1}^{(1)} = e_{t+1}$. Note that $\lim T/P \to 1$ as $\pi = \infty$. Therefore

$$\begin{aligned}
\hat{M}_P &= P^{-1} \sum_{t=R}^{T} e_{t+1}^{(1)} x_t \\
&\Rightarrow N\left(0, \lim T^2 \left(P^{-2} \frac{1}{T^2} \sum_{t=R}^{T} x_t^2\right) \sigma_e^2\right) \\
&= N\left(0, \sigma_e^2 \int_{\xi}^{1} J_x^c(r)^2 \, \mathsf{d}r\right)
\end{aligned}$$

$$\hat{W}_P = \frac{1}{P} \sum_{t=R}^{T} \left(e_{t+1}^{(1)} x_t - \hat{M}_P\right)^2 = \frac{1}{P} \sum_{t=R}^{T} \left(e_{t+1}^{(1)} x_t\right)^2 - \hat{M}_P^2$$

$$
\begin{aligned}
P^{-1}\hat{W}_P &= P^{-1}\left[\frac{1}{P}\sum_{t=R}^{T}\left(e_{t+1}^{(1)}x_t\right)^2 - \hat{M}_P^2\right] \\
&= \left(\frac{T^2}{P^2}\right)\sigma_e^2\frac{1}{T^2}\sum_{t=R}^{T}x_t^2 + o_p\left(1\right) \\
&\Rightarrow \sigma_e^2\int_{\xi}^{1} J_x^c\left(r\right)^2 \mathsf{d}r
\end{aligned}
$$

$$
CCS_P = \hat{W}_P^{-0.5}\sqrt{P}\hat{M}_P \Rightarrow N\left(0, \frac{\sigma_e^2\int_{\xi}^{1} J_x^c\left(r\right)^2 \mathsf{d}r}{\sigma_e^2\int_{\xi}^{1} J_x^c\left(r\right)^2 \mathsf{d}r}\right) = N\left(0,1\right)
$$

Proposition 7. Under Assumptions 1b and 2c, ENC_P is asymptotically standard normal under $\mathbb{H}_0$.

Proof: Under the null hypothesis and $\{x_{1,t}\} \equiv 0$. Therefore ENC only has A_2 term, whose limiting distribution is

$$\int_{\xi}^{1} \begin{pmatrix} 1 & J_x^c(s) \end{pmatrix} \begin{pmatrix} \xi & \int_{s-\xi}^{s} J_x^c(r)\,\mathsf{d}r \\ \int_{s-\xi}^{s} J_x^c(r)\,\mathsf{d}r & \int_{s-\xi}^{s} J_x^c(r)^2\,\mathsf{d}r \end{pmatrix}^{-1} \times \begin{pmatrix} \int_{s-\xi}^{s} 1\mathsf{d}V_e(r) \\ \int_{s-\xi}^{s} J_x^c(r)\,\mathsf{d}V_e(r) \end{pmatrix} \mathsf{d}V_e(s)$$

Monte Carlo simulations (below) show that ENC is standard normal under $\mathbb{H}_0$ (the null distribution of ENC_P is virtually the same as that of CCS_P).

Table: Rejection frequency under 5% level, $b = 0$ (Without intercept on small model)

		$P = 48$			$P = 1200$		
		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
$\phi = 0$	$R = 60$	0.004	0.032	0.051	0.000	0.048	0.046
$\sigma_e = 1$	$R = 120$	0.014	0.043	0.048	0.000	0.038	0.054
	$R = 240$	0.020	0.051	0.058	0.000	0.032	0.049
$\phi = 0.5$	$R = 60$	0.006	0.039	0.056	0.000	0.036	0.052
$\sigma_e = 1$	$R = 120$	0.009	0.034	0.058	0.000	0.040	0.047
	$R = 240$	0.017	0.036	0.061	0.000	0.034	0.056
$\phi = 0.9$	$R = 60$	0.003	0.032	0.044	0.000	0.045	0.052
$\sigma_e = 1$	$R = 120$	0.008	0.040	0.052	0.000	0.036	0.049
	$R = 240$	0.018	0.040	0.056	0.000	0.035	0.044
$\phi = 0.95$	$R = 60$	0.003	0.038	0.053	0.000	0.055	0.055
$\sigma_e = 1$	$R = 120$	0.012	0.042	0.063	0.000	0.037	0.056
	$R = 240$	0.011	0.032	0.052	0.000	0.031	0.042
$\phi = 0.99$	$R = 60$	0.002	0.027	0.054	0.000	0.045	0.052
$\sigma_e = 1$	$R = 120$	0.005	0.030	0.054	0.000	0.045	0.052
	$R = 240$	0.015	0.037	0.061	0.000	0.030	0.055
$\phi = 1$	$R = 60$	0.002	0.035	0.059	0.000	0.041	0.059
$\sigma_e = 1$	$R = 120$	0.006	0.033	0.056	0.000	0.031	0.045
	$R = 240$	0.009	0.036	0.058	0.000	0.037	0.061

