Global Research Unit
Working Paper #2021-004

How did the asset markets change after the Global Financial Crisis?

Kuang-Liang Chang, National Sun Yat-sen University
Charles Ka Yui Leung, City University of Hong Kong

How did the asset markets change after the Global Financial Crisis?

Kuang-Liang Chang and Charles Ka Yui Leung

This version: March 2021

Abstract

The Global Financial Crisis (GFC) changes the relative economic riskiness and risk-adjusted-performance of different asset markets. While the empirical distribution for stock return shifted to the right and became more concentrated around the mean after the GFC, the real estate market counterparts moved to the left and became more spread out. The economic risk of the OFHEO and Case-Shiller housing indices was smaller than the counterpart of the equity REIT (EREITs) market before the financial crisis, it substantially increased. Also, the economic performance of the OFHEO and Case-Shiller housing indices decreased after the financial crisis. They are below the performance indices of the stock and EREITs markets. The ex-post real estate premium vanishes. If we presume the "best model" to be the same before and after the GFC, we could severely misestimate the risk after the GFC.

Key Words: economic index of riskiness, risk-adjusted-performance index, real estate markets, stock markets
JEL Classifications: C50, G32, R30

Acknowledgements

The authors gratefully acknowledge the comments and valuable suggestions from Man Cho, Yuichiro Kawaguchi, Jin Man Lee, Jay Sa-aadu, the seminar participants at the annual conference of Chinese Society of Housing Studies, Asian Real Estate Society, and international conference of Taiwan Finance Association. Chang thanks the grant from Ministry of Science and Technology of Taiwan (MOST 104-2410-H-415-004) and Leung thanks the City University of Hong Kong for financial support. Joe Ng provides capable research assistance.

Correspondence

Chang, Department of Political Economy, National Sun Yat-sen University, Kaohsiung City, Taiwan (R.O.C.), kuangliangc@mail.nsysu.edu.tw ; Leung, Department of Economics and Finance, City University of Hong Kong, Kowloon Tong, Hong Kong, kycleung@cityu.edu.hk

1. Introduction

This paper focuses on two crucial aspects of the recent Global Financial Crisis (GFC). First, it is the underestimation and mispricing of risk.[1] Second, investors significantly change their portfolios after the GFC.[2] Also, central banks and governments responded with significant policy changes, such as adopting unconventional monetary policies by the major central banks and different "fiscal stimulus packages." With all these changes, the pre-crisis statistical models for the empirical asset return dynamics may or may not provide a good approximation after the GFC.[3]

Therefore, this paper first re-assesses the risk and risk-adjusted return (or "performance") of major assets, such as stock, equity real estate investment trusts (EREITs), and housing markets by utilizing different measures of economic risk and performance.[4] Second, this paper provides a simple method to assess whether statistical models built before the GFC continue to be useful in describing the return dynamics after the GFC. Implicitly, we adopt the statistics literature view that "all models are wrong, but some models are more useful than the others."[5] Hence, a model can lose the usefulness after some dramatic events such as the GFC. We devote a

---

[1] For instance, shortly after the collapse of Lehman Brothers, J. C. Trichet (January 2009), then President of the European Central Bank (ECB), stated that "… *the appropriate identification, assessment and handling of risks in the financial sector are the key issue to be considered most carefully amid the current global financial turmoil.*" (italics added).

[2] For instance, the Wall Street Journal (2012) reported, "*The landscape has changed for the asset management industry in the wake of the global financial crisis*. Investors are skittish after seeing their portfolios shrink in value…. admits Emad Mansour, chief executive officer of the Qatar First Investment Bank (QFIB). *'There is a general lack of interest by investors—both individuals and institutional—to plow money into markets that are directionless*.'" (italics added).

[3] For instance, Leung and Ng (2019) find that in business cycle frequency, many relationships among the housing market variables and macroeconomic variables have changed significantly after the GFC.

[4] Heaton and Lucas (2000) examine several waves of data from the Survey of Consumer Finance. The portfolio weights vary across years and levels of net worth. Stock accounts for 14~33% of the portfolio. Real estate accounts for 30~60% of the portfolio, cash accounts for 5~16% of the portfolio, bonds account for 3~11% of the portfolio, and pension accounts for less than 10% of the portfolio.

[5] See Lv and Liu (2014), among others, for more discussion on this point.

separate section for more discussion on this issue.

Risk and performance measurement are not trivial tasks.[6] Since we cannot directly observe risk, there are many different risk measures, and correspondingly, various estimates of risk-adjusted-performance -- such diversity in measurement matters. For instance, Homm and Pigorsch (2012) investigate performance rankings across 25 hedge funds and find that hedge fund rankings change dramatically when a distinct performance index is used. We choose an economic performance indicator that remains valid for non-normal distribution, as asset returns are typically non-normal.[7]

Economic reasons also drive our concerns for useful statistical models of asset returns dynamics. For both academic researchers and practitioners, statistical models summarize the history, guide future forecasts, and quantify our evaluation of risk and performance. For instance, if the asset return is independent and identically distributed (*i.i.d.*), as some previous studies assume, then the current period return would not help us predict future yield. However, if asset return is positively and serially correlated, a high performance observed in the current period would suggest a high return in the next period. It would thus alter our forecast for the near future.[8] Many "early warning systems" adopted by policymakers also assume some form of intertemporal dependence.

---

[6] The Sharpe ratio is a common and widely accepted measure for comparing the relative performance of asset markets. Aumann and Serrano (2008) consider those specific characteristics are necessary for the right risk measurement. They include continuity, positive homogeneity, subadditivity, and monotonicity. They demonstrate that widely used risk measures such as standard deviation and value-at-risk are inadequate for measuring price risk due to monotonicity failure concerning first-order dominance. Homm and Pigorsch (2012) recommend an economic performance index for estimating a risky asset's performance. That index has vital properties, such as scale invariance, monotonicity concerning first-order and second-order dominance, and generalized continuity.

[7] For instance, see Homm and Pigorsch (2012) and Schulze (2014) for analytically solvable examples. Note that the analytical solutions of the economic risk and performance indices are not always available. Chen et al. (2014), among others, utilize a numerical method to solve the economic index of riskiness for the student-t distribution, left-skewed log-normal distribution, and right-skewed log-normal distribution. They show that the economic index of riskiness is related to the mean, skewness, and kurtosis. Furthermore, Homm and Pigorsch (2012) observe no significant difference in the economic index of riskiness and the economic performance index for parametric specification and non-parametric specification.

[8] See Anand et al. (2015) for more discussion on this point.

Therefore, choosing a model with the "right" type of dependency is essential. This paper would estimate different models, each of which implicitly assumes another form of "intertemporal correlation."

Since housing plays a vital role in the GFC, we include its return dynamics and compare that with other financial assets.[9] Consequently, the highest frequency we can use is monthly data. Thus, this study would complement previous research, which uses higher frequency data and focuses only on financial assets (Mondria and Quintana-Domeque, 2013; Mollah et al. 2016; Lehkonen, 2015).

Constrained by the sample size, we would proceed with relatively simple models that we can estimate more accurately. They include four commonly used specifications, including the auto-regressive (AR) model, AR-GARCH model, Markov-switching AR model, and Markov-switching AR-ARCH model. While these models may be relatively simple, various authors have adopted these specifications to study asset returns' dynamic behaviors: stock, REIT, and real estate.[10] We, therefore, adopt these models in this study to facilitate the comparison. We then conduct a formal model-comparison to select the "best" model from our set of candidate models for the asset returns in each period (pre-crisis and post-crisis). Our empirical strategy allows for the possibility that, for some asset, the "best" model that describes the return dynamics before the GFC may no longer be so in the post-crisis era. As we have briefly reviewed, many changes have taken place in the asset market, and we cannot rule out such a possibility. We will provide more discussion on this point later.

This paper builds on several existing lines of literature. Our focus on the risk and

---

[9] Many authors have discussed this point. Among others, see Hendershott et al. (2010), Malpezzi (2017) for a review of the literature.

[10] For example, Dueker (1997), Maheu and McCurdy (2000), Engle and Patton (2001), Ang and Bekaert (2002), Crawford and Fratantoni (2003), Cotter and Stevenson (2006), Bredin et al. (2007), Miles (2008a, 2008b), Chen et al. (2009), Liow et al. (2009, 2011), Ashley and Patterson (2010), Chang (2010), Tsai et al. (2010), Zhou and Kang (2011), Chang et al. (2012) and Karoglou et al. (2013).

performance of different asset classes is related to the research on the “asset return premium,” including the “equity premium puzzle” highlighted by Mehra and Prescott’s (1985), “real estate premium puzzle” by Shilling (2003).[11] In this paper, we compute different riskiness measures and the risk-adjusted performance of various asset classes before and after the GFC. Therefore we contribute to the discussion on those puzzles.

This paper also builds on emerging literature insights, which compares the dynamic properties of different assets.[12] This paper departs from these studies by including the possibility that different asset returns could be proxied by various statistical models during different times. This paper also relates to the "structural change" literature, which assumes the same statistical model applies, but a change in parameter values has occurred.[13] With some abuse of notations, we may label such parameter change as “structural change in the intensive margin” (SCIM). At the same time, this paper allows for an additional possibility that even the statistical model has changed, which we may label as “structural change in the extensive margin” (SCEM). In this paper, our model selection procedure chooses the "optimal model," and hence differentiate the SCIM from the SCEM.

There is also an ongoing debate on the impacts on the financial market and the real economy of the unconventional monetary policy (UMP) and large scale asset purchase (LSAP) that the major central banks have conducted.[14] Our focus is more on the (potential) changes in the asset return distribution and the statistical models that we can use to approximate the asset return dynamics after the GFC, which would include not

---

[11] For the equity premium puzzle, see Mehra (2006) for a survey. For the real estate premium puzzle, see, for instance, Seiler et al. (1999), Cheng et al. (2010b), and Lin and Liu (2008), and their references to related discussions.

[12] It includes Quan and Titman (1999), Glascock et al. (2000), Liow and Yang (2005), Chang et al. (2011), etc.

[13] See Bai and Perron (1998), Chong (2001), Hansen (2001), Perron (2006), among others.

[14] Among others, see Bernanke (2018), D’Amico and King (2013), Eksi and Tas (2017), Eser and Schwaab (2016), Hamilton (2018), Steeley (2015), Weale and Wieladek (2016), and the reference therein.

only the effects of UMP but also other factors. Hence, this paper is related to that literature but with a different research focus.

The rest of this paper is organized as follows. Section 2 introduces the indices to measure economic risk and economic performance, the models we use to characterize the return processes and the Monte Carlo simulation method. Section 3 reports the different models' forecasting performance and explains their classification into different "equivalent predictive power classes" (EPPC). We also discuss empirical findings. The last section concludes the paper.

2. The economic performance measure and empirical specification

This paper utilizes the economic performance index developed by Aumann and Serrano (2008), Homm and Pigorsch (2012) to reinvestigate financial assets' risk and performance. For each asset during each sampling period, we allow its return dynamics follows one of the four widely used models, which essentially assumes the returns correlate over time in different ways. They are the AR, AR-GARCH, Markov-switching AR, and Markov-switching AR-ARCH models.[15] We adopt the Monte Carlo simulation method (MCSM) to compute asset returns' economic performance, as no closed-form solutions are available given these choices. Since alternative empirical models and alternative performance indices would be developed in the future, and they are unlikely to have closed-form solutions, this paper also demonstrates how a performance comparison can be conducted using the MCSM.

2.1 Economic index of riskiness and economic performance index

Although the Sharpe ratio is a widely accepted method used to compare financial

[15] Among others, see Bollerslev (2009), Chauvet and Piger (2008), and Duprey and Klaus (2017) for a review of the literature.

assets' relative performance and is easily implemented, it lacks monotonic stochastic dominance and axiomatic justifications. Recently, Homm and Pigorsch (2012) propose an alternative economic performance index (EPI), which is defined as follows:

$$\mathrm{EPI} = \frac{E(r_t - r_{f,t})}{AS(r_t - r_{f,t})} \tag{1}$$

where r_t is the nominal return of a risky asset; $r_{f,t}$ is the return of a risk-free asset; $r_t - r_{f,t}$ is the excess return; E denotes the expectation operator; and $AS(r_t - r_{f,t})$ refers to the economic risk index of Aumann and Serrano (2008). Like the Sharpe ratio, the EPI measure is also a *risk-adjusted* index. Aumann and Serrano (2008, Theorem A) prove that for any given gamble g, there is a unique positive number $R(g)$ that satisfies the equation

$$E\left(e^{-g/R(g)}\right) = 1. \tag{2}$$

Aumann and Serrano (2008) prove that their index has several desirable properties. However, we refer readers to the original papers to discuss the EPI and AS indices due to space limitations. We provide some highlights of these discussions in the appendix.

2.2 The time-series dynamics of excess return

This paper employs four widely-adopted models of the time-series dynamics of excess returns. Our first model is the simple autoregressive (AR) model. It is often regarded as an "atheoretical benchmark." Formally, it is given by

$$r_t - r_{f,t} = \theta_0 + \theta_1(r_{t-1} - r_{f,t-1}) + \varepsilon_t \tag{3a}$$

$$\varepsilon_t = \sqrt{\omega} z_t, \qquad z_t \sim N(0,1) \tag{3b}$$

Where r_t is the return of the asset, $r_{f,t}$ is the risk-free return, θ_0 and θ_1 are parameters that describe the autoregressive process; ω is the variance; z_t is a Gaussian random error with a mean of zero and a variance of one.

The second model is the AR-GARCH model, which Bollerslev (1986) developed to capture the fact that the innovation's variance may be time-varying. It is formulated as follows:

$$r_t - r_{f,t} = \theta_0 + \theta_1(r_{t-1} - r_{f,t-1}) + \varepsilon_t \tag{4a}$$

$$\varepsilon_t = \sqrt{h_t} z_t, \qquad z_t \sim N(0,1) \tag{4b}$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{4c}$$

where ω, α, and β are parameters that describe the conditional variance process and conditional variance h_t is a linear function of ε_{t-1}^2 and h_{t-1}. The persistence of conditional variance is captured by the sum of parameters α and β.

The third model captures the idea that the relationship between the current and previous period returns could change across different regimes. More specifically, the Markov Switching Autoregressive (MS-AR) model is an AR specification with two different dynamical processes, and can be specified as follows:

$$r_t - r_{f,t} = \theta_{0,s_t} + \theta_{1,s_t}(r_{t-1} - r_{f,t-1}) + \varepsilon_{t,s_t} \tag{5a}$$

$$\varepsilon_{t,s_t} = \sqrt{\omega_{s_t}} z_t, \qquad z_t \sim N(0,1) \tag{5b}$$

where s_t is a state variable that controls the pattern of regime switches. Constrained by data availability, we assume in this paper that the state variable has two states: 1 and 2. The evolution specification of s_t is assumed to depend on the last period's state variable. A logistic function is used to ensure that the transition probabilities of the state variable are positive values. The four transition probabilities are given by:

$$\mathrm{P}(s_t = 1 | s_{t-1} = 1) = P_{11} = \frac{\exp(p)}{1+\exp(p)} \tag{5c}$$

$$\mathrm{P}(s_t = 2 | s_{t-1} = 1) = P_{12} = \frac{1}{1+\exp(p)} \tag{5d}$$

$$\mathrm{P}(s_t = 1 | s_{t-1} = 2) = P_{21} = \frac{1}{1+\exp(q)} \tag{5e}$$

$$\mathrm{P}(s_t = 2 | s_{t-1} = 2) = P_{22} = \frac{\exp(q)}{1+\exp(q)} \tag{5f}$$

Our fourth model, MS-AR-ARCH, is an extension of the AR-ARCH specification and has the following dynamics:[16]

$$r_t - r_{f,t} = \theta_{0,s_t} + \theta_{1,s_t}(r_{t-1} - r_{f,t-1}) + \varepsilon_{t,s_t} \tag{6a}$$

$$\varepsilon_{t,s_t} = \sqrt{h_{t,s_t}} z_t, \qquad z_t \sim N(0,1) \tag{6b}$$

$$h_{t,s_t} = \omega_{s_t} + \alpha_{s_t} \bar{\varepsilon}_{t-1}^2 \tag{6c}$$

where s_t is a state variable, and $\bar{\varepsilon}_{t-1}$ is the conditional mean of ε_{t-1}, given the information set Ω_{t-2}. The transition probabilities of the state variable are the same as those described in the MS-AR model.

