Comparing Nested Predictive Regression Models with Persistent Predictors*

Yan Ge† and Tae-Hwy Lee‡

November 29, 2014

Abstract

This paper is an extension of Clark and McCracken (CM 2001, 2005, 2009) and Clark and West (CW 2006, 2007). While CM/CW assume that a predictor is weakly stationary, this paper considers a highly persistent predictor with an autoregressive (AR) root local to unity. CW find that the statistic of Diebold-Mariano (1995 DM) is under-sized because the DM statistic tends to be negative under the null hypothesis of the equal predictive ability. We find this under-sized problem in DM becomes even more severe when the predictor is persistent with the AR root closer to unity. We find, however, this problem does not exist in the encompassing (ENC) statistic of CM (2001) even when a predictor follows the Ornstein-Uhlenbeck process. This means that ENC remains correct in size and high in power. The size of ENC is robust to near-unit-root persistence in the predictor. We also examine the statistic of Chao, Corradi, and Swanson (2001 CCS) in comparison with DM and ENC. It is shown that CCS is not robust to the persistent predictor as it becomes severely under-sized when the predictor becomes persistent (even if it has correct size with a weakly stationary predictor) and it loses the power substantially when the predictor is persistent. An application to the predictive regression of the equity premium reveals strong predictive ability of several persistent predictors (such as inflation and interest rate) by ENC, but with little or none from DM and CCS.

Key Words: predictive regression, local to unit root process, Ornstein-Uhlenbeck process, encompassing, size and power, equity premium.

JEL Classification: C53, E37, E27

*We thank Aman Ullah for suggesting to consider persistent predictors for the encompassing test.

†Department of Economics, University of California, Riverside, CA 92521. E-mail: yge001@ucr.edu

‡Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu

1 Introduction

When two nonnested models are compared, Diebold and Mariano (DM 1995) point out that the t-statistic of the mean squared forecast error (MSPE) loss-differential is asymptotically standard normal. However, when two nested models in which a weakly stationary predictor is added in Model 2 are compared, Clark and McCracken (CM 2001, 2005, 2009) point out that the t-statistic of DM behaves quite differently from non-nested case since both the numerator and denominator degenerate, which may result in non-standard normal distribution. They point out that due to the finite sample parameter estimation error (PEE), the DM statistic tends to be negative under the null hypothesis of the equal predictive ability. They correct the negative bias by adding a positive term and propose a test that is equivalent to encompassing (ENC thereafter) test of Nelson (1972) and Harvey, Leybourne and Newbold (1998) and the t-statistic. Under the null hypothesis in which the stationary covariate has no predictive power, CM (2001) show that the test is asymptotically standard normal when the number of in-sample to out-of-sample ratio goes to infinity. In this paper we show that under the null hypothesis, the ENC test restores the asymptotically standard normal when the ratio of the out-of-sample number of forecasts (P) to the in-sample number of observations (R) goes to infinity ($P/R \to \infty$). Under the null hypothesis, the encompassing test also has good power.

This paper extends CM (2001, 2005, 2009) and Clark and West (CW 2006, 2007) from the mean regression with weak stationary predictor to a highly persistent predictor with an AR root local to unity. We compare two nested regression models using the squared-loss function. We show that DM statistic still tends to be negative under the null hypothesis of the equal predictive ability and is more severely undersized if the predictor is a highly persistent predictor. The t-statistic of encompassing test, in which a positive term is added to correct the negative bias of DM, is a robust test and has the correct size under the null hypothesis. We analytically show that the robustness arises from the super consistency property of the additional predictor from Model 2 that follows Ornstein–Uhlenbeck process, thus the convergent rate of the forecast error from Model 2 is faster than that from Model 1 and the asymptotic distribution of ENC statistic has the same asymptotic distribution as shown in CW (2006, 2007). We use Monte Carlo simulation to compare three different statistics and show that when the highly persistent estimator is added in Model

2, the ENC statistic is robust and has the correct size, whereas both DM test and conditional moment test (CCS test) are seriously undersized. An application to the predictive regression of the equity premium reveals strong predictive ability of several persistent predictors (such as inflation and interest rate) by ENC, but with little or none can be seen from DM or CCS.

The paper is organized as follows. Section 2 illustrates the methods of testing out-of-sample Granger-causality in mean using rolling scheme. Section 3 illustrates the asymptotic distribution of the encompassing test with a weak stationary predictor from CM (2001). Section 4 presents the asymptotic distribution of the encompassing test with a weak stationary predictor when the ratio of out-of-sample to in-sample observation is infinite. Section 5 presents the asymptotic distribution of encompassing test with a highly persistent estimator when the ratio of out-of-sample to in-sample observation is infinite. Section 6 is Monte Carlo simulation to examine the finite sample size and power behavior of the DM, ENC and CCS statistics. In Section 7 we present the empirical analysis for Goyal and Welch (2008) in comparing the two nested mean models. Section 8 concludes.

2 Comparing Nested Conditional Mean Models

To test for the out-of-sample predictive ability of x_t for y_{t+1}, we consider the following two nested models with the predictor x_t in Model 2 being local to unit root process:

$$\text{Model 1} \quad : \quad y_{t+1} = x'_{1,t}\beta_{1,t} + e^{(1)}_{t+1} = c_1 + e^{(1)}_{t+1}, \tag{1}$$

$$\text{Model 2} \quad : \quad y_{t+1} = x'_{2,t}\beta_{2,t} + e^{(2)}_{t+1} = c_2 + bx_t + e^{(2)}_{t+1}, \tag{2}$$

where c_i is the constant term for Model i, x_t is the predictor with local to unit autoregressive (AR) root process $x_{t+1} = \phi x_t + v_{t+1}$. We will consider the simple case when $x'_{1,t} = 1$ and $x'_{2,t} = (1\ x_t)'$. Under the null hypothesis, $b = 0$ and $e^{(1)}_{t+1} = e^{(2)}_{t+1}$, denoted as e_{t+1}. At each time t, both c_i and b are estimated with the rolling window of size R up to time t. Therefore

$$\begin{aligned} \hat{c}_{1,t} - c_1 &= B_1(t)\, H_1(t) \\ \left(\hat{c}_{2,t}, \hat{b}_t\right)' - (c_2, b)' &= B_2(t)\, H_2(t) \end{aligned}$$

where $x'_{1,t} = 1$, $x'_{2,t} = (1\ x_t)'$, $q_{i,t} = x'_{i,t}x_{i,t}$ for Model i at time t, and $B_i(t) = \left(R^{-1}\sum_{j=t-R}^{t-1} q_{i,j}\right)^{-1}$, $h_{i,t} = x'_{i,t}e_{t+1}$ and $H_i(t) = R^{-1}\sum_{j=t-R}^{t-1} h_{i,t}$. Let $f^{(1)}_{t+1} = \hat{c}_{1,t}$ be the forecasts for Model 1 and

$f_{t+1}^{(2)} = \hat{c}_{2,t} + \hat{b}_t x_t$ be the forecast for Model 2 at time t and $\hat{e}_{t+1}^{(1)} = y_{t+1} - f_{t+1}^{(1)}$, $\hat{e}_{t+1}^{(2)} = y_{t+1} - f_{t+1}^{(2)}$ be the forecast errors with the squared forecast-error loss

$$L\left(\hat{e}_{t+1}^{(i)}\right) \equiv \left(\hat{e}_{t+1}^{(i)}\right)^2, \qquad i = 1, 2.$$

To test for equal predictive accuracy of the two models, the null hypothesis is

$$\mathbb{H}_0 : \mathbb{E}\left[L\left(\hat{e}_{t+1}^{(1)}\right) - L\left(\hat{e}_{t+1}^{(2)}\right)\right] = 0. \tag{3}$$

Under $\mathbb{H}_0$, x_t does not Granger-cause y_{t+1} in mean and thus $b = 0$. If x_t Granger-causes y_{t+1}, i.e., $b \neq 0$, thus the alternative hypothesis is

$$\mathbb{H}_1 : \mathbb{E}\left[L\left(\hat{e}_{t+1}^{(1)}\right) - L\left(\hat{e}_{t+1}^{(2)}\right)\right] > 0. \tag{4}$$

The Diebold-Mariano square loss differential is defined as

$$\hat{D}_P = P^{-1} \sum_{t=R}^{T} L\left(\hat{e}_{t+1}^{(1)}\right) - L\left(\hat{e}_{t+1}^{(2)}\right), \tag{5}$$

and the adjusted MSFE loss-differential is defined as

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

where R is the number of observations in the rolling windows for the in-sample estimation, P is the number of out-of-sample forecasts, and $R + P = T + 1$. We also compare DM and ENC with CCS test by Chao et al (CCS 2001), which is constructed as follows: Under $\mathbb{H}_0$, $b = 0$, which implies $\mathbb{E}\left(e_{t+1}^{(1)} x_t\right) = 0$. The out-of-sample test statistic of CCS is constructed from

$$\hat{M}_P = P^{-1} \sum_{t=R}^{T} \hat{e}_{t+1}^{(1)} x_t. \tag{6}$$

The three statistics are standardized to form the DM statistic $DM_P \equiv \hat{S}_P^{-0.5} \sqrt{P} \hat{D}_P$, the encompassing statistic $ENC_P \equiv \hat{Q}_P^{-0.5} \sqrt{P} \hat{B}_P$, and the CCS statistic $CCS_P \equiv \hat{W}_P^{-0.5} \sqrt{P} \hat{M}_P$, where $\hat{S}_P, \hat{Q}_P$ and $\hat{W}_P$ are the consistent estimators of $S_P = var\left(\sqrt{P}\hat{D}_P\right)$, $Q_P = var\left(\sqrt{P}\hat{B}_P\right)$, and $W_P = var\left(\sqrt{P}\hat{M}_P\right)$, respectively.

3 Asymptotic Distribution of ENC With a Stationary Predictor (CM 2001)

First, we consider a stationary predictor as in CM (2001, 2005), CW (2006, 2007), and CCS (2001).

Assumption 1. $\{x_t\}$ is a weakly stationary process and $\mathbb{E}(q_{i,t})$ is bounded for all t and $i = 1, 2$. We define $B_i = (\mathbb{E}q_{i,t})^{-1}$ for model $i = 1, 2$.

Let $\pi = \lim_{P,R\to\infty} P/R$ and $\xi = R/T = R/(P+R)$. Note that $\frac{1}{\xi} - 1 \to \pi$. We consider three cases on π :

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 1 and Assumption 2a,

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

under $\mathbb{H}_0$, where $W(s)$ is a Wiener process and $s \in [0, 1]$. When Assumption 2a holds, the RHS of Equation (7) is *not* standard normal.

Proposition 2 (CM 2001). Under Assumption 1 and Assumption 2b,

$$ENC_P \Rightarrow \lim_{\xi\to 1} \frac{\int_\xi^1 \xi^{-1}\left[W(s) - W(s-\xi)\right] \mathrm{d}W(s)}{\sqrt{\int_\xi^1 \xi^{-2}\left[W(s) - W(s-\xi)\right]^2 \mathrm{d}s}} \sim N(0, 1), \tag{7}$$

under $\mathbb{H}_0$, where $W(s)$ is a Wiener process and $s \in [0, 1]$. When Assumption 2b holds, the RHS of Equation (7) is standard normal.

Remark 1. CM (2001) show that when Assumption 2b holds ($\xi \to 1, \pi \to 0$) then ENC_P is asymptotically standard normal. However, CM (2001) did not consider the case when Assumption 2c holds ($\xi \to 0, \pi \to \infty$). In Section 4 below, we consider this case and show that ENC_P is still asymptotically standard normal.

Remark 2: CM (2001) assumes Assumption 1 that the predictor $\{x_t\}$ is weakly stationary and show that ENC_P is asymptotically standard normal under Assumption 2b. In Section 5, we show

that ENC_P is asymptotically standard normal under Assumption 2c when the predictor has a root local to unity.

4 Asymptotic Distribution of ENC with a Stationary Predictor when $P/R \to \infty$

In this section, we will show that the asymptotic distribution of ENC_P under $\mathbb{H}_0$, as shown in Equation (7), is asymptotically standard normal under Assumption 2c (when $\lim_{P,R\to\infty} P/R \to \infty$).

Proposition 3. Under Assumption 1 and Assumption 2c,

$$ENC_P \Rightarrow \lim_{\xi\to 0} \frac{\int_\xi^1 \xi^{-1}\left[W(s)-W(s-\xi)\right]\mathrm{d}W(s)}{\sqrt{\int_\xi^1 \xi^{-2}\left[W(s)-W(s-\xi)\right]^2\mathrm{d}s}} \sim N\left(0,1\right), \tag{8}$$

under $\mathbb{H}_0$.

