Forecast Encompassing and Granger Causality in Predictive Models of Expected Shortfalls and Growth Shortfalls

Yaojue Xu*

November 14, 2021

Latest Version Link

Abstract

Despite the importance of modeling the Expected Shortfalls (ES) in financial risk analysis, a testing framework for Granger Causality (GC) in ES does not exist. This makes the econometric specification of predictive models for ES arbitrary. It is due to the absence of an objective function to evaluate ES, i.e., ES is not "elicitable" (Gneiting, 2011 *JASA*). In this paper, I adopt the concept of "higher-order elicitability" introduced by Fissler and Ziegel (FZ, 2016 *AoS*). While the ES alone is not elicitable, ES is elicitable jointly with the corresponding quantile. In this paper, using the FZ scoring function for a pair of ES and quantile, I develop a new test for GC in ES. My test statistic is based on the forecast encompassing (ENC) statistic for a pair of ES and quantile under the FZ scoring function, which is a "strictly consistent" scoring rule. The ENC statistic for a pair of ES and quantile is based on the martingale difference property of a vector of the first order conditions to minimize the higher order elicitable FZ scoring function. I prove the asymptotic normality of the ENC statistic. Monte Carlo simulations are presented to examine the finite sample behavior of our ENC statistic to test for GC in ES, which shows a proper size and a good power. Finally, I consider two applications, one on Value-at-Risk and its ES of the S&P500 financial returns and another on Growth-at-Risk and Growth Shortfall (GS) of the US GDP growth.

Keywords: Higher-order Elicitability, Consistent scoring function, Value-at-Risk, Expected Shortfall, Granger Causality, Encompassing.

JEL Classification: C53, E37, E27

*Department of Economics, University of California, Riverside, CA 92521. E-mail: yxu122@ucr.edu

1 Introduction

Risk management is important in financial institutions, and choosing a proper risk measure is crucial in financial risk management. Two risk measures are most widely used in financial markets: Value at Risk (VaR) and Expected Shortfall (ES). Basel II in 1996 proposed VaR as a proper risk measure, so VaR has become the standard measure of financial market risk. VaR measures the maximum potential loss of a given portfolio over a certain period at a given confidence level, so VaR is the quantile with a given tail probability. However, A quantile-based VaR has two main drawbacks. First, VaR cannot satisfy the "coherent" since VaR lacks subadditivity. Artzner et al. (1999) introduce coherent risk measures. They define a risk measure to be coherent if the risk measure satisfies translation invariance ($\rho(X + a) = \rho(X) + a$ for all $a \in \mathbb{R}$), subadditivity ($\rho(X+Y) = \rho(X)+\rho(Y)$), positive homogeneity ($\rho(\lambda X) = \lambda\rho(X)$ for all $\rho \geq 0$) and monotonicity ($Y \leq X$ implies $\rho(Y) \leq \rho(X)$). Thus, lack of subadditivity means that the VaR of a portfolio can be larger than the sum of the individual VaRs, which violates the conventional concept that diversity reduces risk. Second, VaR focus on the probability of the losses but not the magnitude of the losses. Thus, VaR might not be an appropriate risk measure in some situations. Basel III in 2013 proposed another risk measure, which is Expected Shortfall. ES is defined as the conditional expectation of the return given that it exceeds the VaR, so more sensitive to the magnitude of extreme losses. Artzner et al. (1999) also prove that ES is "coherent". However, the measure of ES is not "elicitable" (Gneiting, 2011). Elicitable risk measure means there is a "strictly consistent" scoring function for this risk measure. Then I can compare competing forecasts of this risk measure with respect to the consistent score (Fissler & Ziegel, 2016). There is no strictly consistent scoring function for ES. Therefore, it is difficult to evaluate and compare the ES forecasts due to this drawback.

Gneiting (2011) points out that the mean is elicitable, but the variance is not. However, a pair of mean and variance is elicitable. Fissler and Ziegel (2016) introduce a strictly consistent scoring function for VaR and ES, so the pair of VaR and ES is elicitable. The existence of a strictly consistent scoring function for VaR and ES (FZ scoring function) accelerates the development of ES forecast. Patton et al. (2019) propose new dynamic models to forecast VaR and ES by using the FZ scoring function. Taylor (2019) uses the FZ scoring function to forecast VaR and ES based on the asymmetric Laplace distribution. In this paper, in order to test Granger causality, I also use the strictly consistent scoring function for VaR and ES to conduct in-sample estimation and out-of-sample forecasting.

Out-of-sample forecast comparison is widely used in many fields because it is suggested to test Granger Causality, which determines whether some independent variables can predict the dependent variable (Ashley et al., 1980). Many papers in the literature focus on out-of-sample tests for equal accuracy and encompassing (Diebold & Mariano, 1995; Clark & West, 2006, 2007). Out-of-sample test for equal accuracy and encompassing need two models. When the big model includes all the variables in the small model and has one or more other variables that are not in the small model, these two models are called nested models because the big model nests the small model. When these two models include the same variables and each model also has one or more unique variables, these two models are called non-nested models. These two models overlap each other, and no one can nest another.

When two non-nested models are to be compared, Diebold and Mariano (1995) introduce the DM statistic for comparing predictive accuracy. Under the null hypothesis of no difference in predictive accuracy, the test statistic DM is asymptotically N(0,1) distributed. However, Harvey et al. (1997) point out that the DM test performs well for large samples but could oversize for moderate samples. Hence, they modified the DM statistic by correcting the bias (HNL statistic). They use a critical value from the student's t-distribution instead of the standard normal distribution that the

DM test used. However, when there are two nested models to be compared, according to Clark and McCracken (2001), the DM statistic for equal accuracy and the HNL statistic for encompassing are invalid because these statistics cannot converge to the standard normal distribution. Moreover, If there are two nested models, Clark and West (2006) and Clark and West (2007) prove that the DM statistic for mean regression has a downside bias, and this downward bias can be corrected if they add a non-negative adjustment term on it. In this paper, I extend the mean regression from Clark and West (2006, 2007) to ES and show that DM statistics for ES using the FZ scoring function also has a bias under the null hypothesis. I develop the encompassing statistic for Granger Causality in ES. The encompassing statistic performs good in the size and power tests.

I consider two applications in this paper. The first application focuses on the financial market. I want to examine if the Macroeconomic and financial variables Granger Cause the equity premium of S&P 500 in VaR and ES. The second application is devoted to Macroeconomics. Measures of the downside risk are essential in risk management. The increasing number of policymakers has focused on the downside risk in the last decade. The International Monetary Fund (IMF) has recently popularized a risk measure for GDP growth called Growth-at-Risk (GaR). GaR is the worst conditional GDP growth distribution at a given coverage level (5th percentile) depending on financial conditions (Adrian et al., 2019). Moreover, Adrian et al. (2019); Chavleishvili et al. (2021) define a measure of adverse real economic impact to be Growth Shortfall (GS). GS is the expectation of the GDP growth when it is less than GaR. Like ES, the GS is not elicitable, but a pair of GaR and GS is elicitable. Therefore, the FZ scoring function is also a strictly consistent scoring function for GaR and GS. In the second application, I want to check if the GDP growth can be predicted by the financial conditions in GaR and ES by using the FZ scoring function.

The paper is organized as follows. In section 2, I review the definitions, lemmas, and theorems of the elicitability. In section 3, I show that the DM statistic has a bias with nested models, and introduce the encompassing test for Granger-Causality in forecasting ES by using the FZ scoring function. In section 4, I prove the ENC for the FZ scoring function is asymptotically standard normal as the number of out-of-sample forecasts tends to infinity. In section 5, I conduct Monte Carlo simulations to show that the encompassing statistic has good size and power. In section 6, I present empirical analysis. Proofs are presented in the appendix.

2 Elicitability

We denote an observation domain O for $y, x \in \mathrm{O} \subseteq \mathbb{R}^{d_1+d_2}$, $d_1 = \dim(y)$ and $d_2 = \dim(x)$, the conditional distribution $F \equiv F(Y_{t+1}|\mathcal{I}_t)$ for Y_{t+1} given X_t. Let $\mathcal{F}$ be a class of distribution function on the observation domain O, and let A be an action domain, $\gamma \in \mathrm{A}$. We define $\Gamma : \mathcal{F} \to \mathrm{A}$ be a functional. For example, $\Gamma(F(y|x))$ may be $\mathbb{E}(Y|X)$, $\mathrm{Q}(Y|X)$, $\mathrm{Var}(Y|X)$, $\mathrm{Mode}(Y|X)$ or $\mathrm{ES}(Y|X)$, where $\mathbb{E}(Y|X)$ is the conditional mean, $\mathrm{Q}(Y|X)$ is the conditional quantile, $\mathrm{Var}(Y|X)$ is the conditional variance, $\mathrm{Mode}(Y|X)$ is the conditional mode and $\mathrm{ES}(Y|X)$ is the conditional Expected Shortfall. Note that Γ can be a vector of several of these.

Definition 1: (Gneiting, 2011; Fissler & Ziegel, 2016) A scoring function is an $\mathcal{F}-$integrable function $S : A \times O \to \mathbb{R}$. S is said to be $\mathcal{F}$-*consistent* for a functional $\Gamma : \mathcal{F} \to \mathrm{A}$ if $\mathbb{E}_F S(\Gamma(F), Y) \leq \mathbb{E}_F S(\gamma, Y)$ for all $F \in \mathcal{F}$ and for all $\gamma \in \mathrm{A}$. Furthermore, S is *strictly* $\mathcal{F}$-*consistent* for Γ if it is $\mathcal{F}$-consistent for Γ and if $\mathbb{E}_F S(\Gamma(F), Y) = \mathbb{E}_F S(\gamma, Y)$ implies that $\gamma = \Gamma(F)$ for all $F \in \mathcal{F}$ and for all $\gamma \in \mathrm{A}$. □

Definition 2: (Gneiting, 2011; Fissler & Ziegel, 2016) An identification function is an $\mathcal{F}-$integrable function $V : A \times O \to \mathbb{R}^k$. V is said to be an $\mathcal{F}$-*identification function* for a functional $\Gamma : \mathcal{F} \to \mathrm{A} \subseteq \mathbb{R}^k$ if $\mathbb{E}_F V(\Gamma(F), Y) = 0$ for all $F \in \mathcal{F}$. Furthermore, V is a *strict* $\mathcal{F}$-*identification function* for Γ if $\mathbb{E}_F V(\gamma, Y) = 0$ holds if and only if $\gamma = \Gamma(F)$ for all $F \in \mathcal{F}$ and for all $\gamma \in \mathrm{A}$. □

A statistical functional is elicitable if a scoring function exists that the correct forecast of the functional is a unique minimizer of the expected score. We can compare or rank the forecasts of the elicitable functional with their realized scores (Fissler & Ziegel, 2016). Many statistical functionals are 1-elicitable, such as expectation, ratios of expectations, quantiles (Value-at-Risk), and expectiles. However, some functional are not 1-elicitable, such as variance, mode, and Expected Shortfall (Gneiting, 2011). Osband (1985) points out that a non-elicitable functional can be a component of an elicitable functional. For example, variance is not elicitable, but there exists a 2-elicitable functional of mean and variance.

Definition 3: (Fissler & Ziegel, 2016) A functional $\Gamma : \mathcal{F} \to A \subseteq \mathbb{R}^2$ is called 2-elicitable if there exists a strictly $\mathcal{F}-$consistent scoring function for Γ. Then the functional $\Gamma = (\Gamma_1, \Gamma_2) : \mathcal{F} \to \mathrm{A}$ is 2 elicitable. □

According to Fissler and Ziegel (2016), there is a relation among strictly consistent scoring function S, strict identification function V, and a matrix-valued function h. We define a vector-valued $\mathbf{m}$ to be a vector of m_1 and m_2, and a vector-valued $\boldsymbol{\gamma}$ to be a vector of γ_1 and γ_2. Taking the first order conditions of the expected loss function with respect to γ, γ_1 and γ_2 respectively, we can get $\mathbb{E}_F m$, $\mathbb{E}_F m_1$, and $\mathbb{E}_F m_2$ to be

$$\mathbb{E}_F \mathbf{m} \equiv \frac{\partial \mathbb{E}_F S(\boldsymbol{\gamma}, Y)}{\partial \boldsymbol{\gamma}} = \mathbf{h}(\boldsymbol{\gamma}) \mathbb{E}_F \mathbf{V}(\boldsymbol{\gamma}, Y) = 0, \tag{1}$$

$$\mathbb{E}_F m_1 \equiv \frac{\partial \mathbb{E}_F S(\gamma_1, \gamma_2, Y)}{\partial \gamma_1} = h_{11}(\gamma_1, \gamma_2) \mathbb{E}_F V_1(\gamma_1, \gamma_2, Y) + h_{12}(\gamma_1, \gamma_2) \mathbb{E}_F V_2(\gamma_1, \gamma_2, Y) = 0,$$

$$\mathbb{E}_F m_2 \equiv \frac{\partial \mathbb{E}_F S(\gamma_1, \gamma_2, Y)}{\partial \gamma_2} = h_{21}(\gamma_1, \gamma_2) \mathbb{E}_F V_1(\gamma_1, \gamma_2, Y) + h_{22}(\gamma_1, \gamma_2) \mathbb{E}_F V_2(\gamma_1, \gamma_2, Y) = 0,$$

where

$$\mathbf{m} = [m_1, m_2]', \qquad \boldsymbol{\gamma} = [\gamma_1, \gamma_2]',$$

$$\mathbf{V}(\gamma_1, \gamma_2, Y) = \begin{bmatrix} V_1(\gamma_1, \gamma_2, Y) \\ \\ V_2(\gamma_1, \gamma_2, Y) \end{bmatrix},$$

$$\mathbf{h}(\gamma_1, \gamma_2) = \begin{bmatrix} h_{11}(\gamma_1, \gamma_2) & h_{12}(\gamma_1, \gamma_2) \\ \\ h_{21}(\gamma_1, \gamma_2) & h_{22}(\gamma_1, \gamma_2) \end{bmatrix}.$$

Concerning equation (1), the strict identification function V is a function of variable y and functional γ, and the function h is a function of γ only. There is no y in the function of h. Functional forms of V1 and V2 are shown in Eq (3).

Quantile is elicitable, but Expected Shortfall is not elicitable (Gneiting, 2011). Fissler and Ziegel (2016) propose a strictly consistent scoring function of joint Value-at-risk (VaR) and Expected Shortfall (ES) as

$$\begin{aligned} S(\gamma_1, \gamma_2, y) = {} & (1\{y \le \gamma_1\} - \alpha) G_1(\gamma_1) - 1\{y \le \gamma_1\} G_1(y) \\ & + G_2(\gamma_2) \left(\gamma_2 - \gamma_1 + \frac{1}{\alpha} 1\{y \le \gamma_1\} (\gamma_1 - y) \right) - \mathcal{G}(\gamma_2) + a(y), \end{aligned} \tag{2}$$

where γ_1 represents VaR and γ_2 represents ES, $\mathcal{G}' = G_2$, G_1 is increasing function, and G_2 is increasing and convex function. Therefore, a pair of VaR and ES is 2-elicitable. With regard to Fissler and Ziegel (2016), taking the derivatives of the scoring function with respect to γ_1 and γ_2, we have

$$\mathbb{E}_F m_1 = \frac{\partial \mathbb{E}_F S(\gamma_1, \gamma_2, Y)}{\partial \gamma_1} = \left[G_1'(\gamma_1) + \frac{1}{\alpha} G_2(\gamma_2) \right] \mathbb{E}_F (1\{y \le \gamma_1\} - \alpha) = 0,$$

$$\mathbb{E}_F m_2 = \frac{\partial \mathbb{E}_F S(\gamma_1, \gamma_2, Y)}{\partial \gamma_2} = \frac{\gamma_1 G_2'(\gamma_2)}{\alpha} \mathbb{E}_F (1\{y \le \gamma_1\} - \alpha) + G_2'(\gamma_2) \mathbb{E}_F \left(\gamma_2 - \frac{1}{\alpha} y 1\{y \le \gamma_1\} \right) = 0,$$

where the strict identification function is

$$\mathbf{V}(\gamma_1, \gamma_2, y) = \begin{bmatrix} V_1(\gamma_1, \gamma_2, Y) \\ \\ V_2(\gamma_1, \gamma_2, Y) \end{bmatrix} = \begin{bmatrix} 1\{y \leq \gamma_1\} - \alpha \\ \\ \gamma_2 - \frac{1}{\alpha} y 1\{y \leq \gamma_1\} \end{bmatrix}, \tag{3}$$

and the matrix-valued function h is

$$\mathbf{h}(\gamma_1, \gamma_2) = \begin{bmatrix} G_1'(\gamma_1) + \frac{1}{\alpha} G_2(\gamma_2) & 0 \\ \frac{\gamma_1 G_2'(\gamma_2)}{\alpha} & G_2'(\gamma_2) \end{bmatrix}. \tag{4}$$

For the first component of the identification function V, $\mathbb{E}_F \left(1\{y \leq \gamma_1\} - \alpha\right) = 0$ if and only if $\gamma_1 = \mathrm{VaR}_\alpha(y)$. For the second component of V, $\mathbb{E}_F \left(\gamma_2 - \frac{1}{\alpha} y 1\{y \leq \gamma_1\}\right)$ if and only if $\gamma_2 = \mathrm{ES}_\alpha(y)$. Therefore, $\mathbb{E}_F V(\gamma_1, \gamma_2, y) = 0$ if and only if $(\gamma_1; \gamma_2)$ equals the true VaR and ES of y, so the joint functional of VaR (Quantile) and ES is 2-elicitible. It indicates that this scoring function is strictly consistent for a pair of VaR and ES, which means that it can be used for forecasts comparison of VaR and ES.

Patton (2011) states that positive homogeneity of the scoring function is important for forecast comparison. Moreover, Patton and Sheppard (2009) find out that homogeneity of degree zero leads to the higher power of Diebold-Mariano tests in volatility forecast. Patton et al. (2019) denote S_{FZ0} as the scoring function with homogeneity of degree zero. S_{FZ0} is homogeneous of degree zero if and only if $G_1(\gamma) = 0$ and $G_2(\gamma) = -1/\gamma$. For easy to read, let q replace γ_1 and e replace γ_2. Then, the "FZ0" scoring function is

$$S_{FZ0}\left(q_\alpha, e_\alpha, y\right) = S_\alpha\left(q_\alpha, e_\alpha, y\right) = -\frac{1}{\alpha e_\alpha} \mathbf{1}\{y \leq q_\alpha\}(q_\alpha - y) + \frac{q_\alpha}{e_\alpha} + \log(-e_\alpha) - 1. \tag{5}$$

In this paper, we use this FZ scoring function S_{FZ0} to test forecast encompassing in Expected Shortfall.

