INTERNATIONAL REAL ESTATE REVIEW

Losing Track of the Asset Markets: the Case of Housing and Stock

Kuang-Liang Chang
Department of Applied Economics, National Chiayi University, 580 Sinmin Road, Chiayi City 60054, Taiwan. Tel.: 886-5-2732852, E-mail: klchang@mail.ncyu.edu.tw

Nan-Kuang Chen
Department of Economics, National Taiwan University, 21 Shuchow Road, Taipei 10021, Taiwan. Tel.: 886-2-2351-9641 ext. 471, E-mail: nankuang@ntu.edu.tw

Charles Ka Yui Leung*
Department of Economics and Finance, City University of Hong Kong, Kowloon Tong, Hong Kong. Tel.: 852-2788-9604, Fax: 852-2788-8806, E-mail: kycleung@cityu.edu.hk

This paper revisits the relationships among macroeconomic variables and asset returns. Based on recent developments in econometrics, we categorize competing models of asset returns into different "Equivalence Predictive Power Classes" (EPPCs). During the pre-crisis period (1975-2005), some models emphasize that imperfect capital markets outperform an AR(1) for the forecast of housing returns. After 2006, a model that includes both an external finance premium (EFP) and the TED spread "learns and adjusts" faster than competing models. Models that encompass GDP experience a significant decay in predictive power. We also demonstrate that a simulation-based approach is complementary to the EPPC methodology.

---

* Corresponding author

Keywords

Monetary Policy, Financial Market Variables, Univariate Benchmark for A Single-regime (USB), Markov Regime Switching, Forecasting

I'n view of the structural equality of explanation and prediction, it may be said that an explanation ... is not complete unless it might as well have functioned as a prediction."

Carl Hempel, 1942, The Function of General Laws in History.

1. Introduction

This paper contributes to the literature in a number of ways. First, it revisits the relationship between some of the macroeconomic variables and asset prices (housing and stock). On the one hand, relevant information has been summarized in asset prices based on the efficient market hypothesis (EMH), and hence the introduction of macroeconomic variables should not improve our prediction for these asset prices.[2] On the other hand, there is an emerging strand in the literature which suggests that there are non-trivial interactions among macroeconomic variables and asset prices, and therefore including macroeconomic variables would enhance our understanding of asset prices.[3] A simple way to distinguish these two bodies of theories above would be to set asset prices as the dependent variable and introduce the macroeconomic variables as the independent variables, in addition to the inclusion of lagged asset prices. If the coefficients of the macroeconomic variables are statistically insignificant, then the EMH is confirmed. Conversely, if the coefficients of the macroeconomic variables are statistically significant, then the EMH is rejected and the macroeconomic variable-asset price interactions are indeed important.

While such an approach is intuitive and easy to improve, there are several shortcomings. First, if asset price movements indeed generate a wealth effect or collateral effect,[4] then there is a feedback effect from asset prices to

---

[2] Clearly, there are different views about the EMH and ways to test the EMH. Among others, see Fama (1970), Darrat and Glascock (1989), LeRoy (1989).

[3] Among others, see Green (2002), Case, Quigley and Shiller (2005), and Campbell and Cocco (2007) for a discussion of the wealth effect that can be created by asset price fluctuations; Leung (2004), Chen and Leung (2008), Jin et al. (2012) and the reference therein for the interaction between housing prices and macroeconomic variables through the collateral effect.

[4] Since aggregate consumption constitutes almost 70% of the total GDP of the U.S., and many countries target their exports to the USA, the wealth effect generated by asset price swings can have important implications to the economies of both the U.S. and many of its trade partners. The collateral effect refers to the scenario in which continuous declines

macroeconomic variables. Thus, an endogeneity bias may occur. The *minimum models* that can embed the potential feedback among asset prices and macroeconomic variables are the vector autoregressive (*VAR*) models,[5] but comparing VAR models is not always a trivial task.[6] A common practice is to apply the test proposed by Diebold and Mariano (1995), which allows for a bilateral model comparison. In the context of asset price movements, we may want to compare more than two models as the rationale is quite clear. Different theories are "represented" by different econometric models, in the sense of the different variables and different possible relationships among variables that are being highlighted.[7] To compare which "theory" of asset price movements, it is important to employ an appropriate procedure to select the "best performing model" among competing ones. Our second major contribution to the literature thus combines the work on multi-model performance comparison by Mariano and Preve (2012) and the work on the "model confidence set" by Hansen et al. (2011) in order to sequentially "eliminate" less competitive models. As a result, we can categorize models into several "Equivalence Predictive Power Classes" (EPPCs). We may be able to effectively provide indirect evidence on the theories that have the same explanatory powers and those that have lower explanatory power than some of their rivaling theories. Thus, the empirical results obtained herein might provide some reference for the future development of theoretical modeling. The method is very simple, can be applied to any finite number of models and very different contexts, and hence may have some independent interest.[8]

---

in house prices can cause a quick decay of collateral quality and value, thus potentially leading to a credit crunch and subsequent rise in bankruptcy and foreclosures.

[5] Sims (1980b) makes a strong case for why the estimating of a system of equations, especially in the context of a dynamically interacting system, is econometrically more sensible than single equation estimation in a macroeconometric context.

[6] It is well known that under some conditions, we can view a VAR model as a collection of univariate regressions. In fact, some standard computer packages deliver a separate R^2 measure for each of the equations within a multivariate VAR. Therefore, it is possible that one VAR model produces a higher "R^2" in the "stock price equation" and yet a lower "R^2" in the "house price equation" than the other VAR. Yet both the "stock price equation" and "house price equation" are part of a dynamic system and hence it seems that reading the "individual equation R^2" may not be sufficient.

[7] Notice that the reduced form dynamics of many dynamic stochastic general equilibrium (DSGE) models have a VAR representation (for instance, see Kan et al., 2004; Leung, 2014). While identification can be an issue and we might not be able to recover the underlying DSGE model from the estimated VAR, comparing the empirical performance of a collection of VAR models might still provide an indirect test for the capacity of the models to account for the data. In the current context, a VAR at least could provide us with some idea about whether regime-switching is important, the variables that should be included in the DSGE, etc. See Kapetanios et al. (2007), among others, for further discussion on this point. See also Pagan and Robinson (2014) who show that some of the existing DSGE models with imperfect capital markets may not explain the data very well.

[8] For instance, see Kwan et al. (2015) for an application of this model comparison method on structural models.

This paper specifically focuses on U.S. aggregate data during the period of 1975Q2-2012Q1 and studies various versions of the VAR models, with and without regime-switching, which arguably represent different views on the driving force of the asset markets.[9] While we have explained why VAR models should be used, our choice of a regime switching model also seems to be non-controversial. First, the possibility of regime switching in the macroeconomic and financial time series has long been studied and verified.[10] Among others, Chen and Leung (2008) show that in the presence of collateral constraints, bankruptcy possibility and asset price spillover, the relationship between aggregate output and real estate price can be very non-linear (piece-wise continuous with different slopes in different segments) and hence may not be well captured by the widely used linear VAR model. In addition, Chang, Chen and Leung (2011) argue that a regime switching model can be consistent with two stylized facts in the housing market: (1) short-run predictability, which has been repeatedly documented since Case and Shiller (1990), and (2) long-run non-profitability, which is a prerequisite of the long run efficiency of the housing market.[11] The inclusion of monetary policy variables would further justify the use of a regime-switching model.[12]

In this paper, we focus on the return stock price index (SRET) and the house price index (HRET), as well as variables that may affect the two asset returns. Our choice of variables is mainly guided by the previous literature on asset price dynamics and will be explained in further detail in the following section.[13] For a more balanced understanding, we conduct both in-sample forecasting (ISF) and out-of-sample forecasting (OSF) herein. Our OSF takes two different

---

[9] The National Bureau of Economic Research, among others, has also indicated that the economic recession started in the first quarter of 2008. When it will end, however, is still a topic for debate.

[10] Again, the literature is too large to be reviewed here. Among others, see Hamilton (1994), Maheu and McCurdy (2000).

[11] Note that if the true model is a single-regime and short-run predictability holds, then we cannot have long-run non-profitability at the same time. Moreover, there is a large amount of literature on testing the housing market efficiency based on this simple relationship. Among others, see Chang et al. (2012, 2013) for more discussion.

[12] For instance, Sims and Zha (2006) find that the changes in monetary policy "*were of uncertain timing, not permanent, and not easily understood, even today*" and that models which "*treat policy changes as permanent, non-stochastic, transparent regime changes are not useful in understanding this history...*" (Italics added by author of this paper). Casual observation also suggests that the conduct of monetary policy has changed over time along with changes in the chair of the Federal Reserve System and several episodes that dramatically affect inflation and economic activity (such as oil price shocks). Thus it may be appropriate to explicitly allow for regime-switching behavior in a study on asset returns and monetary policy.

[13] For instance, there is a large class of dynamic equilibrium models which suggest that house and stock prices should be correlated. Among others, see Leung (2007), Kwan et al. (2015) and the reference therein.

approaches: conditional expectations and simulation-based methods.[14] The merits of using the former approach have been widely discussed in the previous literature. On the other hand, confidence intervals (CIs) may not be available from the former approach. This is an important issue for the regime switching models, because the system not only receives shocks within a regime, but also experiences stochastic switching from one regime to another. By following Sargent, Williams and Zha (2006), we adopt a simulation-based approach to calculate the median path and the CI, with regular model updating.

There is clearly a large and recent literature on asset pricing.[15] For instance, Maheu, McCurdy and Song (2012) study the daily returns of equity indices for 125 years and focus on the regime-dependence of the transition probability matrices. Grishchenko and Rossi (2012) employ the monthly data in the Consumer Expenditure Survey from 1984 to 2012 to estimate the asset price model. Clark (2011) takes real-time data from 1985Q1 to 2009Q1 and shows that stochastic volatility improves the real time forecasting of macroeconomic variables. There is also a strand of literature that compares the rate of returns of real estate versus the stock market from an investment perspective.[16]

This paper differs from the literature in several dimensions. First, this paper takes a "*dynamical system approach*". Our estimation is multivariate and when regime-switching occurs in our equation, it is actually the *whole dynamical system that switches from one regime to another*. This clearly differs from, and hence complements, some of the literature which take a univariate approach. Second, we focus on aggregate data that start from 1975 and hence, we can cover a longer period of time. Since the official aggregate data on housing is quarterly in frequency, we adjust the frequency of the other variables to quarterly in frequency, and hence our work complements earlier works that focus on the higher frequency movements of asset returns. Third, we focus on the application of classical econometrics. Perhaps more importantly, this paper compares not only how different models predict for the in-sample period (or, the "*pre-crisis period*"), but also how different model performances evolve as we recursively update the parameter estimates with new data in the out-of-sample period (or, the "*post-crisis period*"). To some extent, we examine the ability of different models to "*learn and adopt*" within changing asset markets.

