Linear and Nonlinear Predictability of International Securitized Real Estate Returns: A Reality Check

Authors Juan Cabrera, Tao Wang, and Jian Yang

Abstract This paper examines the short-horizon return predictability of the ten largest international securitized real estate markets, paying special attention to possible nonlinearity-in-mean as well as nonlinearity-in-variance predictability. Although international securitized real estate returns are generally not predictable based on commonly-used statistical criteria, there is much evidence for the predictability based on economic criteria (i.e., direction of price changes and trading rule profitability), which is more often due to nonlinearity-in-mean. The forecast combinations for various models appear to improve the forecasting performance, while the allowance of data-snooping bias using White's reality check substantially mitigates spurious out-of-sample forecasting performance and weakens otherwise overwhelmingly strong predictability. Overall, there is robust evidence for the predictability in many international securitized real estate markets.

The global real estate securities market has grown tremendously over the last decade. Even after the dramatic fall in value during the global financial crises in 2007–2009, the market capitalization of the FTSE EPRA/NAREIT Global Index still amounted to $690 billion by June 2010. The securitized real estate return predictability is not only closely related to market efficiency (Chordia, Roll. and Subrahmanyam, 2005), but also has important implications for practitioners on their optimal portfolio allocations to real estate as an independent asset class in their portfolio management (MacKinnon and Zaman, 2009). However, the return predictability of real estate securities is perhaps one of the most important and controversial topics in the real estate literature (Li and Wang, 1995).

There are different approaches to examining securitized real estate return predictability. One popular approach (e.g., Liu and Mei, 1992, 1994; Li and Wang, 1995; Ling, Naranjo, and Ryngaert, 2000) examines the return predictability of real estate securities using multifactor asset pricing models. Those studies typically find that excess equity REIT returns are far less predictable out-of-sample than in-sample. Another strand focuses on the notion of weak-form market efficiency and explores the securitized real estate return predictability based on past returns. Such a time series approach, as emphasized by Serrano and Hoesli (2010), may create a level playing field to evaluate the predictability of securitized real estate returns in comparison with that of stock returns. More generally, Grinblatt and Moskowitz (2004) point out that past returns contain information about expected returns. Therefore, the use of past returns to predict future performance has attracted many studies.

Many earlier studies along this line often use the autocorrelation test, the unit root test, or the variance ratio test (e.g., Ma, 1990; Nelling and Gyourko, 1998; Kleiman, Payne, and Sahu, 2002). These tests, however, assume linearity and only investigate serial uncorrelatedness rather than a martingale difference.[1] Theoretically, the existence of fads, rational speculative bubbles, or a not-too-complex chaotic process would suggest the possibility of nonlinear predictability but not linear predictability in asset returns (Hsieh, 1991; McQueen and Thorley, 1991). Hence, the tests based on the assumption of linearity may fail to capture predicable nonlinearities in mean and yield misleading conclusions in favor of the martingale hypothesis (Hong and Lee, 2003).

Nevertheless, one exception using nonlinear-in-mean models is Brooks and Tsolacos (2003), who compare the predictability of linear ARMA, linear VAR, and nonlinear neural networks models in five European countries. They conclude that the neural networks model generally makes the most accurate predictions over a one-month horizon. On the other hand, while many earlier studies use nonlinear-in-variance GARCH models to model securitized return volatility (e.g., Jirasakuldech, Campbell, and Emekter, 2009), few explore the usefulness of GARCH models in out-of-sample securitized real estate return predictability, with the notable exception of Serrano and Hoesli (2010). In general, the research on securitized real estate return predictability has not been as thorough as on direct real estate return predictability in terms of using and comparing different forecasting techniques, particularly nonlinear models (Serrano and Hoesli, 2010, 173).[2]

Filling the gap, this study reexamines the return predictability of the ten largest international securitized real estate markets using linear and non-linear models. The study contributes to the literature in the following important aspects. First, this study employs a number of nonlinear models that allow for both potential nonlinearity-in-mean and nonlinearity-in-variance.[3] In particular, some variants of popular nonlinear models on direct real estate market (e.g., Nguyen and Cripps, 2001; Crawford and Fratantoni, 2003; Guirguis, Giannikos, and Anderson, 2005; Miles, 2008a, 2008b; Peterson and Flanagan, 2009; Osland, 2010) and securitized real estate market (e.g., Brooks and Tsolacos, 2003; Serrano and Hoesli, 2010) are all used in this study. While recent studies have used nonlinear models to forecast securitized real estate returns, they either primarily focus on nonlinear-in-variance GARCH models (e.g., Serrano and Hoesli, 2010) or just one particular nonlinear-in-mean model (e.g., Brooks and Tsolacos, 2003). By contrast, the current study employs multiple nonlinear models and essentially uses the model selection approach of Swanson and White (1997), who do not presume any model to be the ''true'' model and thus do not require the specification of a correct model for its valid application. Further addressing the issue that no single technique has been found universally superior in forecasting either securitized or direct real estate returns (Serrano and Hoesli, 2010), the current study uses the forecast combination approach to pool forecasts from various models to improve predictability (e.g., Hong and Lee, 2003; Yang, Su, and Kolari, 2008; Rapach, Strauss, and Zhou, 2010).

Second, the study differs from most previous studies by examining out-of-sample forecasting performance more thoroughly. The out-of-sample forecasting evidence arguably bears directly on predictability and is important to mitigate the concern of in-sample model overfitting (particularly for nonlinear models).[4] This study presents out-of-sample evidence based on economic criteria, in addition to statistical criteria widely used in the literature (e.g., Nguyen and Cripps, 2001). In particular, to the best of our knowledge, no earlier studies have considered the economic criterion as measured by the direction of forecasted price changes in the real estate literature, which have practical value to investors and other decision-makers. From a perspective of decision-making under uncertainty, there exist important circumstances under which this criterion is exactly the right one for maximizing the economic welfare of the forecaster (Leitch and Tanner, 1991; Hong and Lee, 2003). Directional predictability in asset returns also has important implications for market timing and the resulting active asset allocation management. Hence, this study is perhaps the first to comprehensively report evidence on the predictability of the direction of changes for a number of international securitized real estate markets. Also, although trading rule profitability as an economic criterion has been explored in several previous studies (e.g., Ling, Naranjo, and Ryngaert, 2000; Serrano and Hoesli, 2010), it is extended here by exploring trading rule profitability based on multiple nonlinear-in-mean models and their combinations.[5]

Finally, this study further extends previous studies by using White's reality check test (White, 2000) to address the concern of data-snooping bias (i.e., spuriously superior predicative ability of some complex models due to chance). When different forecast models using the same data are compared, it is crucial to take into account the complexity of such models individually, which otherwise may result in misleading inference in favor of more complex models.[6] However, no earlier studies in the real estate literature have addressed the issue, which is shown to be nontrivial in this study.

This study also for the first time explores the return predictability of the largest real estate investment trusts (REITs) in the United States compared to the large cap stock market index, and daily return predictability of U.S. mortgage REITs. The rest of this paper is organized as follows: Section 2 presents econometric methodology; Section 3 describes the data and discusses the empirical results; and finally, Section 4 concludes the paper.