2.3 The Monte Carlo simulation method

Although the four models of asset return dynamics considered in this paper are somehow standard, closed-form solutions for the EPI and economic risk index are unavailable. Therefore, we follow the approach of Chen et al. (2014) to compute the EPI and economic risk index. Formally, we consider the following unconstrained minimization problem:

$$\min \qquad \left(\frac{1}{N}\sum_{t=1}^{N} exp\left(-\frac{r_t - r_{f,t}}{AS(r_t - r_{f,t})}\right) - 1\right)^2 \tag{7}$$

The indices can be obtained through the following steps:

Step 1: Estimate the parameters of the time-series model using the quasi-maximum likelihood estimation method.

Step 2: Simulate 100,000 observations of excess returns through a time-series process where the model parameters are estimated from Step 1.

Step 3: Calculate the economic index of riskiness, and the sample mean using the last 50,000 observations generated from Step 2.[17] Following Chen et al. (2014), the numerical optimization procedure is implemented using the OPTMUM procedure of

---

[16] See also Haas et al. (2004), Kim and Hwang (2018), Liu (2006), among others.

[17] The first 50,000 observations are omitted to remove the possible influence of starting values on simulated data.

the Gauss program.

Step 4: Calculate the economic performance index through the sample mean and the economic index of riskiness obtained from Step 3.

A merit of this method is that it is simulation-based. Hence, the small sample issue is not a concern in our calculation.

3. Data and empirical findings

3.1 Data

This paper investigates the economic index of riskiness and several asset market indices' economic performance at a monthly frequency.[18] The indices include the Standard and Poor's 500 Stock Index, Dow Jones Industrial Average Index, Nasdaq Composite Index, FTSE NAREIT All Equity REITs Price Index, Office of Federal Housing Enterprise Oversight (OFHEO) purchase-only index, and S&P/Case-Shiller U.S. National Home Price Index. As the GFC might have changed the market participants' expectations and investment strategies, we allow for the possibility that the statistical model that works well with a particular asset before the crisis may no longer work well after. In this paper, the pre-crisis period covers 2000:m1-2006:m6. The post-crisis period covers 2009:m1-2019:m12, and the "in-crisis" period is avoided.[19] We also avoid the Russian financial crisis's potential effect in 1998 by starting our sample in 2000.

---

[18] We do not have access to data that more frequent than monthly observations. Studies have examined the joint dynamics of different assets at a quarterly rate (e.g., Chang et al., 2011, 2012, 2013). However, the sample size is relatively small in that case, and the estimation of pre-crisis versus post-crisis sub-sample could be difficult. Therefore this study, which is based on monthly data, naturally complements the previous studies.

[19] There are different classifications of the pre-crisis versus post-crisis periods. See Dungey et al. (2015) for a detailed analysis. Since we employ monthly data, we cannot divide the sample into more than two sub-periods.

Several studies suggest that lending and trading behaviors during the GFC are very different from that before the crisis. Hence, it may be essential to examine the pre-crisis and post-crisis periods separately. Among others, see Afonso et al. (2011), Brunnermeier (2009), Gorton (2010), Ivashina and Scharfstein (2010) for more details.

Our data come from the usual sources. Standard and Poor's 500 Stock Index and the Dow Jones Industrial Average Index are obtained from Datastream. The OFHEO purchase-only Index and the Equity RETIS Index are obtained from the Federal Housing Finance Agency and the REIT.com website. The Nasdaq Composite Index and three-month Treasury Bill interest rate come from the Federal Reserve Bank of St. Louis. The S&P/Case-Shiller Index comes from S&P Dow Jones Indices LLC. The annualized excess return is derived from by subtracting the three-month Treasury Bill rate from the annualized return of a given financial asset.

According to the transaction process, transaction time and transaction cost, the three stock indices and the Equity REITs index are labeled as "high-liquidity assets" in this paper. In comparison, the two national housing indices are considered as "low-liquidity assets." Figures 1 and 2 provide the time-series plots of annualized excess returns for the pre- and post-crisis periods. The time-series patterns for the high- and low-liquidity assets are entirely different. In Figure 1, the annualized excess returns for high-liquidity assets increase significantly around 2003. However, the counterpart of the low-liquidity assets first increase and then decline sharply before the GFC. Figure 2 shows that the high-liquidity assets' returns decrease slightly over time, while the low-liquidity assets' counterparts significantly increase around 2011-2013.

(Figures 1 and 2 about here)

3.2 Empirical findings

In the main text, we use the root mean square error (RMSE) to compare the in-sample forecasting performance across four empirical models.[20] Both the return forecasts and variance forecasts are examined. Table 1 reports the return forecasts and

[20] We also adopt the MAE (mean absolute error) and report the corresponding results in the appendix. Compared to MAE, RMSE tends to “punish more” for significant forecast error and hence might be more suitable for the study of risk in this paper.

variance forecasts for the pre-crisis period.[21] For the post-crisis period, the forecasting performance for return forecasts and variance forecasts is reported in Table 2. The MS-AR model has a smaller RMSE than the other three models in most cases for the pre-crisis period.[22] The MS-AR model seems to be more accurate for return forecasts during the post-crisis period, but it is not always superior for variance forecasts.

(Tables 1-2 about here)

While the root-mean-square errors across models are not the same, the difference might not be statistically significant in differentiating their forecast performance. Combining the results in Hansen et al. (2011), Mariano and Preve (2012), Kwan et al. (2015), Chang et al. (2016) propose a procedure to classify competing models into different equivalent predictive power classes (EPPCs). Models in the same class have the same predictive power, while models in Class 1 have higher predictive power than the models in Class 2. The square loss criterion (SLC) is used here to implement the EPPC procedure.[23] Tables 3 and 4 summarize the relative model performance for the pre- and post-crisis period, respectively. The predictive power classes in the pre- and post-crisis periods are slightly different. Panel A of Table 3 shows that the four models have the same forecasting ability in return forecasts. Panel B of Table 3 shows that the four models can be divided into two classes for three different stock returns. For the SP500 stock return, the MS-AR and MS-ARCH models belong to Class 1, and they produce more accurate forecasts than the AR and AR-GARCH models, which are classified as Class 2. However, for the EREITs, OFHEO housing index, and Case-

[21] See Gray (1996) for more details about the forecasting error of conditional variances.
[22] The AR-ARCH model is selected for two housing returns. For the S&P/Case-Shiller return, the MS-ARCH model does not converge.
[23] We also repeat the analysis with the absolute loss criterion (ALC). Again, SLC tends to impose a more massive penalty on "large error" than ALC. The results are similar and are reported in the appendix.

Shiller housing index, the four specifications' variance forecasts are not statistically different. In other words, for stock returns, models differ not in terms of forecasting the performance, but in terms of predicting the risk in the pre-crisis period. For real estate-related assets, there is no difference in terms of return or variance.

In the post-crisis period, Panel A of Table 4 shows that the four models do not have a statistically significant difference in return forecasts for SP500 returns, Down Jones stock returns, and two housing returns (OFHEO and Case-Shiller). Moreover, the variance forecasts of the four models for three stock returns are not significantly different. Hence, models display no difference in forecasting stock returns in the post-crisis era. For the Case-Shiller return, the MS-AR, and MS-ARCH models produce more accurate forecasts in return (Class 1) than the AR, AR-ARCH models (Class 2). However, for the OFHEO return, the AR, AR-ARCH, and MS-ARCH models predict better than the MS-AR, and the latter is therefore considered Class 2.

(Tables 3-4 about here)

We define the "best model" as the model which delivers the smallest RMSE for Class 1 models. We then use the estimated parameters to simulate each of the asset returns.[24] We then use the Epanechnikov kernel density to find the empirical distribution based on the simulated data, using Silverman's bandwidth (1992). We simulate it for 100,000 periods and then drop the first half (sometimes regarded as the "training period"), using only the second half to "plot" the empirical distribution and other calculations (such as performance measures). Because we use 50,000 periods, "small sample bias" is not a concern.

The empirical distributions of asset returns are plotted in Figures 3. The solid

[24] If the best model is non-stationary, then a second-best model is chosen. If the second best model is non-stationary, a third best model is used to simulate the observations. This paper uses the third-best model for the S&P/Case-Shiller housing return during the pre-crisis period.

(*dotted*) line shows the simulated distribution for the pre-crisis (*post-crisis*) period. The empirical distributions across different asset markets are very different. Besides, its empirical distributions before and after the crisis also differ for any given asset market. For instance, in the pre-crisis period, the SP500 stock return and OFHEO housing return have a bimodal distribution, and the remaining asset markets have a unimodal distribution. In the post-crisis period, the OFHEO housing returns have an asymmetric bimodal distribution, while the high-liquidity markets have an asymmetric unimodal distribution.[25]

(Figures 3 and 4 about here)

Compared with the empirical distribution in the pre-crisis period return, the *post-crisis distributions for the three stock market indices shifts to the right*. The value ranges of the three stock market indices in the post-crisis period are narrower. It means that there is a *decrease in extreme negative returns* and a *reduction in return variation* for financial asset markets. In contrast, the empirical distributions of the OFHEO and Case-Shiller's housing returns *shift to the left* in the post-crisis period. And the *ranges of the distributions are more extensive* in the post-crisis period than in the pre-crisis period, suggesting an *increase in extreme negative return and growth in return variation*. For the EREITs returns, the empirical distribution is more or less symmetric around the mean in both the pre- and post-crisis periods. The observed distribution in the post-crisis period has "a little shorter left tail and a longer right tail fatter left tail" than the pre-crisis period, which also suggests a *decrease in extreme negative return* and an *increase in extra-ordinary positive returns*.[26]

---

[25] Bimodal distribution or "twin peaks" have been studied in the economics literature. For instance, see Quah (1996).

[26] It should be clear that while our sample may be relatively small, the "small sample issue" (SSI) should apply to both pre-crisis and post-crisis sub-samples. SSI by itself would not make "model switching" more or less likely to happen. For instance, MS-AR(1) is the best model to explain the housing return

Notice that Figure 3 is based on the ex-post data. When agents need to make economic decisions in real-time, they *may not be sure of the proper distribution* and would need to *take the uncertainty of variances into considerations*. Figure 4 provides such an estimate. The empirical distribution of the (implied) variance for the post-crisis period is *more concentrated around the mean* than that for the pre-crisis period for all asset returns, which means that the uncertainty of variances becomes less severe post-crisis period. For instance, the "best model" for Down Jones returns in the post-crisis period is an AR specification with constant variance. Hence, Figure 4 shows that its distribution of the variance of return collapses to a vertical line during the post-crisis period. One possible explanation is the active participation of the U.S. government in the asset markets through TARP and QE policies are useful in shortening an otherwise more prolonged recession and reducing the uncertainty of variance (Brookings, 2018; Liang et al., 2018).

Table 5 summarizes different measures of risk for the best performing models. Panels A and B show the variances and the 1% value-at-risk estimates, respectively.[27] Several observations are in order. The values of variance and value-at-risk are more substantial for the more liquid assets than the less liquid ones, regardless of the period. Moreover, *according to the variance and value-at-risk, the risk measure for three stock returns dramatically decreases after the financial crisis*. This observation is consistent with Figures 3(a)-3(c) that the range of stock returns' empirical distributions falls. The observed distributions also move to the right after the GFC, thus decreasing the risk measures.

On the other hand, the variances and value-at-risk measures for two housing

(S&P/Case-Shiller) for both the pre-crisis and post-crisis periods. SSI would not make models that implied “more risk” to be selected. The results of our model selection procedures suggest a decrease in risk in financial asset returns and an increase in housing returns in the post-crisis period at the same time.

[27] The value-at-risk measure at $\alpha\%$ is defined as the αth percentile of the distribution of excess returns multiplied by -1. The larger the value-at-risk, the larger the negative return.

markets increase after the GFC. This observation is consistent with Figures 3(e)-3(f) that the range of the empirical distributions of housing returns increases. The observed distributions of housing returns move to the left after the GFC, thus increasing the risk measures. For the EREITs returns, the variance increases, but the value-at-risk measure decreases after the GFC. The value-at-risk substantially decreases for the three stock returns and EREITs returns after the financial crisis. Specifically, there is a consistent finding that the variance and value-at-risk reduce after the financial crisis for stock assets and housing assets. There is no consistent finding for EREITs returns.[28] Furthermore, the view that real estate assets have smaller risks than high-liquidity assets is confirmed.

(Table 5 about here)

Panels C and D of Table 5 report the AS_N and AS values, respectively.[29] Because the simulated sample mean is negative for the three stock markets, the AS_N and AS are not estimated here.[30] As Figure 3 shows, the empirical distribution of the simulated data is distinct from the normal distribution. Thus, the AS may provide a better measure of the risk. On the other hand, AS_N is widely used, and hence we also offer the AS_N value for reference.

For real estate-related assets, the AS_N value is smaller than the AS value for the EREITs and S&P/Case-Shiller housing returns during the pre-crisis period, suggesting

[28] Our results are different from Zhou and Anderson (2012) due to several reasons. First, they use daily data while we use monthly data. Second, their sampling period is from 1993 to 2009, while we explicitly compare the pre-crisis (before the Global Financial Crisis) period from the post-crisis period. Third, they cover nine REIT market across the nations while we compare different asset markets within the United States.

[29] When the excess return is normally distributed, the AS index coincides with the AS_N index. See the Appendix for more details.

[30] The economic index of riskiness proposed by Aumann and Serrano (2008) operates under the assumption that the ex-ante excess return is non-negative. When the mean of asset returns is negative, the economic index is undefined. The intuition is that if the ex-ante excess return is negative, then the investor does not need to hold this risky asset as a risk-free asset is a better choice. In this paper, we focus on the excess ex-post return, which is possible to be negative.

an underestimation of risk. Simultaneously, the AS_N value is larger than the AS value for the OFHEO purchase-only housing returns, which indicates an overestimation of risk. Furthermore, the two housing markets' risk measure is smaller than that for the EREITs market, irrespective of the adopted economic index.

After the GFC, the risk measure for two housing returns (OFHEO purchase-only and S&P/Case-Shiller) *dramatically increases, regardless of the type of economic index of riskiness adopted*. More specifically, for the OFHEO housing return, the AS increases from 0.625 to 2.192, and the AS_N, increases from 0.791 to 2.150. The AS index for the S&P/Case-Shiller housing returns increases from 1.451 to 9.603. For the AS measure, the economic risk measure for the S&P/Case-Shiller is higher than that for the more liquid assets during the post-crisis period. The S&P/Case-Shiller housing returns show the highest risk, followed by the Nasdaq stock returns, ERIETs, Dow Jones stock returns, SP500 stock returns, and OFHEO purchase-only housing returns. *The housing markets have a more considerable risk than the stock markets and RERITs market.* Finally, we also confirm the literature finding that the risk ranking across assets is sensitive to the choice of risk measures.