Proof: We firstly consider the numerator of equation (8) by dividing $[0,1]$ to n equal segments and let $t=[Ts]$, where $[Ts]$ is the integer part of Ts and $s\in[0,1]$. Since ξ is sufficiently small, we can write $\xi \equiv 1/n = 1+P/R$. We discretize both the numerator and the denominator. Let $\{u_i\}_{i=1}^n$ be a mixing sequence drawn from the standard normal distribution $N\left(0,1\right)$ with $E\left(u\right)=0$ and $\mathrm{var}(u)=1$. Let $V_t=\sum_{i=1}^t u_i$ be the partial sum. Then we have $U_t=\sum_{i=1}^t u_i \sim N(0,t)$ and therefore

$$\frac{U_t}{\sqrt{n}} = \frac{\sum_{i=1}^t u_i}{\sqrt{n}} \equiv U_n\left(s\right) \Rightarrow W(s),$$

where $U_n\left(s\right)$ is a '*cadlag*' function and $W\left(s\right)$ is a Wiener process. Note that

$$\begin{aligned} n^{-1}\sum_{t=1}^n u_{t-1}u_t &= n^{-1}\sum_{t=1}^n U_{t-1}u_t - n^{-1}\sum_{t=1}^n U_{t-2}u_t \\ &\Rightarrow \int_\xi^1 W(s)\mathrm{d}W(s) - \int_\xi^1 W(s-\xi)\mathrm{d}W(s) \\ &= \int_\xi^1 \left[W(s)-W(s-\xi)\right]\mathrm{d}W(s) \end{aligned}$$

Considering the term $\int_\xi^1 \left[W(s)-W(s-\xi)\right]^2\mathrm{d}s$ in the denominator, we have

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

We construct the an AR(1) regression model, regressing $\{u_{t+1}\}$ on $\{u_t\}$:

$$u_{t+1} = \delta u_t + e_t$$

The estimator $\hat{\delta}$ equals $\left(\sum_{t=1}^{n} u_{t-1}u_t\right) / \left(\sum_{t=1}^{n} u_{t-1}^2\right)$ and the variance $\hat{\delta}$ equals $\left(\sum_{t=1}^{n} u_{t-1}^2\right)^{-1}\mathrm{var}(u) = \left(\sum_{t=1}^{n} u_{t-1}^2\right)^{-1}$. Therefore Equation (8) can be approximated by

$$\frac{\int_{\xi}^{1} \xi^{-1} \left[W(s) - W(s-\xi)\right] \mathrm{d}W(s)}{\sqrt{\xi^{-2} \int_{\xi}^{1} \left[W(s) - W(s-\xi)\right]^2 \mathrm{d}s}} \Rightarrow \frac{\sum_{t=1}^{n} u_{t-1}u_t}{\sqrt{\sum_{t=1}^{n} u_{t-1}^2}} \sim N(0,1).$$

□

5 Asymptotic Distribution of ENC with a Persistent Predictor when $P/R \to \infty$

Suppose the predictor x_t in Model 2 follows an AR process $x_{t+1} = \phi x_t + v_{t+1}$ where $\mathbb{E}\left(v_{t+1}^2\right) = \sigma_v^2$. If $|\phi| < 1$, then

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

as $T \to \infty$. Many recent papers generalize the above to the case when ϕ approaches to 1 as the sample size T increases, see Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Giraitis and Phillips (2006), Mikusheva (2007, 2014), Park (2003), Phillips (1987), Phillips and Lee (2013), and Stock (1991). Let $\phi = 1 - c/T$ for some fixed constant $c \geq 0$, $t = [Tr]$, $r \in [0\ 1]$. Let $x_{[Tr]}/\sqrt{T} \Rightarrow J_x^c(r) = \int_0^r e^{(r-s)c} dB_x(s)$ be an Onstein-Uhlenbeck process and B_x is a Brownian motion. If the AR coefficient ϕ is local to unity, then

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

as $T \to \infty$. To consider the persistent predictor we take the local to unit root process in the following Assumption 3.

Assumption 3. $\{x_t\}$ follows an AR process with a root local to unity, $\phi = 1 - c/T$, for some fixed constant $c \geq 0$.

Let $t \equiv [Ts]$ and $\xi \equiv R/T$. Then we have $t/T \to s$ and $(t - R + 1)/T \to (s - \xi)$. Under Assumption 3,

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

as $T \to \infty$. Now, we state the main result, for the numerator of ENC_P, that is $\sqrt{P}\hat{B}_P$.

Proposition 4. Under Assumption 3 and Assumption 2c, we have $\sum_{t=R}^{T} \hat{e}_{t+1}^{(1)} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right) = \sum_{t=R}^{T} e_{t+1} \left(e_{t+1} - \hat{e}_{t+1}^{(1)}\right) + o\left(\xi^{-1}\right)$ under $\mathbb{H}_0$.

Proof: Under the null hypothesis that $b = 0$, $e_{t+1}^{(1)} = e_{t+1}^{(2)} =: e_{t+1}$. Note that

$$\hat{e}_{t+1}^{(i)} = e_{t+1} - x_{i,t}' \left(\hat{\beta}_{i,t} - \beta_{i,t}\right)$$

for Model i. Recall $x_{1,t}' = 1$. For $\hat{B}_P$, the numerator of ENC_P, we decompose

$$\begin{aligned}
& \sum_{t=R}^{T} \hat{e}_{t+1}^{(1)} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right) \\
= & \sum_{t=R}^{T} \left[e_{t+1} - x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_{1,t}\right)\right] \left(e_{t+1} - x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) - e_{t+1} + x_{2,t}' \left(\hat{\beta}_{2,t} - \beta_{2,t}\right)\right) \\
= & \sum_{t=R}^{T} \left[e_{t+1} - x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_{1,t}\right)\right] \left(-x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) + x_{2,t}' \left(\hat{\beta}_{2,t} - \beta_{2,t}\right)\right) \\
= & \sum_{t=R}^{T} e_{t+1} \left[-x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_{1,t}\right)\right] + \sum_{t=R}^{T} e_{t+1} \left[x_{2,t}' \left(\hat{\beta}_{2,t} - \beta_{2,t}\right)\right] \\
& + \sum_{t=R}^{T} \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) x_{1,t} x_{1,t}' \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) - \sum_{t=R}^{T} \left(\hat{\beta}_{1,t} - \beta_{1,t}\right) x_{1,t} x_{2,t}' \left(\hat{\beta}_{2,t} - \beta_{2,t}\right) \\
\equiv & A_1 + A_2 + A_3 + A_4 \qquad (9)
\end{aligned}$$

Lemmas 1-3 show that $A_1 + A_2 + (A_3 + A_4) = O\left(\frac{T}{R}\right) + O\left(\frac{P}{T}\right) + o(1)$. Hence (9) is dominated by A_1 because $\frac{T}{R} \to \infty$ and $\frac{P}{T} \to 1$ under Assumption 2c. □

Lemma 1. Under Assumption 2c, $A_1 \Rightarrow -\sigma_e^2 \xi^{-1} \int_{\xi}^{1} [W(s) - W(s-\xi)] \mathrm{d}V_e(s) = O\left(\xi^{-1}\right) = O\left(\frac{T}{R}\right)$ under $\mathbb{H}_0$.

Proof: Following Lemma A6 of CM (2001), we show

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

□ □

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

Proof: Rewrite

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

where $G_T = \mathrm{diag}\left(T^{0.5}, T\right)$ as before, and for the two bracketed terms in the last line, we have

$$\begin{aligned} G_T^{-1}\left(\sum_{j=t-R+1}^{t} x_{2,j}x'_{2,j}\right)G_T^{-1}/\xi &\Rightarrow \begin{pmatrix} \xi & \int_{s-\xi}^{s} J_x^c\left(r\right)dr \\ \int_{s-\xi}^{s} J_x^c\left(r\right)dr & \int_{s-\xi}^{s}\left(J_x^c\left(r\right)\right)^2 dr \end{pmatrix}/\xi \sim O\left(1\right), \\ G_T^{-1}\sum_{j=t-R+1}^{t} x_{2,j}e_{j+1}/\xi &\Rightarrow \begin{pmatrix} \int_{s-\xi}^{s} 1dV_e\left(r\right) \\ \int_{s-\xi}^{s} J_x^c\left(r\right)dV_e\left(r\right) \end{pmatrix}/\xi \sim O\left(1\right), \end{aligned}$$

where $J_x^c\left(r\right)$ is an Ornstein-Uhlenbeck process, and $V_e\left(r\right)$ is a Wiener process. Hence

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

Therefore $A_2 = \sum_{t=R}^{T} e_{t+1}x'_{2,t}\left(\hat{\beta}_{2,t}-\beta_{2,t}\right) = O\left(1-\xi\right)$. □

Lemma 3. Under Assumption 2c, $A_3 + A_4$ is $o\left(1\right)$ under $\mathbb{H}_0$.

Proof: Let $E_T = \mathrm{diag}\left(T^0, T^{0.5}\right)$, $F_T = \mathrm{diag}\left(T^1, T^{1.5}\right)$, $G_T = \mathrm{diag}\left(T^{0.5}, T^1\right)$, then for any 2×2 matrix K, we have $E_T F_T = G_T G_T$ and

$$E_T \times K \times F_T = G_T \times K \times G_T,$$

because E_T, F_T, G_T are diagonal. Therefore

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

where the second line appears to be the same as the second bracketed right-hand side term in (A7) of Lemma A10 in CM (2001), which shows that the above is $o(1)$ under Assumption 1. However, under Assumption 3, x_t has an AR root local to unity. We show below that the local-to-unit root in x does not affect Lemma A10 of CM (2001). This is because terms involving x can be suitably normalized as follows

$$\begin{aligned} A_3 + A_4 &= \sum\nolimits_{t=R}^{T} H_1'(t) B_1(t) q_{1,t} B_1(t) H_1(t) \\ &\quad - \sum\nolimits_{t=R}^{T} H_1'(t) B_1(t) x_{1,t} \left(x_{2,t}' E_T^{-1}\right) \left[E_T \times R^{-1} B_2(t) \times F_T \times \xi\right] \left[F_T^{-1} \times R H_2(t) / \xi\right] \\ &\equiv \sum\nolimits_{t=R}^{T} H_1'(t) B_1(t) q_{1,t} B_1(t) H_1(t) - \sum\nolimits_{t=R}^{T} H_1'(t) B_1(t) x_{1,t} \ddot{x}_{2,t}' \ddot{B}_2(t) \ddot{H}_2(t). \end{aligned}$$

where

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

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

and

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

Therefore, $\ddot{x}'_{2,t}, \ddot{B}_2(t), \ddot{H}_2(t)$ have the same orders of magnitude as $x'_{2,t}, B_2(t), H_2(t)$ in stationary case of Lemma A10 in CM (2001). Therefore $A_3 + A_4$ is $o(1)$ not only under Assumption 1 but also under Assumption 3. □

Based on Lemmas 1-3 under Assumption 2c and Assumption 3, $\sum_{t=R}^{T} \hat{e}_{t+1}^{(1)} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right) = -\sum_{t=R}^{T} e_{t+1} \left(e_{t+1} - \hat{e}_{t+1}^{(1)}\right) + o(1)$. Hence, this is the encompassing test for the martingale difference model $y_{t+1} = e_{t+1}$ and the constant mean model $y_{t+1} = c + e_{t+1}^{(1)}$, as studied by CW (2006). Under $\mathbb{H}_0$, ENC_P is asymptotically standard normal. Proposition 4 states this result.

Proposition 5. Under Assumption 3 and 2c, $\lim_{\xi \to 0} ENC_P \Rightarrow N(0,1)$ under $\mathbb{H}_0$.