---

[14] We will provide more justifications on why this would be appropriate in later sections.
[15] For instance, Welch and Goyal (2008), among others, argue that most if not all models fail to predict equity premium. Recently, Phillips (2013) proves that the confidence intervals in some of the predictive regressions have *zero converge probability* and the corresponding statistics Q would indicate predictability even when there is none.
The focus of this paper is different. First, we emphasize on the ability of the model to account for both stock and housing prices, rather than the equity premium. Second, we emphasize on how different models "*learn and adopt*" since the 2006 housing price decline. Therefore, we will adopt a different approach, as will be explained in later sections.
[16] Among others, see Ibbotson and Siegel (1984), and Quan and Titman (1997, 1999).

To our knowledge, these features are not emphasized by the previous literature and our paper can supplement that gap.

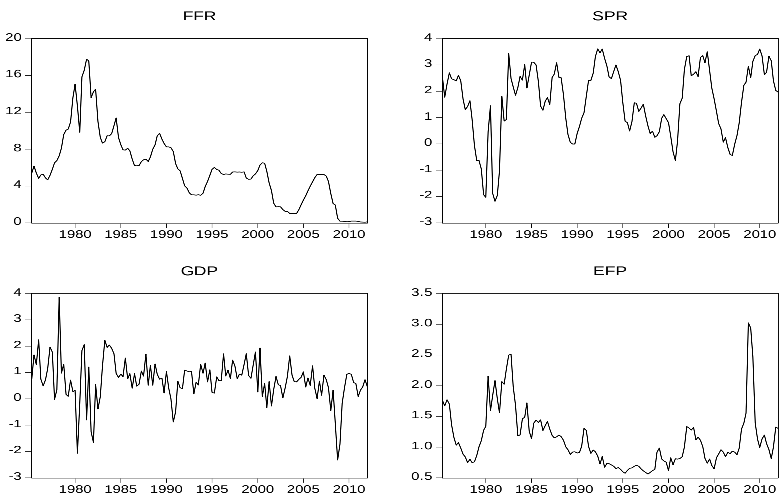
The rest of the paper is organized as follows. In Section 2, the econometric model is described and a statistical summary of the data will be given. In Section 3, the empirical estimation results are presented with the baseline model. In Section 4, forecasting performances are compared across models. Finally, Section 5 concludes.

2. Econometric Analysis

2.1 Data

The analysis of this paper is based on U.S. data that cover the period of 1975Q2-2012Q1. Limited by data availability, the dimensionality constraint of the econometric model, we focus on the returns of the stock price index (SRET) and the house price index (HRET), as stock and house are the most important assets for a typical household in the U.S. Focus is placed on asset return rather than asset prices, because the latter tend to be non-stationary while the former may be mean-reverting. Figure 1 provides a visualization of these variables and clearly demonstrates two stylized facts: (1) the negativity of the housing return in recent years, and (2) the high volatility of stock returns (relative to the housing returns).

Figure 1a Federal Funds Rate (FFR), Term Spread (SPR), Percentage Changes in Gross Domestic Product (GDP), External Finance Premium (EFP)



The previous literature on asset price dynamics guides our selection of the variables that are included, on top of the asset returns. They include: the (3-month) federal funds rates (hereafter *FFR*) which is a measure of the U.S. monetary policy; the term spread (*SPR*) which is a measure of the difference between the long-term and short-term interest rates;[17] *EFP* which is a measure of the degree of credit market imperfect from the perspective of non-financial firms;[18] the TED spread (*TED*), which is a measure of the degree of credit market imperfect from the perspective of the banks;[19] and GDP growth rates (*GDP*).

Figure 1b Market Liquidity (TED), Stock Index Return (SRET), Housing Market Return (HRET)

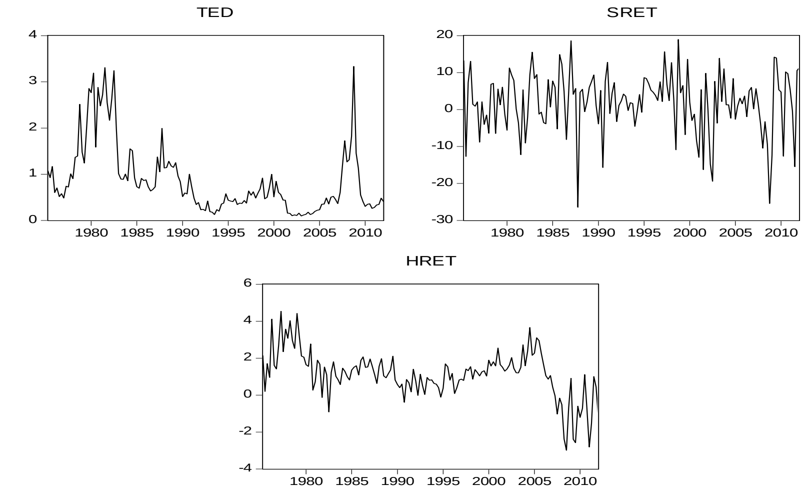
[17] In this paper, SPR is defined as the discrepancy of the long term (10 years) interest rate and the short term (3-month) counterpart.

[18] In this paper, EFP is defined as the spread between high-rating and low-rating corporate bonds (Baa-Aaa).

There are a number of available series that have been used as the measure of external finance premium. Among these are the prime spread (prime loan rate - federal funds rate), corporate bond spread (Baa-Aaa), and high-yield bond spread (Bbb-Aaa). De Graeve (2007) argues that the prime loan spread provides a poor indication of the financing conditions of firms which are typically considered vulnerable to credit market frictions, because it focuses on firms with the highest credit quality, to which financial constraints pertain the least. Gertler and Lown (1999) show that the high-yield bond spread is strongly associated with both general financial conditions and the business cycle (as predicted by the financial accelerator). However, the series only started in the early 1980s. Therefore, we choose the corporate bond spread (Baa-Aaa) as our measure of external finance premium.

[19] In this paper, the TED spread is defined as the difference between the 3-month Eurodollar deposit rate and the 3-month T-bill rate.

The widely-used BBA LIBOR, compiled by the British Bankers' Association, started only from January 1986. Therefore, we replace the 3-month LIBOR rate with the 3-month Eurodollar deposit rate. These two series are highly correlated. Both the corporate bond spread and the 3-month Eurodollar deposit rate are from H.15 statistical release ("Selected Interest Rates") issued by the Federal Reserve Board of Governors.



The choices can be easily justified. For instance, our inclusion of a monetary policy variable in a study on asset price dynamics is consistent with the literature.[20] On the one hand, the monetary policy is widely perceived to be influential to both the stock and housing markets. On the other hand, the Federal Reserve may react to the stock market movement in some instances (e.g. Rigobon and Sack, 2003). In this paper, we follow Sims (1980a) and choose FFR to represent the movement of the monetary policy. Similarly, the GDP growth rate (GDP) seems to be a natural choice for a proxy of "economic fundamental".[21] The SPR is well-known to contain information about future inflation, future real economic activities as well as asset returns.[22] Thus, it may be instructive to include the term structure as a (partly) "*forward-looking variable*" in the regression without taking any stand on the formation of future inflation or interest rate expectation.[23] The *EFP* and the TED spread (*TED*) are

---

[20] For instance, audiences in different occasions have suggested that we adopt the "financial stress index" (FSI). However, the FSI is available only from 1990 onwards, while other variables in our dataset are available since 1975. The use of the FSI will force us to lose much information, including how the system reacts to dramatic events such as the 1987 stock market crash. Furthermore, the FSI may have higher predictive power with higher frequency data, while this paper focuses on the quarterly frequency where GDP data are available.

[21] In contrast, the aggregate consumption may be "too smooth" to account for the movement of stock returns very well. Thus, we use the GDP instead. The literature is too large to be reviewed here. Among others, see Mehra and Prescott (2003).

[22] This statement has been confirmed by data in the U.S. as well as other advanced countries. For a review of more recent literature, see Estrella (2005), Estrella and Trubin (2006), among others.

[23] In the literature on term structures, much effort has been devoted to verify the "expectation hypothesis". However, Collin-Dufresne (2004) shows that there are several

included in the analysis because some recent literature which highlights the role of an imperfect capital market find these variables to be important proxies for the degree of capital market imperfection.[24] While EFP may capture the degree of capital market tightness faced by non-financial firms, the TED spread, which is the difference between the interbank rate and the riskfree rate, can be interpreted as a proxy for the capital market tightness faced by banks. For instance, Kiyotaki and Moore (1997) indicate that in a collateral-constrained economy, the borrowing constraint, which holds at equilibrium, would take the following form:

$$debt \leq E\left(\frac{future-collateral-value}{discount-factor}\right).$$

Clearly, there are other variables that may be important for explaining the asset returns during the sampling period. Unfortunately, not every potentially important variable is available since 1975, and some other variables may not be as useful as they may seem. For instance, it has been suggested to us that the 30 year mortgage rate would improve the performance of our models. In the Appendix, we provide empirical evidence that this may not be the case, at least not for our sampling period. In addition, our "*system approach*" also limits the total number of variables that could be included in the empirical analysis. Thus, the current list tries to balance *economic validity* and *data availability*.

Our choice of the sampling period is also constrained by data availability, as 1975 is the earliest date that the U.S. quarterly data on housing price is available.[25] Given our selected group of VAR models, we examine the in-sample fitting for the period of 1975Q1-2005Q4 and the out-of-sample forecasts for the period beginning with 2006Q1. We choose 2005Q4 as the cut-off point, because the rate of increase in the house price growth that started in the 1990s peaked around the end of 2005.[26] Furthermore, we allow the models

versions of the expectation hypothesis and they are not consistent with one another. Thus, the explicit formulation of the expectation may matter to the final empirical result.

[24] For instance, Jin et al. (2012) have recently shown that the movement of EFP can be related to the housing returns in a DSGE model. In general, as surveyed by Bernanke and Gertler (1995), EFP is perceived as a measure of the "risk premium" and hence a reflection of the credit market conditions that are faced by non-financial firms in the literature. For models that emphasize on the role of imperfect capital markets in the propagation of shocks over the business cycles, see Bernanke, Gertler and Gilchrist (1999), Christiano, Motto and Rostagno (2007), and Davis (2010), among others. However, these papers do not explicitly model housing.

[25] Another merit of choosing 1975 as the starting point is that it also avoids the first oil price crisis, which may be a period of "indeterminacy," especially in respect to the monetary policy, which will make the empirical identification difficult. Among others, see Lubik and Schorfheide (2004) for more analysis on this.

[26] Clearly, there is substantial diversity in terms of when the asset price cycle ends. In this paper, we would experiment different end-dates in the robustness check section and the appendix also discusses the related literature in more details.

to "learn" in later sections in the sense that we would recursively re-estimate the model, for instance, from 1975 to 2006, and use it to predict the asset prices in 2007, and then re-estimate again by using the data from 1975 to 2007 in order to predict the asset prices in 2008, and so on. Thus, the in-sample that we initially choose may not affect our results as much as some of the related studies.