(Table 6 about here)

Table 6 summarizes the performance measures for the best specification. The economic performance measures include the Sharpe ratio, EPI_N, and EPI. As Panel A of Table 6 shows, the Sharpe ratios for the three stock markets are negative before the financial crisis but are positive after the financial crisis. The Sharpe ratios of the three real-estate-related assets (EREITs, OFHEO purchase-only, and S&P/Case-Shiller) decrease after the financial crisis. Panel A of Table 6 also shows that in terms of the Sharpe ratio, the housing returns are higher than the more liquid assets (stock and REIT) only in the pre-crisis period. It suggests that the *asset markets do not compensate for* *the illiquidity of housing in the post-crisis period*. Panels B and C of Table 6 further confirm that, during the pre-crisis period, the two housing market indices outperform the EREITs index. However, after the financial crisis, the illiquid housing does not perform as well as the more liquid assets, irrespective of the economic performance measures. In other words, the illiquidity premium for housing vanishes in ex-post terms. Furthermore, the high- and low-liquidity assets' performance rankings do not change the Sharpe ratio and EPIN. The EPIN and EPI measures show a slight difference in performance ranking.[31]

One may wonder how it can be possible that the less liquid assets deliver lower risk-adjusted returns than other investments in equilibrium. We conjecture that several explanations are possible. Unlike financial assets, housing is traded in a lumpy fashion: we either sell the whole housing unit or not. Macroprudential measures also limit the number of mortgage loans and participation in the housing market (IMF, 2014). At the same time, the Obama government has imposed many urban and housing policies, which may change the incentives of the housing market participants (Agarwal and Varshneya, 2020; DeFilippis, 2016). This paper focuses on estimating and comparing the risk and returns across different asset markets before and after the GFC, leaving lower returns in the housing market issue for future research.

4. How important is the model selection procedure?

We receive a common question from different conferences and seminars: how important is the model selection procedure in the current context? How would it affect our results? To address this question, we conduct the following experiment. Consider an asset i, $i = 1,2,\ldots$ Assume that the best model to explain the return of asset i

[31] Furthermore, the performance rankings of the high- and low-liquidity assets do not change the Sharpe ratio and EPI_N. The EPI_N and EPI measures show a slight difference in performance ranking.

during the pre-crisis period is $j(i)$. Notice that if the best model in the pre-crisis period maintains the best model in the post-crisis period, there is no need for model selection. We simple re-estimate the parameters with the post-crisis period data of asset i. For future reference, we label this re-estimated model as $j'(i)$. Thus, $j(i)$ and $j'(i)$ differ only in parameter values, but not functional forms.

On the other hand, we have already selected the best model for asset i during the post-crisis period in the previous section, based on Kwan et al. (2015), Chang et al. (2016). For future reference, we label that model as $k(i)$. If model $k(i)$ and $j'(i)$ are the same, our model selection procedure simply confirms that the same model maintains its supremacy over the other models. For instance, MS-AR(1) is the best model to explain the SP500, Nasdaq, and EREITs returns for both the pre-crisis and post-crisis periods. On the other hand, in the previous section, we have seen that $k(i)$ and $j'(i)$ can be different models. For instance, while the best model for OFHEO purchase-only return in the pre-crisis period is MS-AR(1), it is replaced by MS-ARCH in the post-crisis period. Notice that we select the "best model" by minimizing the sum of some (conditional) forecast error. In Figure 5, we plot the distributions of return of the data, and the counterpart implied by $k(i)$ and $j'(i)$. Here are more detailed remarks.

(Figure 5 and 6 about here)

(1) SP500, Nasdaq, and EREITs have the same best model before and after the Global Financial Crisis.

(2) The difference between the best pre-crisis model applies to the post-crisis period, and the best post-crisis model is insignificant for Dow Jones.

(3) For the S&P/Case-Shiller index, the empirical distribution of actual data is a Bimodal distribution. Yet the observed distribution evaluated from the pre-crisis best

model apply to the post-crisis period is a single-peak distribution. Hence, if we insist on the best model in the pre-crisis period continues to be the best model in the post-crisis period, the bias could be enormous. Now, through the model selection procedure, our best model in the post-crisis period also generates a bimodal distribution, which mimics the distribution of the data.

In Figure 6, we apply the same logic to model-implied variance. The distributions generated by $j'(i)$ (i.e., when we assume the same model maintains to be the best before and after the crisis) are, in general, very different from that generated by $k(i)$, the model selected by our model comparison procedure. And in many cases, the distribution implied by $k(i)$ appears to be "closer" to that generated by $j'(i)$. Here are more detailed remarks.[32]

(1) Notice that the "true variance" is unknown. Only the empirical distributions for variances of the best pre-crisis model apply to the post-crisis period, and the best post-crisis model is plotted.

(2) The best model for Nasdaq does not change before and after the GFC.

(3) For EREITs, OFHEO purchase-only, and S&P/Case-Shiller, the pre-crisis best model applies to the post-crisis period has a more considerable variance than the post-crisis best model refers to the post-crisis period. If we ignore the structural change and use the same model found in the pre-crisis period, the "uncertainty about variance" will be overestimated.

(4) For SP500, the risk is more spread out in the post-crisis best model than the pre-crisis counterpart. Thus, the pre-crisis best model applies to the post-crisis period will

---

[32] Notice that as the bandwidth suggested by Silverman (1992) is used, $h = \frac{0.9}{(N)^{1/5}} \min(\sqrt{V(X)}, \frac{Q_3 - Q_1}{1.349})$, where Q_3 is the third quartile and Q_1 is the first quartile. For the MS-AR model, there are two different variances. When the $Q_3 - Q_1 = 0$, the maximum variance replaces the Q_3, and the minimum variance replaces Q_1.

under-estimate the "*uncertainty about variance*."

(5) In Figure 6(b) (Dow Jones), the shape of the empirical distribution of variance generated by the best model in the pre-crisis period and the counterpart in the post-crisis period is very similar. But, the post-crisis best model's empirical distribution applies to the post-crisis period is slightly in the right of that of the pre-crisis best model apply to the post-crisis period. If we ignore the structural change and use the same model found in the pre-crisis period, the "possibility of occurring higher variance" will be underestimated.

In sum, our paper demonstrates that even for the same asset, there is a need to choose the "best model" in different periods. If the best model selection mechanism is ignored, the return and risk can be significantly mismeasured.

5. Conclusions

This paper began with an insightful observation by Mr. J. C. Trichet that an inappropriate measure of risk is crucial. An inadequate economic risk measure can lead to suboptimal investment strategies for individuals and governments' misdirected policy. Several authors have made outstanding contributions in this area. For instance, Aumann and Serrano (2008) develop an economic index of riskiness to evaluate an asset's price risk. Homm and Pigorsch (2012) develop a performance index and show that their index is superior to the usually adopted Sharpe ratio. Recently, in a review of the GFC, Malpezzi (2017) also observes that "…*Many risks are 'fats in the tails,' and anyway rare events do occur*…" (italics added). This paper compares relative risk and relative performance across housing, REITs, and stock markets in light of these contributions. Our models allow for different forms of "fat tails in risk." This paper selects the "best" model for each asset return before and after the GFC based on the recent econometrics literature. We find that the "best" model does change in some cases.

Furthermore, the risk measures and risk-adjusted-performance ranks also significantly change after the GFC. In particular, the empirical distributions of the OFHEO purchase-only housing returns and S&P/Case-Shiller returns shift considerably to the left and become more spread out after the GFC, resulting in an increase in variance and growth in value-at-risk measure. The two housing markets have the same structures of economic riskiness. The economic risk increases for OFHEO purchase-only and S&P/Case-Shiller housing returns after the GFC. The ex-post real estate premium virtually *disappears* after the financial crisis.

This paper can be extended in different directions. We have established that the riskiness and risk-adjusted-performance of the major asset markets change significantly after the Global Financial Crisis. However, we have not investigated the causes and mechanisms. Future research should fill the research gap. We should also expand the analysis to asset markets outside the U.S. One may also explore whether the co-movement among different asset returns changes after the GFC.[33] Some of these directions are being pursued and will hopefully enrich our understanding of the asset markets and financial crises.

[33] Among others, see Leung and Ng (2019) and the reference therein.

References

Afonso, G., Kovner, A., Schoar, A., 2011. Stressed, not frozen: The Federal Funds market in the financial crisis, Journal of Finance, 66, 1109-1139.

Agarwal, S., Varshneya, S., 2020. Financial Crisis and the US Mortgage Markets – A Review, mimeo.

Anand, A., Li, T., Kurosaki, T., Kim, Y.S., 2015. Foster-Hart risk and the too-big-to-fail banks: An empirical investigation. mimeo.

Ang, A, Bekaert, G., 2002. International asset allocation with regime shifts. Review of Financial Studies, 15, 1137-1187.

Ashley, R. A., Patterson, D. M., 2010. A test of the GARCH(1,1) specification for daily stock returns. Macroeconomic Dynamics, 14, 137-144.

Aumann, R. J., Serrano, R., 2008. An economic index of riskiness. Journal of Political Economy, 116, 810-836.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes, Econometrica, 66, 47-78.

Bernanke, B., 2018. The real effects of the financial crisis, Brookings Papers on Economic Activity, forthcoming.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 31, 307-327.

Bollerslev, T., 2009. Glossary to ARCH (GARCH), Duke University, mimeo.

Bredin, D., O'Reilly, G., Stevenson, S., 2007. Monetary shocks and REIT returns. Journal of Real Estate Finance and Economics, 35, 315-331.

Brookings Institution, 2018, Charting the Financial Crisis: U.S. Strategy and Outcomes, mimeo.

Brunnermeir, M., 2009. Deciphering the liquidity and credit crunch 2007-2008. Journal of Economic Perspectives, 23, 77-100.

Chang, K. L., 2010. House price dynamics, conditional higher-order moments, and density forecasts. Economic Modelling, 27, 1029-1039.

Chang, K. L., N. K. Chen, N. K., Leung, C. K. Y., 2011. Monetary policy, term structure and real estate return: Comparing REIT, housing and stock, Journal of Real Estate Finance and Economics, 43, 221-257.

Chang, K. L., N. K. Chen, N. K., Leung, C. K. Y., 2012. The dynamics of housing returns in Singapore: How important are the international transmission mechanisms? Regional Science and Urban Economics, 42, 516-530.

Chang, K. L., N. K. Chen, N. K., Leung, C. K. Y., 2013, In the shadow of the United States: The international transmission effect of asset returns, Pacific Economic Review, 18(1), 1-40.

Chang, K. L., N. K. Chen, N. K., Leung, C. K. Y., 2016. Losing track of the asset markets: The case of housing and stock, International Real Estate Review, 19, 435-492.

Chauvet, M., Piger, J., 2008. A comparison of the real-time performance of business cycle dating methods, Journal of Business and Economic Statistics, 26, 42-49.

Chen, W. S., So, K. P., Lin, M. H., 2009. Volatility forecasting with double Markov switching GARCH models. Journal of Forecasting, 28, 681-697.

Chen, Y. T., Ho, K. Y., Tzeng, L. Y., 2014. Riskiness-minimizing spot-futures hedge ratio. Journal of Banking and Finance, 40, 154-164.

Cheng, P., Lin, Z., Liu, Y., Zhang, Y., 2010a. Illiquidity and portfolio risk of thinly traded assets. Journal of Portfolio Management, 36, 126-138.

Cheng, P., Lin, Z., Liu, Y., Zhang, Y., 2010b. Real estate allocation puzzle in the mixed-asset portfolio: Fact or fiction? Mimeo.

Chong, T. T. L., 2001. Structural Change In AR(1) Models, Econometric Theory, 17(01), 87-155.

Cotter, J., Stevenson, S., 2006. Multivariate modeling of daily REIT volatility. Journal of Real Estate Finance and Economics, 32, 305-325.

Crawford, G. W., Fratantoni, M. C., 2003. Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. Journal of Real Estate Finance and Economics, 31, 223-243.

D'Amico, S., King, T., 2013. Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply, Journal of Financial Economics, 108, 425-448.

DeFilippis, J. ed., 2016. Urban Policy in the Time of Obama, Minneapolis: University of Minnesota Press.

Dueker, M. J., 1997. Markov switching in GARCH processes and mean-reverting stock-market volatility. Journal of Business and Economic Statistics, 15, 26-34.

Dungey, M., Milunovich, G., Thorp, S., Yang, M., 2015. Endogenous crisis dating and contagion using smooth transition structural GARCH, Journal of Banking & Finance, 58, 71-79.

Duprey, T., Klaus, B., 2017. How to predict financial stress? An assessment of Markov switching models, ECB, mimeo.

Eiling, E., Gerard, B., 2015. Emerging Equity Market Comovements: Trends and Macroeconomic Fundamentals, Review of Finance, 19, 1543–1585.

Eksi, O., Tas, B., 2017. Unconventional monetary policy and the stock market's reaction to Federal Reserve policy actions, North American Journal of Economics and Finance, 40, 136-147.

Engle, R. F., Patton, A. J., 2001. What good is a volatility model? Quantitative Finance,

1, 237-245.
Eser, F., Schwaab, B., 2016. Evaluating the impact of unconventional monetary policy measures: Empirical evidence from the ECB's Securities Markets Programme, Journal of Financial Economics, 119, 147-167.
Foster, D. P., Hart, S., 2009. An operational measure of riskiness. Journal of Political Economy, 117, 785-814.
Glascock, J. L., Lu, C., So, R. W., 2000. Further evidence on the integration of REIT, bond, and stock returns, Journal of Real Estate Finance and Economics, 20, 177-194.
Gorton, G., 2010. Information, liquidity, and the (ongoing) panic of 2007. American Economic Review, Papers and Proceedings, 99, 567-572.
Gray, S. F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics, 42, 27-62.

Haas, M., Mittnik, S., Paolella, M., 2004. A New Approach to Markov-Switching GARCH Models, Journal of Financial Econometrics, 2(4), 493–530.
Hamilton, J., 2018. The efficacy of large-scale asset purchases when the short-term interest rate is at its effective lower bound, Brookings Papers on Economic Activity, forthcoming.
Hansen, B., 2001. The New Econometrics of Structural Change: Dating Breaks in U.S. Labour Productivity, Journal of Economic Perspectives, 15(4), 117-128.
Hansen, P., Lunde, A., Nason, J., 2011. The model confidence set. Econometrica, 79(2), 453-497.
Heaton, J., and Lucas, D., 2000, Portfolio choice and asset prices: The importance of entrepreneurial risk, Journal of Finance 55, 1163–1198.
Hendershott, P., Hendershott, R., Shilling, J., 2010. The mortgage finance bubble: causes and corrections, Journal of Housing Research, 19, 1-16.
Homm, U., Pigorsch, C., 2012. Beyond the Sharpe ratio: An application of the Aumann-Serrano index to performance measurement. Journal of Banking and Finance, 36, 2274-2284.
International Monetary Fund IMF, 2014, Staff Guidance Note on Macroprudential Policy—Detailed Guidance on Instruments, mimeo.
Ivashina, V., Scharfstein, D., 2010. Bank lending during the financial crisis of 2008. Journal of Financial Economics, 97, 319-338.
Karoglou, M., Morley, B., Thomas, D., 2013. Risk and Structural instability in US home prices. Journal of Real Estate Finance and Economics, 46, 424-436.
Kim, Y., Hwang, E. 2018. A dynamic Markov regime-switching GARCH model and its cumulative impulse response function, Statistics & Probability Letters, 139, 20-30.