3 Forecast Encompassing Test for Granger Causality in Expected Shortfall

In this section, we develop the forecast encompassing test in the ES. The conditional distribution $F \equiv F(Y_{t+1}|\mathcal{I}_t)$ for Y_{t+1} given X_t. Define $\gamma_1 = q$ for VaR. Define $\gamma_2 = e$ for ES. We have two models. One model is without conditioning on x_t, and the other model is with conditioning on x_t. The unconditional ES of the unconditional distribution $F_1 = F_Y(Y)$ is $\Gamma(F_1)$. The conditional ES of the conditional distribution $F_2 = F_{Y|X}(Y|X)$ is the functional $\Gamma(F_2)$. The two nested models for VaR q_{t+1} are

$$\text{Model 1}: \quad y_{t+1} = c_{1,\alpha} + u^{(1)}_{t+1,\alpha} \equiv x'_{1,t}\beta_{1,\alpha} + u^{(1)}_{\alpha,t+1} \tag{6}$$

$$\text{Model 2}: \quad y_{t+1} = c_{2,\alpha} + b_{1,\alpha}x_t + u^{(2)}_{\alpha,t+1} \equiv x'_{2,t}\beta_{2,\alpha} + u^{(2)}_{\alpha,t+1} \tag{7}$$

where $q^{(1)}_{\alpha,t+1} = c_{1,\alpha} = x'_{1,t}\beta_{1,\alpha}$ and $q^{(2)}_{\alpha,t+1} = c_{2,\alpha} = x'_{2,t}\beta_{2,\alpha}$. ES for Model 1 and Model 2 are defined as $e^{(1)}_{t+1} = \mathbb{E}_F\left[Y_{t+1} \mid Y_{t+1} \leq q^{(1)}_{\alpha,t+1}\right]$ and $e^{(2)}_{t+1} = \mathbb{E}_F\left[Y_{t+1} \mid Y_{t+1} \leq q^{(2)}_{\alpha,t+1}\right]$. The dependent variable y_{t+1} is a scalar random variable. $x_{1,t}$ is a strict subset of $x_{2,t}$. $x'_{1,t} = 1, \beta_{1,\alpha} = c_{1,\alpha}, x'_{2,t} = (1, x_t)\,, \beta_{2,\alpha} = (c_{2,\alpha}, b_\alpha)$. Below the subscript α is omitted for simplicity.

The FZ scoring function is the scoring function S_{FZ0} in Patton et al. (2019), that is

$$S_\alpha\left(q, e, y\right) = -\frac{1}{\alpha e}\mathbf{1}\{y \leq q\}\left(q - y\right) + \frac{q}{e} + \log\left(-e\right) - 1.$$

According to Patton et al. (2019), the first oder conditions of FZ scoring function with respect to VaR q_{t+1} and ES e_{t+1} are

$$\mathbb{E}_F\left[-\frac{1}{\alpha e_{t+1}}\mathbf{1}\{Y_{t+1} \leq q_{t+1}\} + \frac{1}{e_{t+1}}\right] = 0, \tag{8}$$

$$\mathbb{E}_F\left[\frac{1}{\alpha\left(e_{t+1}\right)^2}\mathbf{1}\left\{Y_{t+1} \leq q_{t+1}\right\}\left(q_{t+1} - Y\right) - \frac{q_{t+1}}{e^2_{t+1}} + \frac{1}{e_{t+1}}\right] = 0, \tag{9}$$

which means that $m_1 = -\frac{1}{\alpha e_{t+1}}\mathbf{1}\{y_{t+1} \leq q_{t+1}\} + \frac{1}{e_{t+1}}$ and $m_2 = \frac{1}{\alpha e^2_{t+1}}\mathbf{1}\{y \leq q_{t+1}\}(q_{t+1} - y_{t+1}) -$

$\frac{q_{t+1}}{e_{t+1}^2} + \frac{1}{e_{t+1}}$ are martingale difference sequences. In the last section, we have already defined $m = (m_1, m_2)'$. So, rewriting equations (8) and (9), we have

$$\mathbb{E}_F \mathbf{m} = \mathbf{h}(q_{t+1}, v_{t+1}) \mathbb{E}_F \mathbf{V}(q_{t+1}, v_{t+1}, y_{t+1}) = 0, \tag{10}$$

where the strict identification functions in equation (1) are

$$\mathbf{V}(q_{t+1}, e_{t+1}, y_{t+1}) = \begin{bmatrix} V_1(q_{t+1}, e_{t+1}, y_{t+1}) \\ V_2(q_{t+1}, e_{t+1}, y_{t+1}) \end{bmatrix} = \begin{bmatrix} 1\{y_{t+1} \leq q_{t+1}\} - \alpha \\ e_{t+1} - \frac{1}{\alpha} y_{t+1} 1\{y_{t+1} \leq q_{t+1}\} \end{bmatrix}, \tag{11}$$

and the matrix-valued function h is

$$\mathbf{h}(q_{t+1}, v_{t+1}) = \begin{bmatrix} -\frac{1}{\alpha e_{t+1}} & 0 \\ -\frac{q_{t+1}}{\alpha e_{t+1}^2} & \frac{1}{e_{t+1}^2} \end{bmatrix}. \tag{12}$$

In equation (11), V_1 shows how we define VaR (quantile). We have $\mathbb{E}_F V_1 = 0$, so $\mathbb{E}_F(1\{y_{t+1} \leq q_{t+1}\} - \alpha) = 0$ implies $E_F(1\{y_{t+1} \leq q_{t+1}\}) = \alpha$. V_2 shows how we define ES. We know that $\mathbb{E}_F V_2 = 0$, so $\mathbb{E}_F(e_{t+1} - \frac{1}{\alpha} y_{t+1} 1\{y_{t+1} \leq q_{t+1}\})$ implies $e_{t+1} = \frac{1}{\alpha} \mathbb{E}_F(y_{t+1} 1\{y_{t+1} \leq q_{t+1}\})$.

In order to test equal predictive accuracy of the two nested models, we set the null and alternative hypotheses are

$$\mathbb{H}_0 : \mathbb{E}_F \left[S_\alpha \left(q_{t+1}^{(1)}, e_{t+1}^{(1)}, Y_{t+1} \right) - S_\alpha \left(q_{t+1}^{(2)}, e_{t+1}^{(2)}, Y_{t+1} \right) \right] = 0 \tag{13}$$

$$\mathbb{H}_1 : \mathbb{E}_F \left[S_\alpha \left(q_{t+1}^{(1)}, e_{t+1}^{(1)}, Y_{t+1} \right) - S_\alpha \left(q_{t+1}^{(2)}, e_{t+1}^{(2)}, Y_{t+1} \right) \right] > 0 \tag{14}$$

We set the alternative hypothesis is one side because we expect that the forecasts from Model 2 are better than those from Model 1, as Clark and West (2006) did. When coefficient $b = 0$, x does not Granger cause y. When coefficient $b \neq 0$, x Granger causes y. Define R as the in-sample observations. Define P as the out-of-sample forecasts. The DM statistics based on the FZ loss differential is

$$D = \mathbb{E}_F \left[S_\alpha \left(q_{t+1}^{(1)}, e_{t+1}^{(1)}, Y_{t+1} \right) - S_\alpha \left(q_{t+1}^{(2)}, e_{t+1}^{(2)}, Y_{t+1} \right) \right]$$

$$\hat{D}_{R,P} = P^{-1} \sum_{t=R}^{T} \left[S_\alpha \left(\hat{q}_{t+1}^{(1)}, \hat{e}_{t+1}^{(1)}, Y_{t+1} \right) - S_\alpha \left(\hat{q}_{t+1}^{(2)}, \hat{e}_{t+1}^{(2)}, Y_{t+1} \right) \right]$$

where $T + 1 = R + P$. For mean regression with nested models, Clark and West (2006, 2007) prove that the DM statistic has downward bias and show that adjusted DM statistics (DM statistics plus the adjusted term) obtains zero mean expectation under the null hypothesis, corrects the size, and increases the power. Thus, at first, we attempt to show that the DM statistic based on the FZ loss differential $\hat{D}_{R,P}$ has a non-zero mean under the null hypothesis.

Thereom 1: the DM statistic of the FZ loss differential $\hat{D}_{R,P}$ has a non-zero mean. Thus, $\mathbb{E}\hat{D}_{R,P} \neq 0$ under $\mathbb{H}_0$. □

Proof: In order to show that $\hat{D}_P$ is non-zero, we consider eight cases. (1).$Y_{t+1} > \hat{q}_{t+1}^{(1)}, Y_{t+1} > \hat{q}_{t+1}^{(2)}$ and $\hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$.(2).$Y_{t+1} > \hat{q}_{t+1}^{(1)}, Y_{t+1} > \hat{v}_{t+1}^{(2)}$ and $\hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$.(3).$Y_{t+1} \leq \hat{q}_{t+1}^{(1)}, Y_{t+1} > \hat{q}_{t+1}^{(2)}$ and $\hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$.(4).$Y_{t+1} \leq \hat{q}_{t+1}^{(1)}, Y_{t+1} > \hat{q}_{t+1}^{(2)}$, and $\hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$.(5).$Y_{t+1} > \hat{q}_{t+1}^{(1)}, Y_{t+1} \leq \hat{q}_{t+1}^{(2)}$ and $\hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$.(6).$Y_{t+1} > \hat{q}_{t+1}^{(1)}, Y_{t+1} \leq \hat{q}_{t+1}^{(2)}$, and $\hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$. (7).$Y_{t+1} \leq \hat{q}_{t+1}^{(1)}, Y_{t+1} \leq \hat{q}_{t+1}^{(2)}$ and $\hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$.(8).$Y_{t+1} \leq \hat{q}_{t+1}^{(1)}, Y_{t+1} \leq \hat{v}_{t+1}^{(2)}$ and $\hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$. In these eight cases, we can show that $\hat{D}_{R,P}$ has non-zero mean. See Appendix A2 for detail. □

Due to the bias of DM statistic, we want to develop the encompassing test to show that the encompassing statistic has zero mean. In order to find out the encompassing statistic for the ES, we build a combined FZ scoring function by combining two models, VaR q and ES e, and take the derivative of the expectation of the combined FZ scoring function with respect to the weight λ.

Theorem 2: Under $\mathbb{H}_0$,

$$C \equiv \mathbb{E}_F \left[m_1^{(1)} \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) + m_2^{(1)} \left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)} \right) \right] = 0. \tag{15}$$

$$\hat{C}_{R,P} \equiv P^{-1} \sum_{t=R}^{T} \left[\hat{m}_{1,t+1}^{(1)} \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) + \hat{m}_{2,t+1}^{(1)} \left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)} \right) \right] \xrightarrow{p} C = 0 \tag{16}$$

as $R, P \to \infty$ and $P/R \to \infty$. □

Proof: We combine model 1 and model 2 with weight $1 - \lambda$ and λ respectively. To estimate the expectation of FZ scoring function with combined VaR $q_{t+1}^{(c)}$ and ES $e_{t+1}^{(c)}$ to find out the the encompassing statistic under null hypothesis, we have

$$\lambda = \arg\min_{\lambda} \mathbb{E}_F S_\alpha (q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}),$$

where $q_{t+1}^{(c)} = (1 - \lambda) q_{t+1}^{(1)} + \lambda q_{t+1}^{(2)}$ and $e_{t+1}^{(c)} = (1 - \lambda) e_{t+1}^{(1)} + \lambda e_{t+1}^{(2)}$. We define C to be the first order condition of the expected FZ scoring function with respect to λ, then we have

$$\begin{aligned} C &\equiv \frac{\partial \mathbb{E}_F \left[S_\alpha (q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}) \right]}{\partial \lambda} \\ &= \frac{\partial \mathbb{E}_F S_\alpha (q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1})}{\partial v^{(c)}} \frac{\partial q_{t+1}^{(c)}}{\partial \lambda} + \frac{\partial \mathbb{E}_F S_\alpha (q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1})}{\partial e_{t+1}^{(c)}} \frac{\partial e_{t+1}^{(c)}}{\partial \lambda} \\ &= \mathbb{E} \left[m_1^{(c)} \left(q_{t+1}^{(2)} - q_{t+1}^{(1)} \right) + m_2^{(c)} \left(e_{t+1}^{(2)} - e_{t+1}^{(1)} \right) \right] \\ &= \mathbb{E} \left[m_1^{(1)} \left(q_{t+1}^{(2)} - q_{t+1}^{(1)} \right) + m_2^{(1)} \left(e_{t+1}^{(2)} - e_{t+1}^{(1)} \right) \right] = 0 \end{aligned}$$

under $\mathbb{H}_0$.

Due to the first order condition, C should be 0. We estimate C by $\hat{C}_{R,P}$, so $\hat{C}_{R,P}$ is defined as

$$\hat{C}_{R,P} \equiv P^{-1} \sum_{t=R}^{T} \left[\hat{m}_{1,t+1}^{(1)} \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) + \hat{m}_{2,t+1}^{(1)} \left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)} \right) \right] \xrightarrow{p} C = 0 \tag{17}$$

under $\mathbb{H}_0$, as $R, P \to \infty$ and $P/R \to \infty$

Under the null hypothesis, $\lambda = 0$, we obtain $q_{t+1}^{(c)} = q_{t+1}^{(1)}$ and $e_{t+1}^{(c)} = e_{t+1}^{(1)}$. Thus, we obtain $\mathbb{E}(\hat{C}_{R,P}) = 0$, so $\hat{C}_{R,P} \xrightarrow{p} 0$ as $R, P \to \infty$. See Appendix A2 for details. □

In order to get the encompassing statistic for ES, we consider endowing Model 1 and Model 2 with the weight $1 - \lambda$ and λ respectively, and find out the property of optimal weight $\hat{\lambda}$. Theorem 3 gives the relationship between $\hat{\lambda}$ and $\hat{C}_P$.

Theorem 3: We denote $\lambda = \arg\min_\lambda \mathbb{E}_F S\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right)$. Under $\mathbb{H}_0, \hat{\lambda} \to 0$, $\hat{C}_{R,P} \xrightarrow{p} 0$ as $R, P \to \infty$.

Proof: See Appendix B.

This theorem shows that $\hat{\lambda} \to 0$ and $\hat{C}_{R,P} \xrightarrow{p} 0$ as $R, P \to \infty$ are equivalent.

We compare the DM statistic and the encompassing statistic with the CCS statistic, the test statistic by Chao et al. (2001). Chao et al. (2001) show that the CCS statistic is $\hat{M}_{R,P} = P^{-1} \sum_{t=R}^{T} \hat{u}_{t+1}^{(1)} x_t$. In order to compare the equal predictive accuracy of two nested expectile models, we standardize these three statistics. The DM statistic is $\text{DM}_P \equiv \hat{S}_{R,P}^{-0.5} \sqrt{P} \hat{D}_{R,P}$, where $S_{R,P} = \text{var}\left(\sqrt{P}\hat{D}_{R,P}\right)$ and $S_{R,P} - \hat{S}_{R,P} \xrightarrow{p} 0$. The encompassing statistic is $\text{ENC}_P \equiv \hat{Q}_{R,P}^{-0.5} \sqrt{P} \hat{C}_{R,P}$, where $Q_{R,P} = \text{var}\left(\sqrt{P}\hat{C}_{R,P}\right)$ and $Q_{R,P} - \hat{Q}_{R,P} \xrightarrow{p} 0$. The CCS statistic is $\text{CCS}_P \equiv \hat{W}_{R,P}^{-0.5} \sqrt{P} \hat{M}_{R,P}$, where $W_{R,P} = \text{var}\left(\sqrt{P}\hat{M}_{R,P}\right)$ and $W_{R,P} - \hat{W}_{R,P} \xrightarrow{p} 0$.

4 Asymptotic Normality of the ENC Statistic for ES

In this section, we compare two nested models using encompassing test and model combination, then explore the properties of the two methodologies.

4.1 Encompassing Test for Equal Predictive Accuracy

We derive the asymptotic distribution of ENC_P. To show ENC_P to be asymptotical standard normality, we simplify the ENC_P statistic.

$$\text{ENC}_P = Q_P^{-1/2}\sqrt{P}\hat{C}_{R,P} = \frac{\sum_t c_t}{\sqrt{\sum_t c_t^2 - P\hat{C}_{R,P}^2}},$$

where

$$\hat{m}_{1,t+1} = -\frac{1}{\alpha\hat{e}_{t+1}^{(1)}}\mathbf{1}\{Y_{t+1} \leq \hat{q}_{t+1}^{(1)}\} + \frac{1}{\hat{e}_{t+1}^{(1)}}$$

$$\hat{m}_{2,t+1} = \frac{1}{\alpha\left(\hat{e}_{t+1}^{(1)}\right)^2}\mathbf{1}\left\{Y_{t+1} \leq \hat{q}_{t+1}^{(1)}\right\}\left(\hat{q}_{t+1}^{(1)} - Y_{t+1}\right) - \frac{\hat{q}_{t+1}^{(1)}}{\left(\hat{e}_{t+1}^{(1)}\right)^2} + \frac{1}{\hat{e}_{t+1}^{(1)}}.$$

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

In order to obtain the limiting distribution in Theorem 4, we need some notations. We define $g\left(u_{t+1}^{(i)}\right) = \left[\alpha - 1\left(u_{t+1}^{(i)} < 0\right)\right]$. The score $k_{i,t}(\beta_{i,t})$ is the first order condition of expected FZ scoring function with respect to $\beta_{i,t}$. Let $h_{i,t+1} = -k_{i,t+1}(\beta_{i,t})$ and $H_i(t+1) - \frac{1}{R}\sum_{t=1}^{R} h_{i,t+1}$. We denote the Hessian to be $B_i^{-1} = \mathbb{E}_F\Lambda(\beta_{i,t})$, which is the second order condition of expected FZ scoring function with respect to $\beta_{i,t}$. Therefore, $B_1 = 1$ and $B_2 = \begin{pmatrix} 1 & 0 \\ 0 & \mathbb{E}_F\Lambda(\beta_{2,t}) \end{pmatrix}^{-1}$. By the central limit theorem, $\sqrt{R}\left[R^{-1}\sum_{s=t-R+1}^{t} h_{2,s}\right] \sim N\left(0, Z^2B_2^{-1}\right)$, where $Z^2B_2^{-1} = \mathbb{E}[k_{i,t}(\beta_{i,t})k_{i,t}(\beta_{i,t})']$.

The following assumptions can be used to obtain the limiting distribution in Theorem 4. These assumptions are only sufficient but not necessary and sufficient.

Assumption 1: The parameter estimates $\hat{\beta}_{i,t}, i = 1, 2, t = R, \ldots, T$, satisfy $\hat{\beta}_{i,t} - \beta_i = B_i(t)H_i(t) = \left(R^{-1}\sum_{j=t-R+1}^{t}\Lambda_{i,j}\right)^{-1}\left(R^{-1}\sum_{j=t-R+1}^{t} h_{i,j}\right)$. □

Assumption 2: Let $U_t = \left[\mathbb{E}k_t(\theta), x_{2,t}' - \mathbb{E}x_{2,t}', h_{2,t}', \text{vec}\left(h_{2,t}h_{2,t}' - \mathbb{E}h_{2,t}h_{2,t}'\right),\right]$. (a) $\mathbb{E}U_t = 0$. (b) Denote $\tilde{U}_t$ to be the vector of U_t. $R^{-1}\text{E}\left(\sum_{j=t-R+1}^{t}\tilde{U}_j\right)\left(\sum_{j=t-R+1}^{t}\tilde{U}_j\right)'$ $= \Omega < \infty$ is p.d. □

Assumption 3: (a) $\mathbb{E}h_{2,t}h_{2,t}' = Z^2 B_2^{-1}$, (b) $\mathbb{E}\left(h_{2,t} \mid h_{2,t-j}, j = 1, 2, \ldots\right) = 0$. □

Assumptions 2 and 3 allow the application of an invariance principle and are sufficient for joint weak convergence of partial sums and averages of these partial sums to Brownian motion and integrals of Brownian motion.

Assumption 4: $\lim_{P,R\to\infty} P/R = \pi = \infty$. □

In order to get accurate out-of-sample forecasts, it is essential to select an optimal in-sample window. However, no consensus opinion has been reached as it depends on the characteristic of the model. Inoue et al. (2017) find the optimal window size by minimizing the conditional mean squared forecast error in a time-varying predictive regression model. They show that the optimal window size satisfies $R = O(T^{2/3})$. Hong et al. (2020) consider minimizing various out-of-sample forecast errors, including the unconditional, conditional, and global mean squared forecast error, respectively. The optimal window size they found is $R = O(T^{4/5})$. The out-of sample window P has a faster divergent rate than in sample window R. Thus, in our model, we set $P/R \to +\infty$.