Table: Rejection frequency under 5% level, $b = 0.1$
(Without intercept on small model)

		$P = 48$			$P = 1200$		
		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
$\phi = 0$	$R = 60$	0.013	0.078	0.100	0.000	0.463	0.935
$\sigma_e = 1$	$R = 120$	0.037	0.103	0.109	0.001	0.595	0.936
	$R = 240$	0.050	0.131	0.109	0.031	0.758	0.938
$\phi = 0.5$	$R = 60$	0.010	0.102	0.119	0.000	0.616	0.978
$\sigma_e = 1$	$R = 120$	0.038	0.129	0.124	0.007	0.753	0.977
	$R = 240$	0.062	0.150	0.132	0.075	0.873	0.982
$\phi = 0.9$	$R = 60$	0.052	0.298	0.342	0.119	1.000	1.000
$\sigma_e = 1$	$R = 120$	0.099	0.356	0.321	0.726	1.000	1.000
	$R = 240$	0.144	0.401	0.335	0.967	1.000	1.000
$\phi = 0.95$	$R = 60$	0.114	0.468	0.518	0.818	1.000	1.000
$\sigma_e = 1$	$R = 120$	0.194	0.552	0.519	0.995	1.000	1.000
	$R = 240$	0.228	0.577	0.495	1.000	1.000	1.000
$\phi = 0.99$	$R = 60$	0.473	0.795	0.811	1.000	1.000	1.000
$\sigma_e = 1$	$R = 120$	0.544	0.831	0.812	1.000	1.000	1.000
	$R = 240$	0.547	0.827	0.791	1.000	1.000	1.000
$\phi = 1$	$R = 60$	0.565	0.844	0.855	1.000	1.000	0.988
$\sigma_e = 1$	$R = 120$	0.701	0.917	0.899	1.000	1.000	1.000
	$R = 240$	0.804	0.935	0.921	1.000	1.000	1.000

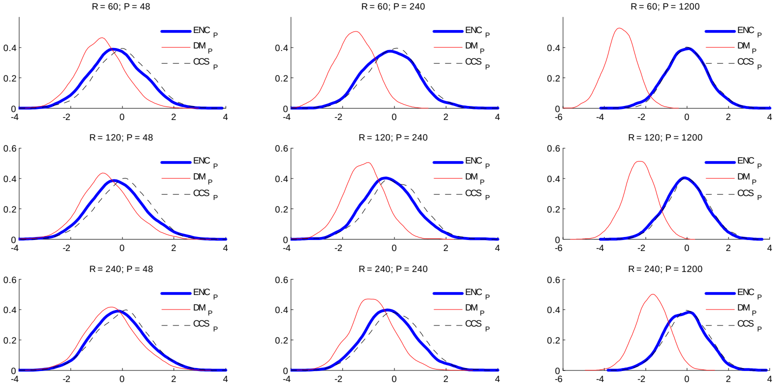

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0$, $b = 0$, $\sigma_e = 1$. No intercept on small model, 2000 Repeats.

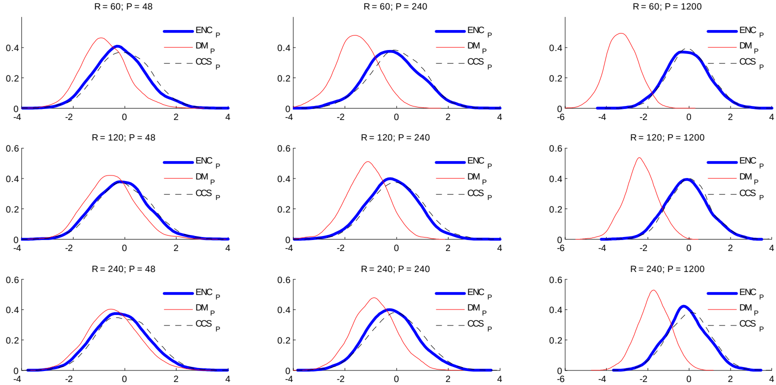

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0$, $b = 0$, $\sigma_e = 0.1$. No intercept on small model, 2000 Repeats.