Exhibit 1 | Summary of Models

Name	Models for $E(Y_t \| I_{t-1})$ and $sign[E(Y_t \| I_{t-1})]$
Benchmark	$E(Y_t \| I_{t-1}) = \mu$
1. AR(d)	$E(Y_t \| I_{t-1}) = \beta_0 + \Sigma_{j=1}^{d} \beta_j Y_{t-j}$
2. EGARCH(p,q)	$E(Y_t \| I_{t-1}) = \mu + \Sigma_{s=1}^{s=k} a_s y_{t-s} + \varepsilon_t$ where $\varepsilon_t \| \Omega_t \sim N(0, h_t^2)$, and $\log(h_t^2) = \omega + \Sigma_i^p \alpha_i \left(\frac{\varepsilon_{t-i}}{h_{t-i}}\right) + \Sigma_j^q \beta_j \log(h_{t-j}^2) + \lambda \left\|\frac{\varepsilon_{t-1}}{h_{t-1}}\right\|$
3. NN(d,q)	$E(Y_t \| I_{t-1}) = \beta_0 + \Sigma_{j=1}^{d} \beta_j Y_{t-j} + \Sigma_{i=1}^{q} \delta_i G(\gamma_{0i} + \Sigma_{j=1}^{d} \gamma_{ji} Y_{t-j})$, $G(z) = (1 + e^{-z})^{-1}$
4. FC(d,L)	$E(Y_t \| I_{t-1}) = \alpha_0(U_t) + \Sigma_{j=1}^{d} \alpha_j(U_t) Y_{t-j}$ where $U_t = Y_{t-1} - L^{-1} \Sigma_{j=1}^{L} Y_{t-j}$
5. NP(k,m)	$E(Y_t \| I_{t-1}) = g(Y_{t-1}, Y_{t-2})$
6. Combined I (1, 3, 4, 5)	AR(d), NN(d,q), FC(d,L) and NP(k,m)
7. Combined II (1–5)	AR(d), GARCH(p,q), NN(d,q), FC(d,L) and NP(k,m)

Notes: The benchmark model is the martingale model. AR(d) is the autoregression model. EGARCH(p,q) is the generalized autoregressive conditional heteroscedasticity model. NN (d,q) is the neural network model. FC is the functional coefficient model of Cai, Fan, and Yao (2000). NP is the nonparametric model estimated by the kernel regression approach. For the NP(k,m) models, the smoothing parameter h is used in nonparametric estimation for minimizing k period out-of-sample.

Econometric Methodology

This study uses the model selection approach with various nonlinear models to explore the possibility that securitized real estate returns are not a martingale, and have the conditional mean dependence in a nonlinear fashion (i.e., nonlinearity-in-mean), and the dependence in second or higher moments (i.e., nonlinearity-in-variance). Although they do represent many of the most popular nonlinear models widely used in the literature thus far, the limited number of the nonlinear models cannot capture all the nonlinearities.

Various models are employed for $E(Y_t|I_{t-1})$, where Y_t is the first difference of securitized real estate market daily closing prices in logarithm and I_{t-1} is the information set available at time $t - 1$. For the benchmark model, the martingale model $Y_t = \mu + \varepsilon_t$ is used. Exhibit 1 lists the various models under study, including the autoregressive model (AR(d)), exponential generalized autoregressive conditional heteroscedasticity model (EGARCH(p,q)), functional coefficient model (FC(d,L)), feedforward artificial neural network (NN(d,q)), nonparametric regression model (NP(k,m)), and some combinations of these models.

The EGARCH Model

The time-varying pattern of asset price volatility is well documented in the literature. This study uses the univariate AR (k)-EGARCH (p,q) model of Nelson (1991) as follows:

$$y_t = \mu + \sum_{s=1}^{s=k} a_s y_{t-s} + \varepsilon_t,$$
$$\varepsilon_t | \Omega_t \sim N(0, h_t^2),$$

and

$$\log(h_t^2) = \omega + \sum_{i}^{p} \alpha_i \left(\frac{\varepsilon_{t-i}}{h_{t-i}}\right) + \sum_{j}^{q} \beta_j \log(h_{t-j}^2) + \lambda \left|\frac{\varepsilon_{t-1}}{h_{t-1}}\right|,$$

where y_t is the return, k is the lag length, $N(0,h_t^2)$ represents the normal density function with mean zero and time varying variance h_t^2, and p and q are lag lengths for the squared residuals and the conditional variance, respectively. The conditional variance of securitized real estate returns, h_t^2, is specified as a linear function of past squared errors and past values of the conditional variance.

The model is estimated using the quasi-maximum likelihood (QML) method. QML parameter estimates can be consistent, even though the conditional log-likelihood function assumes normality while securitized real estate returns may be skewed and leptokurtic. Given a sample of T observations of the return vector, the parameters of the EGARCH model are estimated by maximizing the conditional log-likelihood function:

$$L = \sum_{t=1}^{T} l_t(P) = \sum_{t=1}^{T} (-\log(2\pi) - 0.5\log|H_t| - 0.5\varepsilon_t' H_t^{-1} \varepsilon_t),$$

where P denotes the vector of all the parameters to be estimated. Nonlinear optimization techniques are used to calculate the maximum likelihood estimates based on the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm.

The Functional Coefficient Model

The functional coefficient model includes threshold autoregression models, smooth transition regression, and many other regime switching models as special cases.

It is introduced by Cai, Fan, and Yao (2000) as a new semiparametric nonlinear time series model with time-varying and state-dependent coefficients. The basic model can be specified as:

$$E(Y_t|I_{t-1}) = \alpha_0(U_t) + \sum_{j=1}^{d} \alpha_j(U_t)Y_{t-j},$$

where $\{(Y_t, U_t)'\}$ is a bivariate stationary process. The smoothing variable U_t could be chosen as a function of explanatory variable vector Y_{t-j} or as a function of other variables. In this study, U_t is chosen as the difference between the log securitized real estate price at time $t - 1(p_{t-1})$, and the moving average of the most recent periods L of the log prices at time $t - 1$, $U_t = p_{t-1} - L^{-1} \sum_{j=1}^{L} p_{t-j}$. Traders often use U_t as a buy or sell signal based on its sign, which possibly reveals information on direction of price movements, i.e., the moving average rule. In this study, $L = 200$ is used, which is consistent with a commonly-used moving average rule.

The term $a_j(U_t)$ is estimated using a local linear estimator (when U_t is close to u) by $a_j(U_t) = a_j + b_j(U_t - u)$ (Cai, Fan, and Yao, 2000). The local linear estimator at point u is $\hat{a}_j(u) = \hat{a}_j$, and $\{(\hat{a}_j, \hat{b}_j)\}$ is chosen by minimizing the sum of locally weighted squares defined as follows:

$$\sum_{t=1}^{N} [Y_t - a_j - b_j(U_t - u)]^2 K_h(U_t - u),$$

where h is the smoothing parameter or the bandwidth of the window of the kernel function. h is determined by the modified leave-one-out least square cross-validation method proposed in Cai, Fan, and Yao (2000). $K_h(.)$ is the kernel function. The normal distribution is chosen for the kernel function.

The Artificial Neural Network Model

In forecasting financial time series, artificial neural networks have proven to be useful in capturing nonlinearity-in-mean. Neural networks can well approximate a large class of functions, which is one of its greatest advantages over other commonly-used nonlinear econometric models. In neural network models, many ‘basic’ nonlinear functions can be combined via a multilayer structure. Typically there is one intermediate, or hidden, layer between the inputs and the output. In the model, the explanatory variables simultaneously activate the units in the intermediate layer through some function Ψ and, subsequently, output is produced through some function Φ from the units in the intermediate layer. The basic methodology can be summarized as follows:

$$h_{i,t} = \Psi\left(\gamma_{i0} + \sum_{j=1}^{m} \gamma_{ij} X_{j,t}\right) \quad i = 1,\ldots, q$$

$$Y_t = \Phi\left(\beta_0 + \sum_{i=1}^{q} \beta_i h_{i,t}\right)$$

or

$$Y_t = \Phi\left(\beta_0 + \sum_{i=1}^{q} \beta_i \Psi\left(\gamma_{i0} + \sum_{j=1}^{m} \gamma_{ij} X_{j,t}\right)\right),$$

where $X_{j,t}$ is the input or an independent variable, Y_t is the output or dependent variable, and $h_{i,t}$ is the node or hidden unit in the intermediate or hidden layer. In this study, the lagged dependent variable Y_{t-j} is used as the independent variable $X_{j,t}$. The functions Ψ and Φ can be arbitrarily chosen and still approximate a large class of functions, as long as there are sufficiently large numbers of units in the intermediate layer.