Theorem 4: Let Assumptions 1-4 hold. $\text{ENC}_\text{P} \xrightarrow{d} N(0,1)$, as $P, R \to \infty$, under $\mathbb{H}_0 : \lambda = 0$. □

West (1996) proves that ENC statistic is asymptotically normal under $\pi \geq 0$ for mean regression when two models are non-nested. Clark and McCracken (2001) show that ENC_P statistic is asymptotically standard normal under $\pi = 0$, and ENC_P statistic is not standard normal under $\pi > 0$ when two mean models are nested models. In this theorem, we show that ENC_P statistic is asymptotically standard normality under π tends to ∞. See Appendix C for proof of Theorem 4.

5 Monte Carlo Simulation

5.1 Simulation Design

In order to show the asymptotic distribution of encompassing test for FZ scoring function, we have two DGPs.

In both two DGPs, we set x_t in Model 2 as an AR(1) stationary process, so $x_t = \phi x_{t-1} + v_t$, where $v_t \overset{iid}{\sim} N(0, \sigma_v^2)$. Thus, x has zero mean and variance $\sigma_v^2/(1-\phi)$. The error term $u_{t+1}^{(2)}$ of Model 2 for Value-at-risk (quantile) satisfies

$$\mathbb{E}\left(\alpha - \mathbf{1}\left(u_{t+1}^{(2)} < 0\right) \mid x_t\right) = 0 \tag{18}$$

which means that $\alpha - \mathbf{1}(u_{t+1}^{(2)} < 0)$ is a martingale difference series and the conditional quantile of $v_{t+1}^{(2)}$ given x_t is zero.

In DGP 1, we generate the u_{t+1}^2 following normal distribution. The mean and variance of u_{t+1}^2 satisfy

$$\frac{\mathbb{E}\left(u_{t+1}^{(2)}\right)}{\sqrt{\text{Var}\left(u_{t+1}^{(2)}\right)}} = \frac{-\Phi^{-1}(\alpha)\sigma_u}{\sigma_u} = -\Phi^{-1}(\alpha) \tag{19}$$

We set the $\text{Var}\left(u_{t+1}^{(2)}\right) = \sigma_u^2$, then we can get $\mathbb{E}\left(u_{t+1}^{(2)}\right) = -\Phi^{-1}(\alpha)\sigma_u$. Second, we generate $\{y_{t+1}\}$ from $y_{t+1} = c_2 + bx_t + e_{t+1}^{(2)}$, then we has the mean and variance as follows:

$$\mathbb{E}\left(y_{t+1}\right)=\mathbb{E}\left(c_{12}+b_1x_t+u_{t+1}^{(2)}\right)=c_{12}+0-\Phi^{-1}(\alpha)\sigma_u$$

$$\text{Var}\left(y_{t+1}\right)=\text{Var}\left(c_{12}+b_1x_t+e_{t+1}^{(2)}\right)=b_1^2\sigma_x^2+\sigma_u^2$$

In order to make sure all $\hat{v}_{t+1}$ and $\hat{e}_{t+1}$ to be negative, we set $\mathbb{E}(Y)=-3$, which implies that $c_{12}=-3+\Phi^{-1}(\alpha)\sigma_u$. Third, we find out mean and variance of $u_{t+1}^{(1)}$ as follows:

$$\mathbb{E}\left(u_{t+1}^{(1)}\right)=\mathbb{E}\left(y_{t+1}-c_{11}\right)=\mathbb{E}\left(c_{12}+b_1x_t+u_{t+1}^{(2)}-c_{11}\right)=c_{12}+0-\Phi^{-1}(\alpha)\sigma_u-c_{11}$$

$$\text{Var}\left(u_{t+1}^{(1)}\right)=\text{Var}\left(y_{t+1}-c_{11}\right)=\text{Var}\left(c_{12}+b_1x_t+e_{t+1}^{(2)}-c_{11}\right)=b_1^2\sigma_x^2+\sigma_u^2$$

We set $u_{t+1}^{(1)}$ following normal distribution and should satisfy

$$\frac{\mathbb{E}\left(u_{t+1}^{(1)}\right)}{\sqrt{\text{Var}\left(u_{t+1}^{(1)}\right)}}=\frac{c_{12}-c_{11}-\Phi^{-1}(\alpha)\sigma_u}{\sqrt{b_1^2\sigma_x^2+\sigma_u^2}}=-\Phi^{-1}(\alpha) \tag{20}$$

Simplifying the above equation, $c_{12}-c_{11}=\Phi^{-1}(\alpha)\sigma_u-\Phi^{-1}(\alpha)\sqrt{b_1^2\sigma_x^2+\sigma_e^2}$. Since $c_{12}=-3+\Phi^{-1}(\alpha)\sigma_u$, $c_{11}=-3+\Phi^{-1}(\alpha)\sqrt{b_1^2\sigma_x^2+\sigma_e^2}$. Under the null, when $b_1=0$, $c_{11}=c_{12}$. We consider $\alpha\in\{0.01,0.025,0.05,0.1\},\phi\in\{0,0.95\},\sigma_u\in\{1\},\sigma_v\in\{1\}$, and $b\in\{0,0.1\}$.

In DGP2, we generate y_{t+1} by using GARCH(1,1).

$$y_{t+1}=\sigma_{t+1}u_{t+1}$$

$$\sigma_{t+1}=\omega+\zeta u_t^2+\beta\sigma_t^2+\delta x_t^2$$

$$u_t\sim N(0,1)$$

We set the parameters to be $\zeta=0.05$, $\beta=0.85$, $b\in\{0,0.1\}$, and $\omega=1-\zeta-\beta-\delta$. We make x_t follow $N(0,1)$. According to Patton et al. (2019), Under DGP2, by using standard normal distribution for u_{t+1},

$$\text{VaR}_{\alpha,t+1}=a_\alpha\sigma_{t+1} \qquad a_\alpha=\Phi^{-1}(\alpha)$$

$$\text{ES}_{\alpha,t+1}=b_\alpha\sigma_{t+1} \qquad b_\alpha=-\phi\left(\Phi^{-1}(\alpha)\right)/\alpha$$

In our simulation, we set the number of in-sample observations $R\in\{120,240,480\}$ and out-of-sample forecasts $P \in \{240, 480, 1200\}$. We use rolling windows to estimate θ by minimizing S_{FZ0}. In DGP1, $\theta_1 = c_1$ for Model 1 and $\theta_2 = (c_2, b)'$ for Model 2. Then, we can obtain the one step ahead forecast $q^{(1)}_{t+1} = c_1$ and $e^{(1)}_{t+1} = \mathbb{E}_F(Y_{t+1}|Y_{t+1} \leq q^{(1)}_{t+1})$ for Model 1 and $q^{(2)}_{t+1} = c_2 + bx_t$ and $e^{(2)}_{t+1} = \mathbb{E}_F(Y_{t+1}|Y_{t+1} \leq q^{(2)}_{t+1}$ for Model 2. In DGP2, $\theta_1 = (\omega_1, \zeta_1, \beta_1, a_{1,\alpha}, b_{1,\alpha})$ for Model 1 and $\theta_2 = (\omega_2, \zeta_2, \beta_2, a_{2,\alpha}, b_{2,\alpha}, \delta)$ for Model 2. Then, we can obtain the one step ahead forecast $v^{(1)}_{t+1} = a_{1,\alpha}\sigma^{(1)}_{t+1}$ and $e^{(1)}_{t+1} = b_{1,\alpha}\sigma^{(1)}_{t+1}$ for Model 1 and $q^{(2)}_{t+1} = a_{1=2,\alpha}\sigma^{(2)}_{t+1}$ and $e^{(2)}_{t+1} = b_{2,\alpha}\sigma^{(2)}_{t+1}$ for Model 2. By using those forecast variables, we can obtain $\text{DM}_{R,P}$, $\text{ENC}_{R,P}$, and $\text{CCS}_{R,P}$ statistics. Repeating this procedure 2000 times, we get the asymptotic distributions of $\text{DM}_{R,P}$, $\text{ENC}_{R,P}$, and $\text{CCS}_{R,P}$, and the size and power of these three statistics.

5.2 Simulation result on encompassing test

Figures 1-2 and tables 1-2 show the Monte Carlo simulation in DGP1 for asymptotic distribution and the size and powers of $\text{DM}_{R,P}$, $\text{ENC}_{R,P}$, and $\text{CCS}_{R,P}$ statistics under different α, b, and ϕ. Table 1 shows the result of the size of the test under $b = 0$ and $\phi = 0$. For different in-sample observations R, out-of-sample forecasts P, and α, the result demonstrates that $\text{DM}_{R,P}$ is much less than 5% under the 5% nominal level, implying that DM_P has a downward bias under $\mathbb{H}_0$. However, the sizes of $\text{ENC}_{R,P}$ and $\text{CCS}_{R,P}$ are good under the 5% nominal level. Table 2 shows the result of the power of the test under $b = 0.1$ and $\phi = 0.95$. It illustrates that $\text{ENC}_{R,P}$ has the highest power and $\text{DM}_{R,P}$ has the higher power than $\text{CCS}_{R,P}$ for different in-sample observations R, out-of-sample forecasts P and α.

Figure 1 shows the asymptotic distribution of $\text{DM}_{R,P}$, $\text{ENC}_{R,P}$, and $\text{CCS}_{R,P}$ statistics under $\mathbb{H}_0$ when $\alpha = 0.05$. These figures demonstrate that $\text{DM}_{R,P}$ has a negative mean and high kurtosis, which implies that $\text{DM}_{R,P}$ has a downward bias under $\mathbb{H}_0$. However, the asymptotic distributions of $\text{ENC}_{R,P}$ and $\text{CCS}_{R,P}$ are close to the standard normal distribution. Figure 2 shows the asymptotic distribution of $\text{DM}_{R,P}$, $\text{ENC}_{R,P}$ and $\text{CCS}_{R,P}$ statistic under $\alpha = 0.05$, $b = 0.1$ and $\phi = 0, 0.95$.

From these figures, we can see that the asymptotic distribution of $\text{CCS}_{R,P}$ move to the left, so CCS_{P} has the lowest mean compared to $\text{DM}_{R,P}$ and $\text{ENC}_{R,P}$. Moreover, the asymptotic distribution of $\text{DM}_{R,P}$ and $\text{ENC}_{R,P}$ move to the right, and the means of $\text{DM}_{R,P}$ are lower than means of $\text{ENC}_{R,P}$.

6 Empirical Analysis

6.1 Empirical Analysis of VaR and ES for Monthly Equity Premium

In this subsection, we use the data from Welch and Goyal (2008) to check the predictive variable in VaR and ES for Equity Premium. We choose to use monthly data from January 1926 to December 2019, which contains 1128 observations. There is a total of 15 Macroeconomic and financial variables.

We build two nested models for VaR and ES to test if these 15 variables Granger cause the equity premium respectively. For Model 1, we find out the VaR and ES by using the previous R in-sample observations of the equity premium and then use estimated VaR and ES to predict the future equity premium at time $R+1$. For Model 2, we estimate the constant term and covariance term using the previous R in-sample observations of equity premium and an independent variable and then use these estimated coefficients and the independent variable to forecast the future equity premium. In this empirical work, the forecasts begin 40 years after the data are available, and the out-of-sample forecasts begin 1965. We choose $\alpha \in \{0.01, 0.025, 0.05, 0.1\}$.

Tables 5 - 6 show the ENC statistics for 15 Macroeconomic and financial variables under four different α. In Table 5, forecasts begin 40 years after the data are available, which means the number of in-sample observations $R = 480$. From this table, when $\alpha = 0.01$, we can see that the dividend payout ratio is the only variable that Granger Cause the equity premium in VaR and ES at a 10% significance level. When we choose $\alpha = 0.025$, the earning price ratio, stock variance, and default yield spread Granger Cause the equity premium on VaR and ES respectively at a 5% significance level. Moreover, the equity premium can be predicted by the dividend yield, treasury-bill rate, default return spread, and consumption, wealth, income ratio in VaR and ES at $\alpha = 0.05$, respectively. When $\alpha = 0.1$, seven variables Granger Cause the equity premium in VaR and ES: dividend price ratio, dividend yield, earning price ratio, dividend payout ratio, default return spread, long term return, and consumption, wealth, income ratio.

Table 6 demonstrates the ENC statistics for 15 variables when out-of-sample forecasts begin 1965. From this table, when $\alpha = 0.01$, we can see that the term spread is the only variable that Granger Cause the equity premium in VaR and ES at a 10% significance level. The long term return and earning price ratio Granger Cause the equity premium in VaR and ES at a 1% significance level when $\alpha = 0.025$ and $\alpha = 0.05$ respectively. Moreover, when $\alpha = 0.1$, the equity premium can be predicted by the dividend price ratio, dividend yield, earning price ratio, default return spread, and consumption, wealth, income ratio. From Tables 5 - 6, we can see that the decreasing number of variables can Granger Cause the equity premium when we decrease α. Thus, fewer variables have the predictive ability of the equity premium in VaR and ES when we focus on extreme losses. Compared to forecasts beginning 1965, the equity premium can be predicted by more variables at different α when forecasts begin 40 years after the data are available. Thus, more variables can Granger Cause the equity premium when the number of in-sample observations becomes small.

6.2 Empirical Analysis of GaR and GS for Quarterly GDP Growth

Measures of the downside risk are essential in risk management. The increasing number of policymakers has focused on the downside risk in the last decade. The International Monetary Fund (IMF) has recently popularized a risk measure for GDP growth called Growth-at-Risk (GaR). GaR is the worst conditional GDP growth distribution at a given coverage level (5th percentile) depending on financial conditions (Adrian et al., 2019). Moreover, Adrian et al. (2019); Chavleishvili et al. (2021) define a measure of adverse real economic impact to be Growth Shortfall (GS). GS is the expectation of the GDP growth when it is less than GaR. In order to check if financial conditions have the predictive ability of the Macroeconomics risk, we choose to use the National Financial Conditions Index (NFCI), which is computed by the Federal Reserve Bank of Chicago. The NFCI is calculated from a dynamic factor model with 105 financial variables to estimate the weekly US financial conditions. Financial conditions are tighter than average when NFCI is positive, while financial conditions are looser than average when NFCI is negative.

In this subsection, we use the data from Adrian et al. (2019) to check if the NFCI can predict GDP growth on GaR and GS. The data are Quarterly from 1973Q1 to 2019Q4. There is a total of 188 observations. We consider GDP growth as our y_{t+h} and NFCI as our x_t. The test is an h-step out-of-sample Granger Causality test. We choose $h \in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$. Forecasts of GaR and GS begin 20 years after data are available, which means $R = 80$. The rolling window estimation is used in this test. We also use the FZ scoring function for a pair of GaR and GS to examine the forecast encompassing test for Granger Causality on GaR and GS.

Figure 5 shows the time series of the real GDP growth and the NFCI. This figure is the same as Figure 2 in Adrian et al. (2019). Due to the properties of the GDP growth and the NFCI, there is a negative relationship between these two variables. In the regular periods, when GDP growth is positive, GDP growth is more volatile than the NFCI. During the National Bureau of Economic Research (NBER) recession dates, the GDP growth reaches negative outcomes, and the NFCI reaches positive outcomes. Figure 5 demonstrates the negative relationship between GDP growth and NFCI. Thus, deteriorations in financial conditions coincide with the decreases in GDP growth. Then, we want to use the forecast encompassing test to examine if the NFCI Granger Causes the GDP growth on GaR and GS by using the FZ loss function.

Table 7 displays the ENC statistics and p-values of the Granger Causality test for various h when $\alpha = 0.05$. This table shows that the NFCI Granger Causes the four quarters ahead (one year) GDP growth on GaR and GS at the 10% significance level. It implies that the downside risk will increase in four quarters when the financial conditions become tighter in the current quarter. However, the NFCI cannot Granger Cause GDP growth in the other quarters ahead.

Moreover, we want to explore if there are significant changes in $\hat{C}_{R,t}$ in different periods. Section 2 states that the encompassing statistic $\text{ENC}_{R,P}$ can be obtained by standardizing the $\hat{C}_{R,t}$. Thus, we plot $\hat{C}_{R,t}$ for various of h to check the change in $\hat{C}_{R,t}$ in the out-of-sample period when $\alpha = 0.05$. Figure 6 shows the time series of $\hat{C}_{R,t}$, $\hat{C}_{1,R,t}$, and $\hat{C}_{2,R,t}$, where $\hat{C}_{1,R,t} = \hat{m}_{1,t}\left(G\hat{a}R_{2,t} - G\hat{a}R_{1,t}\right)$ is the GaR term in the $\hat{C}_{R,t}$, and $\hat{C}_{2,R,t} = \hat{m}_{2,t}\left(\hat{GS}_{2,t} - \hat{GS}_{1,t}\right)$ is the GS term in the $\hat{C}_{R,t}$. From Figure 6, we can see that there are pikes of $\hat{C}_{R,t}$, $\hat{C}_{1,R,t}$, and $\hat{C}_{2,R,t}$ for various h during the NBER recession dates. The spike becomes smaller when the quarter ahead h becomes larger. When h is zero, current quarter GDP growth and NFCI are used. Then the spike is more considerable. The spike becomes negligible as the number of quarters ahead h goes up.

Furthermore, we check if there are differences between the forecast GaR without the NFCI and with the NFCI. Figure 7 demonstrates the changes of the $G\hat{a}R_{1,R,t}$, $G\hat{a}R_{2,R,t}$, $\hat{GS}_{1,R,t}$, and $\hat{GS}_{2,R,t}$ in the out-of-sample period, where $G\hat{a}R_{1,R,t}$ and $G\hat{a}R_{2,R,t}$ are the forecast GaR without the NFCI and with the NFCI, and $\hat{GS}_{1,R,t}$ and $\hat{GS}_{2,R,t}$ are the forecast GS without the NFCI and with the NFCI. This figure displays that the difference between $G\hat{a}R_{1,R,t}$ and $G\hat{a}R_{2,R,t}$ and the difference between $\hat{GS}_{1,R,t}$ and $\hat{GS}_{2,R,t}$ become small during the NBER recession dates when the number of quarters ahead h increases from 0 to 8. For the NBER recession dates from 2007Q4 to 2009Q2, the difference between $G\hat{a}R_{1,R,t}$ and $G\hat{a}R_{2,R,t}$ and the difference between $\hat{GS}_{1,R,t}$ and $\hat{GS}_{2,R,t}$ are large when $h = 0, 1, 2, 3, 4$. However, the difference between $G\hat{a}R_{1,R,t}$ and $G\hat{a}R_{2,R,t}$ is very small, and the difference between $\hat{GS}_{1,R,t}$ and $\hat{GS}_{2,R,t}$ is almost zero when $h = 5, 6, 7, 8$.

7 Conclusion

In this paper, I develop a new forecast encompassing test for Granger Causality for VaR and ES. The strictly consistent scoring function for VaR and ES is from Fissler and Ziegel (2016) and Patton et al. (2019). I show that the DM statistic has a bias for the FZ scoring function with two nested models under the null hypothesis. I exhibition that the encompassing statistic (ENC) has zero mean under the null hypothesis with no Granger Causality. I prove that the ENC statistic is asymptotically standard normal. Monte Carlo simulation demonstrates that the ENC statistic has good size and has the highest power under the alternative hypothesis $\mathbb{H}_1$ in finite samples compared to DM and CCS statistics. I demonstrate that ENC statistic improves the predictive accuracy compared with DM test. I consider two applications. The first empirical analysis of VaR and ES illustrates some Macroeconomic and financial variables Granger Cause the equity premium in VaR and ES at different α. The second empirical analysis of GaR and GS shows that the NFCI Granger Causes the four quarters ahead GDP growth on GaR and GS when α is 0.05.

References

Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, *109*(4), 1263–89.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Ashley, R., Granger, C. W., & Schmalensee, R. (1980). Advertising and aggregate consumption: An analysis of causality. *Econometrica*, 1149–1167.