Kwan, Y. K., Leung, C. K. Y. and Dong, J., 2015. Comparing consumption-based asset pricing models: the case of an Asian city, Journal of Housing Economics, 28, 18-41.

Lehkonen, H., 2015. Stock Market Integration and the Global Financial Crisis, Review of Finance, 19, 2039–2094.

Leung, C. K. Y., Ng, C. Y. J., 2019, Macroeconomic Aspects of Housing. In Hamilton, J., Dixit, A., Edwards, S., Judd, K. (eds), Oxford Research Encyclopedia of Economics and Finance, Oxford University Press.

Liang, N., McConnell, M. M., and Swagel, P., 2018, Evidence on Outcomes, Brookings Institution, mimeo.

Lin, Z., Liu, Y., 2008. Real estate returns and risk with heterogeneous investors. Real Estate Economics, 36, 753-776.

Liow, K. H., Chen, Z., Liu, J., 2011. Multiple regimes and volatility transmission in securitized real estate markets. Journal of Real Estate Finance and Economics, 42, 295-328.

Liow, K. H., Ho, K. H. D., Ibrahim, M. F., Chen, Z., 2009. Correlation and volatility dynamics in international real estate securities markets. Journal of Real Estate Finance and Economics, 39, 202-223.

Liow, K. H., Yang, H., 2005. Long-term co-memories and short-run adjustment: securitized real estate and stock markets. Journal of Real Estate Finance and Economics, 31, 283-300.

Liu, J. C., 2006. Stationarity of a Markov-Switching GARCH Model, Journal of Financial Econometrics, 4(4), 573–593.

Lv, J., Liu, J., 2014. Model selection principles in misspecified models. Journal of the Royal Statistical Society, Series B, 76, 141-167.

Maheu, J. M., McCurdy, T. H., 2000. Identifying bull and bear markets in stock returns. Journal of Business and Economic Statistics, 18, 100-112.

Malpezzi, S., 2017. Residential real estate in the U.S. financial crisis, the Great Recession, and their aftermath, Taiwan Economic Review, 45, 5-56.

Mariano, R. and D. Preve, 2012, Statistical tests for multiple forecast comparison, Journal of Econometrics, 169, 123-130.

Mehra, R., 2006. The equity premium puzzle: A review, Foundations and Trends in Finance, 2, 1-81.

Mehra, R., Prescott, E. C., 1985. The equity premium: a puzzle. Journal of Monetary Economics 15, 145-161.

Miles, W., 2008a. Boom-Bust cycles and the forecasting performance of linear and nonlinear models of house prices. Journal of Real Estate Finance and Economics, 36, 249-264.

Miles, W., 2008b. Volatility clustering in U.S. home prices. Journal of Real Estate Research, 30, 73-90.

Mollah, S., Shahiduzzaman Quoreshi, A.M.M., Zafirov, G., 2016. Equity market contagion during global financial and Eurozonecrises: Evidence from a dynamic correlation analysis, Journal of International Financial Markets, Institutions & Money, 41, 151-167.

Mondria, J. and C. Quintana□Domeque, 2013. Financial Contagion and Attention Allocation, Economic Journal, 123, 429–454.

Perron, P., 2006. Dealing with structural breaks. In Patterson, K., Mills, T. C. (Eds.), Palgrave Handbook of Econometrics, Vol. 1:Econometric Theory, Palgrave MacMillan, Basingstoke, Hampshire, UK, 278–352.

Quah, D. T., 1996. Twin Peaks: Growth and Convergence in Models of Distribution Dynamics, Economic Journal, 106, 1045-1055.

Quan, D., Titman, S., 1999. Do real estate prices and stock prices move together? An international analysis, Real Estate Economics, 27, 183-207.

Riddiough, T. 2021. Pension Funds and Private Equity Real Estate: History, Performance, Pathologies, Risks, mimeo.

Rosenblatt, M. 1952. Remarks on multivariate transformation. Annals of Mathematical Statistics, 23, 470-472.

Schulze, K., 2014. Existence and computation of the Aumann-Serrano index of riskiness and its extension. Journal of Mathematical Economics, 50, 219-224.

Seiler, M. J., Webb, J. R., Myer, N. F. C., 1999. Diversification issues in real estate investment. Journal of Real Estate Literature, 7, 163-179.

Shilling, J., 2003. Is there a risk premium puzzle in real estate? Real Estate Economics, 31, 501-525.

Silverman, B. W., 1992. Density Estimation for Statistics and Data Analysis. London: Chapman & Hall.

Steeley, J., 2015.The side effects of quantitative easing: Evidence from the UK bond market, Journal of International Money and Finance, 51, 303-336.

Taylor, J. B., 2009. Getting Off Track: How Government Actions and Interventions, Caused, Prolonged and Worsened the Financial Crisis. Stanford, CA: Hoover Institution Press.

Trichet, J. C., 2009. (Under-)Pricing of risks in the financial sector, speech, ECB.

Tsai, I. C., Chen, M. C., Ma, T., 2010. Modelling house price volatility states in the UK by switching ARCH models. Applied Economics, 42, 1145-1153.

Wall Street Journal, 2016. Global Financial Crisis spurs evolution in the asset

management industry, 4th August.
Weale, M., Wieladek, T., 2016. What are the macroeconomic effects of asset purchases? Journal of Monetary Economics, 79, 81-93.
Zhou, J., Anderson, R. I., 2012. Extreme risk measures for international REIT market. Journal of Real Estate Finance and Economics, 45, 152-170.
Zhou, J., Kang, Z., 2011. A comparison of alternative forecast models of REIT volatility. Journal of Real Estate Finance and Economics, 42, 275-294.

Table 1 RMSE for Pre-crisis period

	AR(1)	AR(1)-GRACH(1,1)	MS-AR(1)	MS-ARCH(1)
Panel A: return forecasts				
SP500	6.106	6.106	*5.788*	5.800
Dow Jones	*5.676*	5.676	5.688	5.834
Nasdaq	12.810	12.963	*12.249*	12.628
EREITs	5.894	5.906	*5.589*	5.701
OFHEO Purchase-only	0.636	0.656#	*0.564*	0.636
S&P/Case-Shiller	0.627	0.656#	*0.543*	---
Panel B:variance forecasts				
SP500	55.708	55.691	*41.982*	42.309
Dow Jones	45.763	45.783	*41.968*	47.940
Nasdaq	317.318	293.258	*241.141*	269.894
EREITs	57.348	58.593	47.868	*42.696*
OFHEO Purchase-only	2.333	2.450#	*1.934*	2.398
S&P/Case-Shiller	1.775	1.940#	*1.487*	---

Note: # means that the AR(1)-ARCH(1) model is employed. --- means that the MS-ARCH(1) model cannot be estimated. The lowest RMSE is in *Italic*.

Table 2 RMSE for Post-crisis period

	AR(1)	AR(1)-GRACH(1,1)	MS-AR(1)	MS-ARCH(1)
Panel A: return forecasts				
SP500	5.257	5.261	*5.229*	5.543
Dow Jones	4.989	5.002	*4.987*	5.004
Nasdaq	6.239	6.244	*6.227*	6.229
EREITs	6.790	6.859	*6.347*	6.803
OFHEO Purchase-only	0.749	0.773#	0.751	*0.700*
S&P/Case-Shiller	0.694	0.694#	0.670	*0.658*
Panel B:variance forecasts				
SP500	48.174	*47.713*	48.359	53.128
Dow Jones	*41.496*	41.744	41.551	42.510
Nasdaq	69.090	69.012	*68.982*	69.013
EREITs	111.770	112.047	*79.630*	109.080
OFHEO Purchase-only	3.393	3.451#	3.414	*1.648*
S&P/Case-Shiller	3.164	3.203#	*1.986*	3.067

Note: # means that the AR(1)-ARCH(1) model is employed.

Table 3 Summary of EPPC for Pre-crisis period in terms of SLC

	Class 1	Class 2
Panel A: return forecasts		
SP500	All models	\
Dow Jones	All models	\
Nasdaq	All models	\
EREITs	All models	\
OFHEO Purchase-only	All models#	\
S&P/Case-Shiller	All models#	\
Panel B: variance forecasts		
SP500	MS-AR, MS-ARCH	AR, AR-GARCH
Dow Jones	AR,AR-GARCH,MS-AR	MS-ARCH
Nasdaq	MS-AR,MS-ARCH	AR,AR-GARCH
EREITs	All models	\
OFHEO Purchase-only	All models#	\
S&P/Case-Shiller	All models#	\

Notes: SLC refers to square loss criterion. The predictive power of Class 1 is better than that of Class 2. # means that the AR(1)-ARCH(1) model is employed.

Table 4 Summary of EPPC for Post-crisis period in terms of SLC

	Class 1	Class 2
Panel A: return forecasts		
SP500	AR,AR-GARCH,MS-AR	MS-ARCH
Dow Jones	AR,AR-GARCH,MS-AR	MS-ARCH
Nasdaq	All models	\
EREITs	All models	\
OFHEO Purchase-only	AR,AR-ARCH,MS-ARCH	MS-AR
S&P/Case-Shiller	MS-AR,MS-ARCH	AR,AR-ARCH
Panel B: variance forecasts		
SP500	All models	\
Dow Jones	All models	\
Nasdaq	All models	\
EREITs	All models	\
OFHEO Purchase-only	All models#	\
S&P/Case-Shiller	AR,MS-AR,MS-ARCH	AR-ARCH

Notes: SLC refers to square loss criterion. The predictive power of Class 1 is better than that of Class 2. # means that the AR(1)-ARCH(1) model is employed.

Table 5 Summary of risk measures for the best specification

	Pre-crisis	Post-crisis
Panel A: Variance		
SP500	291.507	77.963
Dow Jones	149.744	71.226
Nasdaq	1285.749	111.895
EREITs	129.655	678.510
OFHEO Purchase-only	6.648	12.899
S&P/Case-Shiller	1.587#	8.505
Panel B: Value-at-Risk (probability=1%)		
SP500	37.729	7.953
Dow Jones	29.881	8.731
Nasdaq	101.410	10.325
EREITs	20.908	8.928
OFHEO Purchase-only	1.704	3.981
S&P/Case-Shiller	6.219#	6.751
Panel C: AS_N		
SP500	---	3.408
Dow Jones	---	3.340
Nasdaq	---	3.699
EREITs	3.920	16.183
OFHEO Purchase-only	0.791	2.150
S&P/Case-Shiller	0.083#	8.530
Panel D: AS		
SP500	---	3.029
Dow Jones	---	3.200
Nasdaq	---	3.993
EREITs	9.686	3.308
OFHEO Purchase-only	0.625	2.192
S&P/Case-Shiller	1.451#	9.603

Notes: --- means that the risk measure cannot be calculated. # means that the third best model is employed due to stationarity.

Table 6 Summary of performance measures for the best specification

	Pre-crisis	Post-crisis
Panel A: Sharpe ratio		
SP500	-0.306	1.295
Dow Jones	-0.120	1.263
Nasdaq	-0.489	1.430
EREITs	1.452	0.805
OFHEO Purchase-only	1.629	0.835
S&P/Case-Shiller	7.575#	0.171
Panel B: EPI_N		
SP500	---	3.356
Dow Jones	---	3.192
Nasdaq	---	4.090
EREITs	4.218	1.295
OFHEO Purchase-only	5.310	1.395
S&P/Case-Shiller	114.749#	0.058
Panel C: EPI		
SP500	---	3.776
Dow Jones	---	3.332
Nasdaq	---	3.788
EREITs	1.707	6.337
OFHEO Purchase-only	6.722	1.368
S&P/Case-Shiller	6.576#	0.052

Notes: --- means that the performance measure cannot be calculated. # means that the third best model is employed due to stationarity.

Figures

Figure 1. Annualized excess returns for Pre-Crisis period (2000m1-2006m6)

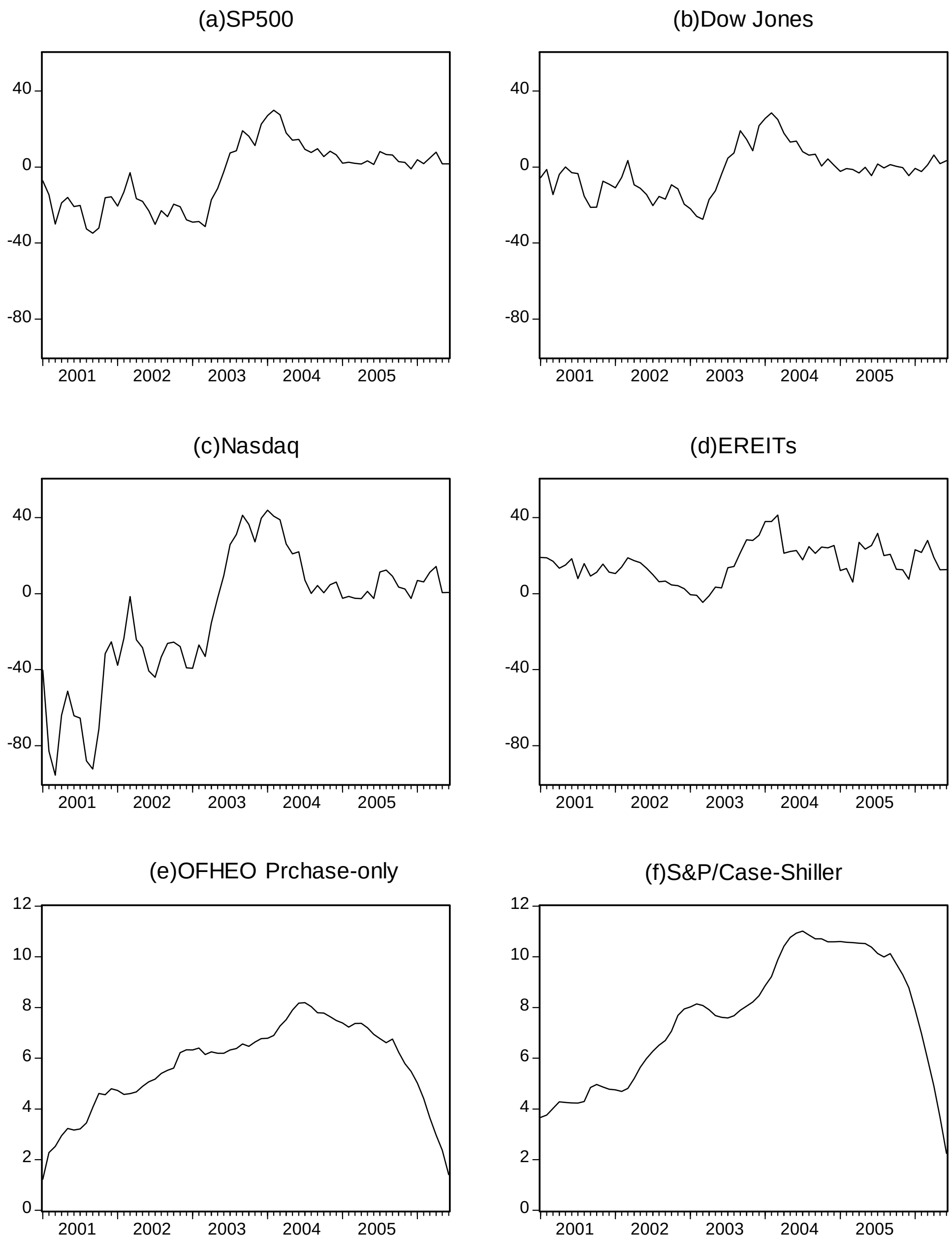


Figure 2 Annualized excess returns for Post-Crisis period
(2009m1-2019m12)