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

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

where $A_1 \Rightarrow -\sigma_e^2 \xi^{-1} \int_\xi^1 [V_e(s) - V_e(s-\xi)]\mathrm{d}V_e(s)$, $c_{t+1} = -e_{t+1}\left(e_{t+1} - \hat{e}_{t+1}^{(1)}\right)$ and $\hat{c}_P = P^{-1}\sum_{t=R}^{T} c_{t+1}$ $= P^{-1}A_1$. The denominator follows from Lemma 4. Therefore, ENC_P is asymptotically standard normal under $\mathbb{H}_0$ from Proposition 3. □

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

Proof: Following Lemma A11 of CM (2001), we have

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

where line 3 follows from Lemma 1 for $P\hat{c}_P = A_1 = O\left(\xi^{-1}\right)$. □

6 Monte Carlo Simulation

We compare the three statistics by changing $\{x_t\}$ from stationary process to Ornstein-Uhlenbeck process and use the data generating process in Model 1 and Model 2 as follows: the additional variable x_t in Model 2 has the AR process: $x_t = \phi x_{t-1} + z_t$, where z_j is i.i.d, following $N(0,1)$, and $\mathbb{E}(z_t|x_{t-1}) = 0$, The error term $e_{t+1}^{(2)} \sim N\left(0, \sigma_e^2\right)$. We set $c_2 = 1$, $b \in \{0.0, 0.1, 1.0\}$, $\phi \in \{0, 0.5, 0.9, 0.95, 0.99, 1\}$, and $\sigma_e \in \{0.1, 1.0\}$. Model 1 is estimated by regressing $\{y_j\}_{j=t-R+1}^{t}$ on constant term to obtain $\hat{c}_{1,t}$, where $t = R, \ldots, T$. Model 2 is estimated by regressing $\{y_j\}_{j=t-R+1}^{t}$ on $\{1, x_{j-1}\}_{j=t-R+1}^{t}$ to obtain $\left(\hat{c}_{2,t}\ \hat{b}_t\right)$. The forecast errors from the two models are $\hat{e}_{t+1}^{(1)} = y_{t+1} - \hat{c}_{1,t}$ and $\hat{e}_{t+1}^{(2)} = y_{t+1} - \hat{c}_{2,t} - \hat{b}_t x_t$ over the forecast evaluation period at $t = R, \ldots, T$. The number of observations for the rolling windows for estimation are chosen from $R \in \{60, 120, 240\}$. Let $P = T - R + 1 \in \{48, 240, 1200\}$. From these, we compute the three statistics DM_P, ENC_P, and CCS_P. All above tests are repeated 2000 times to find out the Monte Carlo distributions of DM_P, ENC_P, and CCS_P, and to compute their size and power.

The Tables 1 shows the size of test with additional covariate from stationary process to local to Ornstein-Uhlenbeck process. We see that DM statistics are undersized for large P/R ratio for all

cases. The CCS statistic has the correct size for stationary process covariate but as ϕ increases from 0 to 0.99, the size distorts downward. ENC test is robust, having correct size for all ϕ ranging from 0 to 0.99. Table 2 and 3 show the power of test. We see that as ϕ increases, the powers approach to 1 dramatically since higher ϕ implies higher singal-to-noise ratio. Figures 1-12 show the Monte Carlo distributions of ENC_P, DM_P, and CCS_P statistics. Under $\mathbb{H}_0$, we can see that for size of test, if the predictor is a stationary process, both ENC_P and CCS_P have correct size. However, when the predictor has a near unit root, such as $\phi = 0.95$ or 0.99, both DM_P and CCS_P lower the size whereas ENC_P still has the correct size. DM_P and CCS_P have correct sizes for small P/R but suffer an undersize problem for large P/R. Under $\mathbb{H}_1$, all the tests have good powers.

Tables 1-3 About Here

Figures 1-12 About Here

7 Application

We apply the three statistics to the Goyal and Welch study (2008) and construct nested models to test if covariates such as dividend-yield ratio (DY), dividend-price ratio (DP), long term rate of yield (LTY) and inflation (INFL) Granger-causes the equity premium. The dividend-yield ratio at time t is defined as the most recent dividend at t divided by stock price at time t, the dividend-price ratio at time t is defined as the most recent dividend at $t-1$ divided by stock price at time t. The explanation of other variables are available from the homepage of A. Goyal. The two nested models are as shown in (10) and (11) below

$$\text{Model 1} \quad : \quad y_{t+1} = c_1 + e_{t+1}^{(1)}, \tag{10}$$

$$\text{Model 2} \quad : \quad y_{t+1} = c_2 + bx_t + e_{t+1}^{(2)}, \tag{11}$$

where y_{t+1} is the equity premium and x_t is the covariate. We use monthly data ranging from 1926 to 2011, containing 1032 observation for all four models. In Model 1, we only have a constant term, therefore at time t, we predict the future equity premium by solely using the historical average of previous R observations of the equity premium from time $t-R+1$ to t. In Model 2, we use the 1-lag covariate to forecast the equity premium in the next month. See Goyal and Welch (2008) for more on data descriptions. We intend to check if two nested models have the same predictive

accuracy. We use the rolling window scheme of the window size R starting from the 15% of the total observations is $T = 1032$ to the 85% of T. The in-sample observation R ranges from $R = 155$ ($R/P = 155/877$) to $R = 877$ ($R/P = 877/155$). The red line represents DM_P under different allocation of R and P. The blue line and dotted line represent ENC_P statistic and CCS_P statistic respectively.

We conclude that (1) The ENC statistics always have higher statistics than DM test. (2) DP and DY figures show that ENC test is significant with small R/P ratio. Intuitively, for R/P, we are unable to account for the y_{t+1} by solely using the previous y up to time t since there is not sufficient information available, therefore we need to exploit the property of additional variable x. In this way, x has predictive power for ENC test; however when R/P is large, we can predict y_{t+1} using previous information of y up to time t, which weakens the predictive power of x. (3) LTY figure shows that long-term yield has predictive power for equity premium for all R ranging from 150 to 877 using ENC test, which can not directly been observed from DM or CCS statistic. (4) INFL figure shows that CCS test is seriously undersized if we use inflation rate as a predictor unless the number of in-sample observations exceed 600, since the statistic is lower than -1.645, whereas ENC test shows that the inflation rate has predictive power when the number of in-sample observations is below 600.

Figure 13 About Here

8 Conclusions

This paper extends the work of Clark and West (2006, 2007) from nested mean model with weak stationary predictor to nested mean model with a highly persistent predictor. CM (2001, 2005, 2009) and CW(2006, 2007) found that the DM statistic tends to be negative under the null hypothesis of the equal predictive ability because of parameter estimation error. We find that the DM is even more severely undersized when the predictor is highly persistent with the AR root closer to unity. We find that the ENC test is robust as it remains the correct size under the null hypothesis of the equal predictive ability, also it has high power under alternative. We show that the highly persistent predictor following the Ornstein-Uhlenbeck process implies the convergent rate of the estimator from Model 2 faster than that from a model with stationary predictor and

the ENC test can shown to be asymptotically standard normal when the ratio of out-of-sample to in-sample observation is infinite. By using Monte-Carlo simulation, we see that ENC test is robust and has the correct size under null hypothesis whereas DM or CCS are severely under-sized even if CCS is shown to have the correct size in CM (2001, 2005, 2009) and CW(2006, 2007) where a weak stationary predictor is used in Model 2. An application to the predictive regression of the equity premium reveals strong predictive ability of several persistent predictors (such as inflation and interest rate) by ENC, but with little or none can be seen from DM and CCS.

References

Bobkoski, M.J. (1983), "Hypothesis testing in nonstationary time series", Unpublished PhD thesis, Dept. of Statistics, University of Wisconsin, Madison.

Cavanagh, C. (1985), "Roots local to unity, Unpublished Manuscript", Harvard University.

Chan, N. H. and Wei, C. Z. (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *Annals of Statistics* 15(3), 1050-1063.

Chao, J. C., V. Corradi and N. Swanson (2001). "An Out of Sample Test for Granger Causality." *Macroeconomic Dynamics* 5(4): 598-620.

Clark, T. E. and M. McCracken (2001). "Tests of equal forecast accuracy and encompassing for nested models." *Journal of Econometrics* 105(1): 85-110.

Clark, T. E. and M. McCracken (2005). "Evaluating Direct Multistep Forecasts." *Econometric Reviews* 24(4): 369-404.

Clark, T. E. and M. McCracken (2009). "Combining Forecasts from Nested Models." *Oxford Bulletin of Economics and Statistics* 71(3): 303-329.

Clark, T. E. and K. D. West (2006). "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis." *Journal of Econometrics* 135(1-2): 155-186.

Clark, T. E. and K. D. West (2007). "Approximately normal tests for equal predictive accuracy in nested models." *Journal of Econometrics* 138(1): 291-311.

Corradi, V. and N. R. Swanson (2002). "A consistent test for nonlinear out of sample predictive accuracy." *Journal of Econometrics* 110(2): 353-381.

Diebold, F.X. (2015), "Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold-Mariano tests", forthcoming, *Journal of Business and Economic Statistics*.

Diebold, F. X. and R. S. Mariano (1995). "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13(3): 253-263.

Giacomini, R. and H. White (2006). "Tests of Conditional Predictive Ability." *Econometrica* 74(6): 1545-1578.

Giraitis, L. and Phillips, P. C. B. (2006), "Uniform Limit Theory for Stationary Autoregression", *Journal of Time Series Analysis* 27(1), 51-60.

Goyal, A. and Welch, I. (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies* 21: 1455-1508.

Harvey, D. S., S. J. Leybourne and P. Newbold (1998). "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics* 16(2): 254-259.

Mikusheva, A. (2007), "Uniform inference in autoregressive models," Econometrica 75(5), 1411-1452.

Mikusheva, A. (2014), "Second Order Expansion of the t-statistic in AR(1) Models," *Econometric Theory*, forthcoming.

Newey, W. K. and J. L. Powell (1987). "Asymmetric Least Squares Estimation and Testing." *Econometria* 55(4): 819-847.

Park, J.Y. (2003): Weak Unit Roots, Unpublished Manuscript, Rice University.

Phillips, P. C. B. (1987). "Towards a unified asymptotic theory for autoregression." *Biometrika* 74(3): 535–547.

Phillips, P. C. B. and J. H. Lee (2013). "Predictive Regression Under Various Degrees of Persistence and Robust Long-Horizon Regression." *Journal of Econometrics* 177(2): 250-264.

Stock, J. H. (1991). "Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series." *Journal of Monetary Economics* 28: 435-459.

West, K. D. (1996). "Asymptotic Inference about Predictive Ability." *Econometrica* 64(5): 1067-1084.

Table 1: Size of test under 0.05 nominal size.