Chao, J., Corradi, V., & Swanson, N. R. (2001). Out-of-sample tests for granger causality. *Macroeconomic Dynamics*, *5*(4), 598–620.

Chavleishvili, S., Engle, R. F., Fahr, S., Kremer, M., Manganelli, S., & Schwaab, B. (2021). The risk management approach to macro-prudential policy.

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

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

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and osband's principle. *The Annals of Statistics*, *44*(4), 1680–1707.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, 489–500.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

Hong, Y., Sun, Y., & Wang, S. (2020). Selection of an optimal rolling window in time-varying predictive regression. *Cornell University Working Paper*.

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, *196*(1), 55–67.

Osband, K. (1985). *Providing Incentives for Better Cost Forecasting* (Unpublished doctoral dissertation). University of California, Berkeley.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Patton, A. J., & Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In *Handbook of financial time series* (pp. 801–838). Springer.

Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, *211*(2), 388–413.

Phillips, P. C. B. (1991). A shortcut to lad estimator asymptotics. *Econometric Theory*, 450–463.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics*, *37*(1), 121–133.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 1067–1084.

Appendix A

A1: Proof of Theorm 1.

DM statistics of FZ loss-differential can be rewrite as follows:

$$\begin{aligned}\hat{D}_P =& P^{-1}\sum_{t=R}^{T}\Bigg[-\frac{1}{\alpha\hat{e}_{t+1}^{(1)}}\mathbf{1}\left\{Y_{t+1}\le\hat{q}_{t+1}^{(1)}\right\}\left(\hat{q}_{t+1}^{(1)}-Y_{t+1}\right)+\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}}+\log\left(-\hat{e}_{t+1}^{(1)}\right)\\&+\frac{1}{\alpha\hat{e}_{t+1}^{(2)}}\mathbf{1}\left\{Y_{t+1}\le\hat{q}_{t+1}^{(2)}\right\}\left(\hat{q}_{t+1}^{(2)}-Y_{t+1}\right)-\frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}}-\log\left(-\hat{e}_{t+1}^{(2)}\right)\Bigg]\end{aligned}$$

In order to prove that $\hat{D}_P$ is non-zero at each t, we denote $\hat{D}_P = P^{-1}\sum_{t=R}^{T} d_t$ and show d_t is non-zero at each t.

(1). $Y_{t+1} > \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} > \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$

$$d_t = \underbrace{\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}}}_{>0}-\underbrace{\frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}}}_{<0}+\underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right)-\log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{<0}$$

(2). $Y_{t+1} > \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} > \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$

$$d_t = \underbrace{\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}}}_{>0}-\underbrace{\frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}}}_{<0}+\underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right)-\log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{>0}$$

(3). $Y_{t+1} \le \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} > \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$

$$\begin{aligned}d_t &= P^{-1}\sum_{t=R}^{T}\left[-\frac{\hat{q}_{t+1}^{(1)}-Y_{t+1}}{\alpha\hat{e}_{t+1}^{(1)}}+\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}}+\log\left(-\hat{e}_{t+1}^{(1)}\right)-\frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}}-\log\left(-\hat{e}_{t+1}^{(2)}\right)\right]\\&=\underbrace{\frac{(\alpha-1)\,\hat{q}_{t+1}^{(1)}+Y_{t+1}}{\alpha\hat{e}_{t+1}^{(1)}}}_{<0}-\underbrace{\frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}}}_{<0}+\underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right)-\log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{<0}<0\end{aligned}$$

(4). $\quad Y_{t+1} \leq \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} > \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$

$$d_t = P^{-1} \sum_{t=R}^{T} \left[-\frac{\hat{q}_{t+1}^{(1)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(1)}} + \frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}} + \log\left(-\hat{e}_{t+1}^{(1)}\right) - \frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}} - \log\left(-\hat{e}_{t+1}^{(2)}\right) \right]$$

$$= \underbrace{\frac{(\alpha-1)\,\hat{q}_{t+1}^{(1)} + Y_{t+1}}{\alpha \hat{e}_{t+1}^{(1)}}}_{<0} \underbrace{- \frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}}}_{<0} + \underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right) - \log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{>0}$$

(5). $\quad Y_{t+1} > \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} \leq \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$

$$d_t = P^{-1} \sum_{t=R}^{T} \left[\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}} + \log\left(-\hat{e}_{t+1}^{(1)}\right) + \frac{\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}} - \frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}} - \log\left(-\hat{e}_{t+1}^{(2)}\right) \right]$$

$$= \underbrace{\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}}}_{>0} + \underbrace{\frac{(1-\alpha)\,\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}}}_{>0} + \underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right) - \log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{<0}$$

(6). $\quad Y_{t+1} > \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} \leq \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$

$$d_t = P^{-1} \sum_{t=R}^{T} \left[\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}} + \log\left(-\hat{e}_{t+1}^{(1)}\right) + \frac{\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}} - \frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}} - \log\left(-\hat{e}_{t+1}^{(2)}\right) \right]$$

$$= \underbrace{\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}}}_{>0} + \underbrace{\frac{(1-\alpha)\,\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}}}_{>0} + \underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right) - \log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{>0} > 0$$

(7). $\quad Y_{t+1} \leq \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} \leq \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} > \hat{e}_{t+1}^{(2)}$

$$d_t = P^{-1} \sum_{t=R}^{T} \left[-\frac{\hat{q}_{t+1}^{(1)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(1)}} + \frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}} + \log\left(-\hat{e}_{t+1}^{(1)}\right) + \frac{\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}} - \frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}} - \log\left(-\hat{e}_{t+1}^{(2)}\right) \right]$$

$$= \underbrace{\frac{(\alpha-1)\,\hat{q}_{t+1}^{(1)} + Y_{t+1}}{\alpha \hat{e}_{t+1}^{(1)}}}_{<0} + \underbrace{\frac{(1-\alpha)\,\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}}}_{>0} + \underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right) - \log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{<0}$$

(8). $\quad Y_{t+1} \leq \hat{q}_{t+1}^{(1)}, \quad Y_{t+1} \leq \hat{q}_{t+1}^{(2)}, \quad \text{and} \quad \hat{e}_{t+1}^{(1)} < \hat{e}_{t+1}^{(2)}$

$$d_t = P^{-1} \sum_{t=R}^{T} \left[-\frac{\hat{q}_{t+1}^{(1)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(1)}} + \frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}} + \log\left(-\hat{e}_{t+1}^{(1)}\right) + \frac{\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}} - \frac{\hat{q}_{t+1}^{(2)}}{\hat{e}_{t+1}^{(2)}} - \log\left(-\hat{e}_{t+1}^{(2)}\right) \right]$$

$$= \underbrace{\frac{\hat{q}_{t+1}^{(1)}}{\hat{e}_{t+1}^{(1)}}}_{>0} + \underbrace{\frac{(1-\alpha)\,\hat{q}_{t+1}^{(2)} - Y_{t+1}}{\alpha \hat{e}_{t+1}^{(2)}}}_{>0} + \underbrace{\left[\log\left(-\hat{e}_{t+1}^{(1)}\right) - \log\left(-\hat{e}_{t+1}^{(2)}\right)\right]}_{>0}$$

Thus, we find out$\mathbb{E}\hat{D}_P \neq 0,$ so the DM statistics has a bias.

A2: Proof of Theorem 2.

We estimate the FZ scoring function with combined error term to find out the the encompassing statistic. We have

$$\lambda = \arg\min_{\lambda} \mathbb{E}S_\alpha\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right)$$

where $q_{t+1}^{(c)} = (1-\lambda)\, q_{t+1}^{(1)} + \lambda q_{t+1}^{(2)}$ and $e_{t+1}^{(c)} = (1-\lambda)\, e_{t+1}^{(1)} + \lambda \hat{e}_{t+1}^{(2)}$. First, we rewrite the expectation of the FZ scoring function with combined functionals $q_{t+1}^{(c)}$ and $e_{t+1}^{(c)}$ as follows:

$$\begin{aligned}
\mathbb{E}_F S_\alpha\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right) &= \int_{-\infty}^{+\infty} -\frac{1}{\alpha e_{t+1}^{(c)}}\mathbf{1}\{Y_{t+1} \le q_{t+1}^{(c)}\}(q_{t+1}^{(c)} - Y_{t+1})\mathrm{d}F_t(Y_{t+1}) + \frac{q_{t+1}^{(c)}}{e_{t+1}^{(c)}} + \log(-e_{t+1}^{(c)}) - 1 \\
&= \int_{-\infty}^{q_{t+1}^{(c)}} -\frac{1}{\alpha e_{t+1}^{(c)}}(q_{t+1}^{(c)} - Y_{t+1})\mathrm{d}F_t(Y_{t+1}) + \frac{q_{t+1}^{(c)}}{e_{t+1}^{(c)}} + \log(-e_{t+1}^{(c)}) - 1 \\
&= \int_{-\infty}^{0} \frac{1}{\alpha e_{t+1}^{(c)}} Z_{t+1}\mathrm{d}F_t(Z_{t+1} + q_{t+1}^{(c)}) + \frac{q_{t+1}^{(c)}}{e_{t+1}^{(c)}} + \log(-e_{t+1}^{(c)}) - 1 \\
&= Z_{t+1}F_t(Z_{t+1} + q_{t+1}^{(c)})|_{-\infty}^{0} - \int_{-\infty}^{0} \frac{1}{\alpha e_{t+1}^{(c)}} F_t(Z_{t+1} + q_{t+1}^{(c)})\mathrm{d}Z_{t+1} \\
&\quad + \frac{q_{t+1}^{(c)}}{e_{t+1}^{(c)}} + \log(-e_{t+1}^{(c)}) - 1 \\
&= -\frac{1}{\alpha e_{t+1}^{(c)}} \int_{-\infty}^{0} F_t(Z_{t+1} + q_{t+1}^{(c)})\mathrm{d}Z_{t+1} + \frac{q_{t+1}^{(c)}}{e_{t+1}^{(c)}} + \log(-e_{t+1}^{(c)}) - 1
\end{aligned}$$

Taking derivative of expectation of the FZ scoring function with combined functionals $q^{(c)}$ and $e^{(c)}$ with respect to combined functionals $q^{(c)}$ and $e^{(c)}$, we have

$$\begin{aligned}
m_{1,t+1}^{(c)} &= \frac{\partial \mathbb{E}_F\left[S_\alpha\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right)\right]}{\partial q_{t+1}^{(c)}} \\
&= -\frac{1}{\alpha e_{t+1}^{(c)}} \int_{-\infty}^{0} f_t(Z_{t+1} + q_{t+1}^{(c)})\mathrm{d}Z_{t+1} + \frac{1}{e_{t+1}^{(c)}} \\
&= -\frac{1}{\alpha e_{t+1}^{(c)}} \int_{-\infty}^{q_{t+1}^{(c)}} f_t(Y_{t+1})\mathrm{d}Y_{t+1} + \frac{1}{e_{t+1}^{(c)}} \\
&= -\frac{1}{\alpha e_{t+1}^{(c)}} F_t(q_{t+1}^{(c)}) + \frac{1}{e_{t+1}^{(c)}} \\
&= \mathbb{E}_F\left[-\frac{1}{\alpha e_{t+1}^{(c)}}\mathbf{1}\{Y_t \le q_{t+1}^{(c)}\} + \frac{1}{e_{t+1}^{(c)}}\right]
\end{aligned}$$

$$
\begin{aligned}
m_{2,t+1}^{(c)} &= \frac{\partial \mathbb{E}_F\left[S_\alpha\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right)\right]}{\partial e_{t+1}^{(c)}} \\
&= \frac{1}{\alpha\left(e_{t+1}^{(c)}\right)^2}\int_{-\infty}^{+\infty} \mathbf{1}\{Y_{t+1} \leq q_{t+1}^{(c)}\}(q_{t+1}^{(c)} - Y_{t+1})\mathrm{d}F_t(Y_{t+1}) - \frac{q_{t+1}^{(c)}}{\left(e_{t+1}^{(c)}\right)^2} + \frac{1}{e_{t+1}^{(c)}} \\
&= \mathbb{E}_F\left[\frac{1}{\alpha\left(e_{t+1}^{(c)}\right)^2}\mathbf{1}\left\{Y_{t+1} \leq \hat{q}_{t+1}^{(c)}\right\}\left(q_{t+1}^{(c)} - Y_{t+1}\right) - \frac{q_{t+1}^{(c)}}{\left(e_{t+1}^{(c)}\right)^2} + \frac{1}{\hat{e}_{t+1}^{(c)}}\right]
\end{aligned}
$$

Taking derivative of combined functionals $q^{(c)}$ and $e^{(c)}$ with respect to weight λ, we obtain

$$
\begin{aligned}
\frac{\partial q_{t+1}^{(c)}}{\partial \lambda} &= \left(q_{t+1}^{(2)} - q_{t+1}^{(1)}\right) \\
\frac{\partial e_{t+1}^{(c)}}{\partial \lambda} &= \left(e_{t+1}^{(2)} - e_{t+1}^{(1)}\right)
\end{aligned}
$$

In order to find the optimal weight λ^*, we take the first order condition of the expected FZ scoring function with respect to λ.

F.O.C w.r.t λ:

$$
\begin{aligned}
C &\equiv \frac{\partial \mathbb{E}_F\left[S_\alpha\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right)\right]}{\partial \lambda} \\
&= \frac{\partial \mathbb{E}_F S_\alpha\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right)}{\partial q_{t+1}^{(c)}}\frac{\partial q_{t+1}^{(c)}}{\partial \lambda} + \frac{\partial \mathbb{E}_F S_\alpha\left(q_{t+1}^{(c)}, e_{t+1}^{(c)}, Y_{t+1}\right)}{\partial e_{t+1}^{(c)}}\frac{\partial e_{t+1}^{(c)}}{\partial \lambda} \\
&= \mathbb{E}_F\left[m_{1,t+1}^{(1)}\left(q_{t+1}^{(2)} - q_{t+1}^{(1)}\right) + m_{2,t+1}^{(c)}\left(e_{t+1}^{(2)} - e_{t+1}^{(1)}\right)\right] \\
&= \mathbb{E}_F\left[m_{1,t+1}^{(1)}\left(q_{t+1}^{(2)} - q_{t+1}^{(1)}\right) + m_{2,t+1}^{(2)}\left(e_{t+1}^{(2)} - e_{t+1}^{(1)}\right)\right] = 0 \qquad \text{under } \mathbb{H}_0
\end{aligned}
$$

$$
\begin{aligned}
\hat{C}_P &= P^{-1}\sum_{t=R}^{T}\left[\hat{m}_{1,t+1}^{(c)}\left(\hat{q}_{t+1}^{(1)} - \hat{q}_{t+1}^{(2)}\right) + \hat{m}_{2,t+1}^{(c0}\left(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)}\right)\right] \\
&= P^{-1}\sum_{t=R}^{T}\left[\hat{m}_{1,t+1}^{(1)}\left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)}\right) + \hat{m}_{2,t+1}^{(2)}\left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)}\right)\right] \\
&\xrightarrow{p} C = 0 \quad \text{ under } \mathbb{H}_0 \quad \text{ as } R, P \to \infty \text{ and } P/R \to \infty.
\end{aligned}
$$

Under $\mathbb{H}_0, \lambda = 0$, then we have $q_{t+1}^{(c)} = q_{t+1}^{(1)}$ and $e_{t+1}^{(c)} = e_{t+1}^{(1)}$. Thus, $\mathbb{E}(\hat{C}_P) = 0$, so $\hat{C}_P \xrightarrow{p} 0$, as $R, P \to \infty$ and $R, P \to \infty$.

Appendix B: Proof of Theorem 3

We have the combined VaR $q^{(c)}_{t+1} = (1-\lambda)q^{(1)}_{t+1} + \lambda q^{(2)}_{t+1}$ and ES $e^{(c)}_{t+1} = (1-\lambda)e^{(1)}_{t+1} + \lambda e^{(2)}_{t+1}$. The expected loss function from combining forecast is

$$\mathbb{E}_F S\left(q^{(c)}_{t+1}, e^{(c)}_{t+1}, Y_{t+1}\right) = \mathbb{E}_F\left[-\frac{1}{\alpha e^{(c)}_{t+1}}\mathbf{1}\{Y_{t+1} \le q^{(c)}_{t+1}\}\left(q^{(c)}_{t+1} - Y_{t+1}\right) + \frac{q^{(c)}_{t+1}}{e^{(c)}_{t+1}} + \log\left(-e^{(c)}_{t+1}\right) - 1\right] \tag{21}$$

Taking the first order condition of Equation (21) with respect to λ, we have

$$C^{(c)} \equiv \nabla_\lambda \mathbb{E}S\left(q^{(c)}_{t+1}, e^{(c)}_{t+1}, Y_{t+1}\right) = \mathbb{E}\left[m^{(c)}_{1,t+1}\left(q^{(2)}_{t+1} - q^{(1)}_{t+1}\right) + m^{(c)}_{2,t+1}\left(e^{(2)}_{t+1} - e^{(1)}_{t+1}\right)\right] = 0$$

To show the necessary condition, if $\lambda = 0$, then $\hat{q}^{(c)}_{t+1} = \hat{q}^{(1)}_{t+1}$ and $\hat{e}^{(c)}_{t+1} = \hat{e}^{(1)}_{t+1}$, so that $\hat{m}^{(c)}_{1,t+1} = \hat{m}^{(1)}_{1,t+1}$, $\hat{m}^{(c)}_{2,t+1} = \hat{m}^{(1)}_{2,t+1}$. we have

$$\hat{C}^{(c)}_P \equiv P^{-1}\sum_{t=R}^{T}\left[\hat{m}^{(1)}_{1,t+1}\left(q^{(2)}_{t+1} - q^{(1)}_{t+1}\right) + \hat{m}^{(1)}_{2,t+1}\left(e^{(2)}_{t+1} - e^{(1)}_{t+1}\right)\right] \xrightarrow{p} 0$$

as shown in Theorem 2 under $\mathbb{H}_0$. To show the sufficient condition, taking the second order condition of Equation (21) with respect to λ, we have

$$\begin{aligned}
&\frac{\partial^2 \mathbb{E}_F S\left(q^{(c)}_{t+1}, e^{(c)}_{t+1}, Y_{t+1}\right)}{\partial \lambda^2} = \frac{\partial C^{(c)}}{\partial \lambda} \\
=&\mathbb{E}_F\left[\frac{1}{\alpha\left(e^{(c)}_{t+1}\right)^3}\left[\alpha - 1\{Y_{t+1} \le q^{(c)}_{t+1}\}\right]\left[e^{(c)}_{t+1}\left(q^{(2)}_{t+1} - q^{(1)}_{t+1}\right) + 2\left(e^{(2)}_{t+1} - e^{(1)}_{t+1}\right)\right]\left(e^{(2)}_{t+1} - e^{(1)}_{t+1}\right)\right. \\
&\left. -\frac{2}{\left(e^{(c)}_{t+1}\right)^3}\left[e^{(c)}_{t+1} - \frac{1}{\alpha}Y_{t+1}1\{Y_{t+1} \le q^{(c)}_{t+1}\}\right]\left(e^{(2)}_{t+1} - e^{(1)}_{t+1}\right)^2 + \frac{1}{\left(e^{(c)}_{t+1}\right)^2}\left(e^{(2)}_{t+1} - e^{(1)}_{t+1}\right)^2\right] \\
=&\mathbb{E}_F\left[\frac{1}{\left(e^{(c)}_{t+1}\right)^2}\left(e^{(2)}_{t+1} - e^{(1)}_{t+1}\right)^2\right] > 0,
\end{aligned}$$

where $\mathbb{E}_F\left[\alpha - 1\{Y_{t+1} \le q^{(c)}_{t+1}\}\right] = 0$ and $\mathbb{E}_F\left[e^{(c)}_{t+1} - \frac{1}{\alpha}Y_{t+1}1\{Y_{t+1} \le q^{(c)}_{t+1}\}\right] = 0$. Therefore there is a unique value $\hat{\lambda} \to 0$ such that $\hat{C}^{(c)}_P \to 0$. $\square$