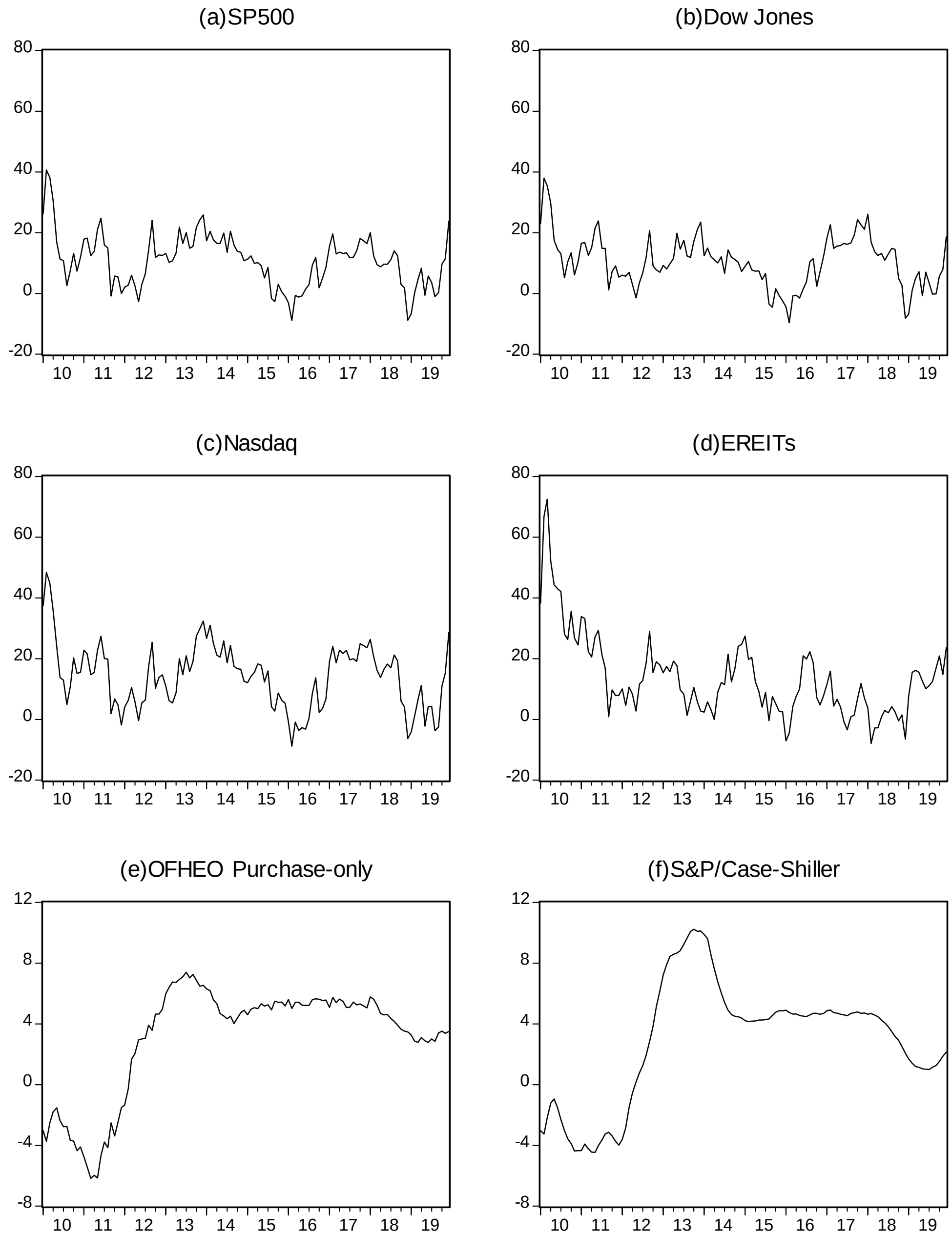
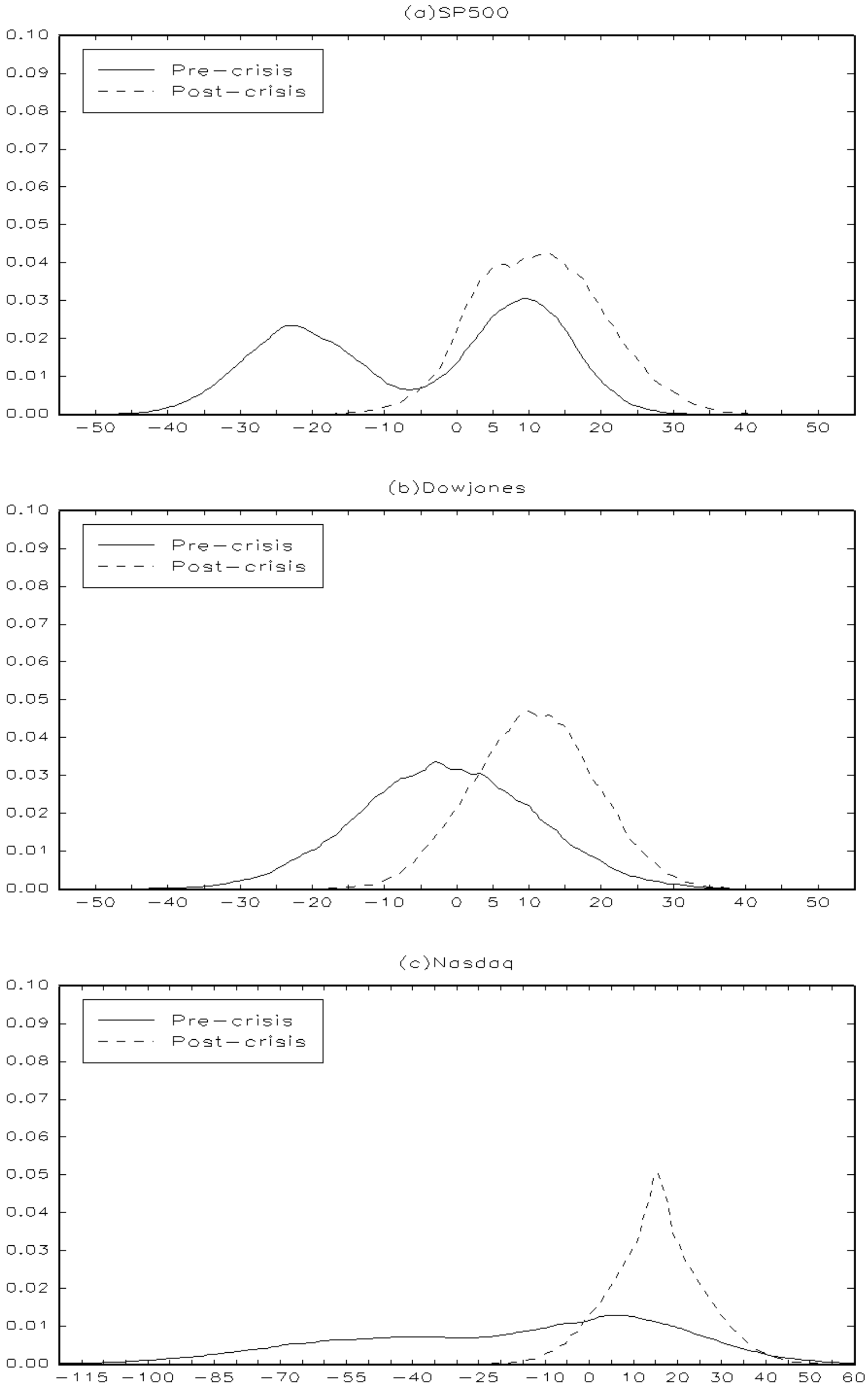
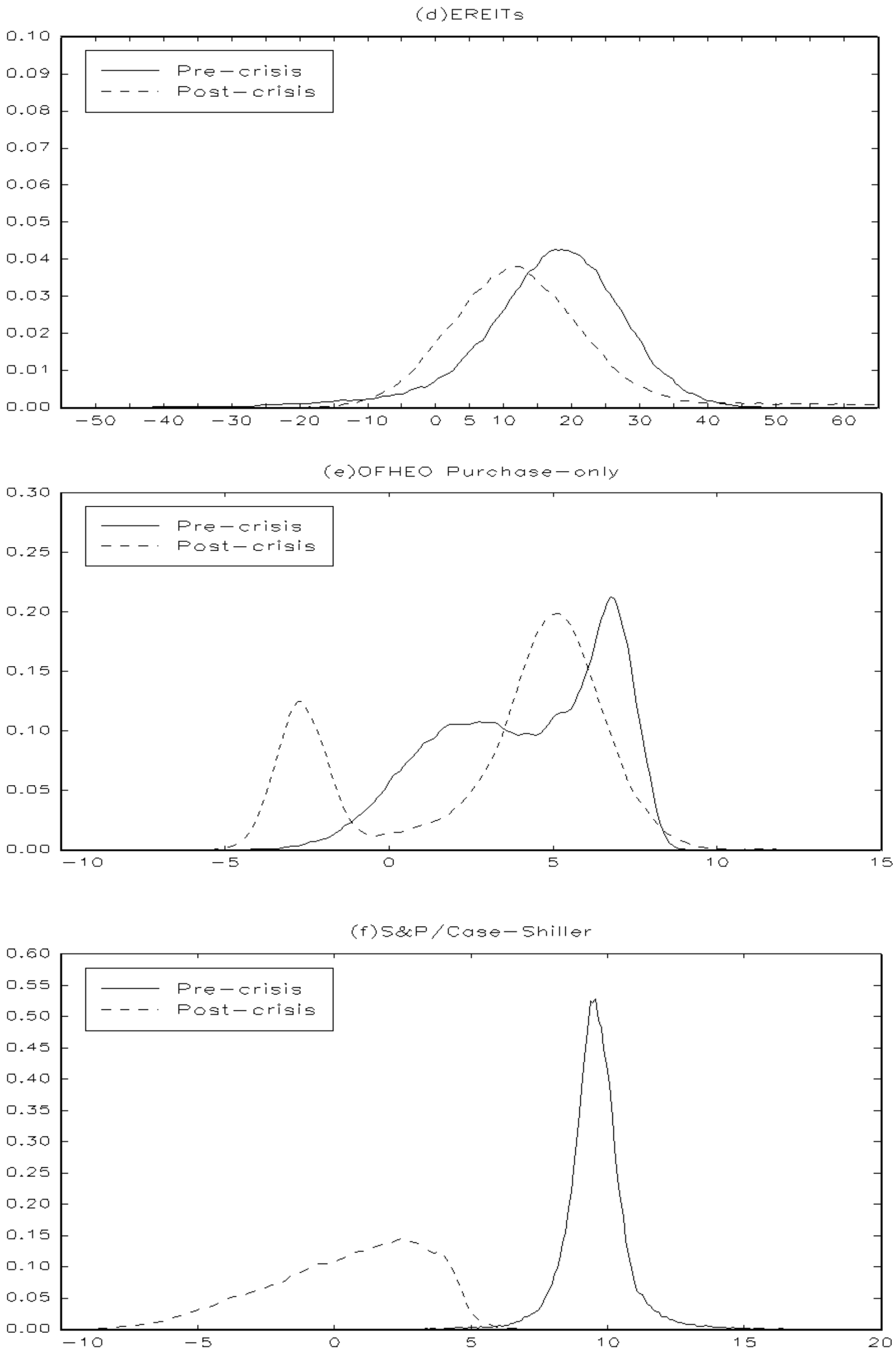


Figure 3. Empirical distributions for returns in terms of minimum RMSE

Note: The third best model is employed for the S&P/Case-Shiller return in the Pre-crisis period because the best and second best models are non-stationary.



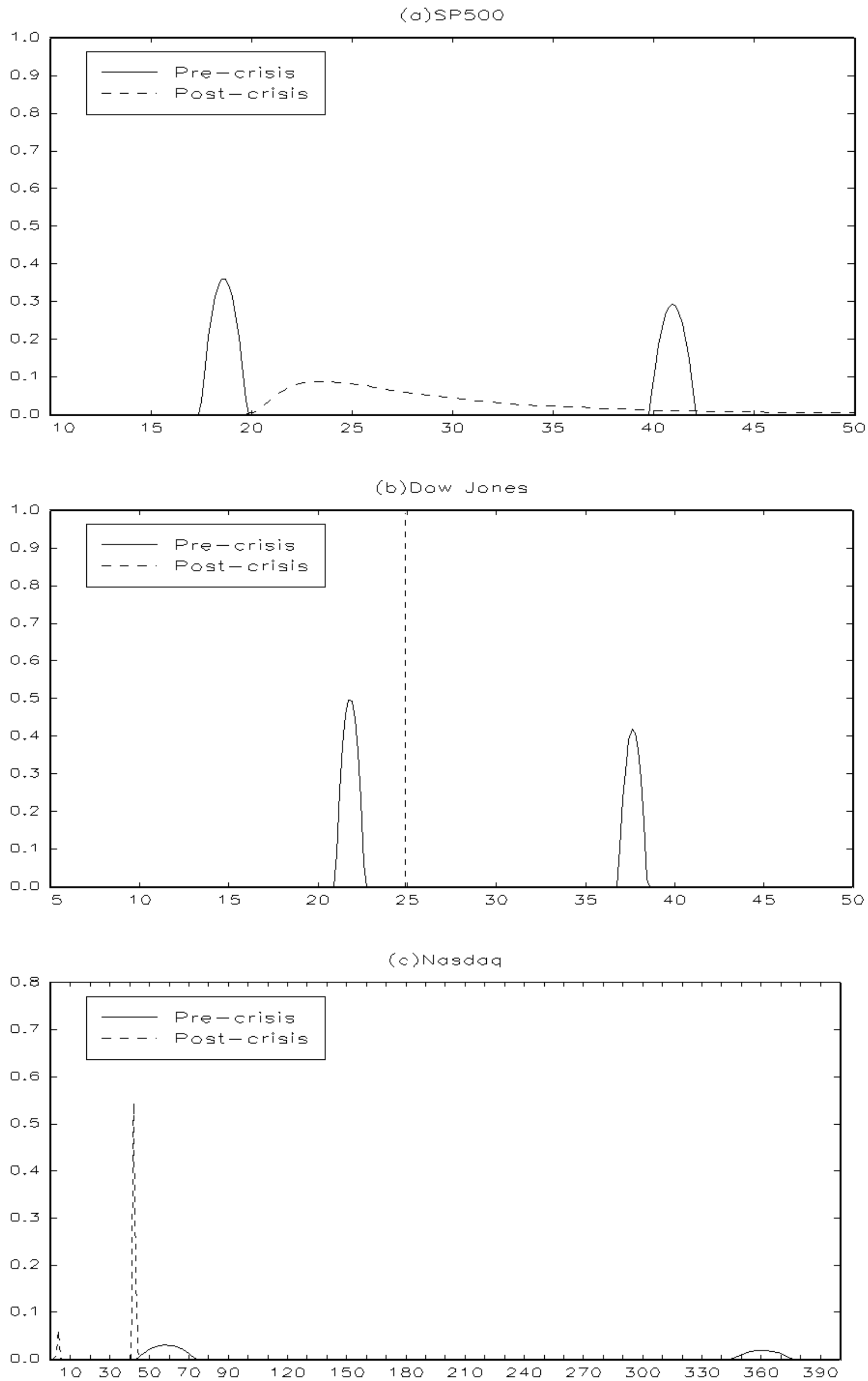
(continued)