For compatibility with the quarterly house price index provided by the Office of Federal Housing Enterprise Oversight (OFHEO), variables that are originally available on a monthly basis are transformed into quarterly. The S&P 500 stock price index is obtained from DataStream. We compute the stock and housing returns by using the growth rates of the stock and housing price indexes respectively. The data on real GDP originate from the Department of Commerce, Bureau of Economic Analysis. The federal funds rate is taken from the H.15 statistical release ("Selected Interest Rates") issued by the Federal Reserve Board of Governors. As for the SPR, we follow Estrella and Trubin (2006) by choosing the difference between the ten-year treasury bond yield and three-month T-bill rate, both of which are released by the Federal Reserve Board of Governors. Since constant maturity rates are available only after 1982 for 3-month T-bills, we use the three-month T-bill rate from the secondary market expressed on a bond-equivalent basis.[27]

While these time series have all been studied in the literature, it may nevertheless be instructive to present some "stylized facts" before any formal modeling. Table 1 is a statistical summary of the variables in the data. The stock returns have a higher mean than the housing returns (1.9 versus 1.1), and an even larger volatility (8.2 versus 1.3). The simple correlation coefficients in Table 2 show that some variables are indeed highly correlated. For instance, the correlation between the FFR and SPR is *-0.6*, FFR and TED is *0.7*, and EFP and TED is *0.6*. On the other hand, some other correlations are close to *zero*. Thus, it is not clear which model will perform better a priori. To facilitate the comparison, models are constructed in certain ways. As shown in Table 3a, for instance, Models A to E would have FFR involved, which can highlight the potential role of monetary policies in asset return dynamics.[28] Models F to H differ from the previous ones as the monetary policy variable FFR is replaced by a financial market variable. Thus, Model F can be interpreted as Model C with FFR replaced by EFP, Model G as Model E with FFR replaced by EFP, and Model H as Model E with FFR replaced by SPR. Thus, a comparison of

---

[27] The 3-month T-bill rate from the secondary market provided by the Federal Reserve System is on a discount basis. We follow Estrella and Trubin (2006) by converting the three-month discount rate (r^d) to a bond-equivalent rate (r): $r = \left[\left(365 * r^d / 100\right) / \left(360 - 91 * r^d / 100\right)\right] * 100$. They argue that this spread provides an accurate and robust measure in predicting U.S. real activity over long periods of time.
[28] For the purpose of parsimony and model comparison, we set the lag period of all models as one (p=1). It turns out that most models with one lag period have the lowest AIC value, compared to models that have more than one lag period. Details are available upon request.

Models A to E on the one hand, and the Models F to H on the other hand would provide some information on the importance of monetary policies in asset price dynamics. Table 3b also shows how we choose the AR(1) as the univariate benchmark for a single-regime (*USB*). Obviously, if, for instance, econometric models that include EFP as a variable outperform other econometric models, it would provide indirect support that the firm financing problem is important for explaining asset market returns. On the other hand, if models that include the TED spread outperform other models, this would suggest that the interbank market is important in explaining the asset market returns. Thus, by investigating the performance of different models, we would be informed about which economic channels may be more important.

Table 1 Statistical Summary of Federal Funds Rate, Term Spread, Growth Rate of the Gross Domestic Product, External Finance Premium, Market Liquidity, Stock Index Return and Housing Market Return (1975Q2-2012Q1)

	FFR	SPR	GDP	EFP	TED	SRET	HRET
Mean	5.808	1.634	0.699	1.125	0.867	1.910	1.123
Median	5.417	1.789	0.744	0.997	0.605	2.436	1.206
Maximum	17.780	3.611	3.859	3.023	3.333	18.952	4.533
Minimum	0.073	-2.182	-2.328	0.560	0.097	-26.431	-2.989
Std. Dev.	3.806	1.343	0.809	0.473	0.750	8.204	1.277
Skewness	0.785	-0.707	-0.536	1.596	1.643	-0.787	-0.490
Kurtosis	3.838	3.006	6.270	5.914	5.216	3.995	4.735
Observations	148.000	148.000	148.000	148.000	148.000	148.000	148.000

Note: FFR denotes the federal funds rate, SPR denotes the term spread, GDP means the growth rate of the gross domestic product, EFP means the external finance premium, TED means the market liquidity, SRET means the stock index return, and HRET means the housing market return.

Table 2 Correlation Coefficients (1975Q2-2012Q1)

	FFR	SPR	GDP	EFP	TED	SRET	HRET
FFR	1.000	-0.608	0.031	0.290	0.722	0.016	0.266
SPR		1.000	0.050	0.097	-0.410	0.011	-0.251
GDP			1.000	-0.360	-0.250	0.148	0.209
EFP				1.000	0.638	-0.050	-0.192
TED					1.000	-0.130	-0.015
SRET						1.000	-0.007
HRET							1.000

Table 3a List of Models

Model	Model Structure	Variable
A	Linear	FFR, SPR, TED, EFP, GDP, SRET, HRET
B	Two-regime	FFR, GDP, SRET, HRET
C	Two-regime	FFR, SPR, SRET, HRET
D	Two-regime	FFR, EFP, SRET, HRET
E	Two-regime	FFR, TED, SRET, HRET
F	Two-regime	EFP, SPR, SRET, HRET
G	Two-regime	EFP, TED, SRET, HRET
H	Two-regime	SPR, TED, SRET, HRET
USB	Linear	Univariate benchmark for a single-regime (i.e. AR(1))

Note: (unless specified, all variables refer to quarterly data) FFR - Federal funds rate; SPR - term spread, which is equal to (10-year bond rate – FFR); TED - TED spread, which is equal to (3-month Eurodollar deposit rate - 3-month T-bill rate), a measure of market liquidity; EFP - External finance premium, which is equal to corporate bond spread (Baa-Aaa), a measure of external finance premium; GDP - GDP growth rate; SRET - stock market return; HRET - housing market return.

Table 3b Choosing Univariate Benchmark for A Single-Regime (USB) Model

	AIC		SBC	
	AR(1)	**AR(2)**	**AR(1)**	**AR(2)**
SRET	7.0774	7.0902	7.1381	7.1712
HRET	2.7589	2.7464	2.8196	2.8274

Note: AIC refers to the Akaike information criterion. SBC refers to the Schwartz Bayesian information criterion.

(1) For SRET, in terms of both the AIC and SBC, the AR(1) model performs better than AR(2).

(2) For HRET, in terms of the AIC, AR(2) is marginally better (2.746 versus 2.759). However, in terms of the SBC, AR(1) is marginally better (2.820 versus 2.827). For parsimony purposes, we choose AR(1) specification.

2.2 Econometric Model

This paper takes a dynamical system approach in the sense that we estimate VAR models which include both asset returns and other macroeconomic and financial variables and allow them to interact with one another. Some justifications have been discussed and we simply re-organize them here. First, much of the literature uses a univariate approach and hence this paper, which uses an alternative approach, is complementary. Second, it is well-known that when the regressors are not distinguishable from the integrated processes, the conclusions about return predictability could be altered.[29] Under the VAR approach, this issue would become less severe because it is less likely that the whole vector follows a unit root process than in the case of an individual variable. Furthermore, as we have argued in the introduction, the reduced form dynamics of some of the dynamic stochastic general equilibrium (*DSGE*) models actually have a VAR representation. Hence, testing the ability of a class of VAR models may provide an indirect test for a class of DSGE models. Thus, the results in this paper could have some implications for the future development of DSGE models. In addition, while some of the existing literature tends to take one of the returns as given and use its movement to explain the other return, the VAR approach naturally allows for dynamic interactions between the asset returns (housing and stock) and other variables, as well as the feedback effects among asset returns. In other words, the VAR approach avoids assigning one of the asset returns as an "exogenous variable", which could lead to potential endogeneity bias.[30]

Our econometric model is a regime-switching VAR, with lag length p for a (vector) process y_t :

[29] Among others, see Torous et al. (2004), Cochrane (2001).

[30] Among others, see Sims (1980a, b) for more discussion on these issues and the potential biases that could be eliminated by using the VAR method.

$$A_0(s_t)y_t = \gamma(s_t) + \sum_{i=1}^{p} A_i(s_t)y_{t-i} + u_t(s_t), \tag{1}$$

where we allow for all parameters, including the intercept and autoregressive coefficients, and covariance matrix of stochastic terms to be contingent on the unobservable state variable $s_t \in S$. The regime-dependent coefficients capture possible nonlinearities or time variation in the lag structure of the model. The stochastic volatility allows for possible heteroskedasticity of the stochastic terms.

The variable of interest $y_t = (y_{1,t},...,y_{m,t})'$ is an $m \times 1$ vector. The stochastic intercept term $\gamma(s_t) = (\gamma_1(s_t),...,\gamma_m(s_t))'$ captures the difference in the intercept under different states. $A_0(s_t)$ is an $m \times m$ state-dependent matrix that measures the contemporaneous relationship between variables, and the econometric identification of the model is obtained through restrictions on $A_0(s_t)$. In addition, $A_k(s_t)$ is an $m \times m$ matrix with each element state-dependent $a_k^{(ij)}(s_t)$, $i, j = 1,...,m$, $k = 1,..., p$. The stochastic error term u_t will be explained below.

The corresponding reduced form of the above model can be obtained by pre-multiplying (1) by $A_0^{-1}(s_t)$, which yields:

$$y_t = \delta(s_t) + \sum_{i=1}^{p} \Phi_i(s_t)y_{t-i} + \varepsilon_t(s_t), \tag{2}$$

where $\delta(s_t) = A_0^{-1}(s_t)\gamma(s_t)$, $\Phi_k(s_t) = A_0^{-1}(s_t)A_k(s_t)$, $\varepsilon_t(s_t) = A_0^{-1}(s_t)u_t(s_t)$, $k = 1,..., p$. $\Phi_k(s_t)$ is an $m \times m$ matrix with state-dependence for each element $\phi_k^{(ij)}(s_t)$, $i, j = 1,...,m$, $k = 1,..., p$. We further define

$$\delta(s_t) \equiv c + \alpha(s_t),$$

which will be explained below. The vector of the stochastic error term $\varepsilon_t(s_t)$ can be further expressed as:

$$\varepsilon_t(s_t) = A_0^{-1}(s_t)u_t(s_t) = \Lambda(s_t)H^{1/2}v_t(s_t),$$

where H is an $m \times m$ diagonal matrix with diagonal elements σ_j^2, $j = 1,...,m$, $\Lambda(s_t)$ is an $m \times m$ diagonal matrix with diagonal elements $\lambda_j(s_t)$, $j = 1,...,m$,

$$\Lambda(s_t) = \begin{bmatrix} \lambda_1(s_t) & 0 & \dots & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_m(s_t) \end{bmatrix},$$

which captures the difference in the intensity of the volatility, and $v_t(s_t)$ is a vector of standard normal distribution, $v_t(s_t) \sim N(0, \Sigma(s_t))$, where the covariance matrix is given by

$$\Sigma(s_t) = \begin{bmatrix} 1 & r_{21}(s_t) & \dots & r_{m1}(s_t) \\ r_{12}(s_t) & 1 & \dots & r_{m2}(s_t) \\ \vdots & \vdots & \ddots & \vdots \\ r_{1m}(s_t) & r_{2m}(s_t) & \dots & 1 \end{bmatrix}. \tag{3}$$

We also include an "atheoretical benchmark," which is AR(q), i.e. an order-q auto-regressive process that is labelled as the USB. This is clearly motivated by the "EMH", which conjectures that all "relevant information" has been reflected in the current (and potentially previous) period price whereby additional variables, such as those provided by the VAR, are considered insignificant. This could be the case in stock return. On the other hand, the housing market is often accused of not being as efficient as the stock market, and hence the housing market prediction can be improved with additional variables, including the stock return. Hence, a comparison of the model performance with USB not only provides a form of an "efficient market test", but also an indirect test of the "cross-market informational spillover".[31] For the linear VAR model, we include all 7 variables. For the regime-switching models, we can only afford to include 4 variables, which is a much shorter list than the linear model, as we need to estimate parameters in each regime, plus the transition probabilities. By design, this puts regime-switching models in a *disadvantageous position*. If, however, the regime-switching models still outperform the widely used linear model, then this suggests that regime changes may indeed be very important in the data. In addition, if models with certain variable(s) consistently outperform alternatives, then such variable(s) may be important to take into account in asset return movements. Thus, it may be important to include different combinations of our listed variables in different models and test for the performance of those models.