Appendix C: Proof of Theorem 4

Under $\mathbb{H}_0, u_{t+1}^{(1)} = u_{t+1}^{(2)} = u_{t+1}$. We define $g\left(u_{t+1}^{(i)}\right) = \left[\alpha - 1\left(u_{t+1}^{(i)} < 0\right)\right]$ and the score $k_{i,t}(\beta_{i,t})$ as follows:

$$\begin{aligned}
k_{i,t+1}(\beta_{i,t}) =& \frac{\partial L\left(Y_{t+1}, q_{t+1}(\beta_{i,t}), e_{t+1}(\beta_{i,t}); \alpha\right)}{\partial \beta_{i,t}} \\
=& \nabla q_{t+1}(\beta_{i,t})' \frac{1}{\alpha e_{t+1}(\beta_{i,t})} g\left(u_{t+1}^{(i)}\right) \\
& - \nabla e_{t+1}(\beta_{i,t})' \frac{1}{\alpha e_{t+1}(\beta_{i,t})^2} \left[g\left(u_{t+1}^{(i)}\right) q_{t+1}(\beta_{i,t}) + (\alpha - g\left(u_t^{(i)}\right)) Y_{t+1} + \alpha e_{t+1}(\beta_{i,t})\right]
\end{aligned}$$

Then, let $h_{i,t+1}$ and $H_i(t+1)$ to be following.

$$\begin{aligned}
h_{i,t} =& - k_{i,t+1}\left(\beta_{i,t}\right) \\
=& \nabla q_{t+1}(\beta_{i,t})' \frac{1}{-\alpha e_{t+1}(\beta_{i,t})} g\left(u_{t+1}^{(i)}\right) \\
& + \nabla e_{t+1}(\beta_{i,t})' \frac{1}{\alpha e_{t+1}(\beta_{i,t})^2} \left[g\left(u_{t+1}^{(i)}\right) q_{t+1}(\beta_{i,t}) + (\alpha - g\left(u_{t+1}^{(i)}\right)) Y_{t+1} + \alpha e_{t+1}(\beta_{i,t})\right]
\end{aligned}$$

$$\begin{aligned}
H_i(t+1) =& - \frac{1}{R} \sum_{t=1}^{R} k_{i,t+1}\left(\beta\right) \\
=& \frac{1}{R} \sum_{t=1}^{R} \Bigg\{ \nabla q_{t+1}(\beta_{i,t})' \frac{1}{-\alpha e_{t+1}(\beta_{i,t})} g\left(u_{t+1}^{(i)}\right) \\
& + \nabla e_{t+1}(\beta_{i,t})' \frac{1}{\alpha e_{t+1}(\beta_{i,t})^2} \left[g\left(u_{t+1}^{(i)}\right) q_{t+1}(\beta_{i,t}) + (\alpha - g\left(u_{t+1}^{(i)}\right)) Y_{t+1} + \alpha e_{t+1}(\beta_{i,t})\right] \Bigg\}
\end{aligned}$$

We denote the Hessian to be

$$\begin{aligned}
B_i^{-1} =& \mathbb{E}_F \Lambda(\beta_{i,t}) \\
=& \frac{\partial \mathbb{E}_F\left[k_{i,t}(\beta_{i,t})\right]}{\partial \beta_{i,t}} \\
=& \mathbb{E}\left[\frac{f_{t+1}\left(q_{t+1}(\beta_{i,t})\right)}{-\alpha e_{t+1}(\beta_{i,t})} \nabla q_{t+1}(\beta_{i,t})' \nabla q_t(\beta_{i,t}) + \frac{1}{e_{t+1}(\beta_{i,t})^2} \nabla e_{t+1}(\beta_{i,t})' \nabla e_{t+1}(\beta_{i,t})\right]
\end{aligned}$$

Therefore, $B_1 = 1$ and $B_2 = \begin{pmatrix} 1 & 0 \\ 0 & \mathbb{E}_F \Lambda(\beta_{2,t}) \end{pmatrix}^{-1}$. Let J denote the selection matrix (1,0) such that $Jh_{2,t} = h_{1,t}$. Let sup $_t$ denote sup $_{t-R+1 \leq \cdot \leq t}$. And matrix A and C will defined in Lemma 3 and

$\tilde{h}_{2,t} = Z^{-1}A'CB_2^{1/2}h_{2,t}$, $\tilde{H}_2(t) = Z^{-1}A'CB_2^{1/2}H_2(t)$. U_t and $\tilde{U}_t$ will defined in Assumption 2. By the central limit theorem, $\sqrt{R}\left[R^{-1}\sum_{s=t-R+1}^{t} h_{2,s}\right] \sim N\left(0, Z^2B_2^{-1}\right)$, where $Z^2B_2^{-1} = \mathbb{E}[k_t(\theta)k_t(\theta)']$. We use $\sum_t$ instead of $\sum_{j=t-R+1}^{t}$ and use $\hat{q}_{t+1}^{(i)}$ and $\hat{e}_{t+1}^{(i)}$ instead of $q_{t+1}(\hat{\beta}_{i,t})$ and $e_{t+1}(\hat{\beta}_{i,t})$ for simplicity.

Lemma 1: (Clark & McCracken, 2001) Let Assumptions 1-4 hold.
For i=1,2, $\sum_t H_i'(t)B_i(t)h_{i,t+1}h_{i,t+1}'B_j(t)H_j'(t) = \sum_t H_i'(t)B_i\mathbb{E}(h_{i,t+1}h_{i,t+1}')B_jH_j'(t) + o_p(t)$. □

Proof: Adding and subtracting $\mathbb{E}(h_{i,t+1}h_{i,t+1}')$,

$$\begin{aligned}&\sum_t H_i'(t)B_i(t)h_{i,t+1}h_{i,t+1}'B_j(t)H_j(t)\\ =&\sum_t H_i'(t)B_i\mathbb{E}(h_{i,t+1}h_{i,t+1}')B_jH_j(t) + \sum_t H_i'(t)B_i\left(h_{i,t+1}h_{i,t+1}' - \mathbb{E}(h_{i,t+1}h_{i,t+1}')\right)B_jH_j(t).\end{aligned}$$

Now, we need to show the second term is $o_p(1)$.

$$\begin{aligned}&\sum_t H_i'(t)B_i\left(h_{i,t+1}h_{i,t+1}' - \mathbb{E}(h_{i,t+1}h_{i,t+1}')\right)B_jH_j(t)\\ =&T^{-1/2}\sum_t(\frac{T}{t})^2\left[T^{-1/2}\sum_{s=1}^{t}h_{j,s}'B_j \otimes T^{-1/2}\sum_{s=1}^{t}h_{i,s}'B_i\right]\text{vec}\left[T^{-1/2}\left(h_{i,t+1}h_{i,t+1}' - \mathbb{E}(h_{i,t+1}h_{i,t+1}')\right)\right]\end{aligned}$$

$\sum_t(\frac{T}{t})^2\left[T^{-1/2}\sum_{s=1}^{t}h_{j,s}'B_j \otimes T^{-1/2}\sum_{s=1}^{t}h_{i,s}'B_i\right]\text{vec}\left[T^{-1/2}\left(h_{i,t+1}h_{i,t+1}' - \mathbb{E}(h_{i,t+1}h_{i,t+1}')\right)\right]$ is $Op(1)$ follows from Assumption 2, Corollary 29.19 of Davidson (1994) and Theorem 3.1 of Hansen (1992). The proof is complete. □

Lemma 2: Let Assumptions 1-3 hold. (a) Let $-J'B_1J + B_2 = M$ and $B_2^{-1/2}MB_2^{-1/2} = Q$, then Q is idempotent. (b). Define matrix $A = (0,1)'$ and $C = I_{2\times 2}$ Then $Q = CAA'C$. □

Proof: (a).

$$\begin{aligned}
Q &\equiv B_2^{-1/2} M B_2^{-1/2} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & \mathbb{E}_F \Lambda(\beta_{2,t}) \end{pmatrix}^{1/2} \begin{pmatrix} 1 & 0 \\ 0 & \mathbb{E}_F \Lambda(\beta_{2,t}) \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & \mathbb{E}_F \Lambda(\beta_{2,t}) \end{pmatrix}^{1/2} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},
\end{aligned}$$

so Q is an idempotent matrix with rank 1.

(b). Lemma A4(b) of Clark and McCracken (2001). □

Lemma 3: (Clark & McCracken, 2001) Let Assumptions 1-4 hold.

$$\sum_t \tilde{H}_2'(t) \tilde{h}_{2,t+1} \xrightarrow{d} \xi^{-1} \int_\xi^1 \left[W(s) - W(s-\xi) \right] \mathrm{d}W(s),$$

where $\xi = R/T$ and $\sum_t$ denotes $\sum_{t=R+1}^T$ (similarly hereinafter). □

Proof:

$$\begin{aligned}
\sum_t \tilde{H}_2'(t) \tilde{h}_{2,t+1} &= \sum_t \left(\frac{1}{R} \sum_{j=t-R+1}^{t} \tilde{h}_{2,j}' \right) \tilde{h}_{2,t+1} \\
&= \sum_t \left(\frac{1}{R} \sum_{j=1}^{t} \tilde{h}_{2,j}' - \frac{1}{R} \sum_{j=1}^{t-R} \tilde{h}_{2,j}' \right) \tilde{h}_{2,t+1} \\
&= \frac{T}{R} \left[\sum_t \left(T^{-1/2} \sum_{j=1}^{t} \tilde{h}_{2,j}' - T^{-1/2} \sum_{j=1}^{t-R} \tilde{h}_{2,j}' \right) \left(T^{-1/2} \tilde{h}_{2,t+1} \right) \right]
\end{aligned}$$

According to the continues mapping theorem and Corollary 29.19 of Davidson (1994), $T^{-1/2} \sum_{j=1}^{t} \tilde{h}_{2,j}' \Longrightarrow W(s)$, $T^{-1/2} \sum_{j=1}^{t-R} \tilde{h}_{2,j}' \Longrightarrow W(s-\xi)$ and $T^{-1/2} \tilde{h}_{2,t+1} \Longrightarrow \mathrm{d}W(s)$. Since $T/R = \xi^{-1}$, the proof is complete. □

Lemma 4: (Clark & McCracken, 2001) Let Assumptions 1-4 hold.

$$\sum_t \tilde{H}_2'(t)\tilde{H}_2(t) \xrightarrow{d} \xi^{-2}\int_{\xi}^{1}\left[W(s)-W(s-\xi)\right]^2 \mathrm{d}s.$$

□

Proof:

$$\begin{aligned}
\sum_t \tilde{H}_2'(t)\tilde{H}_2(t) &= \sum_t \left(\frac{1}{R}\sum_{j=t-R+1}^{t}\tilde{h}_{2,j}'\right)'\left(\frac{1}{R}\sum_{j=t-R+1}^{t}\tilde{h}_{2,j}'\right)\\
&= \sum_t \left(\frac{1}{R}\sum_{j=1}^{t}\tilde{h}_{2,j}' - \frac{1}{R}\sum_{j=1}^{t-R}\tilde{h}_{2,j}'\right)'\left(\frac{1}{R}\sum_{j=1}^{t}\tilde{h}_{2,j}' - \frac{1}{R}\sum_{j=1}^{t-R}\tilde{h}_{2,j}'\right)\\
&= \frac{1}{T}\left(\frac{T}{R}\right)^2\sum_t \left(T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j}' - T^{-1/2}\sum_{j=1}^{t-R}\tilde{h}_{2,j}'\right)'\left(T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j}' - T^{-1/2}\sum_{j=1}^{t-R}\tilde{h}_{2,j}'\right)
\end{aligned}$$

According to the continues mapping theorem and Corollary 29.19 of Davidson (1994), $T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j}' \Longrightarrow W(s)$, $T^{-1/2}\sum_{j=1}^{t-R}\tilde{h}_{2,j}' \Longrightarrow W(s-\xi)$, $T^{-1/2}\tilde{h}_{2,t+1} \Longrightarrow \mathrm{d}W(s)$ and $1/T \Longrightarrow ds$. Since $T/R = \xi^{-1}$, the proof is complete. □

Lemma 5: Let Assumptions 1-4 hold.

$$\begin{aligned}
&\sum_t \left[\hat{m}_{1,t+1}\left(\hat{q}_{t+1}^{(2)}-\hat{q}_{t+1}^{(1)}\right)+\hat{m}_{2,t+1}\left(\hat{e}_{t+1}^{(2)}-\hat{e}_{t+1}^{(1)}\right)\right]\\
&= \sum_t \left[\left(-\frac{1}{\alpha\hat{e}_{t+1}^{(1)}}\mathbf{1}\{Y_{t+1}\leq \hat{v}_{t+1}^{(1)}\}+\frac{1}{\hat{e}_{t+1}^{(1)}}\right)\left(\hat{q}_{t+1}^{(2)}-\hat{q}_{t+1}^{(1)}\right)\right.\\
&\quad \left.+\left(\frac{1}{\alpha\left(\hat{e}_{t+1}^{(1)}\right)^2}\mathbf{1}\left\{Y_{t+1}\leq \hat{q}_{t+1}^{(1)}\right\}\left(\hat{q}_{t+1}^{(1)}-Y_{t+1}\right)-\frac{\hat{q}_{t+1}^{(1)}}{\left(\hat{e}_{t+1}^{(1)}\right)^2}+\frac{1}{\hat{e}_{t+1}^{(1)}}\right)\left(\hat{e}_{t+1}^{(2)}-\hat{e}_{t+1}^{(1)}\right)\right]\\
&= Z^2\sum_t \tilde{H}_2'(t)\tilde{h}_{2,t+1}+o_p(1)
\end{aligned}$$

□

Proof:

$$\sum_t \left[\left(-\frac{1}{\alpha \hat{e}_{t+1}^{(1)}} \mathbf{1}\{Y_{t+1} \leq \hat{q}_{t+1}^{(1)}\} + \frac{1}{\hat{e}_{t+1}^{(1)}} \right) \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) \right.$$

$$\left. + \left(\frac{1}{\alpha \left(\hat{e}_{t+1}^{(1)} \right)^2} \mathbf{1}\left\{ Y_{t+1} \leq \hat{q}_{t+1}^{(1)} \right\} \left(\hat{q}_{t+1}^{(1)} - Y_{t+1} \right) - \frac{\hat{q}_{t+1}^{(1)}}{\left(\hat{e}_{t+1}^{(1)} \right)^2} + \frac{1}{\hat{e}_{t+1}^{(1)}} \right) \left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)} \right) \right]$$

$$= \sum_t \left\{ \frac{1}{\alpha \hat{e}_{t+1}^{(1)}} \left(\alpha - \mathbf{1}\{Y_{t+1} \leq \hat{q}_{t+1}^{(1)}\} \right) \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) \right.$$

$$\left. - \frac{1}{\alpha \left(\hat{e}_{t+1}^{(1)} \right)^2} \left[\left(\alpha - \mathbf{1}\left\{ Y_{t+1} \leq \hat{q}_{t+1}^{(1)} \right\} \right) \hat{q}_{t+1}^{(1)} + \mathbf{1}\left\{ Y_{t+1} \leq \hat{q}_{t+1}^{(1)} \right\} Y_{t+1} - \alpha \hat{e}_{t+1}^{(1)} \right] \left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)} \right) \right\}$$

$$= \sum_t \left\{ \frac{1}{\alpha \hat{e}_{t+1}^{(1)}} g\left(\hat{u}_{t+1}^{(1)} \right) \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) \right.$$

$$\left. - \frac{1}{\alpha \left(\hat{e}_{t+1}^{(1)} \right)^2} \left[g\left(\hat{u}_{t+1}^{(1)} \right) \hat{q}_{t+1}^{(1)} + \left(\alpha - g\left(\hat{u}_{t+1}^{(1)} \right) \right) Y_{t+1} - \alpha \hat{e}_{t+1}^{(1)} \right] \left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)} \right) \right\}$$

Taking Taylor Expansion and using generalized function by Galfand and Shilov, every term in the last equation becomes

$$\frac{1}{\alpha \hat{e}_{t+1}^{(1)}} = \frac{1}{\alpha e_{t+1}^{(1)}} - \frac{\nabla e_{t+1}^{(1)\prime}}{\alpha \left(e_{t+1}^{(1)} \right)^2} \left(\hat{\beta}_1 - \beta_1 \right) + O_p\left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right)$$

$$g\left(\hat{u}_{t+1}^{(1)} \right) = g\left(u_{t+1}^{(1)} \right) + \delta\left(u_{t+1}^{(1)} \right) \left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)} \right) + O_p\left(\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)} \right)^2 \right)$$

$$\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} = q_{t+1}^{(2)} + \nabla q_{t+1}^{(2)} \left(\hat{\beta}_{2,t} + \beta_2 \right) + O_p\left(\left(\hat{\beta}_{2,t} - \beta_2 \right)^2 \right) - q_{t+1}^{(1)} - \nabla q_{t+1}^{(1)} \left(\hat{\beta}_{1,t} - \beta_1 \right) + O_p\left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right)$$

$$\frac{1}{\alpha \left(\hat{e}_{t+1}^{(1)} \right)^2} = \frac{1}{\alpha \left(e_{t+1}^{(1)} \right)^2} + \frac{-2 \nabla e_{t+1}^{(1)\prime}}{\alpha \left(e_{t+1}^{(1)} \right)^3} \left(\hat{\beta}_{1,t} - \beta_1 \right) + O_p\left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right)$$

$$
\begin{aligned}
g\left(\hat{u}_{t+1}^{(1)}\right)\hat{q}_{t+1}^{(1)} &= g\left(u_{t+1}^{(1)}\right)q_{t+1}^{(1)} + g\left(\hat{u}_{t+1}^{(1)}\right)\hat{q}_{t+1}^{(1)} - g\left(u_{t+1}^{(1)}\right)q_{t+1}^{(1)} \\
&= g\left(u_{t+1}^{(1)}\right)q_{t+1}^{(1)} + g\left(\hat{u}_{t+1}^{(1)}\right)\hat{q}_{t+1}^{(1)} - g\left(\hat{u}_{t+1}^{(1)}\right)q_{t+1}^{(1)} + g\left(\hat{u}_{t+1}^{(1)}\right)q_{t+1}^{(1)} - g\left(u_{t+1}^{(1)}\right)q_{t+1}^{(1)} \\
&= g\left(u_{t+1}^{(1)}\right)q_{t+1}^{(1)} + g\left(\hat{u}_{t+1}^{(1)}\right)\left(\hat{q}_{t+1}^{(1)} - q_{t+1}^{(1)}\right) + \left(g\left(\hat{u}_{t+1}^{(1)}\right) - g\left(u_{t+1}^{(1)}\right)\right)q_{t+1}^{(1)} \\
&= g\left(u_{t+1}^{(1)}\right)q_{t+1}^{(1)} + g\left(\hat{u}_{t+1}^{(1)}\right)\left(\hat{q}_{t+1}^{(1)} - q_{t+1}^{(1)}\right) + \delta\left(u_{t+1}^{(1)}\right)\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right)q_{t+1}^{(1)} \\
&\quad + O_p\left(\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right)^2\right)q_{t+1}^{(1)}
\end{aligned}
$$