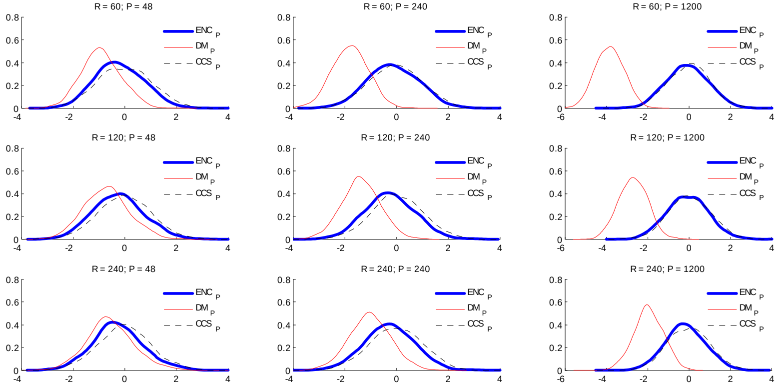

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0.99$, $b = 0$, $\sigma_e = 1$. No intercept on small model, 2000 Repeats.

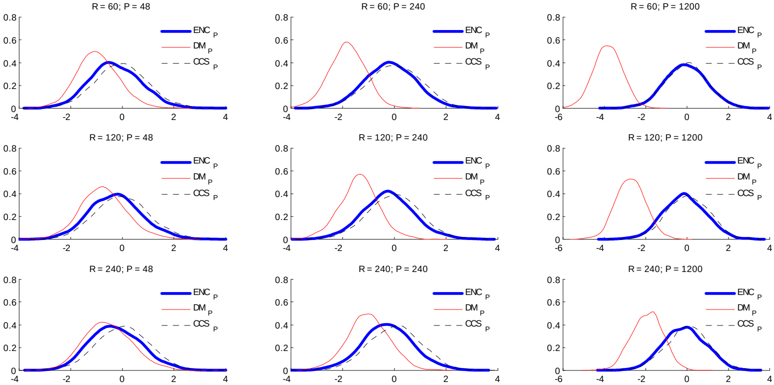

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0.99$, $b = 0$, $\sigma_e = 0.1$. No intercept on small model, 2000 Repeats.

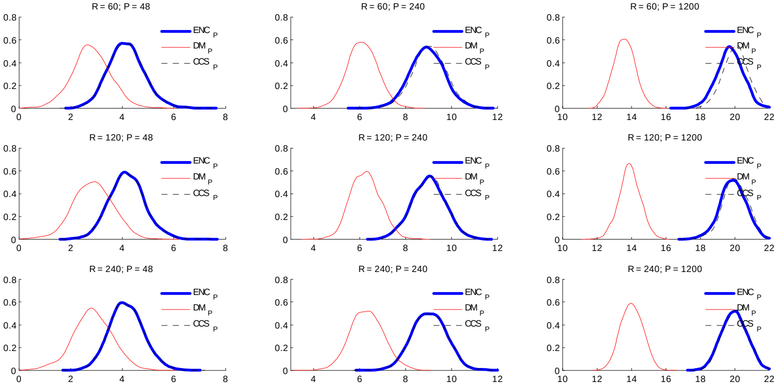

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 0.1$, $\sigma_e = 0.1$. No intercept on small model, 2000 Repeats.

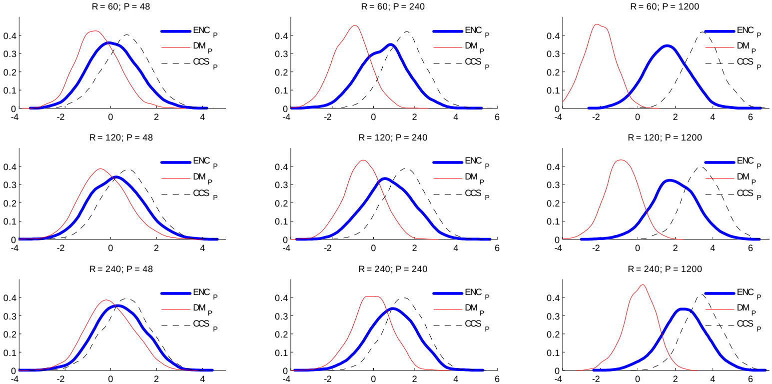

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 0.1$, $\sigma_e = 1$. No intercept on small model, 2000 Repeats.