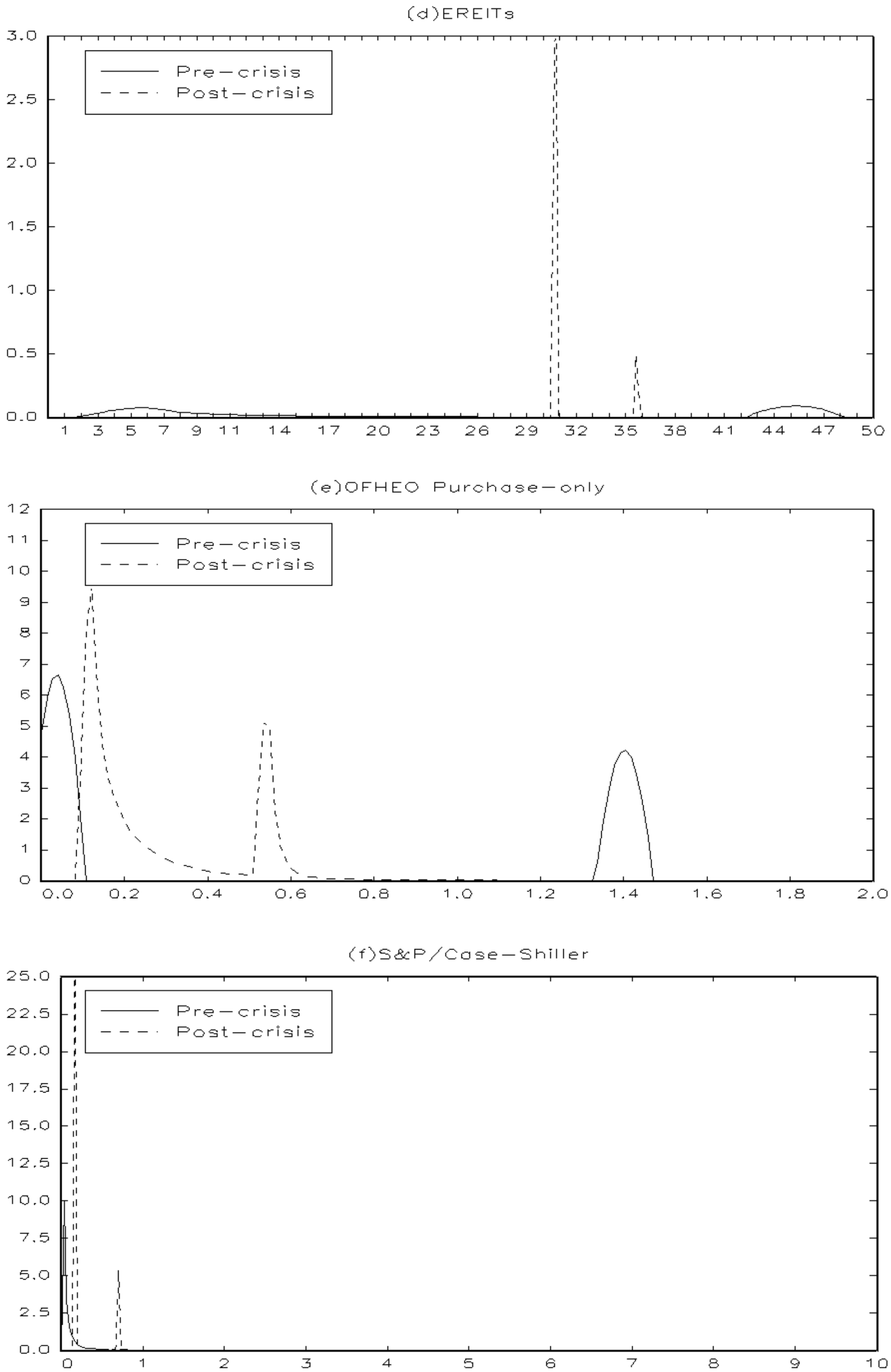
(d)EREITs
Pre−crisis
Post−crisis
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
−50 −40 −30 −20 −10 0 5 10 20 30 40 50 60
(e)OFHEO Purchase−only
Pre−crisis
Post−crisis
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10 −5 0 5 10 15
(f)S&P/Case−Shiller
Pre−crisis
Post−crisis
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10 −5 0 5 10 15 20

Figure 4. Empirical distributions for variances in terms of minimum RMSE

Note: The third best model is employed for the S&P/Case-Shiller return in the Pre-crisis period because the best and second best models are non-stationary.

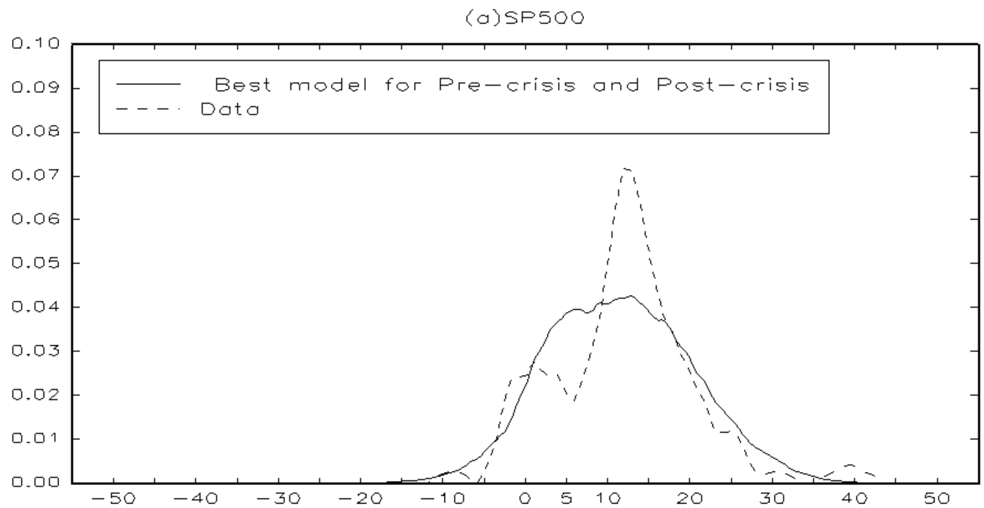


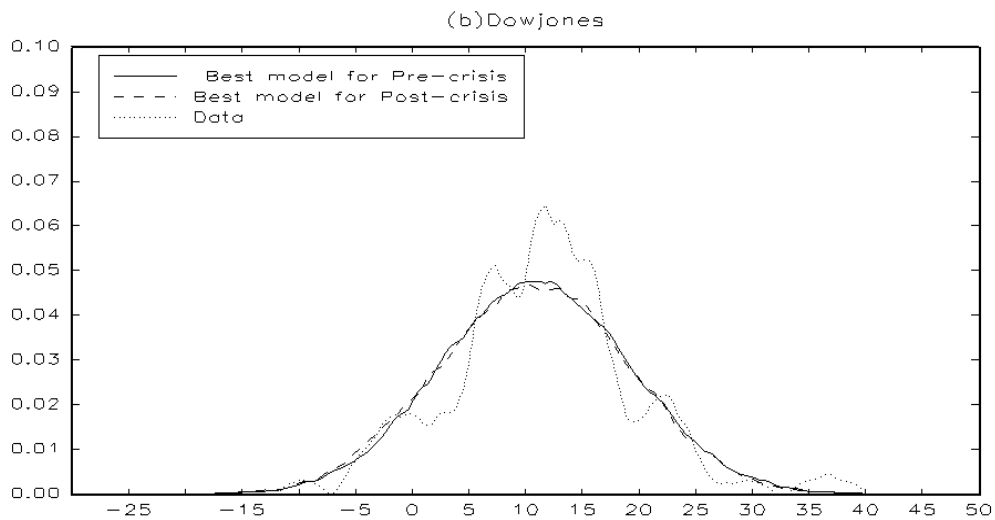
(continued)

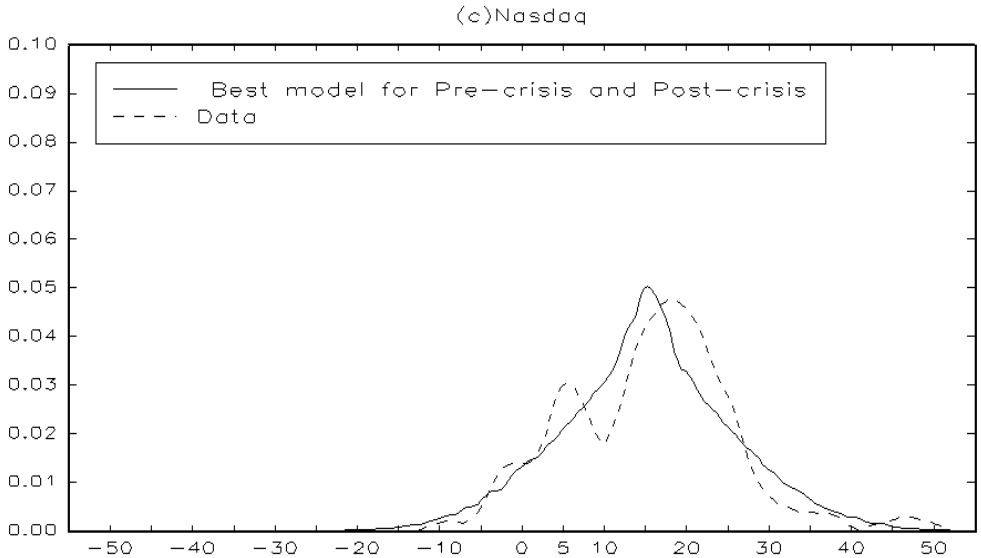


(d)EREITs
Pre−crisis
Post−crisis
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1 3 5 7 9 11 14 17 20 23 26 29 32 35 38 41 44 47 50
(e)OFHEO Purchase−only
Pre−crisis
Post−crisis
12
11
10
9
8
7
6
5
4
3
2
1
0
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0
(f)S&P/Case−Shiller
Pre−crisis
Post−crisis
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0 1 2 3 4 5 6 7 8 9 10

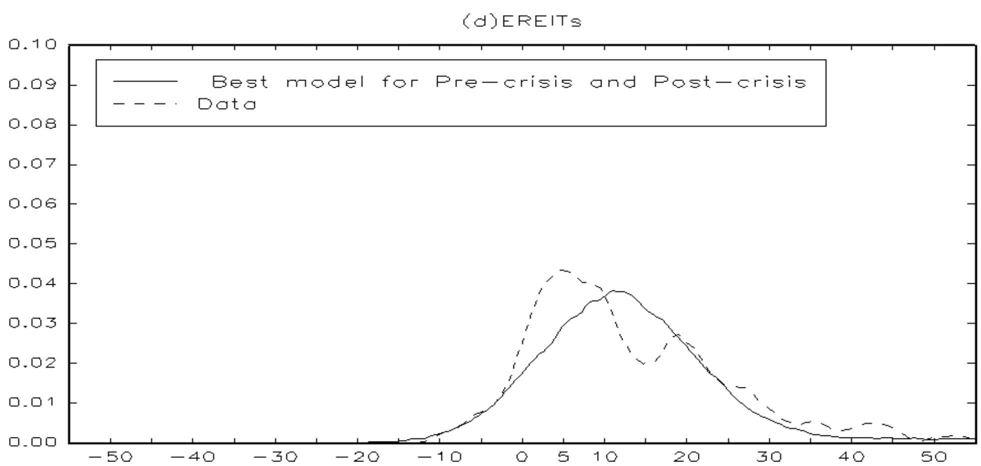
Figure 5 Empirical distributions for returns of the data after GFC, the best performing models in the Post-crisis period, and Post-crisis period.

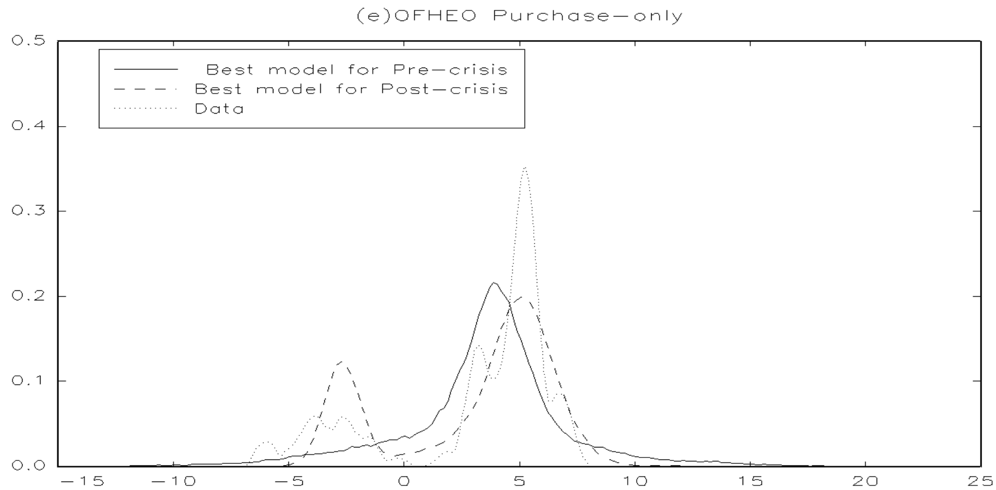





(c)Nasdaq
Best model for Pre-crisis and Post-crisis
Data
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-50 -40 -30 -20 -10 0 5 10 20 30 40 50

(continued)


(d)EREITs
Best model for Pre-crisis and Post-crisis
Data
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-50 -40 -30 -20 -10 0 5 10 20 30 40 50


(e)OFHEO Purchase−only
Best model for Pre−crisis
Best model for Post−crisis
Data
0.5
0.4
0.3
0.2
0.1
0.0
−15
−10
−5
0
5
10
15
20
25