Following the literature (e.g., Gencay, 1998, 1999; Hong and Lee, 2003; Yang, Su, and Kolari, 2008), the single layer feed forward neural networks model is used in this study. This type of model is the most basic but perhaps the most commonly-used neural network model in economic and financial applications. In this model, input variables are connected to multiple nodes (or hidden units), while at each node they are weighted (differently) and transformed by the same activation function Ψ. Furthermore, the output of each node is weighted by β_i, summed and transformed by a second activation function Φ.

Coefficients for the neural network model, NN(d,q) model, are estimated using nonlinear least squares via the Newton-Raphson algorithm. The logistic function for Ψ and the identity function for Φ are employed, which is the common practice in the literature (e.g., Gencay, 1998, 1999; Hong and Lee, 2003; Yang, Su, and Kolari, 2008). The specification for the model is:

$$E(Y_t|I_{t-1}) = \beta_0 + \sum_{j=1}^{d} \beta_j Y_{t-j} + \sum_{i=1}^{q} \delta_i G\left(\gamma_{0i} + \sum_{j=1}^{d} \gamma_{ji} Y_{t-j}\right),$$

where $G(z) = (1 + e^{-z})^{-1}$, I_{t-1} is the information set available at $t - 1$, and Y_t is the dependent variable (i.e., securitized real estate returns).

The Nonparametric Kernel Regression Model

In general, nonlinearities in the conditional means may be complicated and cannot be expressed explicitly. Nonparametric regression provides a way to estimate the model without specifying functional forms. Similar to Harvey (2001), the well-known kernel regression (with some improvements on bandwidth selection to maximize the forecasting power) is used for estimation and forecasting.

A nonparametric regression model can be generally expressed as:

$$E(Y_t|I_{t-1}) = g(Y_{t-1}, Y_{t-2},..., Y_{t-j}).$$

$g(.)$ can be estimated by local linear regression, as mentioned above for the functional coefficient model. In fact, for $y_t = \{y_{t-1}, y_{t-2},..., y_{t-j}\}$, $g(.)$ can be approximated locally by a linear function $g(Y) = a + (Y - y)'b$ or can be approximated $g(y)$ locally by a constant function $g(y) = a$ (i.e., the local constant estimator). This study uses the local constant estimator for the nonparametric kernel regression model, which is relatively simple to implement and has been widely used in applied research. The local constant estimator has also drawn most theoretical attention and thus has clear theoretical properties for estimation and inference of nonparametric models. The local constant estimator at point y is given by $g(y) = \hat{a}$, and $\hat{a}$ minimizes the sum of local weighted squares:

$$\sum_{t=1}^{N} [Y_t - a]^2 \prod_{s=1}^{j} K_{hs}(Y_{t-s} - y_{t-s}),$$

where $K_{h,s}$ is the univariate kernel function, $\Pi_{s=1}^{j} K_{hs}(Y_{t-s} - y_{t-s})$ is the product kernel, and the smoothing parameter $h = (h_1,..., h_j)$ is chosen by the leave-one-out cross-validation procedure. It is well-known that h is the most important parameter in nonparametric estimation, as an inappropriately chosen h will result in poor in-sample and out-of-sample predictions. Existing nonparametric regression models tend to use h, which minimizes the in-sample sum of squared errors to forecast the next-period value based on previous in-sample data. While this h is optimal for in-sample data, it may not be the best h for out-of-sample forecasting.

Following Yang, Su, and Kolari (2008), to find the best h for out-of-sample forecasting, a modified approach is used for the choice of h. For example, suppose that there are data points of x_1 to x_{100} and want to forecast x_{101}. The traditional approach is to find the best h^* to minimize the in-sample sum of squared errors of these 100 data points (x_1 to x_{100}), and then use h^* and these data points (i.e., x_1 to x_{100}) to forecast x_{101}. Yang, Su, and Kolari (2008) propose the following modified nonparametric forecasting method. Here h and the data points of x_1 to x_{80} are used to forecast x_{81}, data points of x_2 to x_{81} to forecast x_{82},... , data points of x_{20} to x_{99} to forecast x_{100}. The h^* is identified that minimizes the sum of squared errors of out-of-sample forecast of points x_{81} to x_{100} and this h^* and data points x_{21} to x_{100} are used to make the final forecast of x_{101}. There are two parameters to establish in this procedure: (1) the out-of-sample evaluation length k is set to be equal to 20 ($\hat{x}_{81}$ to $\hat{x}_{100}$) in the example, and (2) the regression length m is set to be equal to 80 in the example. Therefore, the model is denoted as NP(k,m), where the parameters (k,m) are key to the forecasting performance of the model. To check the robustness of the choice of evaluation length, different evaluation lengths are explored, and it appears that its impact is not substantial. Therefore, the results are based on one particular combination in the exhibits presented below.

Combined Models

It has been argued that, as the pattern of time series changes can vary over time and may not follow a simple data generating process, no single forecasting model performs well for all time periods and under all different criteria. To improve forecasts over individual models, forecast combination has been applied in previous studies. For example, Rapach, Strauss, and Zhou (2010) on the U.S. stock market, Yang, Cabrera, and Wang (2010) on international exchange-traded fund (ETF) markets, and Hong and Lee (2003) and Yang, Su, and Kolari (2008) on foreign currency markets consistently show that forecast combinations can improve return forecast accuracy over a single model. In order to improve the predictability result, this study follows Hong and Lee (2003) and Yang, Su, and Kolari (2008) to combine several forecasting models. More specifically, for the Combined I model, forecasts from the AR(1), NN(1,5), FC(1,200), and NP(200,400) models are pooled and the average of those forecasts is used to predict the conditional mean of price changes. For the Combined II model, forecasts from the AR(1), EGARCH(1,1), NN(1,5), FC(1,200), and NP(200,400) models are pooled and the average used to forecast the conditional mean of price changes. Based on these predictions, the evaluation criteria can be applied.

Data

The data consist of daily returns for the U.S. and the other nine largest international securitized real estate markets and are obtained from Datastream. For the U.S. securitized real estate market, the Equity REIT, Mortgage REIT, and the top 50 REIT indexes are included; the international securitized real estate markets include Australia, France, Germany, Japan, the Netherlands, Sweden, United Kingdom, Hong Kong, and Singapore. The time period covered for the U.S. market is from December 31, 1999 to December 31, 2008. The international market indices span from January 1, 1994 to December 31, 2008. Similar to many earlier studies (e.g., Bond, Karolyi, and Sanders, 2003; Yang, Kolari, and Zhu, 2005; Serrano and Hoesli, 2010), the daily data employed in this study are taken from the EPRA/NAREIT Global Real Estate indices jointly developed and published by the European Real Estate Association (EPRA) and National Association of Real Estate Investment Trusts (NAREIT). These indices are constructed on a consistent basis across countries from the share prices of companies with greater than $US 200 million listed capitalization that derive at least 60% of their income from property investment related activities. Thus, the aim of these indices is to reflect property investment, which is primarily for the purposes of obtaining income, while companies engaged in construction and similar activities are excluded. As discussed in Bond, Karolyi, and Sanders (2003) and Yang, Kolari, and Zhu (2005), despite various limitations to the international real estate data, the EPRA/NAREIT Global Real Estate Indices can be considered to represent general trends in all eligible real estate stocks worldwide and has quickly become a benchmark index for reference in the financial markets.

Empirical Results

A rolling estimation technique is used to generate out-of-sample forecasts. To illustrate, suppose there are N observations in the sample, where $N = R + P$. At time t, a rolling sample of size R observations is used to produce a one-step-ahead forecast, $\hat{Y}_{t+1}$. Therefore, a sequence of P one-step-ahead forecasts can be generated to evaluate each of the models under consideration. As pointed out in Swanson and White (1997), the rolling technique can further allow for the (potentially nonlinear) relation between the current and past returns to evolve across time. Thus, the rolling estimation of various nonlinear models extends the earlier work of Guirguis, Giannikos, and Anderson (2005), among others.