$$
\begin{aligned}
\left(\alpha - g\left(\hat{u}_{t+1}^{(1)}\right)\right)Y_{t+1} &= \left(\alpha - g(u_{t+1}^{(1)}) - g(\hat{u}_{t+1}^{(1)}) + g(u_{t+1}^{(1)})\right)Y_{t+1} \\
&= \left(\alpha - g\left(u_{t+1}^{(1)}\right)\right)Y_{t+1} - \left(g\left(\hat{u}_{t+1}^{(1)}\right) - g\left(u_{t+1}^{(1)}\right)\right)Y_{t+1} \\
&= \left(\alpha - g\left(u_{t+1}^{(1)}\right)\right)Y_{t+1} - \left[\delta\left(u_{t+1}^{(1)}\right)\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right) + O_p\left(\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right)^2\right)\right]Y_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
\alpha\hat{e}_{t+1}^{(1)} &= \alpha\left(e_{t+1}^{(1)} + \nabla e_{t+1}^{(1)}\left(\hat{\beta}_{1,t} - \beta_1\right) + O_p\left(\left(\hat{\beta}_{1,t} - \beta_1\right)^2\right)\right) \\
&= \alpha e_{t+1}^{(1)} + \alpha\left(\nabla e_{t+1}^{(1)}\left(\hat{\beta}_{1,t} - \beta_1\right) + O_p\left(\left(\hat{\beta}_{1,t} - \beta_1\right)^2\right)\right)
\end{aligned}
$$

$$\sum_t \left[\left(-\frac{1}{\alpha \hat{e}^{(1)}_{t+1}} \mathbf{1}\{Y_{t+1} \le \hat{q}^{(1)}_{t+1}\} + \frac{1}{\hat{e}^{(1)}_{t+1}} \right) \left(\hat{q}^{(2)}_{t+1} - \hat{q}^{(1)}_{t+1} \right) \right.$$

$$\left. + \left(\frac{1}{\alpha \left(\hat{e}^{(1)}_{t+1} \right)^2} \mathbf{1}\left\{ Y_{t+1} \le \hat{q}^{(1)}_{t+1} \right\} \left(\hat{q}^{(1)}_{t+1} - Y_{t+1} \right) - \frac{\hat{q}^{(1)}_{t+1}}{\left(\hat{e}^{(1)}_{t+1} \right)^2} + \frac{1}{\hat{e}^{(2)}_{t+1}} \right) \left(\hat{e}^{(1)}_{t+1} - \hat{e}^{(1)}_{t+1} \right) \right]$$

$$= \sum_t \left\{ \left[\underbrace{\frac{1}{\alpha e^{(1)}_{t+1}}}_{A_1} - \underbrace{\frac{\nabla e^{(1)\prime}_{t+1}}{\alpha \left(e^{(1)}_{t+1} \right)^2} \left(\hat{\beta}_{1,t} - \beta_1 \right) + O_p \left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right)}_{A_2} \right] \times \right.$$

$$\left[\underbrace{g \left(u^{(1)}_{t+1} \right)}_{B1} + \underbrace{\delta \left(u^{(1)}_{t+1} \right) \left(\hat{u}^{(1)}_{t+1} - u^{(1)}_{t+1} \right) + O_p \left(\left(\hat{u}^{(1)}_{t+1} - u^{(1)}_{t+1} \right)^2 \right)}_{B2} \right] \times$$

$$\left[\underbrace{\nabla q^{(2)}_{t+1} \left(\hat{\beta}_{2,t} - \beta_2 \right) - \nabla q^{(1)}_{t+1} \left(\hat{\beta}_{1,t} - \beta_1 \right)}_{C_1} + \underbrace{O_p \left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right) + O_p \left(\left(\hat{\beta}_{2,t} - \beta_1 \right)^2 \right)}_{C_2} \right]$$

$$+ \left[\underbrace{\frac{1}{-\alpha \left(e^{(1)}_{t+1} \right)^2}}_{D_1} + \underbrace{\frac{2\nabla e^{(1)\prime}_{t+1}}{\alpha (e^{(1)}_{t+1})^3} \left(\hat{\beta}_{1,t} - \beta_1 \right) - O_p \left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right)}_{D_2} \right] \times$$

$$\left[\underbrace{g \left(u^{(1)}_{t+1} \right) q^{(1)}_{t+1} + \left(\alpha - g \left(u^{(1)}_{t+1} \right) \right) Y_{t+1} - \alpha e^{(1)}_{t+1}}_{E_1} \right.$$

$$+ \underbrace{g \left(\hat{u}^{(1)}_{t+1} \right) \left(\hat{q}^{(1)}_{t+1} - q^{(1)}_{t+1} \right) + \delta \left(u^{(1)}_{t+1} \right) \left(\hat{u}^{(1)}_{t+1} - u^{(1)}_{t+1} \right) q^{(1)}_{t+1} + O_p \left(\left(\hat{u}^{(1)}_{t+1} - u^{(1)}_{t+1} \right)^2 \right) q^{(1)}_{t+1}}_{E_2}$$

$$\left. \underbrace{- \left(\delta \left(u^{(1)}_{t+1} \right) \left(\hat{u}^{(1)}_{t+1} - u^{(1)}_{t+1} \right) + O_p \left(\left(\hat{u}^{(1)}_{t+1} - u^{(1)}_{t+1} \right)^2 \right) \right) Y_{t+1} - \alpha \left(\nabla e^{(1)}_{t+1} \left(\hat{\beta}_{1,t} - \beta_1 \right) + O_p \left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right) \right) \right]}_{E_2} \times$$

$$\left. \left[\underbrace{\nabla e^{(2)}_{t+1} \left(\hat{\beta}_{2,t} - \beta_2 \right) - \nabla e^{(1)}_{t+1} \left(\hat{\beta}_{1,t} - \beta_1 \right)}_{F_1} + \underbrace{O_p \left(\left(\hat{\beta}_{1,t} - \beta_1 \right)^2 \right) + O_p \left(\left(\hat{\beta}_{2,t} - \beta_2 \right)^2 \right)}_{F_2} \right] \right] \right\}$$

$$
\begin{aligned}
AA &= A_1 B_1 C_1 + D_1 E_1 F_1 \\
&= \sum_t \frac{1}{\alpha e_{t+1}^{(1)}} g\left(u_{t+1}^{(1)}\right)\left(\nabla q_{t+1}^{(2)}\left(\hat{\beta}_{2,t}-\beta_2\right) - \nabla q_{t+1}^{(1)}\left(\hat{\beta}_{1,t}-\beta_1\right)\right) \\
&\quad - \frac{1}{\alpha\left(e_{t+1}^{(1)}\right)^2}\left(g\left(u_{t+1}^{(1)}\right) q_{t+1}^{(1)} + \left(\alpha - g\left(u_{t+1}^{(1)}\right)\right) Y_{t+1} - \alpha e_{t+1}^{(1)}\right)\left(\nabla e_{t+1}^{(2)}\left(\hat{\beta}_{2,t}-\beta_2\right) - \nabla e_{t+1}^{(1)}\left(\hat{\beta}_{1,t}-\beta_1\right)\right) \\
&= \sum_t \Bigg[\nabla q_{t+1}^{(2)\prime} \frac{1}{\alpha e_{t+1}^{(1)}} g(u_{t+1}^{(1)})(\hat{\beta}_{2,t}-\beta_2) - \nabla q_{t+1}^{(1)\prime} \frac{1}{\alpha e_{t+1}^{(1)}} g(u_{t+1}^{(1)})(\hat{\beta}_{1,t}-\beta_1) \\
&\quad - \nabla e_{t+1}^{(2)\prime} \frac{1}{\alpha (e_{t+1}^{(1)})^2}\left(g(u_{t+1}^{(1)}) q_{t+1}^{(1)} + \left(\alpha - g(u_{t+1}^{(1)})\right) Y_{t+1} - \alpha e_{t+1}^{(1)}\right)(\hat{\beta}_{2,t}-\beta_2) \\
&\quad + \nabla e_{t+1}^{(1)} \frac{1}{\alpha (e_{t+1}^{(1)})^2}\left(g(u_{t+1}^{(1)}) q_{t+1}^{(1)} + \left(\alpha - g(u_{t+1}^{(1)})\right) Y_{t+1} - \alpha e_{t+1}^{(1)}\right)(\hat{\beta}_{1,t}-\beta_1)\Bigg] \\
&= \sum_t \Bigg\{\left[\nabla q_{t+1}^{(2)\prime} \frac{1}{\alpha e_{t+1}^{(1)}} g(u_{t+1}^{(1)}) - \nabla e_{t+1}^{(2)\prime} \frac{1}{\alpha (e_{t+1}^{(1)})^2}\left(g(u_{t+1}^{(1)}) q_{t+1}^{(1)} + \left(\alpha - g(u_{t+1}^{(1)})\right) Y_{t+1} - \alpha e_{t+1}^{(1)}\right)\right](\hat{\beta}_{2,t}-\beta_2) \\
&\quad - \left[\nabla q_{t+1}^{(1)\prime} \frac{1}{\alpha e_{t+1}^{(1)}} g(u_{t+1}^{(1)}) - \nabla e_{t+1}^{(1)} \frac{1}{\alpha (e_{t+1}^{(1)})^2}\left(g(u_{t+1}^{(1)}) q_{t+1}^{(1)} + \left(\alpha - g(u_{t+1}^{(1)})\right) Y_{t+1} - \alpha e_{t+1}^{(1)}\right)\right](\hat{\beta}_{1,t}-\beta_1)\Bigg\} \\
&= \sum_t \left[-h_{1,t+1}' B_1(t) H_1(t) + h_{2,t+1}' B_2(t) H_2(t)\right] \\
&= \sum_t \left[-h_{1,t+1}' B_1 H_1(t) + h_{2,t+1}' B_2 H_2(t)\right] + o_p(1) \\
&= \sum_t \left[-h_{2,t+1}' J' B_1 J H_2(t) + h_{2,t+1}' B_2 H_2(t)\right] + o_p(1) \\
&= \sum_t \left[h_{2,t+1}' M H_2(t)\right] + o_p(1) \\
&= \sum_t \left[h_{2,t+1}' B_2^{1/2} B_2^{-1/2} M B_2^{-1/2} B_2^{1/2} H_2(t)\right] + o_p(1) \\
&= \sum_t \left[h_{2,t+1}' B_2^{1/2} Q B_2^{1/2} H_2(t)\right] + o_p(1) \\
&= \sum_t \left[h_{2,t+1}' B_2^{1/2} C A A' C B_2^{1/2} H_2(t)\right] + o_p(1) \\
&= Z^2 \sum_t \tilde{H}_2'(t) \tilde{h}_{2,t+1} + o_p(1)
\end{aligned}
$$

Now we need to show that AB, AC, AD, AE, AF, AG, AH are lower order than AA. The order of each element is showed below.

$$A_1 = O_p(1), \qquad A_2 = O_p\left(\frac{1}{\sqrt{R}}\right), \qquad B_1 = O_p(1), \qquad B_2 = O_p\left(\frac{1}{R}\right),$$
$$C_1 = O_p\left(\frac{1}{\sqrt{R}}\right), \qquad C_2 = O_p\left(\frac{1}{R}\right), \qquad D_1 = O_p(1), \qquad D2 = O_p\left(\frac{1}{\sqrt{R}}\right),$$
$$E_1 = O_p(1), \qquad E_2 = O_p\left(\frac{1}{\sqrt{R}}\right), \qquad F_1 = O_p\left(\frac{1}{\sqrt{R}}\right), \qquad F_2 = O_p\left(\frac{1}{R}\right),$$

Besides $\sum_t \delta(u_{t+1}^{(1)})(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)})q_{t+1}^{(1)}$, the order of other elements are straight forwards. For the order of $\sum_t \delta(u_{t+1}^{(1)})(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)})q_{t+1}^{(1)}$, following Phillips (1991), we can get

$$P^{-1/2} \sum_t \left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right) q_{t+1}^{(1)} \xrightarrow{d} N(0, Q)$$

and

$$\begin{aligned} & P^{-1} \sum_t \delta\left(u_{t+1}^{(1)}\right)\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right) q_{t+1}^{(1)} \\ \xrightarrow{p} & \lim_{P\to\infty} P^{-1} \sum_t E\left(\delta\left(u_{t+1}^{(1)}\right)\right)\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right) q_{t+1}^{(1)} = f(0)Q \end{aligned}$$

Thus, we have

$$\sum_t \delta\left(u_{t+1}^{(1)}\right)\left(\hat{u}_{t+1}^{(1)} - u_{t+1}^{(1)}\right) q_{t+1}^{(1)} = Pf(0)O_p(\frac{1}{R}) = O_p\left(\frac{P}{R}\right)$$

By using the order of each element, the order of each term in the $\sum_t(\hat{m}_{1,t}(\hat{q}_{t+1}^{(1)} - \hat{q}_{t+1}^{(2)}) + \hat{m}_{2,t}(\hat{e}_{t+1}^{(1)} - \hat{e}_{t+1}^{(2)})]$ can be proved as following.

$$
\begin{aligned}
AA &= \sum_t A_1 B_1 C_1 + D_1 E_1 F_1 \\
&= \sum_t \left[O_p(1) O_p(1) O_p\left(\frac{1}{\sqrt{R}}\right) + O_p(1) O_p(1) O_p\left(\frac{1}{\sqrt{R}}\right) \right] = O_p\left(\frac{P}{\sqrt{R}}\right))\\
AB &= \sum_t A_1 B_1 C_2 + D_1 E_1 F_2 \\
&= \sum_t \left[O_p(1) O_p(1) O_p\left(\frac{1}{R}\right) + O_p(1) O_p(1) O_p\left(\frac{1}{R}\right) \right] = O_p\left(\frac{P}{R}\right)\\
AC &= \sum_t A_1 B_2 C_1 + D_1 E_2 F_1 \\
&= \sum_t \left[O_p(1) O_p\left(\frac{1}{R}\right) O_p\left(\frac{1}{\sqrt{R}}\right) + O_p(1) O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{\sqrt{R}}\right) \right] = O_p\left(\frac{P}{R}\right)\\
AD &= \sum_t A_1 B_2 C_2 + D_1 E_2 F_2 \\
&= \sum_t \left[O_p(1) O_p\left(\frac{1}{R}\right) O_p\left(\frac{1}{R}\right) + O_p(1) O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{R}\right) \right] = O_p\left(\frac{P}{R^{3/2}}\right)\\
AE &= \sum_t A_2 B_1 C_1 + D_2 E_1 F_1 \\
&= \sum_t \left[O_p\left(\frac{1}{\sqrt{R}}\right) O_p(1) O_p\left(\frac{1}{\sqrt{R}}\right) + O_p\left(\frac{1}{\sqrt{R}}\right) O_p(1) O_p\left(\frac{1}{\sqrt{R}}\right) \right] = O_p\left(\frac{P}{R}\right)\\
AF &= \sum_t A_2 B_1 C_2 + D_2 E_1 F_2 \\
&= \sum_t \left[O_p\left(\frac{1}{\sqrt{R}}\right) O_p(1) O_p\left(\frac{1}{R}\right) + O_p\left(\frac{1}{\sqrt{R}}\right) O_p(1) O_p\left(\frac{1}{R}\right) \right] = O_p\left(\frac{P}{R^{3/2}}\right)\\
AG &= \sum_t A_2 B_2 C_1 + D_2 E_2 F_1 \\
&= \sum_t \left[O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{R}\right) O_p\left(\frac{1}{\sqrt{R}}\right) + O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{\sqrt{R}}\right) \right] = O\left(\frac{P}{R^{3/2}}\right)\\
AH &= \sum_t A_2 B_2 C_2 + D_2 E_2 F_2 \\
&= \sum_t \left[O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{R}\right) O_p\left(\frac{1}{R}\right) + O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{\sqrt{R}}\right) O_p\left(\frac{1}{R}\right) \right] = O_p\left(\frac{P}{R^2}\right)
\end{aligned}
$$

Therefore, all AB, AC, AD, AE, AF, AG, AH are lower order than AA. The proof is complete. □

Lemma 6: Let Assumptions 1-4 hold.

$$
\begin{aligned}
&\sum_t \left[\hat{m}_{1,t+1}\left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)}\right) + \hat{m}_{2,t+1}\left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)}\right)\right]^2 - P\bar{C}^2 \\
&= \sum_t \left[\left(-\frac{1}{\alpha \hat{e}_{t+1}^{(1)}}\mathbf{1}\{Y_{t+1} \le \hat{q}_{t+1}^{(1)}\} + \frac{1}{\hat{e}_{t+1}^{(1)}}\right)\left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)}\right)\right. \\
&\quad \left. + \left(\frac{1}{\alpha\left(\hat{e}_{t+1}^{(1)}\right)^2}\mathbf{1}\left\{Y_{t+1} \le \hat{q}_{t+1}^{(1)}\right\}\left(\hat{q}_{t+1}^{(1)} - Y_{t+1}\right) - \frac{\hat{q}_{t+1}^{(1)}}{\left(\hat{e}_{t+1}^{(1)}\right)^2} + \frac{1}{\hat{e}_{t+1}^{(1)}}\right)\left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)}\right)\right]^2 - P\bar{C}^2 \\
&= Z^4 \sum_t \tilde{H}_2'(t)\tilde{H}_2(t) + o_p(1)
\end{aligned}
$$

□

Proof:

$$
\begin{aligned}
&\sum_t \left[\hat{m}_{1,t+1}\left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)}\right) + \hat{m}_{2,t+1}\left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)}\right)\right]^2 \\
&= \sum_t \left[\left(-\frac{1}{\alpha \hat{e}_{t+1}^{(1)}}\mathbf{1}\{Y_{t+1} \le \hat{q}_{t+1}^{(1)}\} + \frac{1}{\hat{e}_{t+1}^{(1)}}\right)\left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)}\right)\right. \\
&\quad \left. + \left(\frac{1}{\alpha\left(\hat{e}_{t+1}^{(1)}\right)^2}\mathbf{1}\left\{Y_{t+1} \le \hat{q}_{t+1}^{(1)}\right\}\left(\hat{q}_{t+1}^{(1)} - Y_{t+1}\right) - \frac{\hat{q}_{t+1}^{(1)}}{\left(\hat{e}_{t+1}^{(1)}\right)^2} + \frac{1}{\hat{e}_{t+1}^{(1)}}\right)\left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)}\right)\right]^2 \\
&= \sum_t \left[-h_{1,t+1}' B_1(t) H_1(t) + h_{2,t+1}' B_2(t) H_2(t)\right]^2 \\
&= \sum_t \left[-h_{1,t+1}' B_1 H_1(t) + h_{2,t+1}' B_2 H_2(t)\right]^2 + o_p(1) \\
&= \sum_t H_1'(t) B_1' h_{1,t+1} h_{1,t+1}' B_1 H_1(t) + \sum_t H_2'(t) B_2' h_{2,t+1} h_{2,t+1}' B_2 H_2(t) \\
&\quad - 2\sum_t H_1'(t) B_1' h_{1,t+1} h_{2,t+1} B_2 H_2(t) + o_p(1) \\
&= \sum_t H_1'(t) B_1' \mathbb{E}\left(h_{1,t+1} h_{1,t+1}'\right) B_1 H_1(t) + \sum_t H_2'(t) B_2' \mathbb{E}\left(h_{2,t+1} h_{2,t+1}'\right) B_2 H_2(t) \\
&\quad - 2\sum_t H_1'(t) B_1' \mathbb{E}\left(h_{1,t+1} h_{2,t+1}'\right) B_2 H_2(t) + o_p(1) \\
&= Z^2 \sum_t H_1'(t) B_1 H_1(t) + Z^2 \sum_t H_2'(t) B_2 H_2(t) - 2Z^2 \sum_t H_1'(t) B_2 H_2(t) + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= Z^2 \sum_t H_2'(t) J' B_1 J H_2(t) + Z^2 \sum_t H_2'(t) B_2 H_2(t) - 2Z^2 \sum_t H_2'(t) J' B_1 J H_2(t) + o_p(1) \\
&= Z^2 \sum_t H_2'(t) \left[-J' B_1 J + B_2 \right] H_2(t) + o_p(1) \\
&= Z^2 \sum_t H_2'(t) M H_2(t) + o_p(1) \\
&= Z^2 \sum_t H_2'(t) B_2^{1/2} B_2^{-1/2} M B_2^{-1/2} B_2^{1/2} H_2(t) + o_p(1) \\
&= Z^2 \sum_t H_2'(t) B_2^{1/2} C A A' C B_2^{1/2} H_2(t) + o_p(1) \\
&= Z^4 \sum_t \tilde{H}_2'(t) \tilde{H}_2(t) + o_p(1)
\end{aligned}
$$

Lemma 3 implies that $\bar{C} = O_p(P^{-1})$, so $P\bar{C}^2 = o_p(1)$. Thus, the proof is completed. □

Theorem 2: Let Assumptions 1-4 hold. $\text{ENC}_P \xrightarrow{d} N(0,1),$ as $P, R \to \infty,$ under $\mathbb{H}_0 : \lambda = 0.$ □

Proof: In order to show that ENC_P is asymptotically standard normal, we need to show

$$
\begin{aligned}
\text{ENC}_P &= Q_P^{-1/2} \sqrt{P} \hat{C}_P \\
&= \frac{\sum_t \hat{m}_{1,t+1} \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) + \hat{m}_{2,t+1} \left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)} \right)}{\sqrt{\sum_t \left[\hat{m}_{1,t+1} \left(\hat{q}_{t+1}^{(2)} - \hat{q}_{t+1}^{(1)} \right) + \hat{m}_{2,t+1} \left(\hat{e}_{t+1}^{(2)} - \hat{e}_{t+1}^{(1)} \right) \right]^2 - P\bar{C}^2}} \\
&\Longrightarrow \frac{\xi^{-1} \int_\xi^1 \left[W(s) - W(s-\xi) \right] \mathrm{d}W(s)}{\sqrt{\xi^{-2} \int_\xi^1 \left[W(s) - W(s-\xi) \right]^2 \mathrm{d}s}} \xrightarrow{d} N(0,1)
\end{aligned}
$$

We divided $[0,1]$ to $n = [1/\xi] = [T/R]$ equal segments and let $t = [sn]$, where $[\cdot]$ denote the integer part and $s \in [0,1]$. Let $\{\nu_i\}_{i=1}^n$ is mixing sequence with $\mathbb{E}(\nu) = 0$ and $Var(\nu) = 1$. We denote $V_t = \sum_{i=1}^t \nu_i$ to be the partial sum, then $V_t = \sum_{i=1}^t \nu_i \sim N(0,t)$. Thus,

$$
\frac{V_t}{\sqrt{n}} = \frac{\sum_{i=1}^t \nu_i}{\sqrt{n}} = V_n(s) \Rightarrow W(s),
$$

where $V_n(s)$ is a CADLAG process and $V_n(s)$ is a Wiener process.