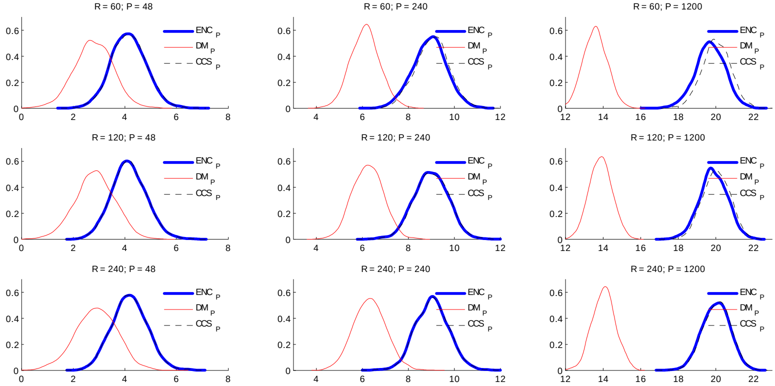

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 1$, $\sigma_e = 1$. No intercept on small model, 2000 Repeats.

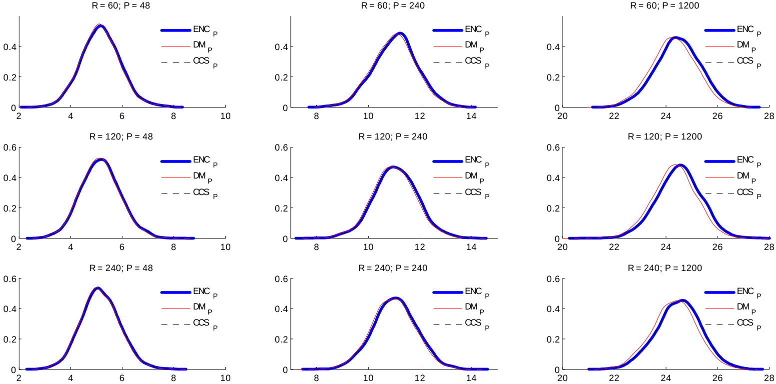

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 1$, $\sigma_e = 0.1$. No intercept on small model, 2000 Repeats.

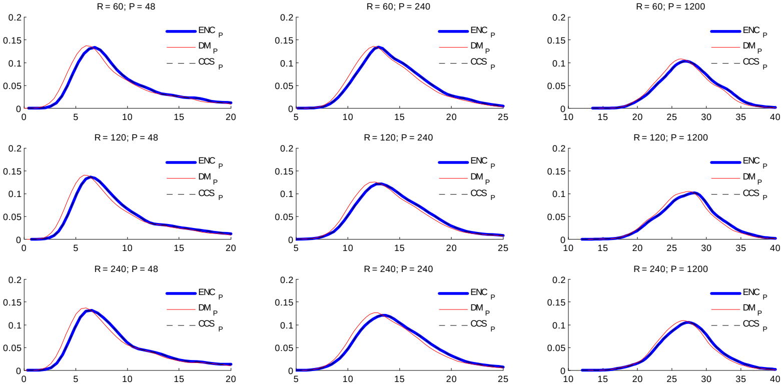

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 0.1$, $\sigma_e = 0.1$. No intercept on small model, 2000 Repeats.

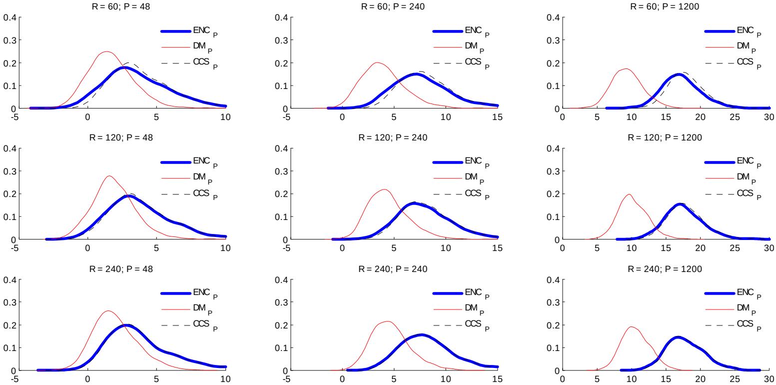

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 0.1$, $\sigma_e = 1$. No intercept on small model, 2000 Repeats.