Four forecasting evaluation criteria are applied to evaluate out-of-sample forecasts of the models relative to the benchmark martingale model. These four evaluation criteria are as follow:

$$MSFE = P^{-1} \sum_{t=R}^{N} (Y_{t+1} - \hat{Y}_{t+1})^2,$$

$$MAFE = P^{-1} \sum_{t=R}^{N} |Y_{t+1} - \hat{Y}_{t+1}|,$$

$$MFTR = P^{-1} \sum_{t=R}^{N} sign(\hat{Y}_{t+1})Y_{t+1},$$

and

$$MCFD = P^{-1} \sum_{t=R}^{N} 1[sign(\hat{Y}_{t+1})sign(Y_{t+1}) > 0],$$

where $sign(.)$ denotes $sign(\hat{Y}_{t+1}) = 1$ if $\hat{Y}_{t+1} \geq 0$ and $sign(\hat{Y}_{t+1}) = -1$ if $\hat{Y}_{t+1} < 0$.

Following Hong and Lee (2003) and Yang, Su, and Kolari (2008), in addition to the commonly-used statistical criteria, mean squared forecast error, and mean squared absolute error (MSFE and MAFE), two economic criteria are also employed: mean forecast trading return (MFTR)and mean correct forecast direction (MCFD). Both economic criteria can be particularly informative to profit-maximizing investors. Because asset returns are volatile, forecast errors can be quite large from period to period. Thus, the statistical accuracy of forecasts (as measured by MSFE and MAFE) may not necessarily imply economic accuracy in the sense of maximizing investor profits. However, economic accuracy might be more relevant, as investors may base their trading decisions on maximizing profits rather than minimizing forecasting errors. Furthermore, accurate forecasts of the direction of price changes may be even more important to investors than the magnitude of the changes, as they can be easily translated into profits. In sum, it is also desirable to compute economic measures (e.g., MFTR and MCFD) of forecast accuracy (e.g., Leitch and Tanner, 1991; Hong and Lee, 2003), and the use of multiple criteria in this study provides a more comprehensive perspective on the predictability of securitized real estate returns.

As mentioned above, an adequately large number of observations are important to having efficient estimates of the model parameters, and thus the size of R must be reasonably large. The size of P must be also large enough to detect the differences in forecasting performance across models. Given the number of observations in the data ($N = 2{,}349$ and $N = 3{,}913$ for U.S. and international securitized real estate markets, respectively), an appropriate or balanced choice for R can be expressed by the ratio $R{:}P = 2{:}1$.[7]

Exhibits 2–5 report the results for the U.S. markets and Exhibits 6–9 report the results for the international markets. Each exhibit contains one of the forecasting evaluation criteria in the order presented above. For example, Exhibit 2 reports the out-of-sample forecast results using the MSFE for the three U.S.-based securitized real estate indices, as well as the S&P 500 index. All forecast results are based on an $R{:}P$ ratio of 2:1. Each table also contains the two distinct p-values: P_1 and P_2 based on the White's (2000) reality check test, which addresses the dangerous practice of data snooping or data re-usage for the purpose of inference. Specifically, White (2000) constructs a method for testing the

Exhibit 2 | Forecast Evaluation Results for U.S. Markets: MSFE

	Top 50	Equity	Mortgage	S&P 500
Benchmark	8.734	8.547	8.593	0.840
AR(1)	0.976	0.977	0.995	0.986
P_1	0.100	0.110	0.320	0.050
P_2	0.080	0.090	0.290	0.050
EGARCH(1,1)	1.001	1.000	1.001	0.999
P_1	0.680	0.470	0.800	0.220
P_2	0.080	0.090	0.290	0.050
NN(1,5)	1.147	1.066	1.037	1.019
P_1	1.000	0.970	0.960	0.870
P_2	0.360	0.340	0.640	0.230
FC(1,200)	3.858	11.990	2.220	0.989
P_1	0.980	0.980	0.990	0.210
P_2	0.690	0.690	0.830	0.390
NP(200,400)	1.047	1.037	1.112	1.011
P_1	0.810	0.770	0.980	0.900
P_2	0.790	0.770	0.920	0.420
Combined I	0.972	0.954	0.988	0.978
P_1	0.070	0.072	0.127	0.014
P_2	0.760	0.652	0.861	0.176
Combined II	0.970	0.951	0.984	0.979
P_1	0.060	0.059	0.040	0.018
P_2	0.750	0.637	0.818	0.176

Notes: (1) The data are daily data from December 31, 1999 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are various models under consideration. For the benchmark model, the MSFEs are in levels ($\times 10^4$). For all other models, they are MSFE ratios relative to that of the benchmark model. The smaller the MSFE, the better the predictive ability of a model.

hypothesis that the best model encountered during a specification search has no predictive superiority over the benchmark model.[8] Thus, the test permits for data snooping to be undertaken with some degree of confidence that one will not mistake results generated by chance for genuinely "good" results. In this study, P_1 is the bootstrap *p*-value for comparing a single model to the benchmark model, which is the martingale model $Y_t = \mu + \varepsilon_t$. P_2 is the bootstrap reality check *p*-value for comparing the *k* models to the benchmark model, as it is the bootstrap reality check *p*-value for the null hypothesis that the best of the first k models has

Exhibit 3 | Forecast Evaluation Results for U.S. Markets: MAFE

	Top 50	Equity	Mortgage	S&P 500
Benchmark	1.726	1.699	1.714	0.635
AR(1)	0.999	0.998	1.006	0.999
P_1	0.410	0.380	0.900	0.410
P_2	0.460	0.410	0.870	0.410
EGARCH(1,1)	1.001	1.001	1.001	0.999
P_1	0.910	0.880	0.940	0.150
P_2	0.550	0.420	0.980	0.490
NN(1,5)	1.059	1.028	1.034	1.008
P_1	1.000	0.990	0.990	0.860
P_2	0.780	0.720	0.990	0.740
FC(1,200)	1.313	2.943	2.918	7.876
P_1	1.000	1.000	1.000	0.570
P_2	0.880	0.870	0.990	0.830
NP(200,400)	1.035	1.020	1.036	1.007
P_1	0.980	0.920	0.990	0.870
P_2	0.930	0.920	1.000	0.890
Combined I	0.996	0.994	1.003	0.994
P_1	0.280	0.195	0.716	0.085
P_2	0.850	0.819	0.999	0.448
Combined II	0.994	0.992	1.000	0.994
P_1	0.200	0.158	0.491	0.076
P_2	0.820	0.782	0.968	0.448

Notes: (1) The data are daily data from December 31, 1999 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are various models under consideration. For the benchmark model, the MAFEs are in levels ($\times 10^2$). For all other models, they are MAFE ratios relative to that of the benchmark model. The smaller the MAFE, the better the predictive ability of a model.

no superior predictive ability over the benchmark model. Thus, the last P_2 value (in the last row of the table) checks if the best of all the models under consideration has superior predictive ability over the martingale model. Obviously, the difference between each P_1 and the last P_2 gives an estimate of data-snooping bias.