		$P = 48$			$P = 240$			$P = 1200$		
$Repeat = 2000$		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
	$R = 60$	0.008	0.040	0.051	0.000	0.030	0.045	0.000	0.042	0.058
$\phi = 0$, $\sigma_e = 0.1$	$R = 120$	0.010	0.026	0.052	0.001	0.023	0.044	0.000	0.042	0.051
	$R = 240$	0.023	0.038	0.048	0.006	0.033	0.053	0.000	0.031	0.043
	$R = 60$	0.005	0.031	0.063	0.000	0.032	0.048	0.000	0.046	0.046
$\phi = 0$, $\sigma_e = 1$	$R = 120$	0.014	0.034	0.060	0.001	0.033	0.056	0.000	0.026	0.045
	$R = 240$	0.026	0.043	0.055	0.004	0.035	0.049	0.000	0.035	0.049
	$R = 60$	0.009	0.039	0.055	0.001	0.030	0.049	0.000	0.041	0.039
$\phi = 0.1$, $\sigma_e = 0.1$	$R = 120$	0.015	0.031	0.055	0.001	0.024	0.052	0.000	0.030	0.046
	$R = 240$	0.021	0.039	0.058	0.003	0.029	0.055	0.000	0.033	0.059
	$R = 60$	0.007	0.027	0.056	0.001	0.032	0.043	0.000	0.041	0.052
$\phi = 0.1$, $\sigma_e = 1$	$R = 120$	0.016	0.035	0.052	0.003	0.026	0.050	0.000	0.028	0.038
	$R = 240$	0.024	0.039	0.058	0.005	0.029	0.052	0.000	0.025	0.056
	$R = 60$	0.008	0.034	0.069	0.000	0.032	0.048	0.000	0.041	0.055
$\phi = 0.5$, $\sigma_e = 0.1$	$R = 120$	0.012	0.036	0.061	0.001	0.029	0.046	0.000	0.033	0.054
	$R = 240$	0.023	0.039	0.059	0.003	0.032	0.052	0.000	0.035	0.051
	$R = 60$	0.008	0.030	0.056	0.000	0.035	0.045	0.000	0.044	0.050
$\phi = 0.5$, $\sigma_e = 1$	$R = 120$	0.014	0.038	0.046	0.002	0.030	0.056	0.000	0.031	0.047
	$R = 240$	0.019	0.037	0.056	0.005	0.026	0.056	0.000	0.023	0.064
	$R = 60$	0.009	0.033	0.053	0.000	0.035	0.039	0.000	0.042	0.044
$\phi = 0.9$, $\sigma_e = 0.1$	$R = 120$	0.013	0.040	0.049	0.002	0.035	0.051	0.000	0.042	0.048
	$R = 240$	0.026	0.041	0.051	0.005	0.025	0.050	0.000	0.035	0.057
	$R = 60$	0.007	0.030	0.053	0.000	0.039	0.047	0.000	0.048	0.041
$\phi = 0.9$, $\sigma_e = 1$	$R = 120$	0.013	0.042	0.055	0.001	0.032	0.055	0.000	0.032	0.046
	$R = 240$	0.020	0.034	0.058	0.005	0.029	0.054	0.001	0.032	0.052
	$R = 60$	0.006	0.032	0.047	0.001	0.032	0.040	0.000	0.046	0.038
$\phi = 0.95$, $\sigma_e = 0.1$	$R = 120$	0.010	0.034	0.050	0.001	0.032	0.045	0.000	0.047	0.049
	$R = 240$	0.021	0.029	0.055	0.006	0.030	0.044	0.000	0.034	0.043
	$R = 60$	0.005	0.032	0.046	0.001	0.032	0.041	0.000	0.045	0.038
$\phi = 0.95$, $\sigma_e = 1$	$R = 120$	0.010	0.034	0.050	0.001	0.032	0.044	0.000	0.047	0.049
	$R = 240$	0.018	0.029	0.055	0.006	0.030	0.044	0.000	0.034	0.043
	$R = 60$	0.005	0.034	0.041	0.000	0.032	0.018	0.000	0.040	0.006
$\phi = 0.99$, $\sigma_e = 0.1$	$R = 120$	0.010	0.039	0.047	0.001	0.026	0.030	0.000	0.042	0.021
	$R = 240$	0.022	0.039	0.053	0.003	0.034	0.051	0.000	0.034	0.034
	$R = 60$	0.007	0.032	0.051	0.000	0.035	0.014	0.000	0.034	0.007
$\phi = 0.99$, $\sigma_e = 1$	$R = 120$	0.015	0.035	0.055	0.000	0.028	0.035	0.000	0.040	0.017
	$R = 240$	0.022	0.044	0.052	0.003	0.030	0.046	0.000	0.041	0.032
	$R = 60$	0.003	0.027	0.037	0.000	0.037	0.009	0.000	0.036	0.001
$\phi = 1$, $\sigma_e = 0.1$	$R = 120$	0.012	0.030	0.052	0.000	0.026	0.018	0.000	0.038	0.002
	$R = 240$	0.029	0.045	0.063	0.001	0.026	0.041	0.000	0.033	0.007
	$R = 60$	0.002	0.030	0.042	0.000	0.035	0.009	0.000	0.037	0.000
$\phi = 1$, $\sigma_e = 1$	$R = 120$	0.010	0.033	0.052	0.001	0.031	0.014	0.000	0.038	0.001
	$R = 240$	0.016	0.031	0.056	0.001	0.023	0.037	0.000	0.029	0.005

Notes: The table shows the size of DM_P, ENC_P and CCS_P test under 5% nominal size from Monte-Carlo Simulation of 2000 times.

Table 2: Power of test under 0.05 nominal size. $b = 0.1$

		$P = 48$			$P = 240$			$P = 1200$		
$Repeat = 2000$		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
	$R = 60$	0.950	1.000	0.052	1.000	1.000	0.027	1.000	1.000	0.020
$\phi = 0$, $\sigma_e = 0.1$	$R = 120$	0.957	1.000	0.041	1.000	1.000	0.044	1.000	1.000	0.036
	$R = 240$	0.942	1.000	0.051	1.000	1.000	0.045	1.000	1.000	0.038
	$R = 60$	0.017	0.084	0.055	0.005	0.222	0.043	0.002	0.603	0.038
$\phi = 0$, $\sigma_e = 1$	$R = 120$	0.040	0.116	0.062	0.023	0.278	0.056	0.031	0.734	0.042
	$R = 240$	0.062	0.138	0.052	0.062	0.308	0.050	0.121	0.832	0.038
	$R = 60$	0.955	1.000	0.105	1.000	1.000	0.235	1.000	1.000	0.645
$\phi = 0.1$, $\sigma_e = 0.1$	$R = 120$	0.956	1.000	0.115	1.000	1.000	0.249	1.000	1.000	0.718
	$R = 240$	0.948	1.000	0.125	1.000	1.000	0.268	1.000	1.000	0.755
	$R = 60$	0.019	0.088	0.064	0.006	0.227	0.071	0.001	0.596	0.087
$\phi = 0.1$, $\sigma_e = 1$	$R = 120$	0.030	0.112	0.057	0.021	0.263	0.066	0.024	0.739	0.090
	$R = 240$	0.069	0.133	0.055	0.062	0.308	0.065	0.122	0.838	0.097
	$R = 60$	0.980	1.000	0.782	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.5$, $\sigma_e = 0.1$	$R = 120$	0.978	1.000	0.788	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	0.978	1.000	0.810	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	0.020	0.106	0.105	0.005	0.265	0.193	0.003	0.743	0.571
$\phi = 0.5$, $\sigma_e = 1$	$R = 120$	0.061	0.153	0.107	0.030	0.352	0.231	0.072	0.861	0.591
	$R = 240$	0.065	0.158	0.091	0.077	0.398	0.211	0.226	0.927	0.635
	$R = 60$	0.998	1.000	0.972	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.9$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.989	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	0.999	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	0.070	0.316	0.300	0.144	0.772	0.782	0.485	1.000	1.000
$\phi = 0.9$, $\sigma_e = 1$	$R = 120$	0.136	0.400	0.365	0.331	0.873	0.852	0.928	1.000	1.000
	$R = 240$	0.168	0.424	0.361	0.447	0.932	0.887	0.991	1.000	1.000
	$R = 60$	1.000	1.000	0.913	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.95$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.953	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.984	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	0.128	0.444	0.371	0.328	0.927	0.882	0.922	1.000	1.000
$\phi = 0.95$, $\sigma_e = 1$	$R = 120$	0.200	0.540	0.454	0.618	0.975	0.952	0.998	1.000	1.000
	$R = 240$	0.264	0.599	0.524	0.726	0.987	0.970	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.713	1.000	1.000	0.942	1.000	1.000	1.000
$\phi = 0.99$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.736	1.000	1.000	0.950	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.827	1.000	1.000	0.971	1.000	1.000	1.000
	$R = 60$	0.237	0.633	0.411	0.699	0.986	0.782	1.000	1.000	0.998
$\phi = 0.99$, $\sigma_e = 1$	$R = 120$	0.426	0.765	0.495	0.936	0.998	0.885	1.000	1.000	1.000
	$R = 240$	0.527	0.814	0.623	0.967	1.000	0.937	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.873	1.000	1.000	0.991	1.000	1.000	1.000
$\phi = 1$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.849	1.000	1.000	0.958	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.863	1.000	1.000	0.932	1.000	1.000	0.997
	$R = 60$	0.270	0.671	0.563	0.783	0.996	0.878	1.000	1.000	0.981
$\phi = 1$, $\sigma_e = 1$	$R = 120$	0.529	0.821	0.665	0.968	1.000	0.916	1.000	1.000	1.000
	$R = 240$	0.672	0.888	0.737	0.990	1.000	0.904	1.000	1.000	0.996

Notes: The table shows the power of DM_P, ENC_P and CCS_P test under 5% nominal size from Monte-Carlo Simulation of 2000 times.

Table 3: Power of test under 0.05 nominal size. $b = 1$

		$P = 48$			$P = 240$			$P = 1200$		
$Repeat = 2000$		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
	$R = 60$	1.000	1.000	0.049	1.000	1.000	0.030	1.000	1.000	0.013
$\phi = 0$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.049	1.000	1.000	0.041	1.000	1.000	0.030
	$R = 240$	1.000	1.000	0.052	1.000	1.000	0.042	1.000	1.000	0.037
	$R = 60$	0.945	1.000	0.046	1.000	1.000	0.027	1.000	1.000	0.020
$\phi = 0$, $\sigma_e = 1$	$R = 120$	0.947	1.000	0.052	1.000	1.000	0.043	1.000	1.000	0.032
	$R = 240$	0.950	1.000	0.045	1.000	1.000	0.047	1.000	1.000	0.044
	$R = 60$	1.000	1.000	0.133	1.000	1.000	0.348	1.000	1.000	0.878
$\phi = 0.1$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.135	1.000	1.000	0.397	1.000	1.000	0.933
	$R = 240$	1.000	1.000	0.167	1.000	1.000	0.427	1.000	1.000	0.946
	$R = 60$	0.957	1.000	0.107	1.000	1.000	0.212	1.000	1.000	0.616
$\phi = 0.1$, $\sigma_e = 1$	$R = 120$	0.948	1.000	0.115	1.000	1.000	0.258	1.000	1.000	0.718
	$R = 240$	0.949	1.000	0.111	1.000	1.000	0.274	1.000	1.000	0.746
	$R = 60$	1.000	1.000	0.939	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.5$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.944	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.952	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	0.982	1.000	0.764	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.5$, $\sigma_e = 1$	$R = 120$	0.977	1.000	0.790	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	0.966	1.000	0.791	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.979	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.9$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.975	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.9$, $\sigma_e = 1$	$R = 120$	0.999	1.000	0.989	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.934	1.000	1.000	0.999	1.000	1.000	1.000
$\phi = 0.95$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.962	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.989	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.913	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 0.95$, $\sigma_e = 1$	$R = 120$	1.000	1.000	0.953	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.984	1.000	1.000	1.000	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.721	1.000	1.000	0.946	1.000	1.000	1.000
$\phi = 0.99$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.739	1.000	1.000	0.950	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.831	1.000	1.000	0.972	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.696	1.000	1.000	0.941	1.000	1.000	1.000
$\phi = 0.99$, $\sigma_e = 1$	$R = 120$	1.000	1.000	0.755	1.000	1.000	0.944	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.829	1.000	1.000	0.968	1.000	1.000	1.000
	$R = 60$	1.000	1.000	0.885	1.000	1.000	0.992	1.000	1.000	1.000
$\phi = 1$, $\sigma_e = 0.1$	$R = 120$	1.000	1.000	0.858	1.000	1.000	0.960	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.868	1.000	1.000	0.932	1.000	1.000	0.997
	$R = 60$	1.000	1.000	0.872	1.000	1.000	0.991	1.000	1.000	1.000
$\phi = 1$, $\sigma_e = 1$	$R = 120$	1.000	1.000	0.850	1.000	1.000	0.958	1.000	1.000	1.000
	$R = 240$	1.000	1.000	0.863	1.000	1.000	0.932	1.000	1.000	0.997

Notes: The table shows the power of DM_P, ENC_P and CCS_P test under 5% nominal size from Monte-Carlo Simulation of 2000 times.