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

We generate an AR(1) regression to show the asymptotical standard normality of ENC_P.

$$
\nu_{t+1} = \zeta\nu_t + \epsilon_t
$$

The estimator $\hat{\zeta} = \sum_{t=1}^{n}\nu_{t-1}\nu_t / \sum_{t=1}^{n}\nu_{t-1}^2$ and $Var(\hat{\zeta}) = (\sum_{t=1}^{n}\nu_{t-1}^2)^{-1}Var(\nu) = (\sum_{t=1}^{n}\nu_{t-1}^2)^{-1}$.
By using CLT,

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

□

Table 1: Size of test (DGP1)

		$P = 240$			$P = 480$			$P = 1200$		
Repeat= 2000		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
	$R = 120$	0.005	0.091	0.040	0.003	0.098	0.044	0.001	0.089	0.047
$\alpha = 0.01$	$R = 240$	0.053	0.123	0.035	0.019	0.093	0.056	0.004	0.086	0.040
	$R = 480$	0.070	0.105	0.034	0.026	0.075	0.048	0.009	0.062	0.054
	$R = 120$	0.006	0.078	0.049	0.003	0.079	0.052	0.001	0.088	0.056
$\alpha = 0.025$	$R = 240$	0.026	0.074	0.055	0.008	0.060	0.056	0.001	0.066	0.053
	$R = 480$	0.030	0.064	0.049	0.013	0.058	0.043	0.004	0.054	0.057
	$R = 120$	0.008	0.057	0.054	0.003	0.065	0.045	0.001	0.068	0.052
$\alpha = 0.05$	$R = 240$	0.013	0.055	0.048	0.002	0.054	0.054	0.000	0.050	0.048
	$R = 480$	0.015	0.043	0.059	0.009	0.049	0.059	0.000	0.044	0.054
	$R = 120$	0.003	0.053	0.043	0.002	0.056	0.055	0.000	0.064	0.047
$\alpha = 0.1$	$R = 240$	0.009	0.047	0.052	0.002	0.043	0.054	0.000	0.045	0.054
	$R = 480$	0.014	0.041	0.050	0.007	0.044	0.054	0.001	0.046	0.063

This table show the size of DM_P, ENC_P, and CCS_P test under 5% significance level, $b = 0$, $\phi = 0$.

Table 2: Size of test (DGP2)

		$P = 240$			$P = 480$			$P = 1200$		
Repeat= 2000		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
	$R = 120$	0.029	0.064	0.048	0.011	0.055	0.035	0.003	0.072	0.041
$\alpha = 0.05$	$R = 240$	0.032	0.059	0.096	0.015	0.056	0.053	0.004	0.066	0.040
	$R = 480$	0.040	0.077	0.121	0.025	0.057	0.076	0.009	0.055	0.038

This table show the size of DM_P, ENC_P, and CCS_P test under 5% significance level, $b = 0$, $\phi = 0$.

Table 3: Power of test (DGP1)

Repeat= 2000		$P=240$ DM$_P$	ENC$_P$	CCS$_P$	$P=480$ DM$_P$	ENC$_P$	CCS$_P$	$P=1200$ DM$_P$	ENC$_P$	CCS$_P$
	$R=120$	0.035	0.238	0.007	0.018	0.349	0.002	0.008	0.673	0.000
$\alpha=0.01$	$R=240$	0.119	0.348	0.009	0.112	0.493	0.002	0.076	0.721	0.000
	$R=480$	0.240	0.460	0.028	0.225	0.571	0.003	0.232	0.790	0.000
	$R=120$	0.060	0.406	0.002	0.153	0.500	0.003	0.025	0.840	0.000
$\alpha=0.025$	$R=240$	0.153	0.497	0.003	0.162	0.687	0.000	0.220	0.931	0.000
	$R=480$	0.251	0.568	0.004	0.288	0.758	0.000	0.496	0.954	0.000
	$R=120$	0.104	0.543	0.001	0.125	0.749	0.000	0.165	0.964	0.000
$\alpha=0.05$	$R=240$	0.221	0.664	0.001	0.321	0.854	0.000	0.535	0.986	0.000
	$R=480$	0.313	0.713	0.000	0.445	0.891	0.000	0.758	0.997	0.000
	$R=120$	0.162	0.707	0.001	0.241	0.894	0.000	0.470	0.999	0.000
$\alpha=0.1$	$R=240$	0.333	0.793	0.000	0.480	0.956	0.000	0.854	1.000	0.000
	$R=480$	0.420	0.853	0.000	0.623	0.975	0.000	0.938	1.000	0.000

This table show the power of DM$_P$, ENC$_P$, and CCS$_P$ test under 5% significance level, $b=0.1$, $\phi=0.95$.

Table 4: Power of test (DGP2)

Repeat= 2000		$P=240$			$P=480$			$P=1200$		
		DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P	DM_P	ENC_P	CCS_P
	$R=120$	0.022	0.098	0.045	0.012	0.100	0.136	0.000	0.177	0.657
$\alpha=0.05$	$R=240$	0.071	0.144	0.058	0.066	0.213	0.077	0.060	0.319	0.359
	$R=480$	0.155	0.201	0.064	0.167	0.363	0.068	0.272	0.654	0.247

This table show the power of DM_P, ENC_P, and CCS_P test under 5% significance level, $b=0.1$, $\phi=0.0$.

Table 5: Predicting $\boldsymbol{\gamma}_\alpha = (\text{VaR}_\alpha, \text{ES}_\alpha)'$ of the Conditional Distribution of the Equity Premium

			Forecasts begin 40 years			
	Variable	Data	$\alpha = 0.010$	$\alpha = 0.025$	$\alpha = 0.050$	$\alpha = 0.100$
d/p	Dividend price ratio	1872 - 2019	0.8456	0.0217	-0.9950	-1.7111*
d/y	Dividend yield	1872 - 2019	0.9034	0.2156	-2.0096**	-2.0647**
e/p	Earning price ratio	1872 - 2019	0.1563	-2.1621**	-1.1869	-2.0079**
d/e	Dividend payout ratio	1872 - 2019	-1.8649*	-1.4992	-1.5904	-2.5146**
svar	Stock variance	1885 - 2019	-0.8076	-2.1099**	-1.3688	-1.6135
infl	Inflation	1913 - 2019	0.5453	1.3445	1.6212	-1.2970
dfy	Default yield spread	1919 - 2019	-0.2556	-2.1410**	-0.5588	0.4735
lty	Long term yield	1919 - 2019	0.4369	0.8794	1.3936	-0.96409
tbl	Treasury-bill rate	1920 - 2019	-0.6512	1.7758*	2.6561***	-0.2662
tms	Team Spread	1920 - 2019	1.2146	-0.6777	-1.3864	0.1449
b/m	Book to market	1921 - 2019	-0.4917	0.1936	0.2225	1.0035
dfr	Default return spread	1926 - 2019	-0.6867	0.9984	-1.8607*	-2,9087***
ltr	Long term return	1926 - 2019	-0.7686	-1.3939	-1.3442	-2.5768***
ntis	Net equity expansion	1927 - 2019	-0.6009	-0.0934	0.4599	0.8047
csp	Consumption, wealth, income ratio	1937 - 2002	-0.7636	-1.5200	-2.1509**	-1.8635*

Table 6: Predicting $\boldsymbol{\gamma}_\alpha = (\text{VaR}_\alpha, \text{ES}_\alpha)'$ of the Conditional Distribution of the Equity Premium

			Forecasts begin 1965			
	Variable	Data	$\alpha = 0.010$	$\alpha = 0.025$	$\alpha = 0.050$	$\alpha = 0.100$
d/p	Dividend price ratio	1872 - 2019	1.1461	-0.0489	-0.9532	-2.7234***
d/y	Dividend yield	1872 - 2019	0.7004	0.1788	-1.1910	-2.4216**
e/p	Earning price ratio	1872 - 2019	-1.0353	-2.1651**	-2.7181***	-3.3621***
d/e	Dividend payout ratio	1872 - 2019	-0.9068	-0.5818	0.1831	1.3407
svar	Stock variance	1885 - 2019	-0.4051	-1.6676*	-1.1233	-1.2407
infl	Inflation	1913 - 2019	0.2271	0.4527	-0.9735	-0.6127
dfy	Default yield spread	1919 - 2019	-0.3626	-1.0582	-0.4422	0.5596
lty	Long term yield	1919 - 2019	-0.6788	0.1752	-0.5766	-0.6432
tbl	Treasury-bill rate	1920 - 2019	0.4081	-0.9483	0.7650	1.0696
tms	Team Spread	1920 - 2019	-1.9046*	-0.8820	-1.7015*	-0.7368
b/m	Book to market	1921 - 2019	0.4758	-0.9507	-0.7258	0.8496
dfr	Default return spread	1926 - 2019	-0.9792	-0.5778	-1.6071	-2.1071**
ltr	Long term return	1926 - 2019	-0.7699	-2.9069***	-1.0749	-1.5730
ntis	Net equity expansion	1927 - 2019	0.4822	-1.8989*	0.3957	0.7531
csp	Consumption, wealth, income ratio	1937 - 2002	-0.6788	-1.6012	-1.3447	-3.2380***

Table 7: Predicting $\boldsymbol{\gamma}_\alpha = (\mathrm{GaR}_\alpha, \mathrm{GS}_\alpha)'$ of the Conditional Distribution of the GDP Growth

	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
$\mathrm{ENC}_{R,P}$	-1.4929	-1.2786	-1.5109	-1.4508	-1.8355	-1.4053	-1.2638	-0.8704	-1.1025
P-value	0.1355	0.2010	0.1308	0.1468	0.0664	0.1599	0.2063	0.3841	0.2702

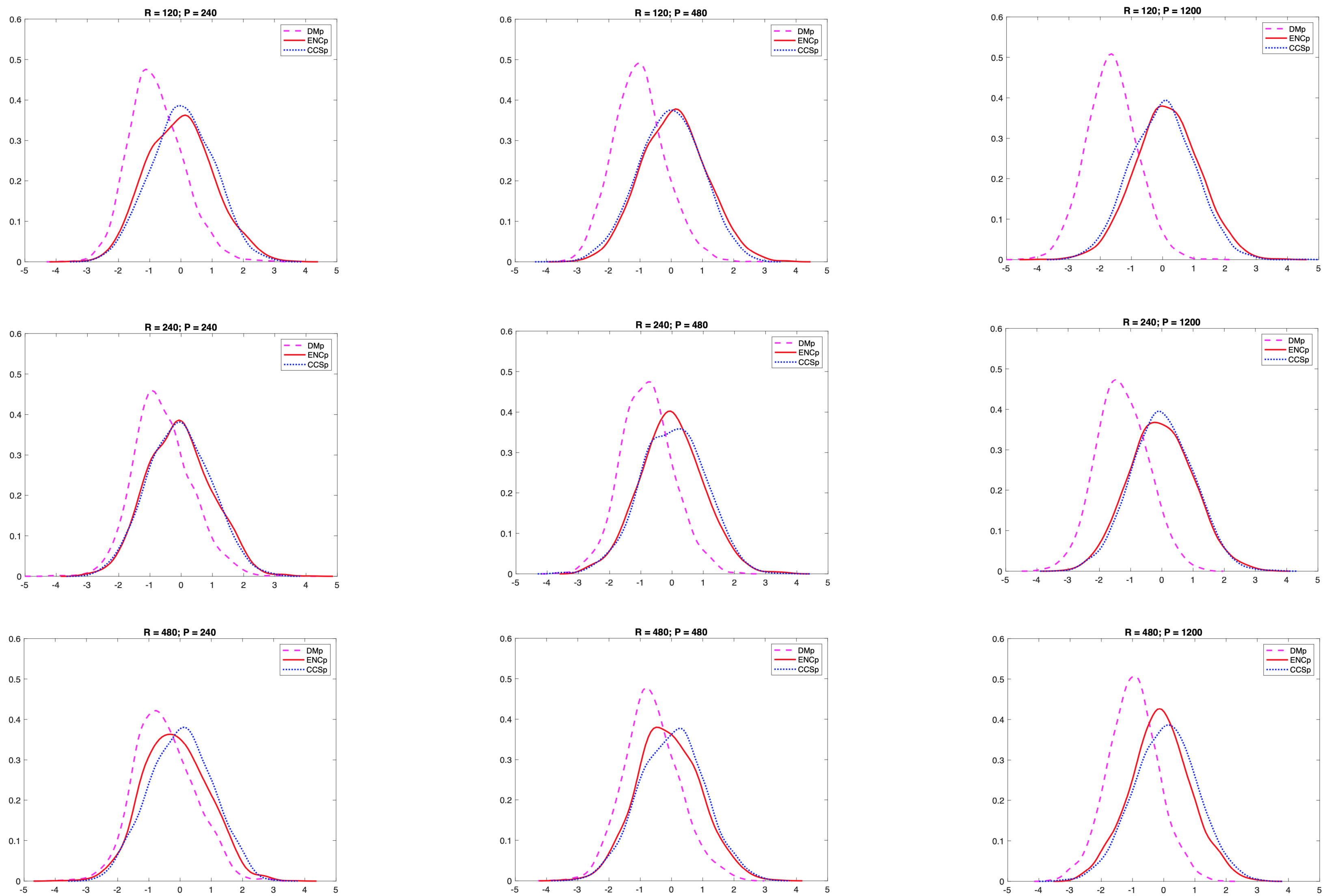

Figure 1: The distribution of D_P, ENC_P, and CCS_P in DGP1, $\phi = 0$, $b = 0$, $\alpha = 0.05$, $\sigma_u = 1$, 2000 Repeats.

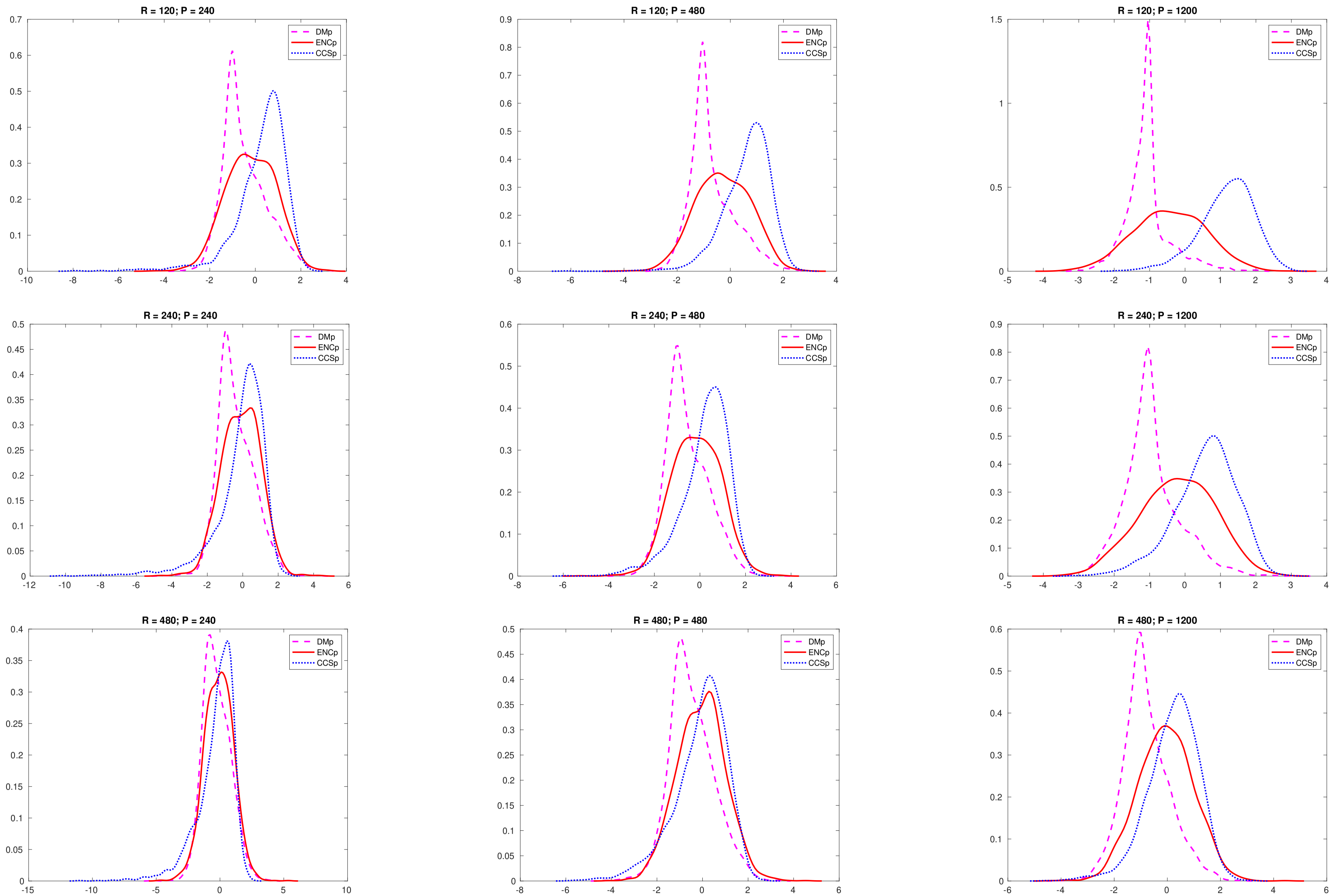

Figure 2: The distribution of D_P, ENC_P, and CCS_P in DGP2, $\phi = 0$, $b = 0$, $\alpha = 0.05$, $\sigma_u = 1$, 2000 Repeats.