Exhibit 1 presents a summary of all the models used in the estimation. Exhibits 2 and 3 report the results for the three U.S.-based securitized real estate indexes and the S&P 500 index using statistical criteria MSFE and MAFE. For the

Exhibit 4 | Forecast Evaluation Results for U.S. Markets: MFTR

	Top 50	Equity	Mortgage	S&P 500
Benchmark	−0.259	−0.224	0.088	−0.018
AR(1)	0.318	0.242	0.207	0.068
P_1	0.000	0.000	0.180	0.020
P_2	0.000	0.000	0.190	0.020
EGARCH(1,1)	−0.135	−0.047	−0.037	−0.029
P_1	0.140	0.080	0.910	0.720
P_2	0.000	0.000	0.270	0.020
NN(1,5)	−0.187	−0.152	0.037	0.009
P_1	0.280	0.320	0.640	0.230
P_2	0.000	0.000	0.370	0.030
FC(1,200)	0.070	−0.143	0.235	0.053
P_1	0.020	0.290	0.100	0.080
P_2	0.000	0.000	0.300	0.060
NP(200,400)	0.296	0.213	0.186	0.010
P_1	0.000	0.010	0.130	0.260
P_2	0.000	0.000	0.320	0.060
Combined I	0.302	0.271	0.255	0.097
P_1	0.000	0.003	0.107	0.006
P_2	0.000	0.001	0.257	0.015
Combined II	0.230	0.204	0.301	0.086
P_1	0.000	0.003	0.022	0.013
P_2	0.000	0.002	0.141	0.015

Notes: (1) The data are daily data from December 31, 1999 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are various models under consideration. The larger the MFTR, the better the predictive ability of a model.

benchmark model, MSFE and MAFE in Exhibits 2 and 3 are numbers based on the formula mentioned previously. For all other models, numbers are in ratios relative to that of the benchmark model. As such, a number less than 1 indicates that the underlying model has a smaller statistical error. For Exhibit 2, the results show that all MSFE ratios for the three nonlinear-in-mean models (NN(1,5), FC(1,200), and NP(200,400)) (except for the S&P 500 index under FC(1,200)) are above 1. Therefore, none of the nonlinear-in-mean models outperforms the benchmark for the securitized real estate returns. These findings are consistent with previous studies (e.g., Hsieh, 1991) that show poor forecasting performance of nonlinear-in-mean models relative to the benchmark martingale models in terms

Exhibit 5 | Forecast Evaluation Results for U.S. Markets: MCFD

	Top 50	Equity	Mortgage	S&P 500
Benchmark	0.468	0.478	0.497	0.495
AR(1)	0.518	0.504	0.478	0.519
P_1	0.010	0.090	0.830	0.110
P_2	0.010	0.090	0.820	0.110
EGARCH(1,1)	0.473	0.484	0.477	0.503
P_1	0.150	0.110	0.890	0.160
P_2	0.010	0.090	0.970	0.110
NN(1,5)	0.466	0.455	0.470	0.504
P_1	0.500	0.830	0.870	0.310
P_2	0.020	0.180	0.980	0.230
FC(1,200)	0.503	0.496	0.482	0.518
P_1	0.070	0.200	0.740	0.180
P_2	0.030	0.250	0.980	0.320
NP(200,400)	0.512	0.499	0.499	0.518
P_1	0.010	0.150	0.470	0.160
P_2	0.040	0.290	0.840	0.360
Combined I	0.514	0.517	0.491	0.547
P_1	0.010	0.015	0.601	0.011
P_2	0.040	0.120	0.848	0.053
Combined II	0.512	0.514	0.497	0.547
P_1	0.000	0.011	0.483	0.013
P_2	0.040	0.120	0.856	0.055

Notes: (1) The data are daily data from December 31, 1999 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are the various models under consideration. The larger the MCFD, the better the predictive ability of a model.

of statistical criteria. The EGARCH(1,1) also shows poor forecasting ability. On the other hand, when evaluated alone, each of the remaining models (AR(1) and the two combinations) in some cases reveals better predictive ability than the benchmark. Note that the Combined II forecasts pool forecasts from all individual models: AR(1), EGARCH(1,1), NN(1,5), FC(1,200), and NP(200,400), while the Combined I forecasts exclude the forecast of the nonlinear-in-variance EGARCH(1,1) model. Based on the MSFE criterion and the P_1 statistics, the AR(1) model beats the benchmark model in all cases numerically, however, statistically, it is only significant for the case of REIT top 50 at the 10% significance level. The Combined I and II models show the most forecasting power

Exhibit 6 | Forecast Evaluation Results for International Markets: MSFE

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Benchmark	1.221	2.278	3.169	3.521	4.861	1.994	2.842	3.356	2.355
AR(1)	1.002	1.003	0.998	0.999	0.995	1.005	1.005	1.003	1.007
P_1	0.890	0.990	0.290	0.450	0.240	0.920	0.850	0.840	0.880
P_2	0.890	0.990	0.280	0.440	0.250	0.910	0.860	0.830	0.900
EGARCH(1,1)	1.000	1.001	1.005	1.002	1.001	1.001	1.008	1.001	1.003
P_1	0.660	0.920	1.000	0.940	0.890	0.820	1.000	0.890	0.950
P_2	0.850	0.980	0.320	0.520	0.250	0.930	1.000	0.950	0.990
NN(1,5)	0.994	1.038	0.991	1.021	1.053	1.046	1.026	1.044	1.063
P_1	0.260	0.980	0.310	0.870	0.940	0.940	0.960	0.830	0.950
P_2	0.280	0.990	0.340	0.740	0.590	0.970	1.000	0.980	1.000
FC(1,200)	4.967	2.104	3.347	1.918	1.762	2.599	4.865	4.497	1.511
P_1	0.910	1.000	0.910	0.950	1.000	0.970	0.960	0.900	0.990
P_2	0.720	1.000	0.680	0.900	0.780	0.990	1.000	0.990	1.000
NP(200,400)	1.001	1.005	1.119	1.000	1.024	1.028	1.011	1.031	1.001
P_1	1.000	0.760	0.900	0.540	0.890	0.890	0.820	0.960	1.000
P_2	0.720	1.000	0.780	0.920	0.860	1.000	1.000	1.000	1.000

Exhibit 6 | (continued)

Forecast Evaluation Results for International Markets: MSFE

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Combined I	0.998	0.998	0.950	0.994	0.989	0.998	0.989	0.998	0.998
P_1	1.000	1.000	0.082	0.246	0.086	1.000	0.200	1.000	1.000
P_2	0.720	1.000	0.687	0.842	0.798	1.000	0.740	1.000	1.000
Combined II	0.998	0.998	0.951	0.993	0.989	0.998	0.988	0.998	0.998
P_1	1.000	1.000	0.092	0.193	0.081	1.000	0.180	1.000	1.000
P_2	0.720	1.000	0.688	0.817	0.793	1.000	0.730	1.000	1.000

Notes: (1) The data are daily data from January 3, 1994 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are various models under consideration. For the benchmark model, the MSFEs are in levels ($\times 10^4$). For all other models, they are MSFE ratios relative to that of the benchmark model. The smaller MSFE, the better the predictive ability of a model.

Exhibit 7 | Forecast Evaluation Results for International Markets: MAFE

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Benchmark	0.747	1.040	1.128	1.245	1.512	0.928	1.112	1.204	1.027
AR(1)	1.002	1.001	0.999	0.996	1.002	1.001	1.004	1.003	1.002
P_1	0.990	0.950	0.280	0.180	0.670	0.720	0.870	0.940	0.830
P_2	0.990	0.940	0.240	0.190	0.670	0.730	0.890	0.950	0.820
EGARCH(1,1)	1.000	1.000	1.005	1.003	1.001	1.000	1.008	1.001	1.001
P_1	0.670	0.640	1.000	0.990	0.980	0.730	1.000	0.960	0.940
P_2	0.870	0.860	0.450	0.190	0.990	0.890	0.990	0.990	0.980
NN(1,5)	1.001	1.015	1.006	1.023	1.018	1.015	1.018	1.007	1.025
P_1	0.590	0.990	0.880	1.000	1.000	1.000	1.000	0.830	1.000
P_2	0.930	0.940	0.650	0.380	1.000	0.950	1.000	0.990	0.990
FC(1,200)	6.694	4.806	4.434	4.017	3.306	5.390	4.495	4.153	4.868
P_1	1.000	1.000	0.990	0.980	1.000	1.000	1.000	0.980	1.000
P_2	0.960	0.970	0.820	0.670	1.000	0.980	1.000	1.000	0.990
NP(200,400)	1.001	1.002	1.019	1.003	1.009	1.012	0.999	1.015	1.001
P_1	1.000	0.630	0.930	0.740	0.870	0.950	0.460	0.990	1.000
P_2	0.960	0.990	0.890	0.700	1.000	0.990	0.930	1.000	0.990

Exhibit 7 | (continued)

Forecast Evaluation Results for International Markets: MAFE

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Combined I	0.998	0.998	0.989	0.996	0.998	0.998	0.994	0.998	0.998
P_1	1.000	1.000	0.098	0.173	0.311	1.000	0.140	1.000	1.000
P_2	0.960	0.990	0.547	0.698	0.789	0.990	0.640	1.000	0.990
Combined II	0.998	0.998	0.990	0.995	0.997	0.998	0.994	0.998	0.998
P_1	1.000	1.000	0.092	0.108	0.261	1.000	0.140	1.000	1.000
P_2	0.960	0.990	0.547	0.655	0.775	0.990	0.640	1.000	0.990

Notes: (1) The data are daily data from January 3, 1994 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are various models under consideration. For the benchmark model, the MAFEs are in levels ($\times 10^2$). For all other models, they are MAFE ratios relative to that of the benchmark model. The smaller the MAFE, the better the predictive ability of a model.