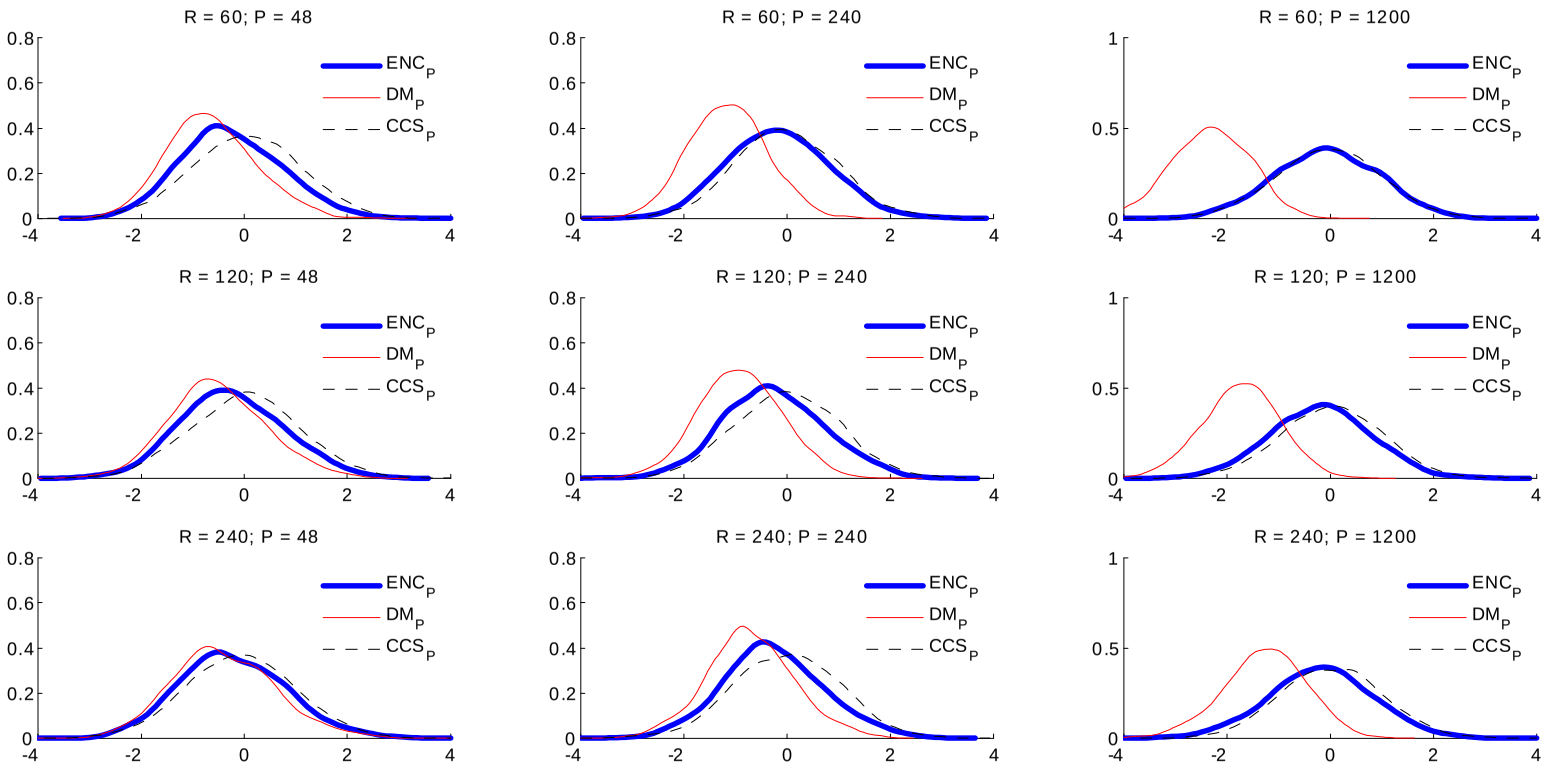

Figure 1: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0$, $b = 0$, $\sigma_e = 1$. 2000 Repeats.

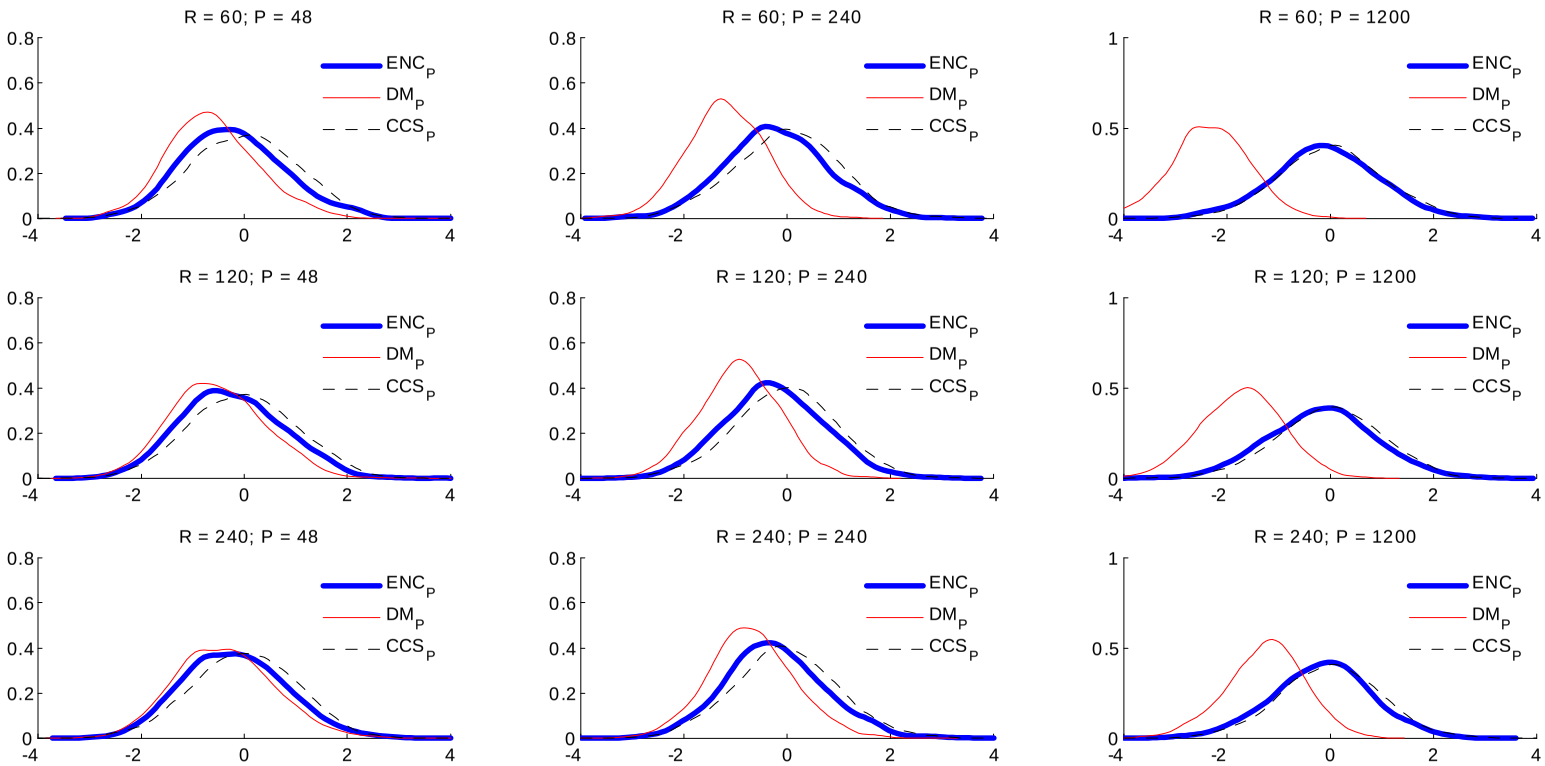

Figure 2: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0$, $b = 0$, $\sigma_e = 0.1$. 2000 Repeats.

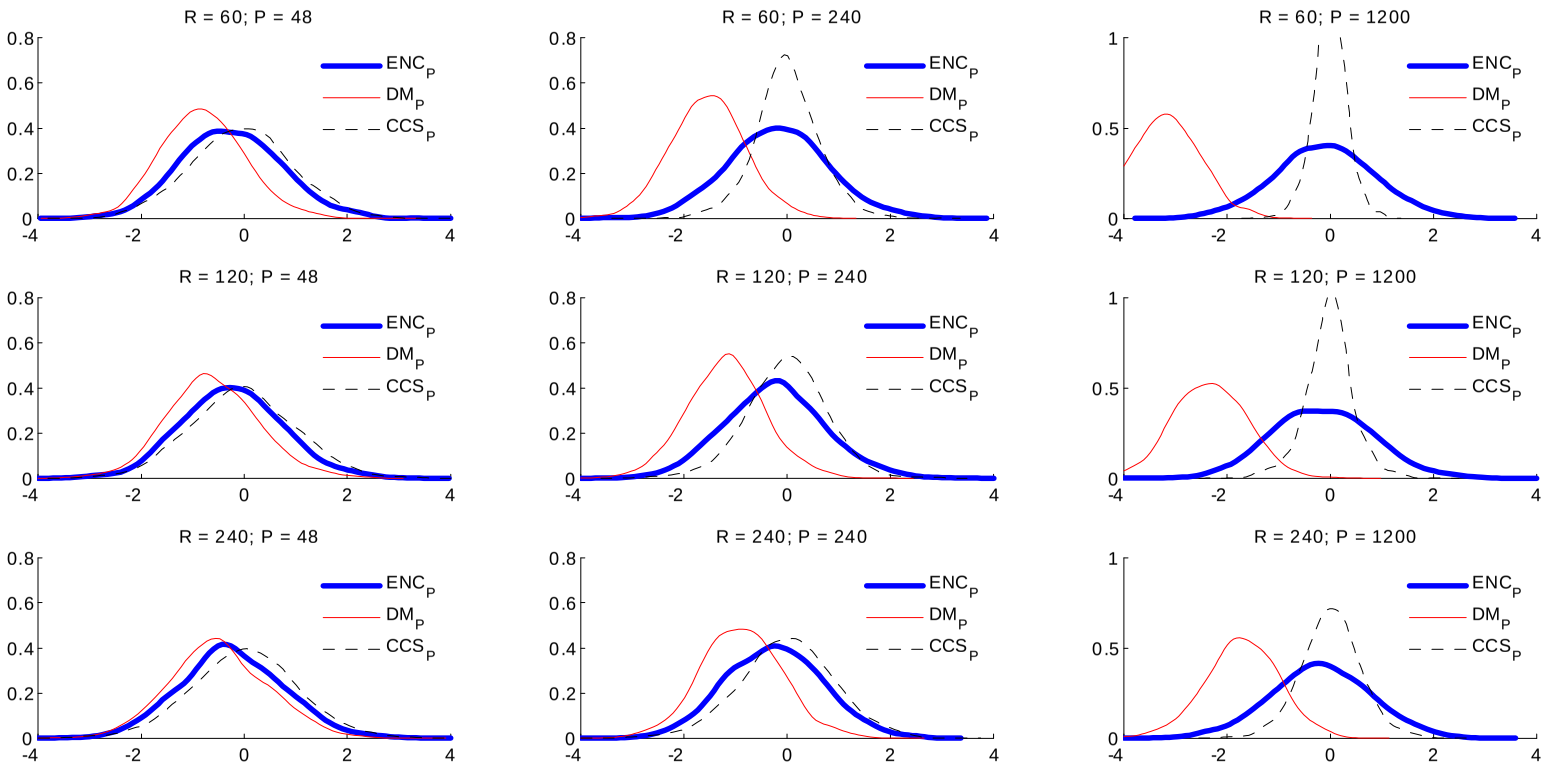

Figure 3: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0.99$, $b = 0$, $\sigma_e = 1$. 2000 Repeats.

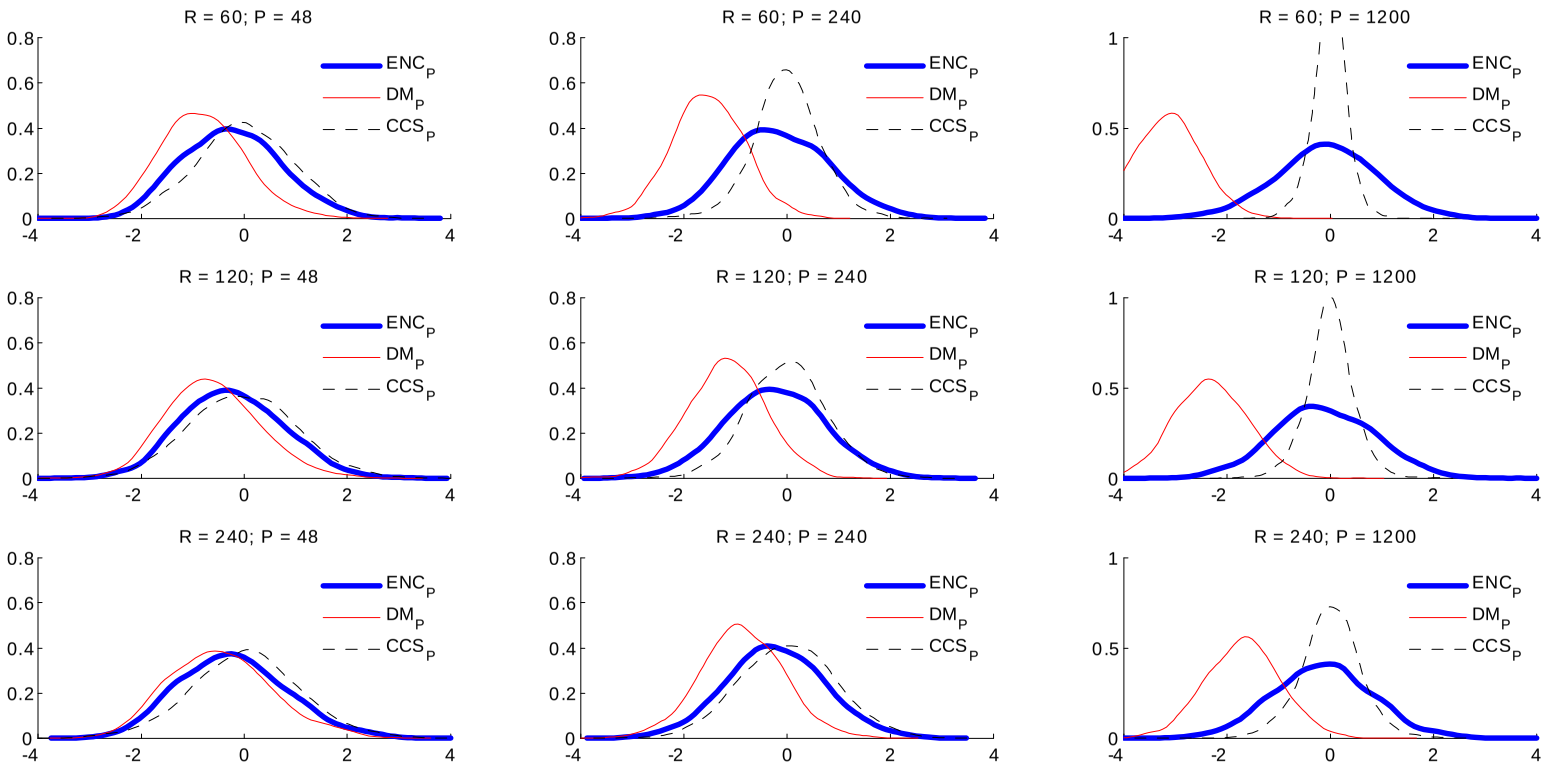

Figure 4: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_0$, $\phi = 0.99$, $b = 0$, $\sigma_e = 0.1$. 2000 Repeats.