[31] Even for the case of aggregate output, it may still be a good idea to use a univariate AR(p) as the benchmark. Among others, see Chauvet and Potter (2012) for more discussion.

2.3 Two-state Markov Process

As the sample size is a severe constraint, we assume that there are only two states, i.e., $s_t \in S = \{1, 2\}$. The procedure for the identification of the regime of the economy for a given period will be discussed below. The Markov switching process relates the probability that regime j prevails in t to the prevailing regime i in $(t-1)$, $\Pr(s_t = j \mid s_{t-1} = i) = p_{ij}$. The transition probabilities are assumed to be constant and the transition matrix is given by:[32]

$$P = \begin{pmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{pmatrix}.$$

Given that the economy can be either in State 1 or 2, the term $\alpha_j(s_t)$, $j = 1,...,m$, defined above, captures the difference in the intercept under different states. For convenience, we set $\alpha_j(1) = 0$ for $s_t = 1$, thus $\alpha_j(2)$ measures the difference in the intercept between States 2 and 1. Furthermore, we set the diagonal element of $\Lambda(s_t)$ at State 1 to unity, i.e., $\lambda_j(1) = 1$, so that if $\lambda_j(2) > 1$, then the volatility in State 2 is greater than that in State 1, and vice versa. Since $v_t(s_t)$ is a vector of standard normal distribution and $\lambda_j(1) = 1$ is set to one, the variance of $y_{j,t}$, $j = 1,...,m$, at State 1 is σ_j^2, and the variance is $(\lambda_j(2))^2 \sigma_j^2$.

2.4 Identification of Regimes

Since the state of the economy cannot be observed, we identify the regime for a given time period by using the smoothed probability approach in Hamilton (1989, 1994), in which the probability of being state s_t at time t is given by $\pi(s_t \mid \Omega_T)$, where $\Omega_T = \{y_1, y_2, ...y_t, ..., y_T\}$. The idea is that we identify the state of the economy from an ex post point of view, and thus the full set of information is utilized. Note that we only allow for two regimes in this paper, i.e., $s_t \in S = \{1, 2\}$. Thus, if $\pi(s_t = j \mid \Omega_T) > 0.5$, then we identify the economy most likely to be in state j, $j = 1, 2$.[33]

2.5 Forecasting

[32] In principle, we could allow the transition probabilities to depend on the observed variables. However, our accessed time series are relatively short and hence we compromise in the modelling choice.

[33] In addition, we follow Francq and Zakoian (2001) to use a spectral radius to determine the stationarity of the regime-switching models. Due to the limits in space, we report the results in the Appendix. We find that most models are stationary. Only Models B and C are marginally non-stationary.

After we estimate all of the above models, we use the calculated probabilities of regime switching for evaluating the forecasting performances of house and stock prices across various models, and then examine both in-sample and out-of-sample forecasting performances. We divide the sample into the in-sample period of 1975Q2-2005Q4, and out-of-sample period of 2006Q1 and afterwards. In any econometric model assessment, the calculation of the CI is important, because it provides a quantitative sense on whether the point estimate of the coefficient and predicted path of variables are far off. Clearly, there is a large body of literature on "predictive regression" in asset pricing which have applied different techniques to construct CIs. Phillips (2014) shows through an analytical method and simulations that some commonly used methods in the literature may be misleading. In particular, Phillips (2014) indicates that “the commonly used Q test is biased towards accepting predictability and associated CIs for the regressor coefficient asymptotically have zero converge probability in the stationary case”. In light of these potential shortfalls, we conduct out-of-sample forecasting (OSF) with two different approaches. The first approach is the conventional conditional moment method. Given the estimation window of 1975Q2-2005Q4 and a forecasting horizon $h = 1,...,4$, the estimated parameters are used to forecast house and stock prices h-step ahead outside the estimation window, by using smoothed transition probabilities. The h-step ahead forecasted value of z_{t+h} based on model i and information at time t, Ω_t, is given by

$$E^i\left(z_{t+h} \mid \Omega_t\right) = \sum_{j=1}^{2} E^i\left[z_{t+h} \mid s_{t+h} = j, \Omega_t\right] \times p\left(s_{t+h} = j \mid \Omega_t\right),$$

where j is the index for the state, and $z_t \in y_t$. The estimation window is then *consecutively updated* with one observation and the parameters are re-estimated. Again the h-step ahead forecasts of house and stock prices are computed outside the new estimation window. The procedure is iterated until the final observation. The forecasts based on this method are basically the h-step ahead conditional expectations of the variable that is being predicted. Most existing (non-Bayesian) works follow this method.

The second approach is the simulation method. A merit of this approach is that we effectively simulate a CI for ourselves and hence do not need to employ other approximation techniques to construct the CI. In a sense, we avoid the critique of Phillips (2014). The idea is simple. We simulate the path of the forecasted values through repeated drawings. The procedure is as follows.

· (Step 1) We estimate the model by using the estimation window of 1975Q2-2005Q4 and obtain the parameters, transition probabilities, and variance-covariance matrix. Given the estimation results, we compute the smoothed probabilities to identify the regime at 2005Q4.

· (Step 2) Given the regime at 2005Q4, we simulate the path of h-step ahead regimes by random drawing, $h = 1,...,4$.[34] Given this particular path of h-step ahead regimes, we can obtain the path of the predicted values of $z_t \in y_t$ from (2).

· (Step 3) We iterate Steps 1 and 2 for 50,001 times to obtain the median of the h-step ahead forecasted values during 2006Q1-2006Q4 and their corresponding CIs.

We then *update* the sample with four more quarterly observations and repeat Steps 1-3, including *re-estimating* the model, in order to simulate the path of predicted values for the subsequent four quarters. This procedure is repeated until the end of our sample.

An advantage of the second approach over the first one is that this method *takes full account of the regime switching model* by determining the path of future regimes by using random drawing, rather than simply taking expectations over transition probabilities. Another advantage is that a CI is *naturally generated*, which enhances the evaluation of the forecasting performance of different models. It should be noted that the regime-switching nature of the model implies that the future forecast is path-dependent and hence the conventional way to construct CIs may not be valid.

To evaluate the performances of the in-sample and out-of-sample forecasts, we follow the literature to compute two widely-used loss functions for $z_t \in y_t$, which are the square loss function, $L\left(e^i_{t+h|t}\right) = \sum\left(e^i_{t+h|t}\right)^2$ and the absolute loss function, $L\left(e^i_{t+h|t}\right) = \sum\left|e^i_{t+h|t}\right|$, where $e^i_{t+h|t}$ denotes the h-step forecast error of model i, $e^i_{t+h|t} \equiv z_{t+h} - E^i\left(z_{t+h} \mid \Omega_t\right)$. For future reference, when we employ the square loss function as a criterion to select models, it is labelled as the "square loss criterion" (SLC). Similarly, when we use the absolute loss function as a criterion to select models, it is labelled as the "absolute loss criterion" (ALC). Clearly, the SLC tends to penalize "big mistakes" more than the ALC. As it will be made clear later, our main conclusions do not depend on which criterion is used. We will then combine these results with our newly proposed model comparison procedure, which will be explained in more detail in the following section.

---

[34] For example, suppose the regime identified at the time 2005Q5 is State 1. We use the transition probabilities p_{11} and p_{12} to generate the state at 2006Q1. Specifically, we draw a value υ from a uniform distribution $U[0,1]$. The state at 2006Q1 is State 1 if $\upsilon \in (0, p_{11})$, and State 2 otherwise. Suppose we have identified the state at 2006Q1 to be State 2, then we use the transition probabilities p_{21} and p_{22} to generate the state at 2006Q2. Therefore, we will be able to simulate the path of h-step ahead regimes.

2.6 Multi-lateral Model Comparison Procedures

On top of computing the differences in the loss functions for each model on the prediction of both assets (housing and stock), we need a statistical procedure to compare whether those differences are significant. Before we explain our procedure, it would be helpful to provide a quick review of the current practice. The existing literature of applied works adopts the Diebold-Mariano test (henceforth DM test) and its variants to assess the "relative performance" of two models in a bilateral manner.[35] While Diebold and Mariano (1995) and Zivot (2004) provide the details, it is nevertheless instructive to outline the test here, as our procedure is closely related to the DM test. In our notations, the DM test is based on the "loss differential" d_t,

$$d_t = L\left(e^1_{t+h|t}\right) - L\left(e^2_{t+h|t}\right)$$

where $L(.)$ is some loss function. Clearly, if the two models have roughly the same predictive power, the expectation of the loss differential will be zero, $E(d_t) = 0$. If, instead, Model 1 predicts better (worse) than Model 2, the expected value of the loss differential will be negative (positive).[36] In practice, d_t is unlikely to be exactly zero. The question is then how we can decide whether Model 1 is in fact "significantly better (or worse)" than Model 2. One of the contributions of the DM test is that it shows that some function of d_t follows the standard normal distribution. Hence, a test statistics can be constructed so that we can scientifically make judgment on models.

In application, researchers may need to choose among many alternatives. Typically, researchers need to repeatedly execute the DM test, and therefore the issue of ordering naturally arises. For instance, consider the case of three competing models, A, B and C. One may compare A against B first, and then compare the "winner" with C. One may also compare A against C first and then compare the winner with B. Do these two slightly different ordering deliver the same final winner? Clearly, as the number of models increase, the number of

---

[35] The DM test has been widely used in the literature. Among others, see Mariano and Preve (2012) for a review of the literature.

[36] The DM statistics will depend on $\overline{d}$, which is an average value of d_t, for different periods t, and the co-variance of d_t and d_{t-j}, $j = 1, 2, \ldots$ As shown by Zivot (2004), other things being equal, if Model 1 which consistently over-predicts in a sub-period and then consistently under-predicts in other sub-periods, it is more likely to obtain not only a lower value of d_t in different periods t, but also a higher value of co-variance between d_t and d_{t-j}, $j = 1, 2, \ldots$ As a result, Model 1 is would be classified as underperforming the alternative model. See Zivot (2004) for more details.

possible ordering significantly increases and the importance of ordering in model comparison might matter.[37]

Some recent development in the literature may help us to address this issue. Hansen et al. (2011) have developed the "model confidence set" (MCS) procedure, in which the idea is to start with a model and put that into the MCS. One then applies the DM test to compare that model with an alternative. If both have the same predictive power in the sense that the "loss differential" d_t is smaller than a certain critical value, then both models will be kept in the MCS; otherwise, only the one with high predictive power will stay in the MCS. We repeat the procedure until we exhaust all of our models. The models in the remaining MCS will therefore have the same predictive power, and by construction, they are better than the models that are not selected into the MCS. Hence, while we still compare models in a bilateral manner, we can still compare any finite set of models.