Exhibit 8 | Forecast Evaluation Results for International Markets: MFTR

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Benchmark	0.004	0.025	−0.026	−0.013	−0.012	−0.003	−0.178	0.029	−0.026
AR(1)	−0.005	0.007	0.038	0.158	0.185	0.041	0.004	−0.005	0.055
P_1	0.650	0.830	0.100	0.010	0.010	0.200	0.010	0.810	0.100
P_2	0.680	0.820	0.130	0.010	0.010	0.200	0.000	0.810	0.080
EGARCH(1,1)	0.007	0.025	−0.143	−0.120	−0.034	−0.017	−0.074	0.026	−0.022
P_1	0.340	0.980	0.950	0.950	0.650	0.860	0.060	0.720	0.390
P_2	0.620	0.500	0.270	0.020	0.010	0.200	0.010	0.870	0.080
NN(1,5)	−0.019	0.006	−0.018	−0.022	−0.046	−0.025	−0.017	0.057	0.016
P_1	0.800	0.650	0.450	0.570	0.660	0.660	0.000	0.300	0.260
P_2	0.770	0.730	0.330	0.020	0.020	0.300	0.010	0.430	0.150
FC(1,200)	0.010	0.094	0.180	0.107	0.040	0.083	0.093	0.077	0.092
P_1	0.440	0.110	0.000	0.050	0.300	0.080	0.000	0.270	0.040
P_2	0.750	0.170	0.000	0.030	0.040	0.130	0.000	0.370	0.060
NP(200,400)	0.009	0.119	0.072	0.113	0.134	0.062	0.099	0.052	0.077
P_1	0.430	0.060	0.060	0.020	0.060	0.110	0.000	0.380	0.060
P_2	0.800	0.110	0.000	0.030	0.040	0.170	0.000	0.440	0.080

Exhibit 8 | (continued)

Forecast Evaluation Results for International Markets: MFTR

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Combined I	0.043	0.082	0.110	0.191	0.179	0.091	0.081	0.089	0.039
P_1	0.060	0.050	0.016	0.001	0.008	0.020	0.000	0.130	0.060
P_2	0.290	0.110	0.001	0.004	0.039	0.140	0.000	0.380	0.080
Combined II	0.044	0.088	0.111	0.193	0.182	0.067	0.059	0.093	0.049
P_1	0.070	0.040	0.003	0.002	0.016	0.020	0.000	0.110	0.020
P_2	0.270	0.110	0.001	0.004	0.039	0.140	0.000	0.360	0.080

Notes: (1) The data are daily data from January 3, 1994 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are various models under consideration. The larger the MFTR, the better the predictive ability of a model.

Exhibit 9 | Forecast Evaluation Results for International Markets: MCFD

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Benchmark	0.506	0.540	0.469	0.497	0.476	0.524	0.426	0.514	0.516
AR(1)	0.496	0.530	0.478	0.535	0.505	0.525	0.479	0.493	0.515
P_1	0.940	0.930	0.240	0.020	0.050	0.460	0.000	0.980	0.500
P_2	0.940	0.920	0.230	0.020	0.050	0.450	0.010	0.980	0.520
EGARCH(1,1)	0.504	0.540	0.454	0.449	0.476	0.524	0.469	0.514	0.517
P_1	0.760	0.000	0.870	0.990	0.480	0.550	0.020	0.750	0.310
P_2	0.880	0.500	0.390	0.040	0.080	0.630	0.010	0.870	0.610
NN(1,5)	0.484	0.523	0.476	0.492	0.473	0.509	0.463	0.512	0.524
P_1	0.970	0.890	0.320	0.600	0.570	0.860	0.020	0.550	0.320
P_2	0.930	0.750	0.510	0.060	0.100	0.740	0.010	0.920	0.490
FC(1,200)	0.503	0.554	0.526	0.523	0.482	0.544	0.534	0.507	0.541
P_1	0.550	0.130	0.000	0.060	0.350	0.090	0.000	0.660	0.050
P_2	0.950	0.270	0.000	0.060	0.110	0.220	0.000	0.940	0.160
NP(200,400)	0.505	0.559	0.510	0.519	0.506	0.536	0.537	0.511	0.536
P_1	0.510	0.070	0.010	0.040	0.020	0.180	0.000	0.600	0.080
P_2	0.960	0.200	0.000	0.060	0.110	0.240	0.000	0.950	0.170

Exhibit 9 | (continued)

Forecast Evaluation Results for International Markets: MCFD

	AU	FR	GR	HK	JP	NE	SG	SW	UK
Combined I	0.466	0.415	0.509	0.539	0.503	0.403	0.531	0.397	0.271
P_1	1.000	1.000	0.006	0.002	0.050	1.000	0.000	1.000	1.000
P_2	0.960	0.230	0.003	0.035	0.114	0.260	0.000	0.960	0.200
Combined II	0.466	0.422	0.508	0.540	0.503	0.397	0.522	0.397	0.275
P_1	1.000	1.000	0.002	0.001	0.041	1.000	0.000	1.000	1.000
P_2	0.970	0.230	0.003	0.032	0.116	0.270	0.000	0.960	0.200

Notes: (1) The data are daily data from January 3, 1994 to December 31, 2008. (2) P_1 is the bootstrap *p*-value for comparing a single model with the martingale model (the benchmark model) using White's (2000) test with 1000 bootstrap replications and a bootstrap smoothing parameter $q = 0.75$. P_2 is the bootstrap reality check *p*-value for comparing *k* models with the martingale model, where the null hypothesis is that the best of the first *k* models has no superior predictive power over the martingale model. (3) AR, EGARCH, NN, FC, and NP are various models under consideration. The larger the MCFD, the better the predictive ability of a model.

as they are able to beat the martingale model in all cases except for the Combined I model for the mortgage REITs, which has a P_1 value of 0.127. Note that the Combined II forecasts perform even better than the Combined I forecasts. The result is suggestive of the importance of using combined forecasting results. However, with allowance of data-snooping bias, the P_2 in the last row suggests that the best forecasting model among the seven models including forecast combinations is no better than the martingale model. Such a drastic change underscores the importance of using the reality check test.

The results obtained using the MAFE as the evaluation criteria (Exhibit 3) are very similar to those for the MSFE. As a matter of the fact, none of the models beat the benchmark statistically. All three nonlinear-in-mean models fail to outperform the martingale model for all the markets. Overall, there is no statistical evidence for the MAFE criterion that nonlinear models outperform the martingale model.