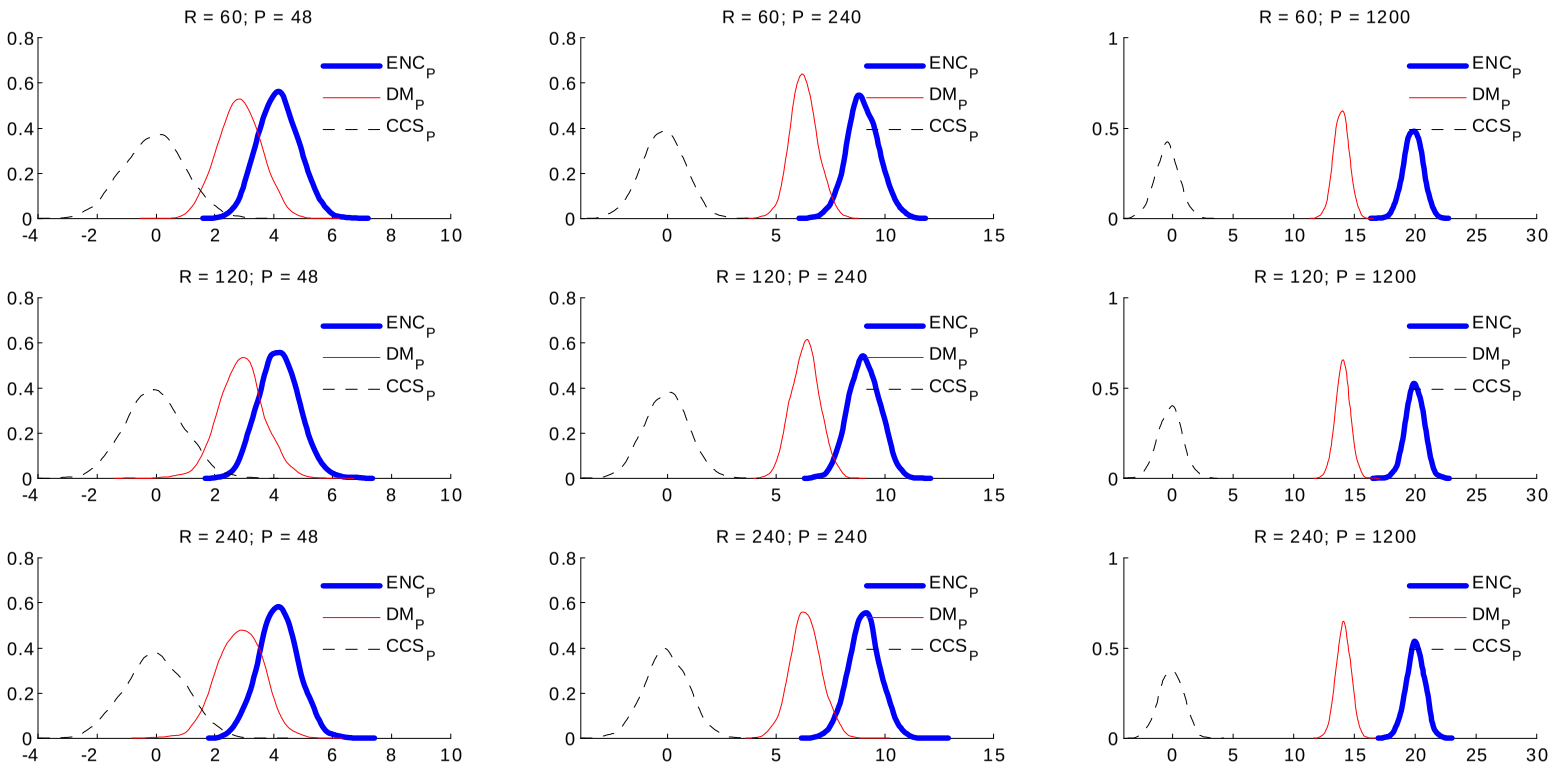

Figure 5: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 0.1$, $\sigma_e = 0.1$. 2000 Repeats.

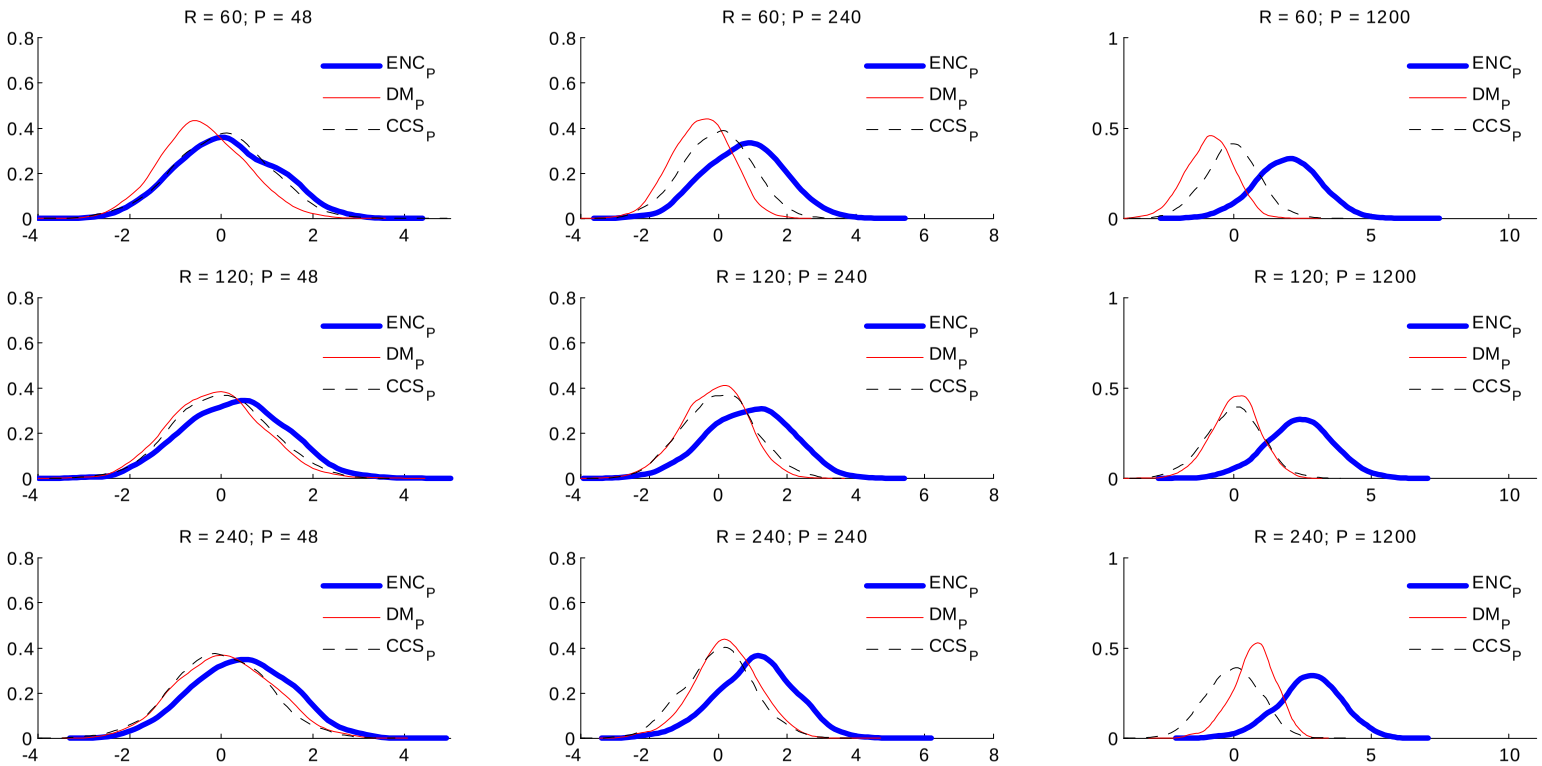

Figure 6: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 0.1$, $\sigma_e = 1$. 2000 Repeats.

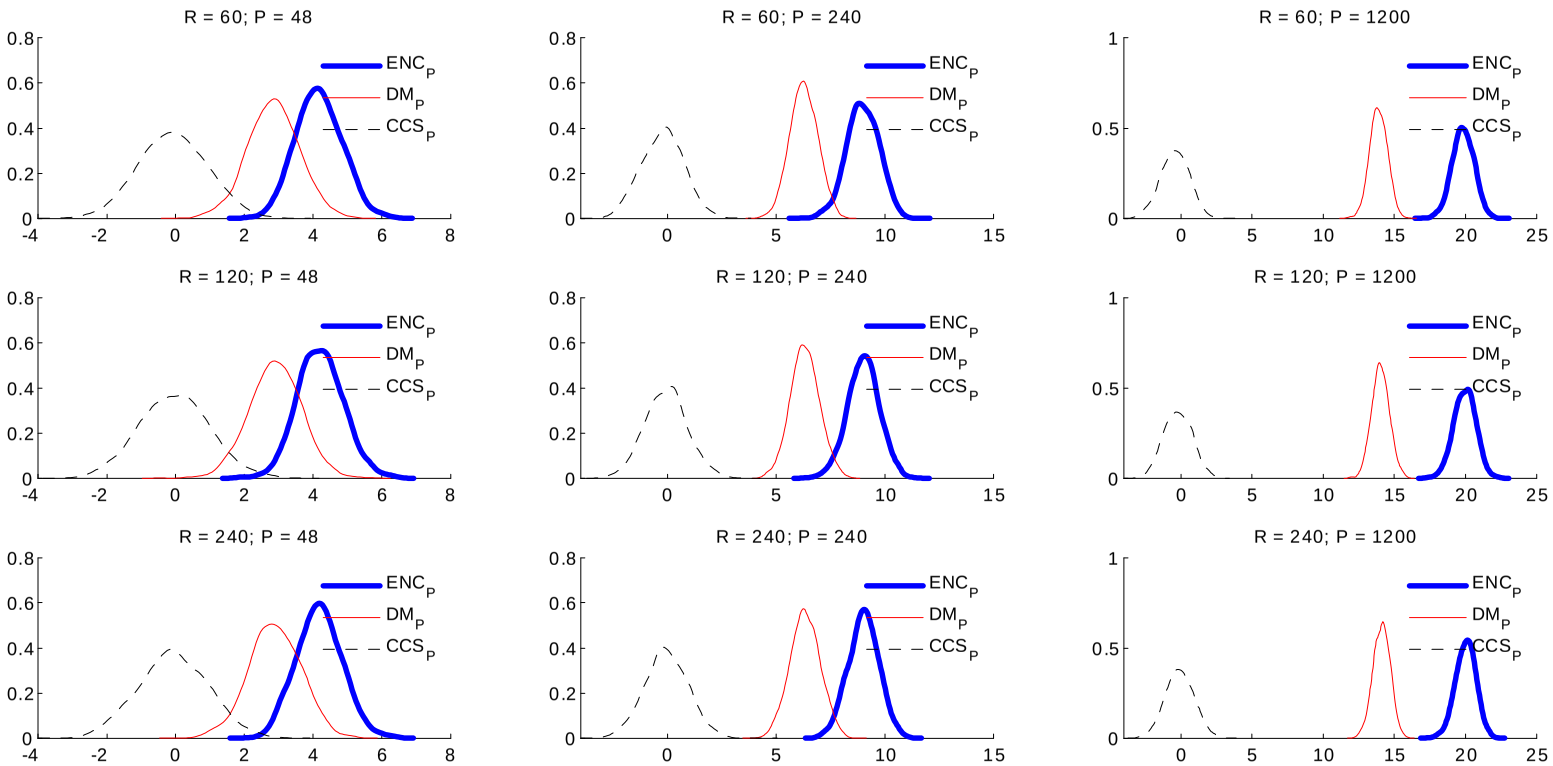

Figure 7: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 1$, $\sigma_e = 1$. 2000 Repeats.