Mariano and Preve (2012), on the other front, generalize the idea of the DM test and *simultaneously compare several models (MP test)*. The idea is simple. Consider the situation with $(K+1)$ models. We then define the " j -th loss differential" $d_{j,t}$

$$d_{j,t} = L\left(e^{j}_{t+h|t}\right) - L\left(e^{j+1}_{t+h|t}\right),\ j = 1,\dots,K$$

where $L(.)$ continues to denote some loss function. We collect these loss differentials in a vector, $d_t = \left\{d_{j,t}\right\},\ j = 1,\dots,K$. We then take its average,

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

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

In this paper, we differentiate competing theories of asset prices, which are represented by different econometrics models. Therefore, it is natural to use the MP test rather than the DM test. Moreover, we need to define a procedure which

---

[37] In fact, the situation is analogous to the old Condorcet Paradox, in which the candidate who wins in every pair-wise situation may not be the winner when all candidates can be selected at one time. In the present context, it means that the order of the comparison of the models would actually affect the final outcome. For more discussion, see Austen-Smith and Banks (2000), among others.

enables us to categorize the models into different "equivalent classes", each of which contains a model with (statistically speaking) the same predictive power. The procedure needs to be "robust" in the sense that the final outcome (i.e. the ranking of different models) would not be affected by the ordering of the models that are first compared. Our procedures are similar to those of Hansen et al. (2011) and the following are the steps.

1. We consider N models that make predictions on the same economic variable. We first rank the models in accordance with a criterion, such as the SLC. Without loss of generality, we assume that according to the chosen criterion, the predictive performance of Model 1 is better than that of Model 2, which in turn, is better than that of Model 3, and so on.

2. We conduct an 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 in accordance with the chosen criterion.

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 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 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 ends. 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 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 ends.

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 in total g classes, then it must be that $N = N_1 + N_2 + ... + N_g$.

10. We repeat the whole process with an alternative model evaluation criterion, e.g. SLC instead of ALC.

Several remarks are in order. First, by definition, N_i, $i = 1, 2, ...$ must be positive. In other words, each "Equivalent Predictive Power Class" (EPPC) is non-empty. In addition, since the models are first ranked in accordance with a given criterion, our elimination procedure is easy to implement. Second, by construction, all models within the same class have the same predictive power, and any model in Class i will always have higher predictive power than any model in Class j, for any i, j, such that $i < j$. Third, it is possible that for the same set of N models, the ranking will vary as the criterion changes (say, from SLC to ALC), and hence the model can be categorized differently. In other words, *the classification of EPPC is criterion-dependent*. In the case that the same set of N models have predictions on several economic variables (for instance, stock and housing returns in this paper), it is also possible that the ranking of models also varies with the variable that we would like to predict. In other words, *EPPC is also variable-dependent*. Note that this procedure still employs conventional criterion, such as the SLC and ALC, and hence facilitates a comparison with the literature. On the other hand, our procedure allows us to compare a large number of competing models based on "fair grounds". As computers become increasingly powerful and data availability improves over time, we believe that comparison among a large number of alternative models may be inevitable and the procedure that we propose here can facilitate such a comparison. As we present our empirical results, these features will become clear.

3. Estimation Results

Limited by data availability, we keep the model as parsimonious as possible. The details of the estimation results for the *whole sampling period* of *1975Q2-2012Q1* are presented in the Appendix, and Table 4a provides a summary. In general, a model that allows for regime switching attains a lower of Akaike information criterion (AIC) value and a higher log-likelihood value. Among all these models, the regime switching model, Model G (EFP, TED, SRET, HRET), has the best goodness of fit, i.e., a significantly lower AIC value than the other models, thus suggesting that credit market frictions and asset returns are indeed significantly inter-related.

For the Markov switching model, recall that we set the volatility at Regime 1 to unity, $\lambda_j(1) = 1$, thus the element $\lambda_j(2)$ measures the relative volatility of Regime 2 over Regime 1. In the Appendix, the figures show that the estimated values of relative volatility $\lambda_j(2)$ are all significantly less than one for $j=1$ and 2, which means that for both federal funds rate and the spread, the volatility in Regime 2 is lower than that in Regime 1. On the other hand, almost all of the $\lambda_3(2)$ and $\lambda_4(2)$ are insignificant, thus suggesting that *for the quarterly stock and housing returns there is no significant difference in volatility across regimes*. Thus, we identify two regimes for this monetary policy tool: a high volatility regime (Regime 1) and a low volatility regime (Regime 2). Table 4b provides a summary of the estimated transition probabilities. It is clear that the regimes are highly persistent, regardless of the models. In particular, most models suggest p_{11} to be close to 0.80 and all models suggest p_{22} to be higher than 0.93. They imply that the expected duration of Regime 1 to be around 1/(1-0.8)= 5.0 quarters and that for Regime 2 is not less than 1/(1-0.93)= 14. 3 quarters.

Table 4a Summary of Goodness of Fit for All Eight Models (1975Q2-2012Q1)

	Model	AIC	SBC
Model A	Single-regime model (FFR, SPR, TED, EFP, GDP, SRET, HRET)	12.1273	13.8284
Model B	Two-regime model (FFR, GDP, SRET, HRET)	13.8681	15.1237
Model C	Two-regime model (FFR, SPR, SRET, HRET)	12.7748	14.0304
Model D	Two-regime model (FFR, EFP, SRET, HRET)	11.0063	12.2619
Model E	Two-regime model (FFR, TED, SRET, HRET)	11.8574	13.1130
Model F	Two-regime model (EFP, SPR, SRET, HRET)	10.6150	11.8706
Model G	Two-regime model (EFP, TED, SRET, HRET)	9.2656	10.5212
Model H	Two-regime model (SPR, TED, SRET, HRET)	11.7468	13.0024
Model USB	Univariate benchmark for a single-regime for (SRET)	7.0774	7.1381
Model USB	Univariate benchmark for a single-regime for (HRET)	2.7589	2.8196

Note: AIC refers to the Akaike information criterion. SBC refers to the Schwartz Bayesian information criterion.

Given the estimated parameters, transition probabilities, and variance-covariance matrices, we estimate the classification of regimes under different models and report the results in Table 5. Basically, these models show similar classifications of the regimes. For periods identified as Regime 1, all of the models include the aftermath of the second oil crisis and the appointment of

Paul Volcker as chairman of the Federal Reserve.[38] Interestingly, when the TED spread (*TED*) is included in Models E, G and H, Regime 1 also includes the stock market crash in 1987, thus suggesting that *TED* picked up the volatility in the credit market after the stock market crash. It is also interesting that under Model E, which is like Model D except that EFP is replaced by the TED spread, the changes in regimes are much more frequent. In general, *models that involve TED experience more regime switching*, thus suggesting higher variability of the risk premium faced by financial intermediations. We also compute the smoothed probabilities for all from Models B to H, as shown in Figure 2. The figure shows the probabilities of the economy being in Regime 1 (high volatility regime) at a given period. Since there are only 2 regimes, the probabilities of being in Regime 2 would be suppressed.

Table 4b Estimated Persistence of Regimes among Models (1975Q2-2012Q1)

	p_{11}	p_{22}
Model B	0.9427	0.9920
Model C	0.9505	0.9914
Model D	0.8060	0.9643
Model E	0.7824	0.9386
Model F	0.8226	0.9492
Model G	0.7952	0.9370
Model H	0.8478	0.9447

Table 5 Identified periods of Regime 1

Model	**Regime 1**		
Model B	1978Q2-1982Q4		
Model C	1979Q4-1986Q2		
Model D	1975Q2-1975Q3	1979Q4-1982Q4	1984Q1-1984Q4
	2001Q1-2002Q1	2008Q4	
Model E	1975Q2-1976Q2	1978Q2-1978Q4	1979Q3-1982Q4
	1984Q2-1984Q4	1987Q2-1987Q4	
	2007Q3-2008Q4	2009Q2	
Model F	1975Q2	1980Q2-1986Q1	2001Q1-2002Q1
	2008Q4	2009Q2-2009Q3	2011Q3-2011Q4
Model G	1975Q2	1976Q2	1978Q4
	1979Q3-1984Q4	1987Q2-1987Q4	2007Q3-2008Q4
	2009Q2-2009Q3		
Model H	1975Q2-1976Q2	1978Q2-1982Q3	1984Q2
	1987Q2-1988Q3	2001Q1-2001Q4	2007Q3-2009Q2

[38] Among others, Goodfriend and King (2005) and Goodfriend (2007) provide a summary of the history of the monetary policies during that period.

Figure 2a Smoothed Probabilities for Model B (FFR, GDP,SRET,HRET) (only the probability of Regime 1 shown)

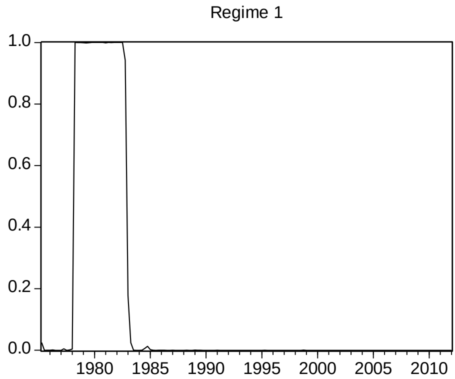


Figure 2b Smoothed Probabilities for Model C (FFR,SPR,SRET,HRET)

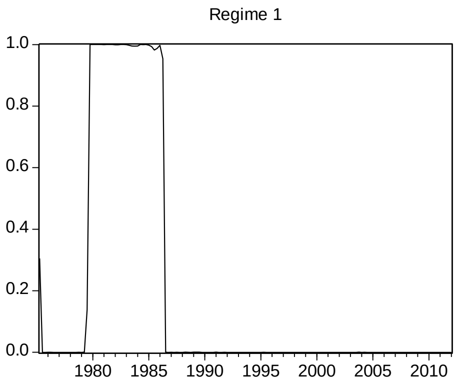


Figure 2c Smoothed Probabilities for Model D (FFR,EFP,SRET,HRET)

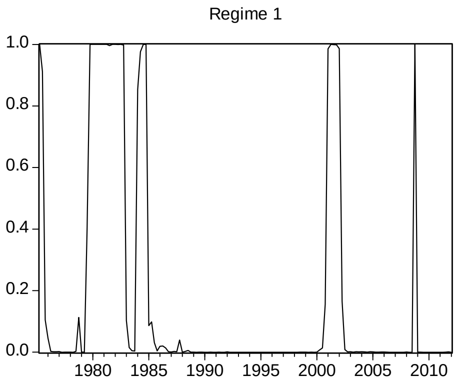


Figure 2d Smoothed Probabilities for Model E (FFR,TED,SRET,HRET)