Exhibits 4 and 5 report the results using the economic criteria for the U.S.-based securitized real estate returns. All results for these two measures are in levels. The meaning of these results is straightforward. The MFTR shows the daily profit (in percentages) generated by the forecasts of the model, and the MCFD shows the percentage of all directional changes correctly predicted by the model. For example, in the case of the top 50 REITs, the AR (1) model generates profit of 0.318% per trading day on average (or equivalently 79.8% per year with 251 trading days) during the out-of-sample period (before allowance for transaction costs) and correctly predicts 51.8% of the directions of changes. which are mostly contributed by the superior performance of the AR(1) model. The results based on the MFTR (Exhibit 4) suggest some evidence of superior predictive ability for the three nonlinear-in-mean models.[9] The NN model does not outperform the benchmark model. However, the FC model is able to beat the predictive power of the benchmark model for the top 50 REITs, mortgage REITs, and the S&P 500 index, while the nonparametric model is able to beat the top 50 REITs and equity REITs, respectively. On the other hand, the results reveal that the AR(1) model generally outperform the benchmark model in all cases, while the EGARCH(1,1) model is not able to improve the forecasts of the martingale model. The numbers from the combined forecasts as well as the reality check test statistic P_2 suggest the superiority of the combined models over the benchmark model for all cases except the mortgage REITs.

Results based on the MCFD criterion in Exhibit 5 are similar to those based on the MFTR in that there is some evidence of a superior predictive ability for the three nonlinear-in-mean models. While the NN model is not able to outperform the benchmark, the FC and nonparametric models are able to outperform the benchmark for the top 50 REIT index. For top 50 and equity REITs, the results reveal that the AR(1), and both combined models are able to improve the forecasts of the martingale model. The numbers from the reality check P_2 in the last row also confirm the superiority of the combined result.

Overall, according to the statistical criteria, there is very limited evidence for predictability based on nonlinear-in-mean models. Only the Combined I and II models for MSFE show some improvement over the benchmark for all three REITs and the S&P 500 index. The results, however, are not strong due to the insignificant reality check of P_2 values. On the other hand, based on economic criteria, there is evidence of predictability for nonlinear-in-mean models, especially for the FC and nonparametric regression models. The case for the superior predictability is also strong when combined models are used.

The results for the ten international securitized real estate markets in Exhibits 6–9 are largely similar to those of the U.S. markets. Using statistical evaluation criteria (see Exhibits 6 and 7), the findings suggest that even without allowance for data-snooping bias, nonlinear-in-mean models cannot outperform the benchmark. Only the Combined I and II models for Japan under MSFE and the Combined I and II models for Germany under MAFE are statistically significant in outperforming the benchmark. Still the results cannot pass the data-snooping reality test. None of the other models outperforms the benchmark.

The economic evaluation criteria in Exhibits 8 and 9 show a completely different picture. There is strong evidence that nonlinear-in-mean models can outperform the benchmark. The NN model outperforms the benchmark for Singapore under both the MFTR and MCFD criteria. The FC model outperforms the benchmark for five countries, Germany, Hong Kong, the Netherlands, Singapore, and U.K. under both criteria, while the nonparametric model outperforms the benchmark for France, Germany, Hong Kong, Japan, Singapore, and U.K. under both criteria. For the AR(1) model, it outperforms the benchmark for Germany, Hong Kong, Japan, Singapore, and U.K. under MFTR, while it outperforms the benchmark for Hong Kong, Japan, and Singapore under MCFD. The Combined I and II models outperform the martingale model for Germany, Hong Kong, Japan, and Singapore under the MCFD criterion. The Combined I and II models also outperform the benchmark for all countries except for Switzerland before allowance for conducting the reality check test under the MFTR criterion. The combined results are quite robust against the data-snooping test, and at least the predictability for five countries (Germany, Hong Kong, Japan, Singapore, and U.K.) remains under either the MFTR or MCFD criterion.

Overall, based on statistical criteria, there is not much evidence of return predictability for the ten international securitized real estate markets. However, based on economic criteria, there remains rather strong evidence after allowance of data-snooping bias that there is return predictability for many of the ten international securitized real estate markets.

Conclusion

This study investigates the out-of-sample predictability of international securitized real estate market returns based on past returns. In addition to a linear model, this study employs several popular nonlinear-in-mean and nonlinear-in-variance models to more comprehensively explore potential nonlinearity in securitized real estate returns. Although international securitized real estate returns are generally not predictable using commonly-used statistical criteria, there is much evidence for the predictability based on two economic criteria: direction of price changes and trading rule profitability. The importance of using economic criteria is consistent with previous studies (e.g., Leitch and Tanner, 1991; Hong and Lee, 2003; Yang, Su, and Kolari, 2008). The securitized real estate return predictability is also largely due to nonlinearity-in-mean, which has not yet been reported in the literature.

The forecast combination for various models, which has not yet been explored in the real estate literature, also appears to improve the forecasting performance for international securitized real estate returns. Further extending the literature, the allowance of data-snooping bias using White's (2000) reality check test weakens otherwise overwhelmingly strong predictability of securitized real estate returns. Overall, there is still much evidence for the predictability in many international securitized real estate markets. Future research should address the data-snooping bias to provide more reliable evidence on forecasting securitized or direct real estate (e.g., housing) price movements.

Endnotes

[1] The terms "random walk" and "martingale" have been interchangeably used in the efficient capital markets literature. However, it is the martingale property (or unpredictability) of security prices that is of essential interest to this huge body of literature (Fama, 1965).

[2] Nelling and Gyourko (1998) also employ a nonparametric runs test to avoid making any arbitrary assumptions about the distribution of REIT returns, which can yield in-sample but not out-of-sample inference. Existing literature in the direct real estate market has addressed the issue of predictability using different nonlinear time-series forecasting techniques (Crawford and Fratantoni, 2003; Guirguis, Giannikos, and Anderson, 2005; Miles, 2008a).

[3] The classification of nonlinearity-in-mean and nonlinearity-in-variance models in this study closely follows Hsieh (1991) and Campbell, Lo, and MacKinlay (1997). For earlier studies demonstrating the usefulness of nonlinear models in forecasting financial markets, see Hsieh (1991) and Gencay (1998) on the U.S. stock market, Gencay (1999), Hong and Lee (2003), and Yang, Su and Kolari (2008) on the foreign exchange market, and Yang, Cabrera, and Wang (2010) on international exchange-traded fund (ETF) markets.

[4] As pointed out in Campbell, Lo, and MacKinlay (1997, 523–24), the problems of overfitting and data-snooping are different but also related. A typical symptom of overfitting is an excellent in-sample fit but poor out-of-sample performance, while data-snooping refers to excellent but spurious out-of-sample performance. The overfitting problem in the real estate literature has been well illustrated by Ling, Naranjo, and Ryngaert (2000) in the context of linear models and Crawford and Fratantoni (2003) in the context of nonlinear models.

[5] The evidence of trading rule profitability is often based on linear models and the U.S. in the securitized real estate literature. An important exception, Serrano and Hoesli (2010) reported trading rule profitability on the same set of markets, but based on a GARCH-type model (in addition to a linear ARMA model), which is designed to capture nonlinearity-in-variance but not nonlinearity-in-mean (Hsieh, 1991).

[6] Timmerman, and White (1999) use White's reality check test to evaluate the performance of numerous technical trading rules in the stock market while Koopman, Jungbacker, and Hol (2005) use the test for evaluating the performance of different volatility models in forecasting the daily variability of the S&P 100 stock index. Hong and Lee (2003), Qi and Wu (2006), and Yang, Su, and Kolari (2008) use White's reality check test to examine the data-snooping issues in trading rules for the foreign exchange market. Yang, Cabrera, and Wang (2010) report that the allowance for data-snooping bias using White's test renders apparent strong return predictability on many ETF markets to be tenuous, undermining an otherwise impressive performance of forecast combinations.

[7] The analysis was also conducted based on the ratio $R{:}P = 1{:}1$, and obtained qualitatively similar results.

[8] A different approach, k-fold cross validation, is used in the field of machine learning to determine how accurately a learning algorithm will be able to predict data that it was not trained on, and could be potentially useful in this situation. Kohavi (1995) provides detailed explanations on the use of the method.