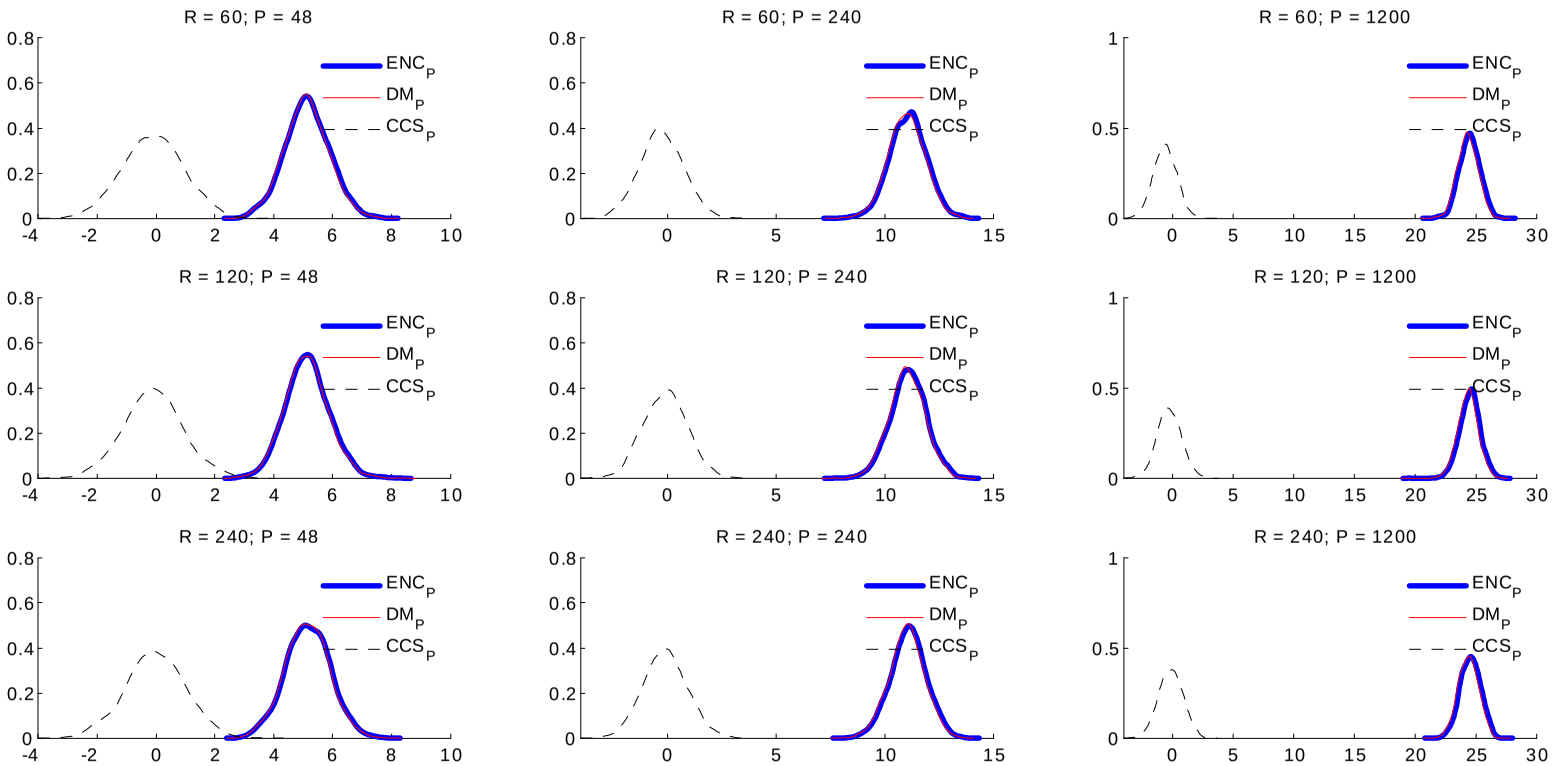

Figure 8: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0$, $b = 1$, $\sigma_e = 0.1$. 2000 Repeats.

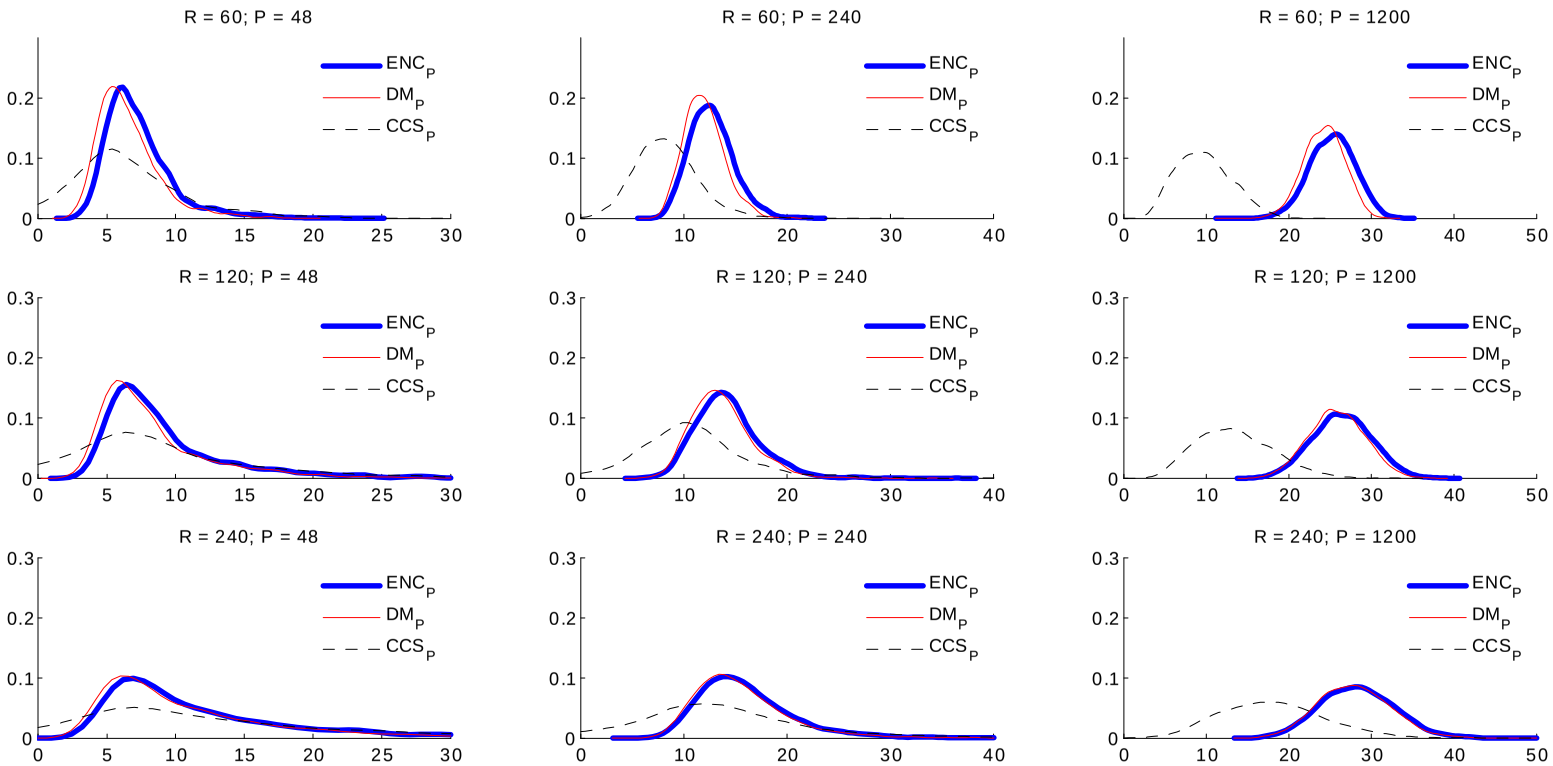

Figure 9: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 0.1$, $\sigma_e = 0.1$. 2000 Repeats.

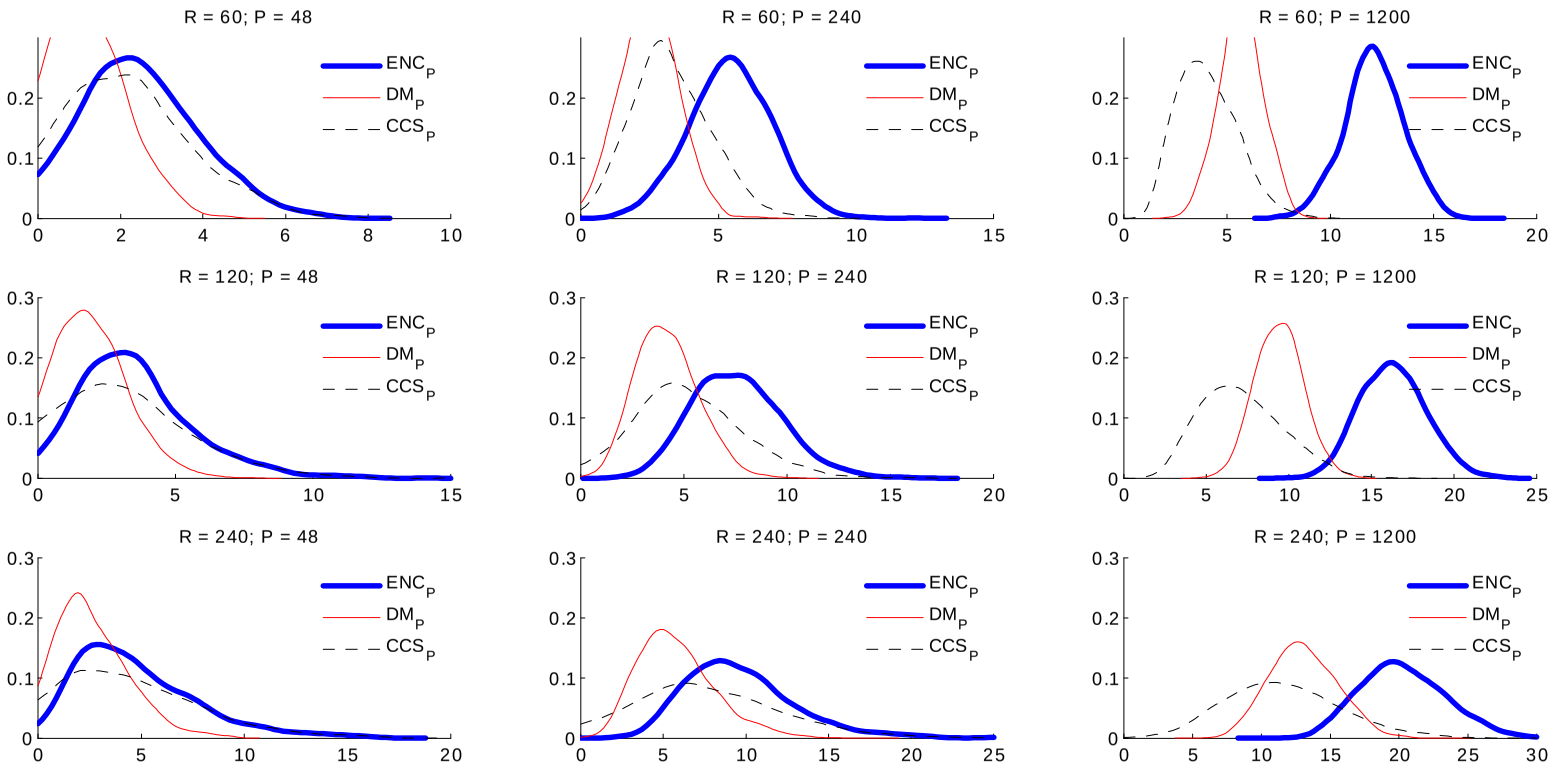

Figure 10: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 0.1$, $\sigma_e = 1$. 2000 Repeats.