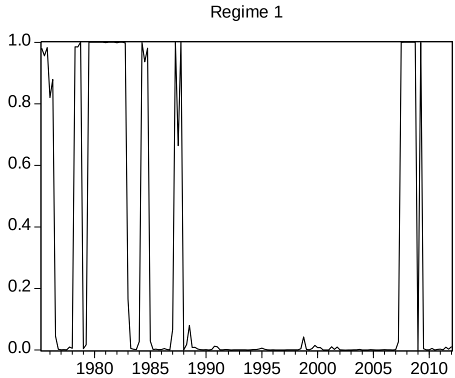


Figure 2e Smoothed Probabilities for Model F (EFP,SPR,SRET,HRET)

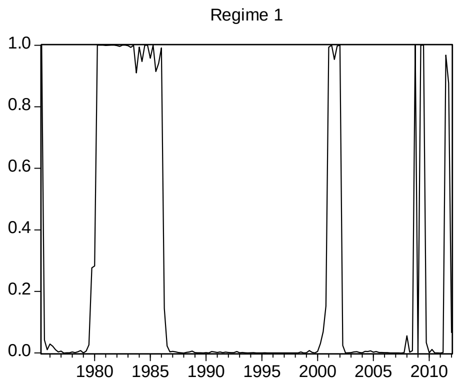


Figure 2f Smoothed Probabilities for Model G (EFP,TED,SRET,HRET)

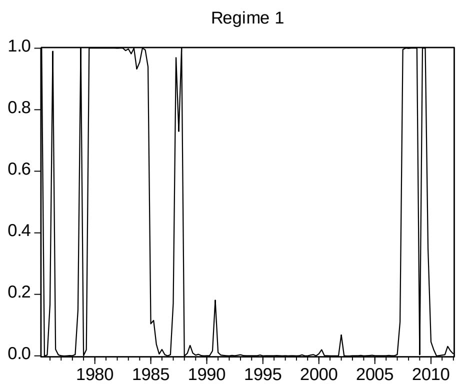
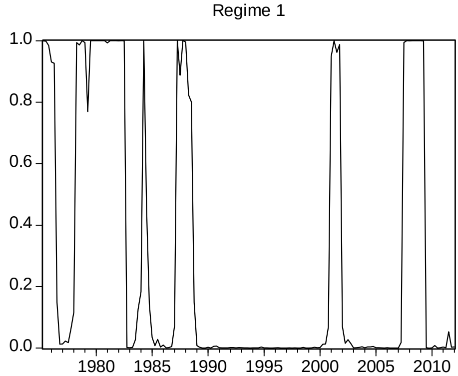


Figure 2g Smoothed Probabilities for Model H (SPR,TED,SRET,HRET)



4. Forecasting

We now proceed to forecast the stock and housing returns. As discussed above, we first conduct ISF for the period of 1975Q2-2005Q4 and then examine the out-of-sample forecasts for the period of 2006Q1-2011Q4, by using the expectations-based and simulation-based methods respectively.

4.1 In-Sample Forecasting

We compute the loss functions based on the SLC and ALC of the in-sample h -step ahead forecasts, $h = 1, ..., 4$, for each variable across all of the models. Several findings are in order. First, as shown in Table 6a, the in-sample forecasts of asset returns are mixed. Model C (FFR, SPR, SRET, HRET) has the best performance for the stock returns. For housing return, however, Model E (FFR, TED, SRET, HRET) outperforms all of the others. Note that both models contain the monetary policy variable, FFR. This is true whether we use the SLC or ALC. On the other hand, it seems that the performances in predicting stock returns across the models are similar. We therefore implement our multi-lateral model comparison procedures and attempt to categorize models into different EPPCs.[39] Our intuition is confirmed by the results shown in Table 6b. Whether we use the SLC or ALC, *we find that all models have the same predictive power in terms of explaining the stock return during the in-sample period*. In particular, no model has a more superior performance than the AR(1) process, which is the USB. This is consistent with the notion that the stock market is very efficient in reflecting all of the relevant information so that adding other variables into the statistical model does not provide any extra predictive power.

---

[39] As a robustness check, we also bilaterally use the conventional DM test and obtain very similar results. The details are in the Appendix.

Table 6a A Summary of In-sample Forecasting Performances (4-Quarter Ahead Forecasts) (1975Q2-2005Q4)

		Stock Return		Housing Return	
		SLC	**ALC**	**SLC**	**ALC**
Model A	Single-regime model (FFR, SPR, TED, EFP, GDP, SRET, HRET)	57.5866	5.6686	0.7577	0.6488
Model B	Two-regime model (FFR, GDP, SRET, HRET)	57.8042	5.6453	0.8304	0.7053
Model C	Two-regime model (FFR, SPR, SRET, HRET)	*57.1605*	*5.6188*	0.7356	0.6691
Model D	Two-regime model (FFR, EFP, SRET, HRET)	58.7039	5.6662	0.7013	0.6364
Model E	Two-regime model (FFR, TED, SRET, HRET)	58.2775	5.6553	*0.6980*	*0.6204*
Model F	Two-regime model (EFP, SPR, SRET, HRET)	60.6847	5.7425	0.7331	0.6502
Model G	Two-regime model (EFP, TED, SRET, HRET)	59.0501	5.7707	0.8295	0.6727
Model H	Two-regime model (SPR, TED, SRET, HRET)	58.2623	5.6366	0.7308	0.6378
USB	Univariate benchmark for a single-regime	57.9966	5.6881	0.8400	0.6726

Note: SLC: Square Loss Criteria; ALC: Absolute Loss Criteria.

Table 6b A Summary of EPPC (In-sample 4-Quarter Ahead Forecasts)

		Class 1	Class 2	Class 3
Stock Return	SLC	All models	/	/
	ALC	All models	/	/
Housing Return	SLC	D, E, F, H	A, C, G	B, USB
	ALC	A, C, D, E, F, H	B, G, USB	/

Note: SLC: Squared Loss Criterion; ALC: Absolute Loss Criterion. Our convention is that if $i > j$, then any model in Class i has less predictive power than any model in Class j.

The situation is very different for housing return. Several observations are in order. First, whether we use the SLC or ALC, Models D, E, F, and H have the same predictive power and are always in Group 1, which means that they are at least as good as the other models. Second, whether we use the SLC or ALC, Model B (FFR, GDP, SRET, HRET) and USB are always in the lowest group. This suggests that *using the monetary policy (FFR) and economic growth rate alone are not sufficient for understanding the housing market, at least for the in-sample period* (1975Q2-2005Q4). *Alternatively, we may say that the information on monetary policy and economic growth has been reflected in the housing return itself*. The fact that AR(1) (which is also USB for the housing return) is always inferior to the four "optimal models" (Models D, E, F, and H) suggests that there is *important information in the financial market that enhances our ability to account for the housing market*. Note, however, that such "cross-market informational spillover" is very subtle. Table 6c also shows that Model G (EFP, TED, SRET, HRET) is always inferior to the four optimal models (Models D, E, F, and H). However, Table 3 shows that Model F is simply Model G with TED spread replaced by the *SPR*, and Model H is simply Model G with the *EFP* replaced by the *SPR*. Does this mean that the SPR is crucial for understanding the housing return during that period? This does not seem to be the case, as both Models D and E, which have the same predictive power as Models F and H, do not contain SPR. Note that the financial variables are correlated and hence some other variables may also contain the information that is relevant for predicting future housing returns. To summarize, our estimations indicate that the *"EMH" does not apply to the housing market during the in-sample period* (1975 to 2005). The data are more consistent with models that emphasize on an imperfect capital market, such as those found in Christiano, Motto and Rostagno (2007), Davis (2010), Jin et al. (2012), among others.

4.2 Out-of-Sample Forecasting via Conditional-Expectation Estimation

We now turn to the OSF of housing and stock returns in 2006Q1, a time when the growth of housing returns began to decline and the sub-prime crisis started to unfold. By following the literature, we first conduct OSF by using conditional-expectations predictions. The Appendix provides details of the out-of-sample h -step ahead forecasts, $h=1,...,4$, for each variable across all models. Tables 6c summarizes the results. In terms of forecasting stock returns, Model H (SPR, TED, SRET, HRET) performs better than the other models, in terms of both the SLC and ALC. In terms of forecasting housing returns, Model D (FFR, EFP, SRET, HRET) performs better than the other models, also in terms of both the SLC and ALC. Naturally, we ask whether the difference is statistically significant. Again, we adopt the same procedure and categorize the models into different EPPCs. Table 6d reports the results, and several observations are in order. In terms of stock return forecasting, we find that: (1) Models B, E, H and USB are always in Group 1, which means that they are at least as good as the other models, and (2) Models D and F are always in the lower group. This is true whether we use the SLC or ALC. The ranking of Models A, C, and G will depend on which criterion is used, thus implying that there are models (especially D and F) which underperform the USB, which is the simple AR(1), in the OSF of stock return. Recall that for the ISF, by using the same set of procedures and same set of models, we have found that all of the models have the same predictive power. In other words, *some models have actually deteriorated relative to the USB* (i.e. AR(1)) *in terms of the ability to predict the stock return*. Note that both Models D and F involve EFP, and this may suggest that the *ability of EFP to track the aggregate stock return after a crisis may not be as good as before the crisis*.

The case of OSF for housing return is perhaps equally interesting. Table 6d clearly shows that regardless whether we use the SLC or ALC, (1) Models C, D, F, G, H and USB are always in Group 1, which means that they have better predictive power, and (2), Models A, B, and E are always in Group 2, which means that their predictive powers are not as good. Note that for the ISF, whether we use the SLC or ALC, Model B (FFR, GDP, SRET, HRET) and USB are always in the lowest group. For OSF, Model B remains in the lowest group, yet the USB is "promoted" to the higher group. This means that whether for the linear VAR with 7 variables (Model A), or a regime-switching VAR, *no model on our list can outperform the simple USB in terms of the OSF of the housing return*. Similar to the case of stock return forecasting, this suggests that some models have deteriorated in terms of forecasting the housing return, at least relative to the simple AR(1) process. In particular, Models A and B are the only models that involve GDP and yet they are always in Group 2, which *suggests that GDP may not be as useful in predicting the house price as before*. Clearly, the model comparison here is far from conclusive and future research can revisit the issue with more rigorous tools.

Table 6c A Summary of Out-of-Sample Forecasting Performances (4-Quarter Ahead Forecasts) (2006Q4-2012Q1)

		Stock Return		Housing Return	
		SLC	ALC	SLC	ALC
Model A	Single-regime model (FFR, SPR, TED, EFP, GDP, SRET, HRET)	126.3311	9.4671	4.8730	1.8844
Model B	Two-regime model (FFR, GDP, SRET, HRET)	119.1706	9.0940	4.8658	1.8975
Model C	Two-regime model (FFR, SPR, SRET, HRET)	122.2287	9.3329	3.7581	1.6619
Model D	Two-regime model (FFR, EFP, SRET, HRET)	127.3973	9.5067	*3.6938*	*1.6136*
Model E	Two-regime model (FFR, TED, SRET, HRET)	115.8617	8.8073	5.0151	1.9273
Model F	Two-regime model (EFP, SPR, SRET, HRET)	133.7992	9.6332	3.8359	1.6712
Model G	Two-regime model (EFP, TED, SRET, HRET)	127.3231	8.8930	4.0599	1.7108
Model H	Two-regime model (SPR, TED, SRET, HRET)	*109.6158*	*8.5835*	4.2041	1.7652
USB	Univariate benchmark for a single-regime	113.6018	8.6529	4.5453	1.8221

Note: SLC: Square Loss Criteria; ALC: Absolute Loss Criteria.