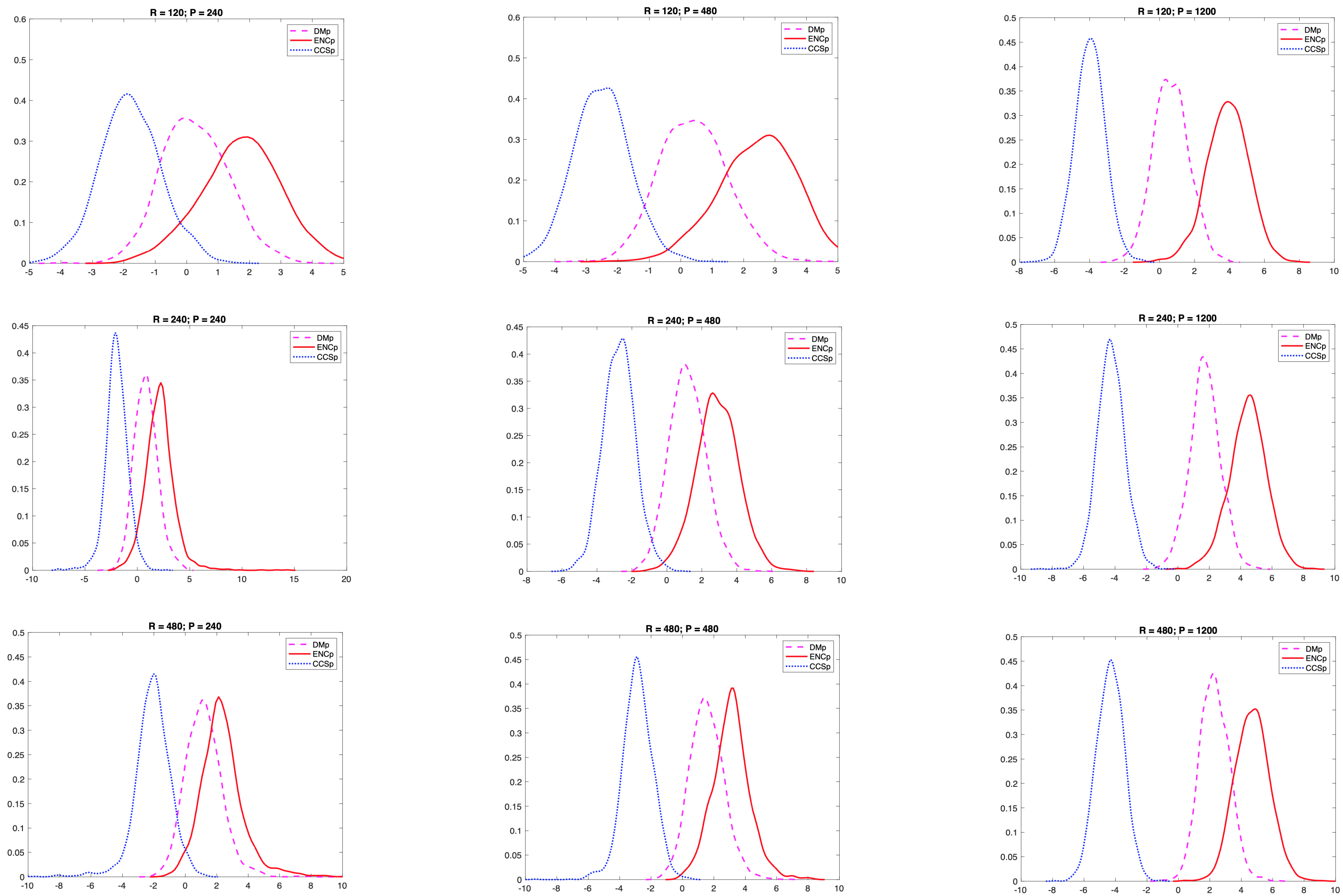

Figure 3: The distribution of DM_P, ENC_P, and CCS_P in DGP1, $\phi = 0.95$, $b = 0.1$, $\alpha = 0.05$, $\sigma_u = 1$, 2000 Repeats.

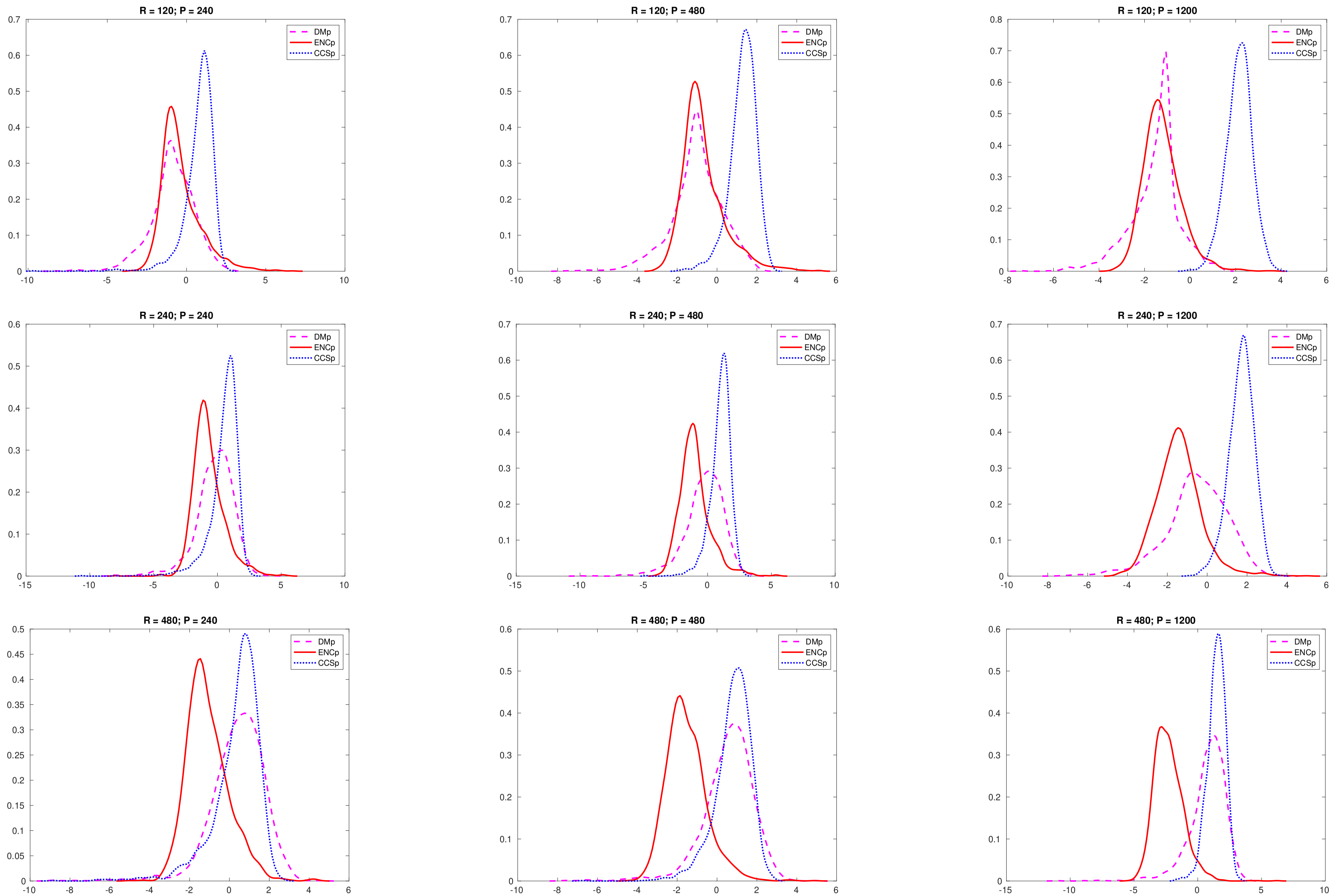

Figure 4: The distribution of DM_P, ENC_P, and CCS_P in DGP2, $\phi = 0$, $b = 0.1$, $\alpha = 0.05$, $\sigma_u = 1$, 2000 Repeats.

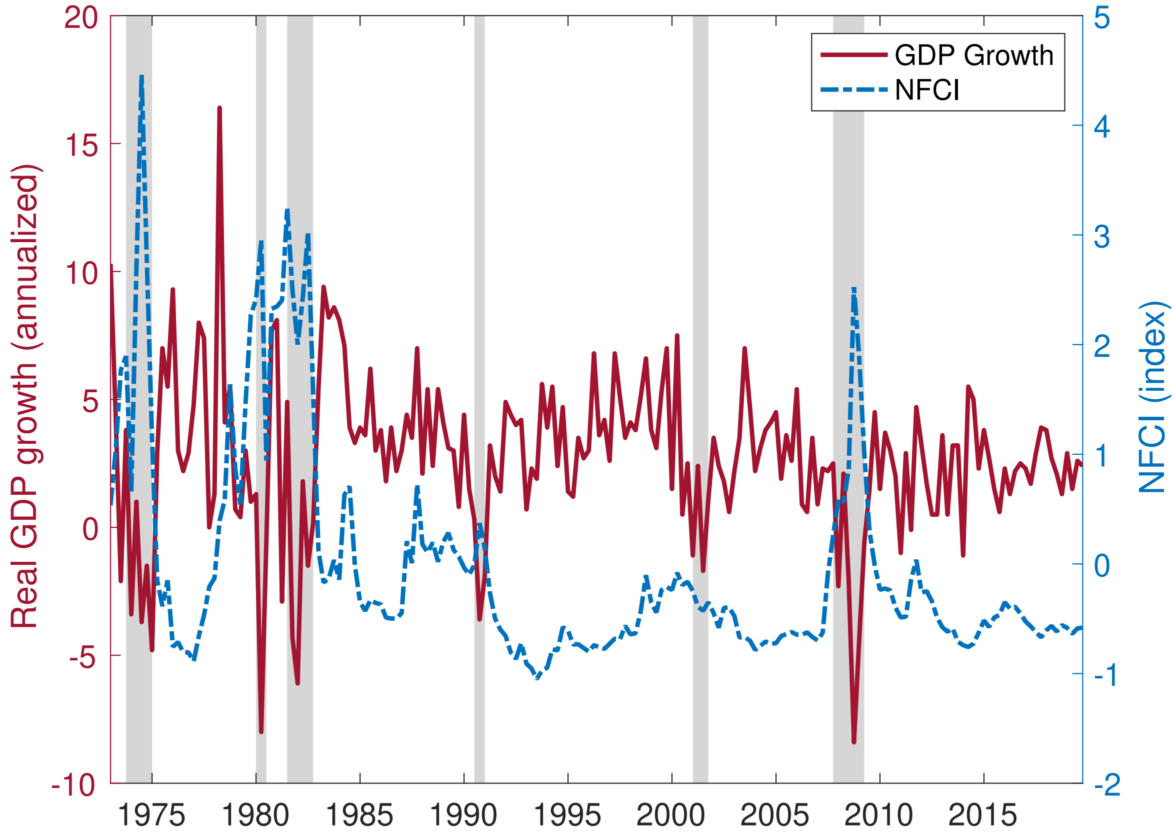

Figure 5: The times series of the GDP growth and the NFCI based on the quarterly data.

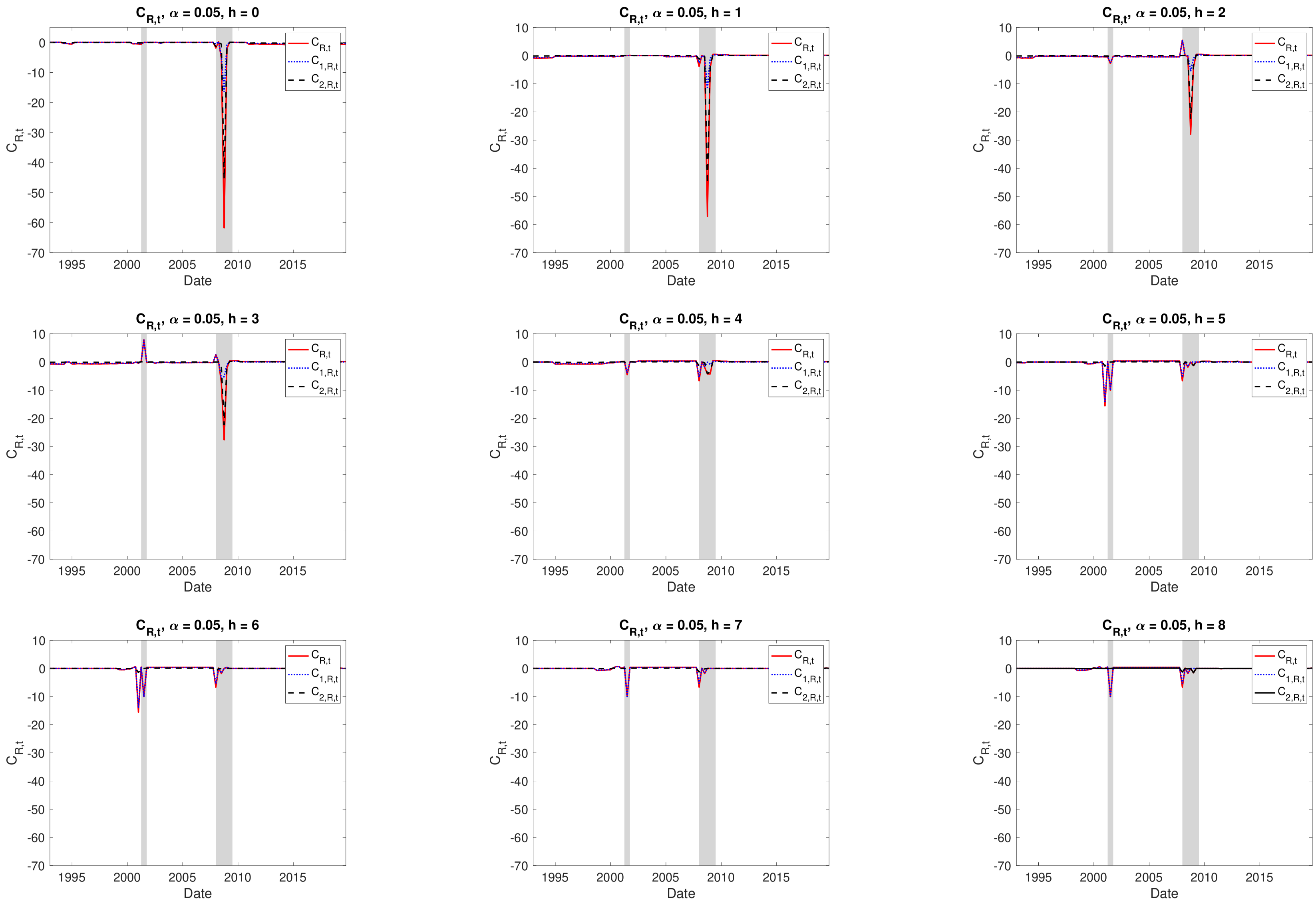

Figure 6: The time series of the $\hat{C}_{R,t}$, $\hat{C}_{1,R,t}$, and $\hat{C}_{2,R,t}$, where $\hat{C}_{1,R,t} = \hat{m}_{1,t}\left(G\hat{a}R_{2,t} - G\hat{a}R_{1,t}\right)$ and $\hat{C}_{2,R,t} = \hat{m}_{2,t}\left(\hat{GS}_{2,t} - \hat{GS}_{1,t}\right)$.

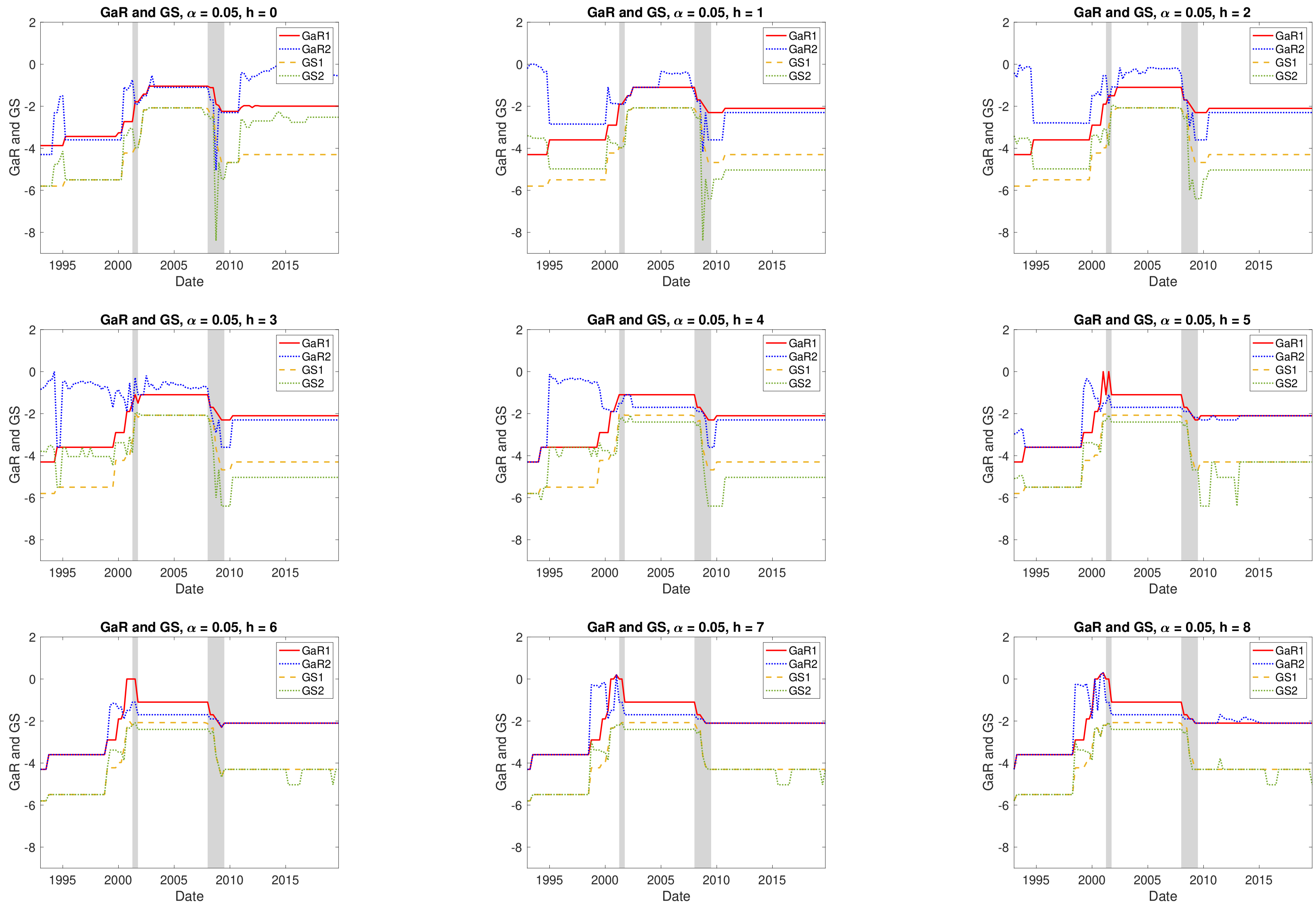

Figure 7: $\hat{GaR}_{1,R,t}$, $\hat{GaR}_{2,R,t}$, $\hat{GS}_{1,R,t}$, and $\hat{GS}_{2,R,t}$, $\alpha = 0.05$.