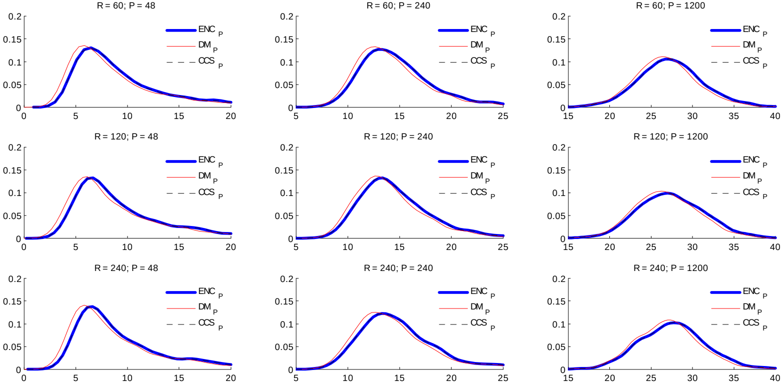

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 1$, $\sigma_e = 1$. No intercept on small model, 2000 Repeats.

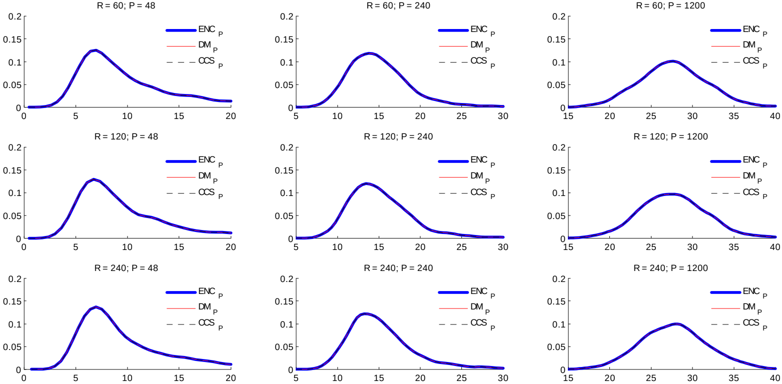

Figure: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 1$, $\sigma_e = 0.1$. No intercept on small model, 2000 Repeats.

Summary

When Model 1 with $a_1 \neq 0$

	distribution under H_0			size		
	DM	ENC	CCS	DM	ENC	CCS
stationary	not N(0,1)	N(0,1)	N(0,1)	bad	good	good
persistent	not N(0,1)	N(0,1)	not N()	bad	good	bad

When Model 1 with $a_1 = 0$

	distribution under H_0			size		
	DM	ENC	CCS	DM	ENC	CCS
stationary	not N(0,1)	N(0,1)	N(0,1)	bad	good	good
persistent	not N(0,1)	N(0,1)	N(0,1)	bad	good	good

9. When the predictor is persistent with a drift

DGP

$$\begin{aligned} y_{t+1} &= a + bx_t + e_{t+1} \\ x_{t+1} &= \mu + \phi x_t + v_{t+1} \\ \phi &= 1 - c/R^\alpha \\ \mu &= \tilde{\mu}/R^\gamma \end{aligned}$$

$a = 1$. $b = 0$.
$e_{t+1} \sim IID\ N(0,1)$
$v_{t+1} \sim IID\ N(0,1)$
$R \in \{50, 100, 200, 400\}$, $P \in \{1600\}$, $R + P = T$
2000 replications

- Unit root (UR) process: $c = 0$.
- Local to unit root (LUR) process: $c > 0$, $\alpha = 1$.
- Moderately integrated unit root (MIUR): $c > 0$, $0 < \alpha < 1$.
- Stationary: $0 < c < 1$, $\alpha = 0$.

Conergence rate of $\hat{b}$ in the predictive regression model
$y_{t+1} = a + bx_t + e_{t+1}$

x	c	α	$\tilde{\mu}$	γ	convergence rate
Stationary	> 0	0	any	NA	$R^{0.5}$
UR	0	NA	0	NA	R
LUR	> 0	1	0	NA	R
MIUR	> 0	$(0\ 1)$	0	NA	$R^{\alpha/2+0.5}$
UR+Drift	0	NA	$\neq 0$	$[0\ \infty)$	$\max\left\{R^{1.5-\gamma}, R\right\}$
LUR+Drift	> 0	1	$\neq 0$	$[0\ \infty)$	$\max\left\{R^{1.5-\gamma}, R\right\}$
MIUR+Drift	> 0	$(0\ 1)$	$\neq 0$	$[0\ \infty)$	$\max\left\{R^{\alpha-\gamma+0.5}, R^{\alpha/2+0.5}\right\}$