Another interesting observation is that, perhaps the *forecasting ability of a model has become more asset-specific for the out-of-sample forecast.*[40] For

[40] Interestingly, this is also the case for structural model comparison. See Kwan et al. (2015), among others, for more details.

instance, while Models D and F are always the inferior models for stock return OSF, they are always in Group 1 for the housing return forecasting. Similarly, while Models B and E are always the "better" models in terms of the OSF for the stock return, they are always the "not-as-good" models in terms of the OSF for housing return. Future research may further investigate this phenomenon of asset-dependent forecasting performance.

Table 6d Summary of EPPC (Out-Of-Sample 4-Quarter Ahead Forecasts)

		Class 1	Class 2
Stock Return	SLC	A, B, C, E, H, USB	D, F, G
	ALC	B, E, G, H, USB	A, C, D, F
Housing Return	SLC	C, D, F, G, H, USB	A, B, E
	ALC	C, D, F, G, H, USB	A, B, E

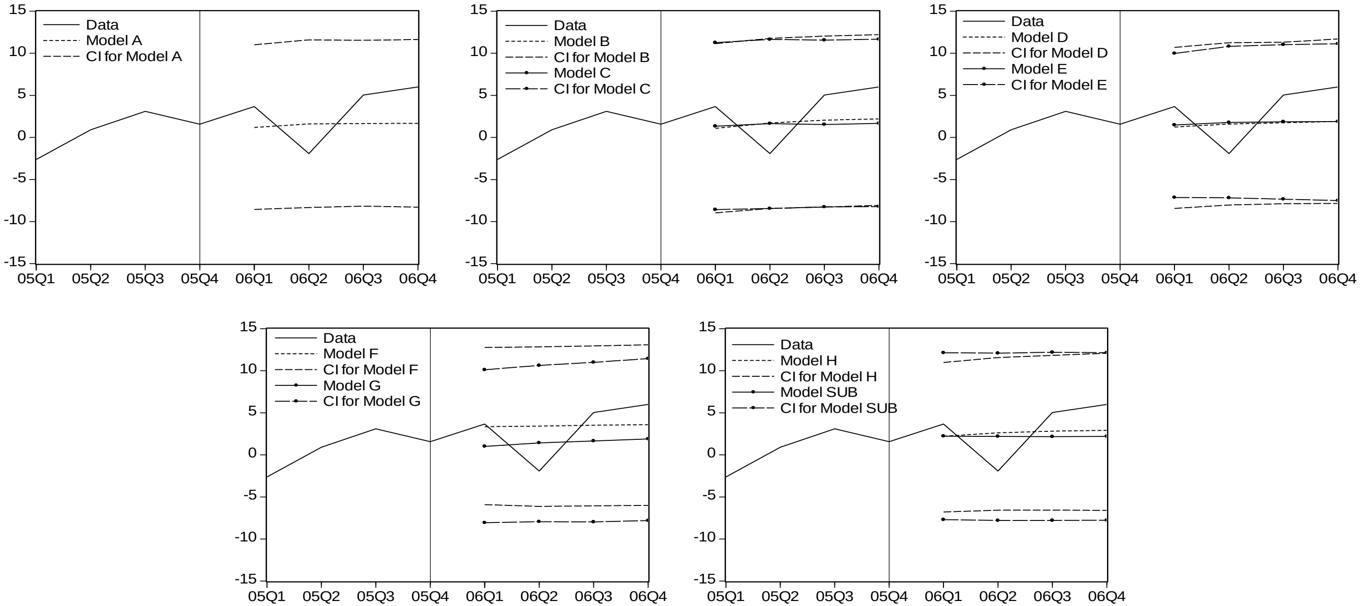
Note: SLC: Square Loss Criterion; ALC: Absolute Loss Criterion. Our convention is that if $i > j$, then any model in Class i has less predictive power than any model in Class j.

4.3 Out-of-Sample Forecasting via Simulation

The results presented in the current section differ in at least two important dimensions from the results presented in the previous section. First, the previous section only provides information on the "relative performance" of different models, as we use different statistical tools and procedures to assess whether some models have more superior predictive power than others. In this section, we assess the "absolute performance" of the different models by using simulation-based forecasting. Second, we aggregate the forecasting performance of each model during the whole out-of-sample period (2006 and afterwards) into some statistics and then compare across the models in the previous section. In the current section, we will compare the forecasting performance of different periods in each year, and then allow the model to be re-estimated with updated data, and then compare again in the subsequent year. Thus, we allow models to "learn and improve" and would like to see which model(s) are more successful in adjusting the parameter with new data and hence provide more accurate forecasting over time.

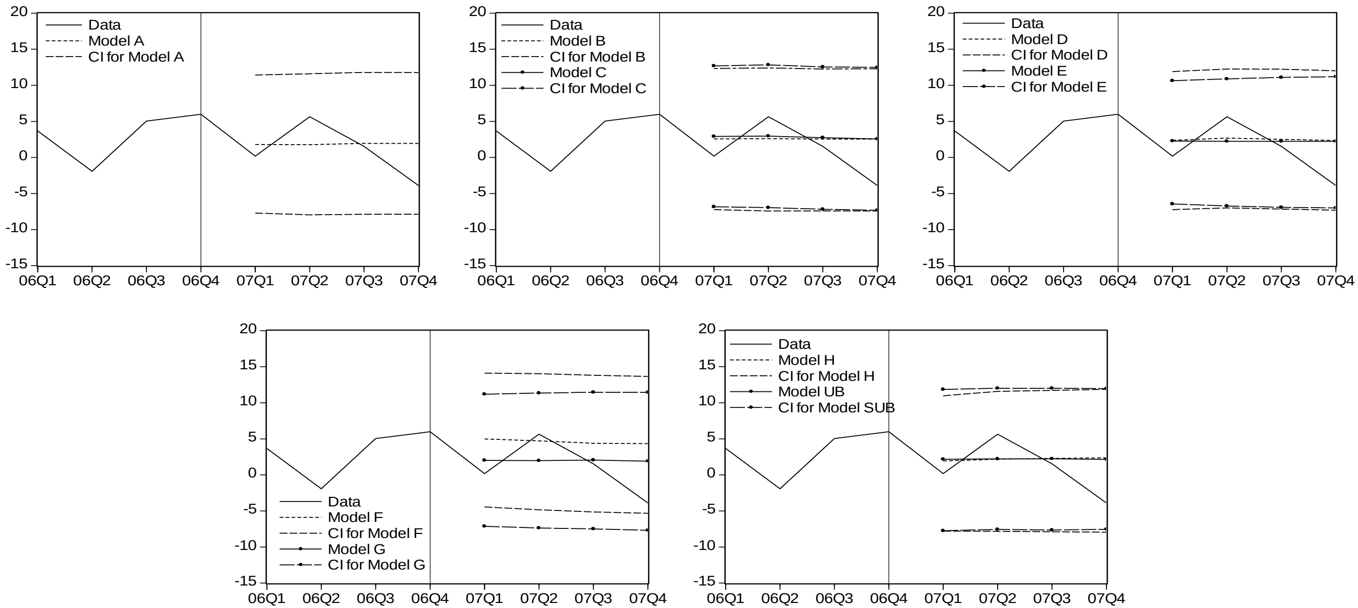
We specifically consider a forecasting window of 4 quarters that starts at 2006Q1, with h-quarter ahead forecasts, $h = 1, \ldots, 4$. After simulating the out-of-sample path of 2006Q1-2006Q4 based on observations up to 2005Q4, the data are updated with four observations and the parameters are re-estimated. The procedure is repeated until we have updated the sample to include all observations from 1975 to 2010 to predict the asset returns in 2011. The purpose of this exercise is to see how the performances of the models change when information is updated. The simulated paths together with their 80-percent CIs are illustrated in Figure 3 for stock return and Figure 4 for housing return. Tables 7a and 7b provide a summary of the performance of the different models.

Figure 3a Simulation-Based Out-of-Sample Forecasts of Stock Returns with 80-Percent Confidence Interval from 2006Q1-2006Q4 based on Information available at 2005Q4



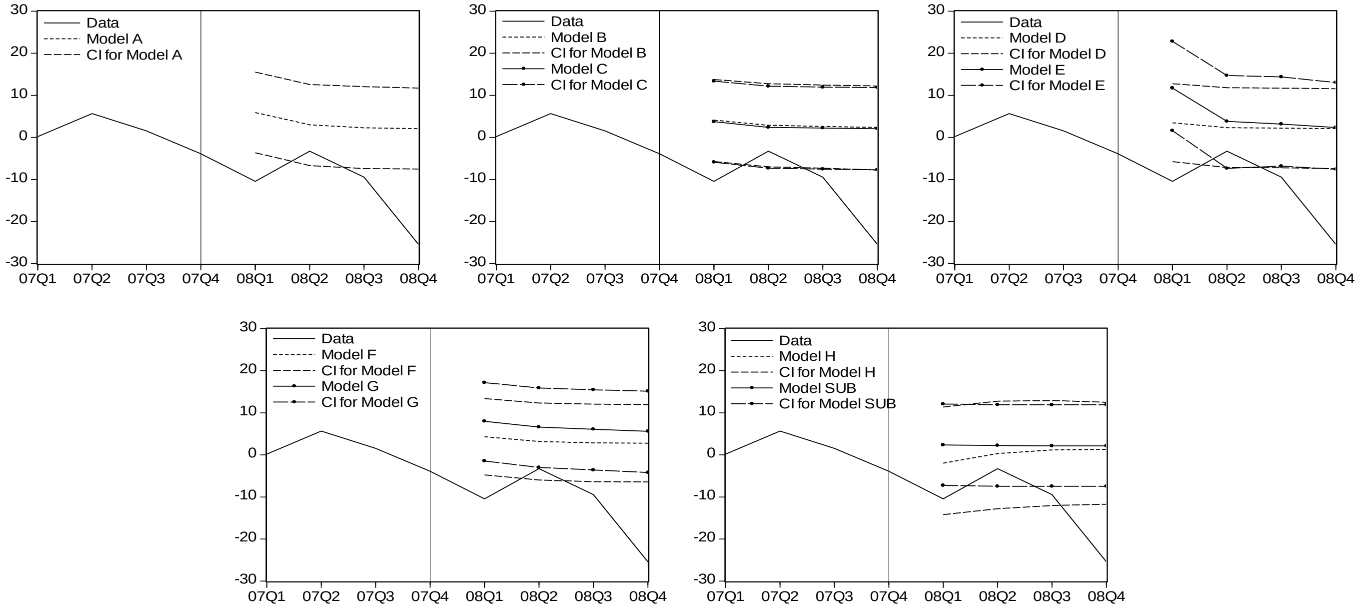
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 3b Simulation-Based Out-of-Sample Forecasts of Stock Returns with 80-Percent Confidence Interval from 2007Q1-2007Q4 based on Information available at 2006Q4