[9] Closely following Fama (1991) and Gencay (1998), transaction costs are not used in the evaluation of trading rule performance in the various models. Although there are surely positive information and trading costs, according to Fama (1991), the research focuses on the more interesting task of laying out the evidence on the adjustment of prices to various kinds of information (such as past returns).

References

Bond, S.A., G.A. Karolyi, and A.B. Sanders. International Real Estate Returns: A Multifactor. Multicounty Approach. *Real Estate Economics*, 2003, 31, 481–500.

Brooks, C. and S. Tsolacos. International Evidence on the Predictability of Returns to Securitized Real Estate Assets: Econometric Models versus Neural Networks. *Journal of Property Research*, 2003, 20:2, 133–55.

Cai, Z., J. Fan, and Q. Yao. Functional-Coefficient Regression Models for Nonlinear Time Series. *Journal of American Statistical Association*, 2000, 95, 941–56.

Campbell, J., A. Lo, and C. MacKinlay. *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press, 1997.

Chordia, T., R. Roll, and A. Subrahmanyam. Evidence on the Speed of Convergence to Market Efficiency. *Journal of Financial Economics*, 2005, 76, 271–92.

Crawford, G. and M. Fratantoni. Assessing the Forecasting Performance of Regime-Switching. ARIMA and GARCH Models of House Prices. *Real Estate Economics*, 2003, 31, 223–43.

Fama, E.F. The Behavior of Stock Market Prices. *Journal of Business*, 1965, 38, 34–105.

——. Efficient Capital Markets: II. *Journal of Finance*, 1991, 46, 1575–1617.

Gencay, R. The Predictability of Security Returns with Simple Technical Trading Rules. *Journal of Empirical Finance*, 1998, 5, 347–59.

——. Linear, Nonlinear and Essential Foreign Exchange Rate Prediction with Simple Trading Rules. *Journal of International Economics,* 1999, 47, 91–107.

Grinblatt, M. and T.J. Moskowitz. Predicting Stock Price Movements from Past Returns: The Role of Consistency and Tax-Loss Selling. *Journal of Financial Economics*, 2004, 71, 541–79.

Guirguis, H.S., C.I. Giannikos, and R.I. Anderson. The U.S. Housing Market: Asset Pricing Forecasts Using Time-Varying Coefficients. *Journal of Real Estate Finance and Economics*, 2005, 30, 33–53.

Harvey, C.R. The Specification of Conditional Expectations. *Journal of Empirical Finance*, 2001, 8, 573–637.

Hong, Y.M. and T.H. Lee. Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models. *Review of Economics and Statistics*, 2003, 85, 1048–62.

Hsieh, D.A. Chaos and Nonlinear Dynamics: Application to Financial Markets. *Journal of Finance*, 1991, 46, 1839–77.

Jirasakuldech, B., R. Campbell, and R. Emekter. Conditional Volatility of Equity Real Estate Investment Trust Returns: A Pre- and Post-1993 Comparison. *Journal of Real Estate Finance and Economics*, 2009, 38, 137–54.

Kleiman, R.T., J.E. Payne, and A.P. Sahu. Random Walks and Market Efficiency: Evidence from International Real Estate Markets. *Journal of Real Estate Research*, 2002, 24, 279–97.

Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. International Joint Conference on Artificial Intelligence, 1995.

Koopman, S.J., B. Jungbacker, and E. Hol. Forecasting Daily Variability of the S&P100 Stock Index Using Historical, Realized and Implied Volatility Measurements. *Journal of Empirical Finance*, 2005, 12, 445–75.

Leitch, G. and E. Tanner. Economic Forecast Evaluation: Profits versus Conventional Error Measures. *American Economic Review*, 1991, 81, 580–90.

Li, Y. and K. Wang. The Predictability of REIT Returns and Market Segmentation. *Journal of Real Estate Research,* 1995, 10, 471–82.

Ling, D.C., A. Naranjo, and M. Ryngaert. The Predictability of Equity REIT Returns: Time Variation and Economic Significance. *Journal of Real Estate Finance and Economics*, 2000, 20, 117–36.

Liu, C.H. and J. Mei. The Predictability of Returns on Equity REITs and Their Co-movement with Other Assets. *Journal of Real Estate Finance and Economics*, 1992, 5, 401–18.

Liu, C.H. and J. Mei. Predictability of Real Estate Returns and Market Timing. *Journal of Real Estate Finance & Economics*, 1994, 8, 115–35.

Ma, C. Mean Reversions in GNMA Returns. *Journal of the American Real Estate and Urban Economics Association*, 1990, 18, 207–26.

MacKinnon, G.H. and A.A. Zaman. Real Estate for the Long Term: The Effect of Return Predictability on Long-Horizon Allocations. *Real Estate Economics*, 2009, 37, 117–53.

McQueen, G. and S. Thorley. Are Stock Returns Predictable? A Test Using Markov Chains. *Journal of Finance*, 1991, 46, 239–63.

Miles, W. Boom-Bust Cycles and the Forecasting Performance of Linear and Non-linear Models of House Prices. *Journal of Real Estate Finance and Economics*, 2008a, 36, 249–64.

——. Volatility Clustering in U.S. Home Prices. *Journal of Real Estate Research*, 2008b, 30, 73–90.

Nelling, E. and J. Gyourko. The Predictability of Equity REIT Returns. *Journal of Real Estate Research*, 1998, 16, 251–68.

Nelson, D. Conditional Heteroscedasticity in Asset Returns: A New Approach. *Econometrica*, 1991, 59, 347–70.

Nguyen, N. and A. Cripps. Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. *Journal of Real Estate Research*, 2001, 22, 313–36.

Osland, L. An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling. *Journal of Real Estate Research*, 2010, 32, 289–320.

Peterson, S. and A.B. Flanagan. Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research*, 2009, 31, 147–64.

Qi, M. and Y. Wu. Technical Trading-Rule Profitability, Data Snooping, and Reality Check: Evidence from the Foreign Exchange Market. *Journal of Money, Credit and Banking,* 2006, 38, 2135–58.

Rapach, D.E., J.K. Strauss, and G. Zhou. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies*, 2010, 23, 821–62.

Serrano, C. and M. Hoesli. Are Securitized Real Estate Returns More Predictable than Stock Returns? *Journal of Real Estate Finance and Economics*, 2010, 41, 170–92.

Sullivan, R., A. Timmermann, and H. White. Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. *Journal of Finance*, 1999, 54, 1647–91.

Swanson, N.R. and H. White. A Model Selection Approach to Real Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics*, 1997, 79, 540–50.

Yang, J., J.F. Cabrera, and T. Wang. Nonlinearity, Data-Snooping, and Stock Index ETF Return Predictability. *European Journal of Operational Research*, 2010, 200, 498–507.

Yang, J., W. Kolari, and G. Zhu. European Public Real Estate Market Integration. *Applied Financial Economics*, 2005, 15, 895–905.

Yang, J., X. Su, and J.W. Kolari. Do Euro Exchange Rates Follow a Martingale? Some Out-of-Sample Evidence. *Journal of Banking and Finance*, 2008, 32, 729–40.

White, H. A Reality Check for Data Snooping. *Econometrica*, 2000, 68, 1097–1126.

We thank Yongmiao Hong and Tae-Hwy Lee for sharing some computer programs used in this study, Qi Li, Ko Wang (the editor), and particularly three anonymous *referees for helpful comments and suggestions. Yang acknowledges financial support from the UC Denver Faculty Research & Scholarship Completion Fund.*

Juan Cabrera, Ramapo College of New Jersey, Mahwah, NJ 07430 or jcabrera@ramapo.edu.

Tao Wang, The City University of New York, Flushing, NY 11367 or tao.wang@qc.cuny.edu.

Jian Yang, University of Colorado Denver, Denver, CO 80202 and Texas A&M University, College Station, TX 77840 or Jian.Yang@ucdenver.edu.