Simulation

Table: Rejection frequency for DM, ENC and CCS test (LUR with $\alpha = 1$)

					$P = 1600$	
	ϕ	μ	R	DM_P	ENC_P	CCS_P
$\tilde{\mu} = 3,\ b = 0,\ c_2 = 1$	0.9	3	50	0.000	0.047	0.000
$c = 5$	0.95	3	100	0.000	0.044	0.000
$\sigma_e = 1,\ \sigma_v = 1$	0.975	3	200	0.000	0.039	0.000
$\alpha = 1,\ \gamma = 0$	0.988	3	400	0.016	0.033	0.000
$\tilde{\mu} = 5,\ b = 0,\ c_2 = 1$	0.9	5	50	0.000	0.046	0.000
$c = 5$	0.95	5	100	0.000	0.044	0.000
$\sigma_e = 1,\ \sigma_v = 1$	0.975	5	200	0.000	0.038	0.000
$\alpha = 1,\ \gamma = 0$	0.988	5	400	0.000	0.032	0.000
$\tilde{\mu} = 10,\ b = 0,\ c_2 = 1$	0.9	10	50	0.000	0.046	0.000
$c = 5$	0.95	10	100	0.000	0.044	0.000
$\sigma_e = 1,\ \sigma_v = 1$	0.975	10	200	0.000	0.038	0.000
$\alpha = 1,\ \gamma = 0$	0.988	10	400	0.000	0.032	0.000

Simulation (Cont)

Table: Rejection frequency for DM, ENC and CCS test (LUR with $\alpha = 1$)

	ϕ	μ	R	DM_P	ENC_P (P = 1600)	CCS_P
$\tilde{\mu} = 3,\ b = 0,\ c_2 = 1$	0.9	0.928	50	0.000	0.046	0.000
$c = 5$	0.95	0.754	100	0.000	0.044	0.000
$\sigma_e = 1,\ \sigma_v = 1$	0.975	0.612	200	0.000	0.039	0.000
$\alpha = 1,\ \gamma = 0.3$	0.988	0.497	400	0.000	0.033	0.000
$\tilde{\mu} = 5,\ b = 0,\ c_2 = 1$	0.9	1.546	50	0.000	0.047	0.000
$c = 5$	0.95	1.256	100	0.000	0.044	0.000
$\sigma_e = 1,\ \sigma_v = 1$	0.975	1.020	200	0.000	0.039	0.000
$\alpha = 1,\ \gamma = 0.3$	0.988	0.829	400	0.000	0.033	0.000
$\tilde{\mu} = 10,\ b = 0,\ c_2 = 1$	0.9	3.092	50	0.000	0.047	0.000
$c = 5$	0.95	2.512	100	0.000	0.044	0.000
$\sigma_e = 1,\ \sigma_v = 1$	0.975	2.040	200	0.000	0.039	0.000
$\alpha = 1,\ \gamma = 0.3$	0.988	1.657	400	0.000	0.033	0.000

Simulation summary

$$
\begin{aligned}
x_t &= \mu + \phi x_{t-1} + v_t \\
\phi &= 1 - c/R^{\alpha} \\
\mu &= \tilde{\mu}/R^{\gamma}
\end{aligned}
$$

	size		
	DM	ENC	CCS
stationary	bad	good	good
LUR+drift	bad	good	bad

Stationary $(c = 5,\ \alpha = 0, \tilde{\mu} \neq 0, \gamma = 0, 0.3)$
LUR+Drift $(c = 5,\ \alpha = 1, \tilde{\mu} \neq 0, \gamma = 0, 0.3)$

10. Conclusions

Conclusions

- **Theory:**
  - DM has the negative asymptotic bias under H_0.
  - The bias-corrected DM is ENC.
  - ENC follows $N(0,1)$ when $\frac{P}{R} \to \infty$, when the predictor is stationary or persistent.
- **Simulation:**
  - DM is undersized.
  - ENC has good size when the predictor is stationary or persistent.
  - CCS is severely under-sized when the predictor is persistent. CCS has proper size when predictor is stationary.
- **Extensions:**
  - When Model 1 $(a_1 = 0)$ is the martingale difference
  - When a predictor is persistent with a drift
- **Application:**
  - ENC shows strong predictive ability of several persistent predictors (such as inflation and interest rate).

Thank you