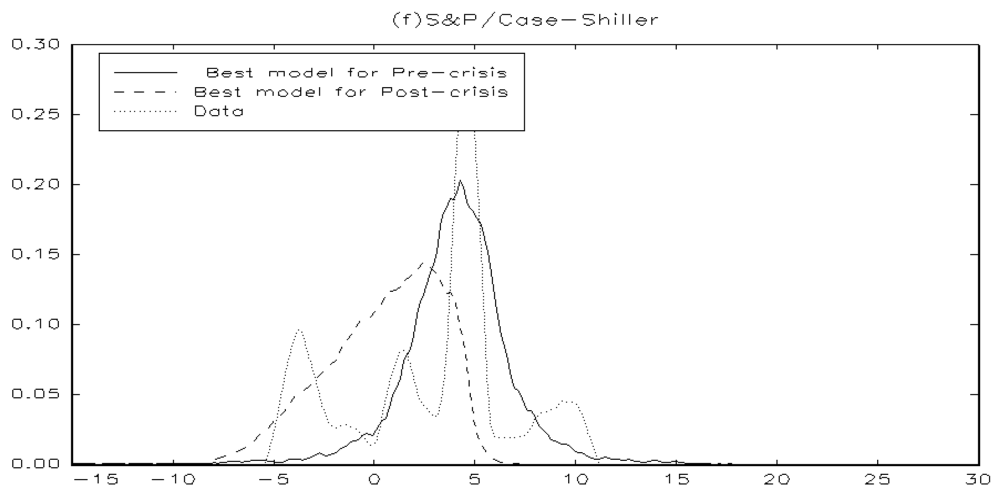
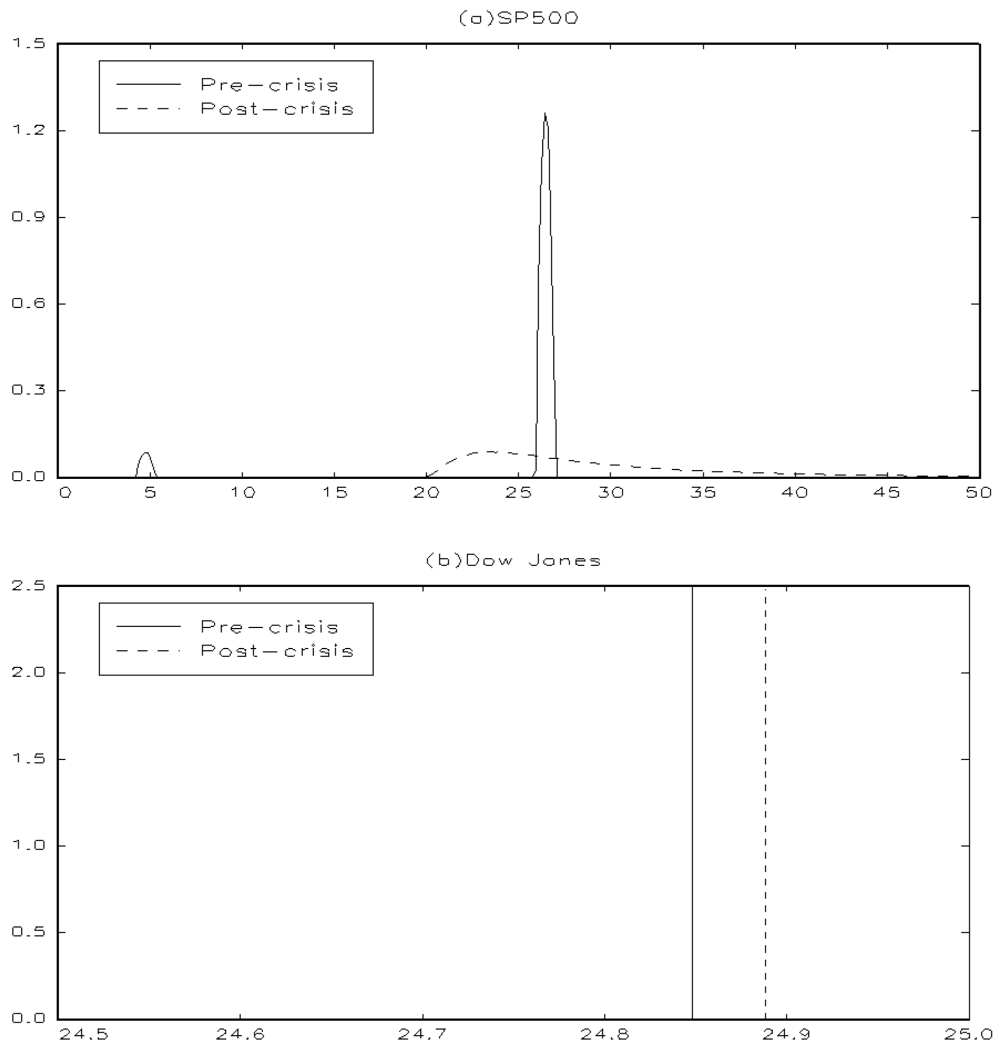
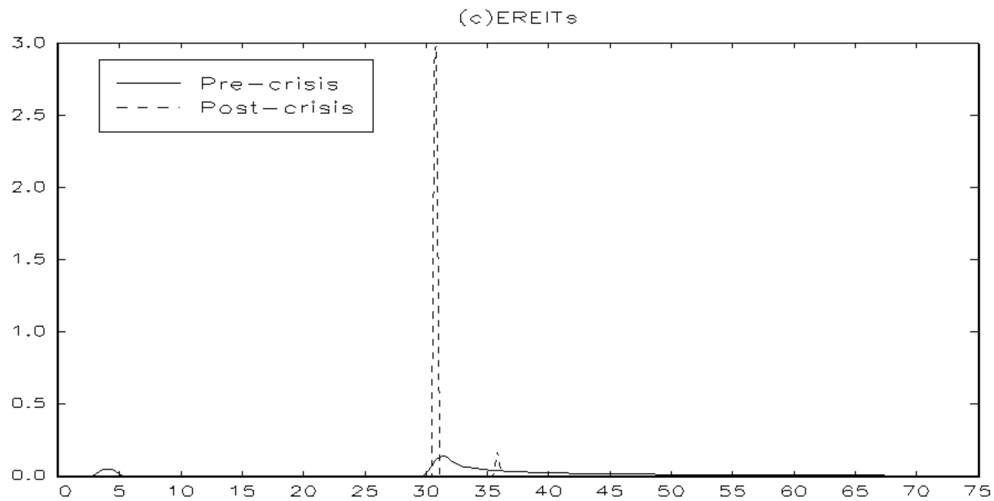
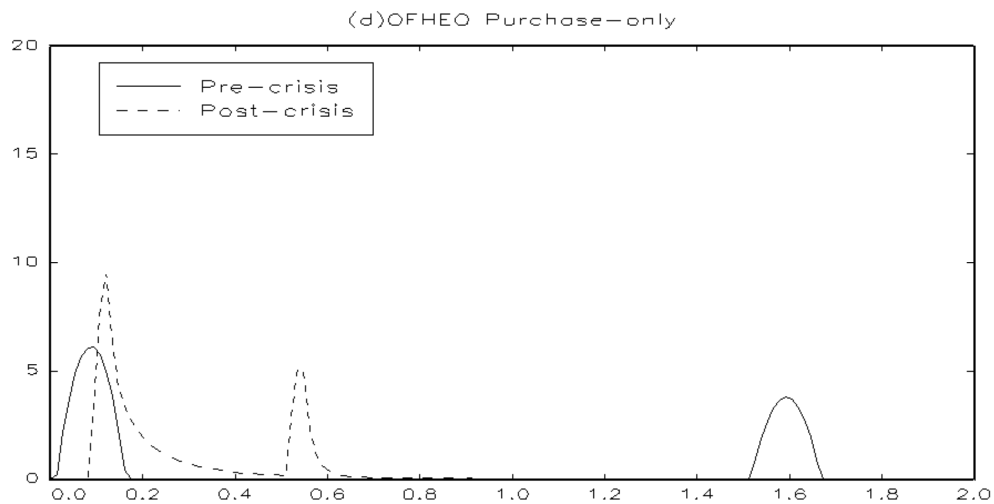
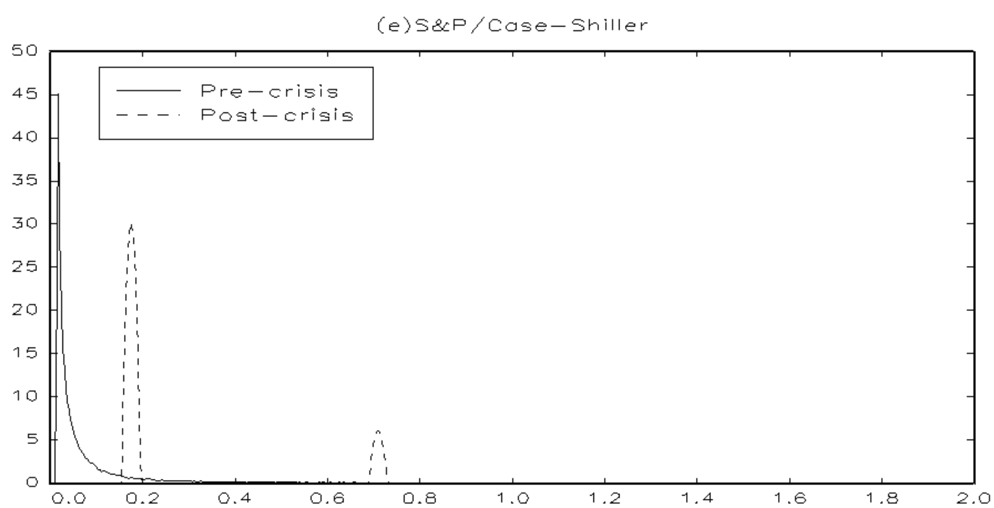

(f)S&P/Case−Shiller
Best model for Pre−crisis
Best model for Post−crisis
Data
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−15
−10
−5
0
5
10
15
20
25
30

Figure 6 Empirical distributions for (implied) variances of the best performing model in the Post-crisis period, and the best Pre-crisis model applies to Post-crisis period




(c)EREITs
Pre-crisis
Post-crisis
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75


(d)OFHEO Purchase-only
Pre-crisis
Post-crisis
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0


(e)S&P/Case-Shiller
Pre-crisis
Post-crisis
50
45
40
35
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0

Appendix

This appendix consists of several parts.

Appendix A provides additional results, including when MAE (instead of RMSE) is used, the classification of EPPC when ALC (instead of SLC) is used, return forecast, and variance forecast.

Appendix B provides more discussion on the economic performance index and the economic risk index.

Appendix C briefly explains the classifications of models into different “equivalent predictive power classes” (EPPC).

Appendix D provides more discussion on the “counter-factual” experiments presented in the text.

Table A-1 MAE for Pre-crisis period

	AR(1)	AR(1)-GRACH(1,1)	MS-AR(1)	MS-ARCH(1)
Panel A: return forecasts				
SP500	4.668	4.667	4.601	4.584
Dow Jones	4.594	4.590	4.557	4.612
Nasdaq	9.392	9.479	9.024	9.152
EREITs	4.560	4.561	4.303	4.423
OFHEO Purchase-only	0.317	0.302#	0.253	0.262
S&P/Case-Shiller	0.356	0.338#	0.270	---
Panel B:variance forecasts				
SP500	39.470	39.416	31.353	31.469
Dow Jones	32.090	31.929	28.551	29.957
Nasdaq	179.397	172.694	153.227	165.642
EREITs	35.350	34.700	31.688	29.790
OFHEO Purchase-only	0.612	0.388#	0.497	0.508
S&P/Case-Shiller	0.553	0.317#	0.388	---

Note: # means that the AR(1)-ARCH(1) model is employed. --- means that the MS-ARCH(1) model cannot be estimated.

Table A-2 MAE for Post-crisis period

	AR(1)	AR(1)-GRACH(1,1)	MS-AR(1)	MS-ARCH(1)
Panel A: return forecasts				
SP500	3.929	3.951	3.893	4.268
Dow Jones	3.840	3.859	3.840	3.848
Nasdaq	4.748	4.770	4.732	4.736
EREITs	4.964	4.968	4.777	4.862
OFHEO Purchase-only	0.416	0.439#	0.420	0.430
S&P/Case-Shiller	0.360	0.359#	0.378	0.314
Panel B:variance forecasts				
SP500	29.969	30.663	30.005	39.1935
Dow Jones	26.128	26.220	26.187	26.935
Nasdaq	42.076	45.221	42.174	42.335
EREITs	51.687	58.281	40.681	49.985
OFHEO Purchase-only	0.781	0.658#	0.735	0.575
S&P/Case-Shiller	0.687	0.416#	0.684	0.477

Note: # means that the AR(1)-ARCH(1) model is employed.

Table A-3 Smmary of EPPC for Pre-crisis period in terms of ALC

	Class 1	Class 2
Panel A: return forecasts		
SP500	All models	\
Dow Jones	All models	\
Nasdaq	All models	\
EREITs	All models	\
OFHEO Purchase-only	AR-ARCH,MS-AR,MS-ARCH	AR
S&P/Case-Shiller	MS-AR	AR, AR-ARCH
Panel B: variance forecasts		
SP500	MS-AR, MS-ARCH	AR, AR-GARCH
Dow Jones	AR-GARCH,MS-AR,MS-ARCH	AR
Nasdaq	All models	\
EREITs	All models	\
OFHEO Purchase-only	All models#	\
S&P/Case-Shiller	AR-ARCH,MS-AR	AR

Notes: ALC refers to absolute loss criterion. The predictive power of Class 1 is better than that of Class 2. # means that the AR(1)-ARCH(1) model is employed.

Table A-4 Summary of EPPC for Post-crisis period in terms of ALC

	Class 1	Class 2
Panel A: return forecasts		
SP500	AR,AR-GARCH,MS-AR	MS-ARCH
Dow Jones	All models	\
Nasdaq	All models	\
EREITs	All models	\
OFHEO Purchase-only	AR	AR-ARCH,MS-AR,MS-ARCH
S&P/Case-Shiller	MS-ARCH	AR,AR-ARCH,MS-AR
Panel B: variance forecasts		
SP500	All models	\
Dow Jones	All models	\
Nasdaq	All models	\
EREITs	AR,MS-AR,MS-ARCH	AR-GARCH
OFHEO Purchase-only	All models#	\
S&P/Case-Shiller	All models#	

Notes: ALC refers to absolute loss criterion. The predictive power of Class 1 is better than that of Class 2. # means that the AR(1)-ARCH(1) model is employed.

Table A-5 Estimation results in terms of RMSE of return forecasts

	Sp500		Dow Jones	
	Pre-crisis	Post-crisis	Pre-crisis	Post-crisis
$\theta_{0,1}$	-10.604**	4.154**	-0.154	-2.477
	(5.197)	(1.808)	(0.733)	(23.905)
$\theta_{0.2}$	2.404**	2.353***		2.141**
	(1.201)	(0.907)		(0.957)
$\theta_{1,1}$	0.531**	-0.309	0.889***	0.034
	(0.213)	(0.424)	(0.059)	(6.315)
$\theta_{1,2}$	0.758***	0.810***		0.802***
	(0.081)	(0.057)		(0.059)
ω_1	40.959***	4.746	32.214***	9.122
	(15.587)	(5.038)	(5.633)	(103.287)
ω_2	18.615***	26.467***		24.848***
	(5.063)	(3.103)		(2.972)
α_1				
α_2				
β_1				
β_2				
p	3.971***	1.413		1.643
	(1.327)	(1.378)		(23.444)
q	4.172	4.346***		7.072
	(2.823)	(1.312)		(31.457)
P_{11}	0.982	0.804		0.838
P_{22}	0.985	0.987		0.999
Log(L)	-205.813	-367.118	-208.239	-363.134

Notes: One asterisks (*), two asterisks (**), and three asterisks (***) refer to significance at 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. Log(L) refers to the logarithm of the likelihood function. Values in parentheses represent standard errors.

Table A-6 Estimation results in terms of RMSE of return forecasts

	Nasdaq		EREITs	
	Pre-crisis	Post-crisis	Pre-crisis	Post-crisis
$\theta_{0,1}$	-17.855	9.861***	-1.730***	34.150
	(11.825)	(3.747)	(0.649)	(71.940)
$\theta_{0.2}$	1.065	2.644**	6.119***	2.389***
	(1.347)	(1.188)	(2.033)	(0.849)
$\theta_{1,1}$	0.669***	0.358	0.976***	0.632
	(0.201)	(0.229)	(0.102)	(0.867)
$\theta_{1,2}$	0.909***	0.823***	0.683***	0.782***
	(0.058)	(0.067)	(0.085)	(0.050)
ω_1	360.409**	3.676	1.931	35.656
	(143.162)	(2.807)	(1.652)	(1712.747)
ω_2	58.424***	42.186***	37.184***	30.723***
	(15.900)	(5.242)	(8.462)	(4.539)
α_1				
α_2				
β_1				
β_2				
p	3.774*	1.316	2.036*	2.820**
	(2.093)	(1.106)	(1.185)	(1.399)
q	4.274	3.830**	3.851**	4.879**
	(1.565)	(1.550)	(1.518)	(2.411)
P_{11}	0.978	0.788	0.884	0.944
P_{22}	0.986	0.979	0.979	0.992
Log(L)	-248.483	-388.404	-202.689	-381.954

Notes: One asterisks (*), two asterisks (**), and three asterisks (***) refer to significance at 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. Log(L) refers to the logarithm of the likelihood function. Values in parentheses represent standard errors.