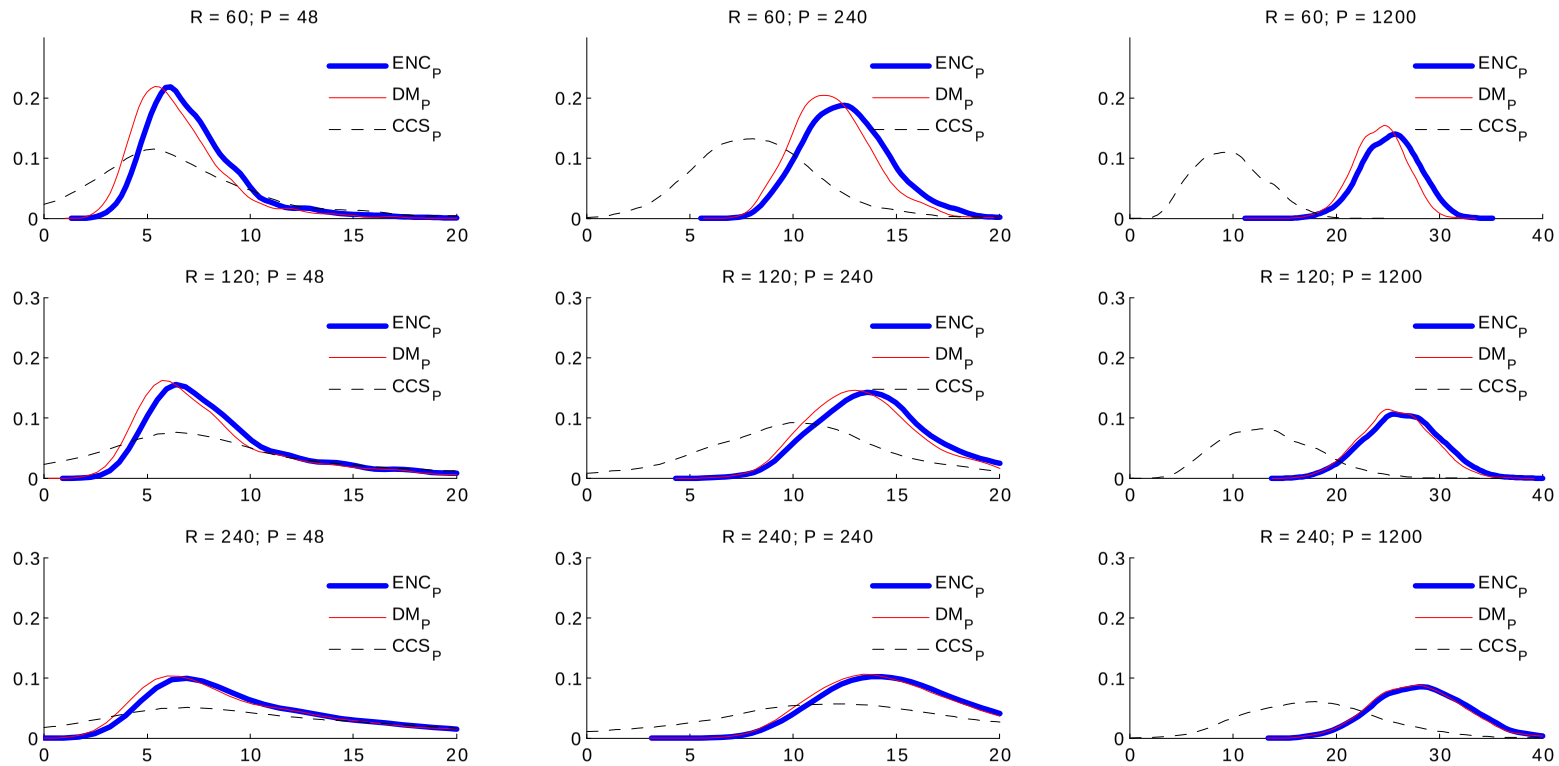

Figure 11: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 1$, $\sigma_e = 1$. 2000 Repeats.

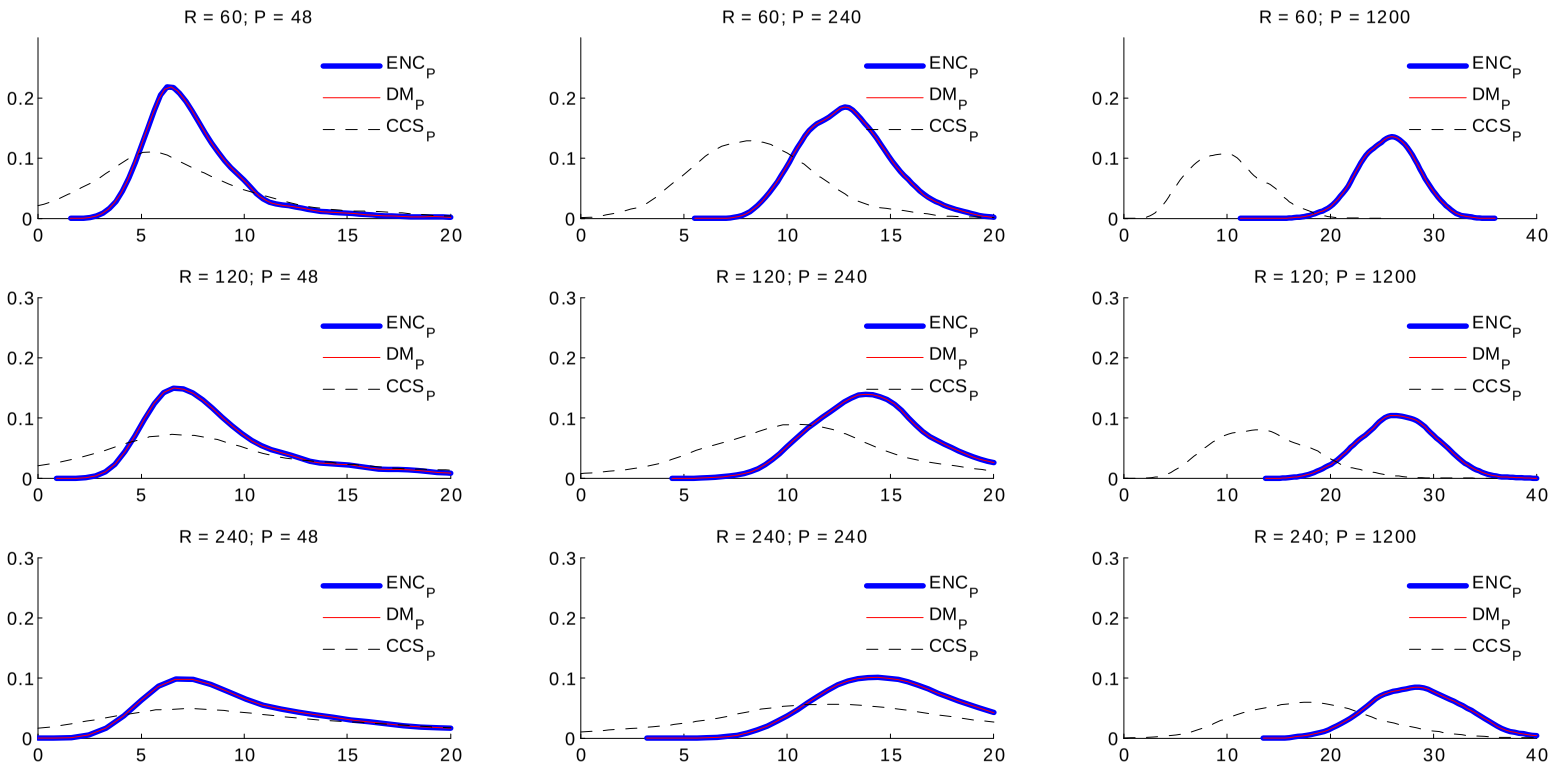

Figure 12: Monte Carlo distribution of ENC (blue line), DM (red line), and CCS (dashed line) under $\mathbb{H}_1$, $\phi = 0.99$, $b = 1$, $\sigma_e = 0.1$. 2000 Repeats.

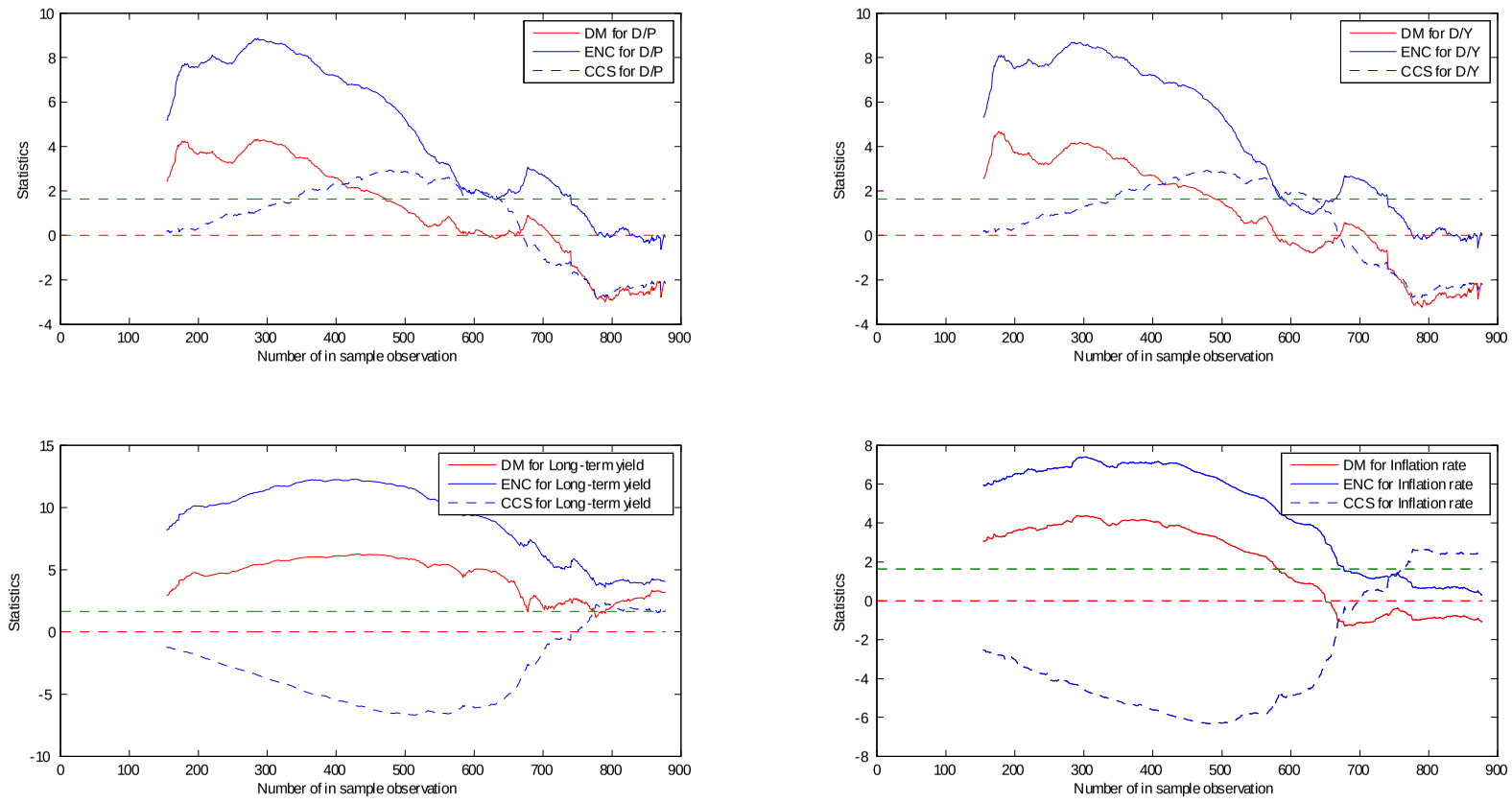

Figure 13: Testing for predictive ability of persistent predictors for equity premium