Table A-7 Estimation results in terms of RMSE of return forecasts

	OFHEO Purchase-only		S&P/Case-Shiller	
	Pre-crisis	Post-crisis	Pre-crisis	Post-crisis
$\theta_{0,1}$	0.393	-2.669***	-2.230	-0.322***
	(1.535)	(0.957)	(2.152)	(0.049)
$\theta_{0.2}$	0.432***	0.269***	0.275***	0.207***
	(0.106)	(0.050)	(0.105)	(0.013)
$\theta_{1,1}$	0.734**	0.020	1.142***	0.963***
	(0.370)	(0.330)	(0.395)	(0.012)
$\theta_{1,2}$	0.940***	0.950***	0.979***	0.955***
	(0.017)	(0.010)	(0.013)	(0.003)
ω_1	1.402	0.533	0.870	0.058**
	(1.209)	(0.478)	(0.856)	(0.027)
ω_2	0.037***	0.107***	0.046***	0.005***
	(0.006)	(0.026)	(0.009)	(0.002)
α_1		0.027		0.123
		(0.406)		(0.266)
α_2		0.487**		99.988×10^{-2}***
		(0.192)		(2.881×10^{-4})
β_1				
β_2				
p	3.054*	4.211**	2.969*	2.127***
	(1.839)	(1.649)	(1.546)	(0.396)
q	3.502**	5.390**	3.508*	2.842***
	(1.782)	(2.423)	(1.871)	(0.405)
P_{11}	0.955	0.985	0.951	0.893
P_{22}	0.971	0.995	0.971	0.945
Log(L)	-12.148	-79.239	-14.494	-8.305

Notes: One asterisks (*), two asterisks (**), and three asterisks (***) refer to significance at 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. Log(L) refers to the logarithm of the likelihood function. Values in parentheses represent standard errors.

Table A-8 Estimation results in terms of RMSE of variance forecasts

	Sp500		Dow Jones	
	Pre-crisis	Post-crisis	Pre-crisis	Post-crisis
$\theta_{0,1}$	-10.604**	1.927**	-4.775	2.098***
	(5.197)	(0.978)	(4.352)	(0.816)
$\theta_{0.2}$	2.404**		1.257	
	(1.201)		(1.402)	
$\theta_{1,1}$	0.531**	0.810***	0.625***	0.804***
	(0.213)	(0.074)	(0.235)	(0.055)
$\theta_{1,2}$	0.758***		0.832***	
	(0.081)		(0.111)	
ω_1	40.959***	9.318	37.648**	24.888***
	(15.587)	(10.250)	(18.876)	(2.803)
ω_2	18.615***		21.794***	
	(5.063)		(5.958)	
α_1		0.160		
		(0.116)		
α_2				
β_1		0.531		
		(0.401)		
β_2				
p	3.971***		3.846*	
	(1.327)		(2.298)	
q	4.172		4.094*	
	(2.823)		(2.458)	
P_{11}	0.982			
P_{22}	0.985			
Log(L)	-205.813	-367.856	-206.293	-363.136

Notes: One asterisks (*), two asterisks (**), and three asterisks (***) refer to significance at 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. Log(L) refers to the logarithm of the likelihood function. Values in parentheses represent standard errors.

Table A-9 Estimation results in terms of RMSE of variance forecasts

	Nasdaq		EREITs	
	Pre-crisis	Post-crisis	Pre-crisis	Post-crisis
$\theta_{0,1}$	-17.855	9.861***	-0.865	34.150
	(11.825)	(3.747)	(0.909)	(71.940)
$\theta_{0.2}$	1.065	2.644**	9.038**	2.389***
	(1.347)	(1.188)	(4.007)	(0.849)
$\theta_{1,1}$	0.669***	0.358	0.946***	0.632
	(0.201)	(0.229)	(0.085)	(0.867)
$\theta_{1,2}$	0.909***	0.823***	0.595***	0.782***
	(0.058)	(0.067)	(0.154)	(0.050)
ω_1	360.409**	3.676	4.371*	35.656
	(143.162)	(2.807)	(2.561)	(1712.747)
ω_2	58.424***	42.186***	45.293*	30.723***
	(15.900)	(5.242)	(26.734)	(4.539)
α_1			0.776	
			(0.618)	
α_2			0.000	
			(0.368)	
β_1				
β_2				
p	3.774*	1.316	4.688***	2.820**
	(2.093)	(1.106)	(1.190)	(1.399)
q	4.274	3.830**	4.180	4.879**
	(1.565)	(1.550)	(3.655)	(2.411)
P_{11}	0.978	0.788	0.991	0.944
P_{22}	0.986	0.979	0.985	0.992
Log(L)	-248.483	-388.404	-202.717	-381.954

Notes: One asterisks (*), two asterisks (**), and three asterisks (***) refer to significance at 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. Log(L) refers to the logarithm of the likelihood function. Values in parentheses represent standard errors.

Table A-10 Estimation results in terms of RMSE of variance forecasts

	OFHEO Purchase-only		S&P/Case-Shiller	
	Pre-crisis	Post-crisis	Pre-crisis	Post-crisis
$\theta_{0,1}$	0.393	-2.669***	-2.230	-2.339***
	(1.535)	(0.957)	(2.152)	(0.881)
$\theta_{0.2}$	0.432***	0.269***	0.275***	0.087
	(0.106)	(0.050)	(0.105)	(0.057)
$\theta_{1,1}$	0.734**	0.019	1.142***	-0.057
	(0.370)	(0.330)	(0.395)	(0.383)
$\theta_{1,2}$	0.940***	0.950***	0.979***	0.985***
	(0.017)	(0.010)	(0.013)	(0.010)
ω_1	1.402	0.533	0.870	0.709
	(1.209)	(0.478)	(0.856)	(0.771)
ω_2	0.037***	0.107***	0.046***	0.174***
	(0.006)	(0.026)	(0.009)	(0.022)
α_1		0.027		
		(0.406)		
α_2		0.487**		
		(0.192)		
β_1				
β_2				
p	3.054*	4.212**	2.969*	3.277**
	(1.839)	(1.649)	(1.546)	(1.583)
q	3.502**	5.390**	3.508*	4.877**
	(1.782)	(2.423)	(1.871)	(2.109)
P_{11}	0.955	0.986	0.951	0.964
P_{22}	0.971	0.995	0.971	0.992
Log(L)	-12.148	-79.239	-14.494	-74.225

Notes: One asterisks (*), two asterisks (**), and three asterisks (***) refer to significance at 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. Log(L) refers to the logarithm of the likelihood function. Values in parentheses represent standard errors.

Appendix B: More discussion on the economic performance index and economic risk index

Although the Sharpe ratio is a widely accepted measure for comparing relative performance among financial assets due to its easy implementation, the lack of monotonic stochastic dominance is its greatest weakness. Instead, Homm and Pigorsch (2012) propose a different economic performance index (EPI), which satisfies both duality and homogeneity axioms, and is defined as follows:

$$\mathrm{EPI} = \frac{E(r_t - r_{f,t})}{AS(r_t - r_{f,t})} \tag{B1}$$

where r_t is the nominal return of a risky asset; $r_{f,t}$ is the return of a risk-free asset; $r_t - r_{f,t}$ is the excess return; E denotes the expectation operator; and $AS(r_t - r_{f,t})$ refers to the economic risk index of Aumann and Serrano (2008). Similar to the Sharpe ratio, the EPI measure is also a risk-adjusted index. The main difference between the EPI and the Sharpe ratio is that the riskiness of a risky asset is proxied by $AS(r_t - r_{f,t})$ for the EPI and by the standard deviation of excess returns for the Sharpe ratio.

Aumann and Serrano (2008) derive the economic index of riskiness in the case of the CARA (constant absolute risk aversion) utility function. According to Theorem B of Aumann and Serrano (2008), the riskiness of an asset can be derived when the expected utility of an investor who holds a risky asset is identical to the expected utility of an investor who does not hold a risky asset. The indifference condition for an investor holding or not holding a risky asset can be expressed as follows:

$$\mathrm{E}\left[-\exp\left(-\alpha\left(W_t + r_t - r_{f,t}\right)\right)\right] = -\exp(-\alpha W_t) \tag{B2}$$

where α is the coefficient of absolute risk aversion and W_t is the initial wealth.

After defining $\alpha = \frac{1}{AS(r_t - r_{f,t})}$, Equation (E2) can be rewritten as follows:

$$\mathrm{E}\left[\exp\left(-\frac{r_t - r_{f,t}}{AS(r_t - r_{f,t})}\right)\right] = 1 \tag{B3}$$

where $AS(r_t - r_{f,t})$ is the economic index of riskiness. This index is negatively related to the coefficient of absolute risk aversion, but does not depend on the wealth level.

When the excess return is normally distributed, Homm and Pigorsch (2012) prove the nonlinear relationship between the economic performance index and the Sharpe ratio. Their relationship has the following form:

$$EPI_N = \frac{E(r_t - r_{f,t})}{AS_N(r_t - r_{f,t})} = 2 \times \left(\frac{E(r_t - r_{f,t})}{\sqrt{V(r_t - r_{f,t})}}\right)^2 = 2 \times SR^2 \tag{B4}$$

where $V(r_t - r_{f,t})$ is the variance of excess return, and SR is the Sharpe ratio.

It is clear from Equation (B4) that EPI_N and SR have a positive relationship. Even though the values for EPI_N and SR differ, the performance ranking of financial assets remains unchanged no matter which performance criterion is used. Evidently, if the assumption of normality is imposed on financial asset returns, the EPI_N can be easily calculated using only the first two moments, rather than the higher moments.

Appendix C: Classification of models into different “equivalent predictive power classes” (EPPC).

This appendix draws heavily on Kwan et al. (2015), Chang et al. (2016).

Consider the situation with $(K+1)$ models. We then define the “ j -th loss differential” $d_{j,t}$

$$d_{j,t} = L\left(e_{t+h|t}^{j}\right) - L\left(e_{t+h|t}^{j+1}\right), \quad j = 1, \ldots, K$$

where $L(.)$ continues to denote some loss function. We collect these loss differential in a vector, $d_t = \{d_{j,t}\}, \quad j = 1, \ldots, K$. We then take average of it,

$$\overline{d} \equiv \frac{1}{P}\sum_{t=1}^{P} d_t \ .$$

Mariano and Preve (2012) (MP hereafter) prove that $P\left(\overline{d}-\mu\right)'\left(\hat{\Omega}\right)^{-1}\left(\overline{d}-\mu\right) \rightarrow \chi_k^2$ (in distribution), where μ is the mean of the distribution, $\hat{\Omega}$ is a consistent estimator of the population variance-covariance matrix Ω, and k is the degree of freedom. Thus, MP test enables us to test whether all models of concern have the same predictive power.

Then we define a procedure which enable us to categorize the models into different “equivalent classes,” each of them contains model with (statistically speaking) the same predictive power. Our procedures are similar to Hansen et al. (2011) and here are the steps.

1. We consider N models that make predictions on the same economic variable. We first rank models according to a criterion, such as SLC. Without loss of generality, we assume that according to the chosen criteria, the predictive performance of Model 1 is better than that of Model 2, which in turn is better than of Model 3, and so on.
2. We conduct the MP test, in which the null hypothesis is that all models have the same predictive power. If the hypothesis is not rejected, then by definition, all N models have the same predictive power on a particular variable according to the

chosen criteria.

3. If the null hypothesis is rejected, then we eliminate the model with the least predictive power. It can be easily identified as the models have been ranked according to the predictive power in Step (1). We then repeat Step (2) until the null hypothesis of equal predictive power is accepted.

4. Assume that in Step (3), there are N_1 models which are found to possess the equal predictive power, $N_1 > 0$. For future reference, they are referred to as *Class 1* among the N models. By construction, there are $(N - N_1)$ models which do not have the same predictive power as the models in Class 1. We now repeat Step (1) on these $(N - N_1)$ models until the null of equal predictive power is not rejected. Assume that there are N_2 models, $N_2 > 0$, in the final list and they are identified as *Class 2*.

5. If $N = N_1 + N_2$, then the procedure stops. The set of models are divided into two classes, and models within each class have the same predictive power. Every model in Class 1 has higher predictive power than any model in Class 2.

6. If, instead, $N > N_1 + N_2$. Again, by construction, there are $(N - N_1 - N_2)$ models which do not have the same predictive power as the models in Class 1 or Class 2. We now repeat Step (1) on these $(N - N_1 - N_2)$ models until the null of equal predictive power is not rejected. Assume that there are N_3 models, $N_3 > 0$, in the final list, and they are identified as *Class 3*.

7. If $N = N_1 + N_2 + N_3$, the procedure stops.

8. If not, we repeat Step (6) and construct *Class 4*.

9. We repeat Step (8) until all N models are categorized into different classes. If there are totally g classes, then it must be that $N = N_1 + N_2 + ... + N_g$.

Appendix D: More discussion on the “counter-factual” experiments presented in the main text.

Recall that in the main text, we conduct a simple experiment. We consider the scenario in which the best model selected in the pre-crisis period to be the best model in the post-crisis period, with parameters re-estimated. We then compare the distribution generated by this “pseudo-best model” with the “truly best model,” which is selected by the formal model selection procedure. In the main text, we presented the graphs. Here we provide more detailed comments, which are divided into two parts. The first part is related to the return forecast and the second is related to the variance forecast. We begin with the return forecast.

(1) SP500 and EREITs have the same best model before and after Global Financial Crisis.

(2) The difference between the pre-crisis best model apply to post-crisis period and the post-crisis best model apply to post-crisis period is insignificant for Dow Jones, Nasdaq and OFHEO index.

(3) For the S&P/Case-Shiller index, the empirical distribution of true data is a Bimodal distribution. The empirical distribution evaluated from the pre-crisis best model apply to post-crisis period is a single-peak distribution. The bias is larger. The empirical distribution evaluated from the post-crisis best model apply to post-crisis period is a two-peak distribution. Its empirical distribution is similar with that of true data. Specifically, if we do not consider the possibility of changing in best model, the “risk of return” and the probability that occurring lower return will be underestimated.

We now present the remarks related to the variance forecast.

(1) Notice that the “true variance” is unknown. Only the empirical distributions for variances of the pre-crisis best model apply to post-crisis period and the post-crisis best model apply to post-crisis period are plotted.

(2) The best model for S&P/Case-Shiller index does not change before and after Global Financial Crisis.

(3) The bandwidth suggested by Silverman (1992) is used.

$$\mathrm{h} = \frac{0.9}{(N)^{1/5}} \min\left(\sqrt{V(X)}, \frac{Q_3 - Q_1}{1.349}\right)$$

where Q_3 is the third quartile and Q_1 is the first quartile.

For the MS-AR model, there are two different variances. When the $Q_3 - Q_1 = 0$, the Q_3 is replaced by the maximum of variance and Q_1 is replaced by the minimum of variance.

(4) For SP500, Dow Jones and EREITs, the pre-crisis best model apply to post-crisis period has larger variance than the post-crisis best model apply to post-crisis period. That is, if we ignore the structural change and use the same model found in the pre-crisis period, the “uncertainty about variance” will be overestimated.

(5) For OFHEO, the risk is more spread out in the post-crisis best model apply to post-crisis period than in pre-crisis best model apply to post-crisis period. The pre-crisis best model applies to post-crisis period will under-estimate the variation of risk.

(6) In Figure 6(c) (Nasdaq), the shape of empirical distribution of variance is very similar for the pre-crisis best model apply to post-crisis period and the post-crisis best model apply to post-crisis period. But, the empirical distribution of the post-crisis best model apply to post-crisis period is slightly in the left of that of pre-crisis best model apply to post-crisis period. That is, if we ignore the structural change and use the same model found in the pre-crisis period, the “occurrence of lower variance” will be underestimated.