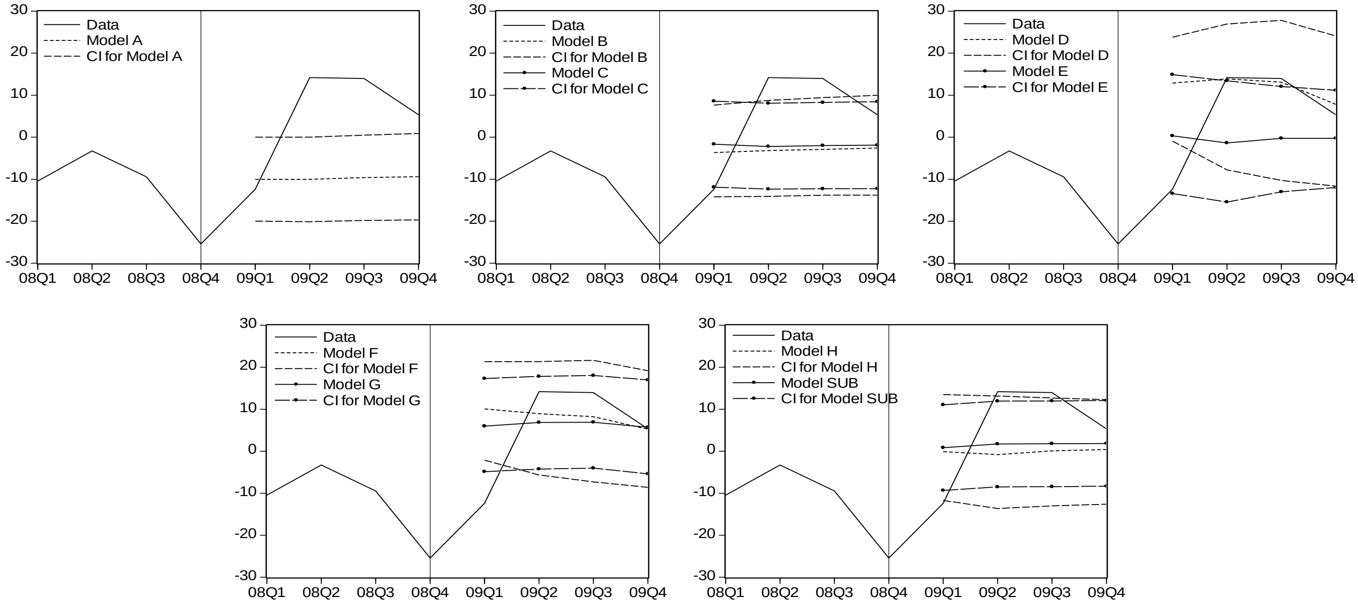
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 3c Simulation-Based Out-of-Sample Forecasts of Stock Returns with 80-Percent Confidence Interval from 2008Q1-2008Q4 based on Information available at 2007Q4



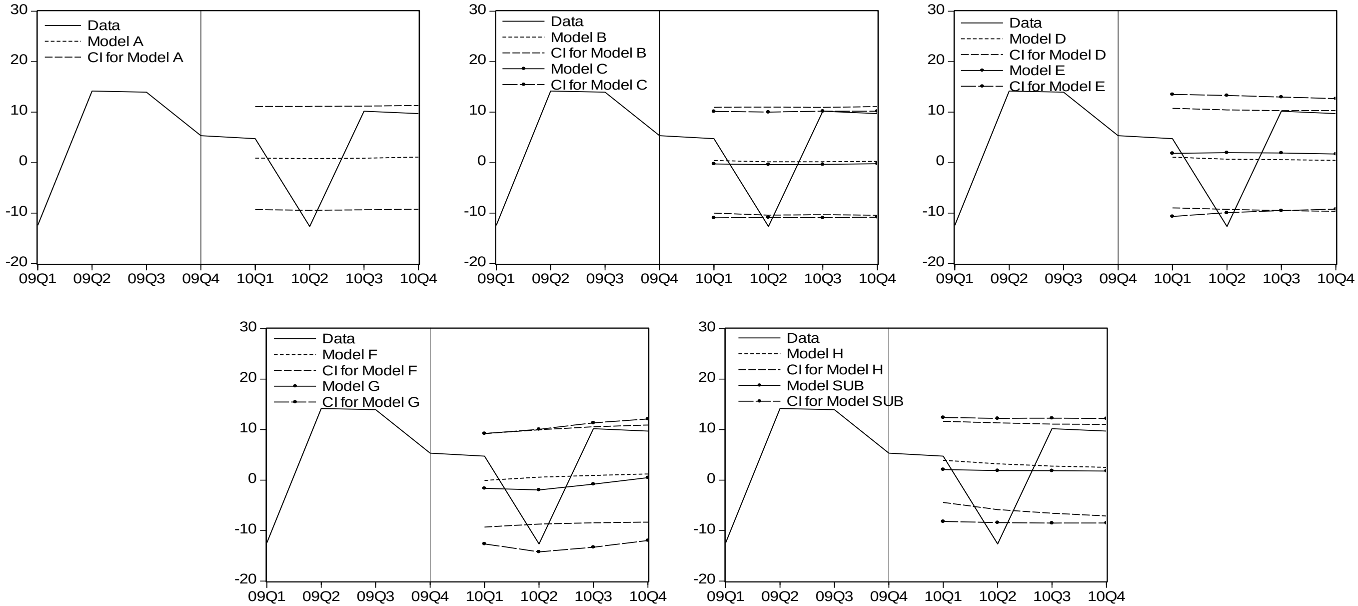
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 3d Simulation-Based Out-of-Sample Forecasts of Stock Returns with 80-Percent Confidence Interval from 2009Q1-2009Q4 based on Information available at 2008Q4



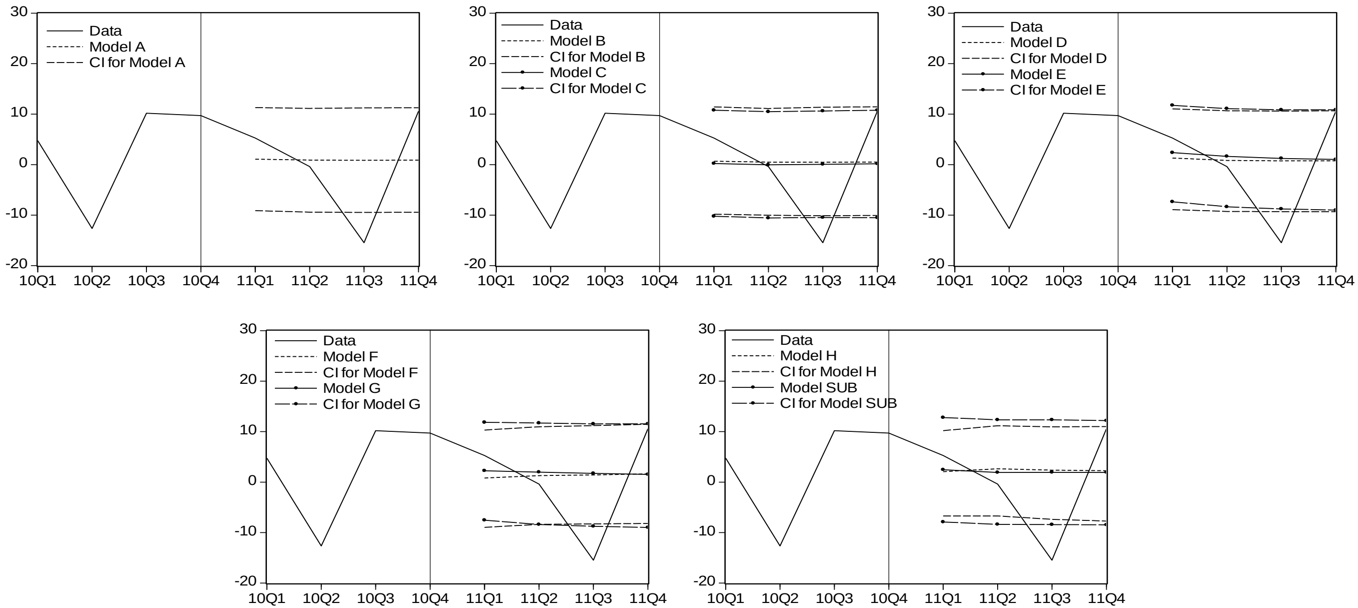
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 3e Simulation-Based Out-of-Sample Forecasts of Stock Returns with 80-Percent Confidence Interval from 2010Q1-2010Q4 based on Information available at 2009Q4



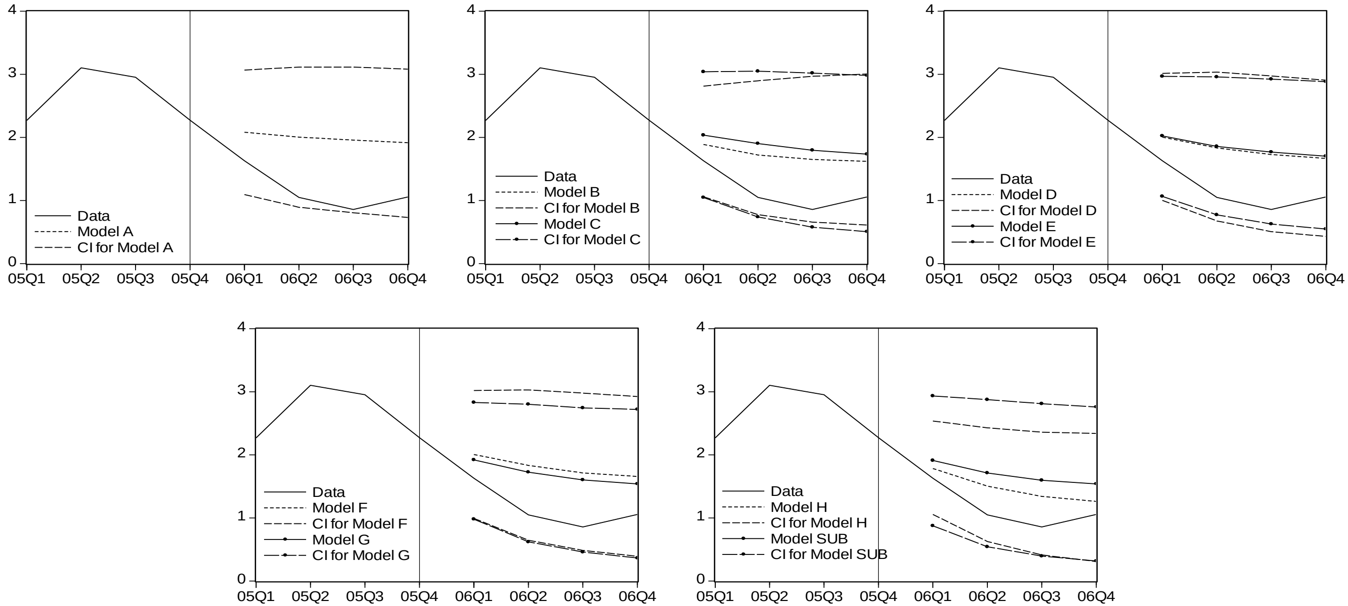
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 3f Simulation-Based Out-of-Sample Forecasts of Stock Returns with 80-Percent Confidence Interval from 2011Q1-2011Q4 based on Information available at 2010Q4



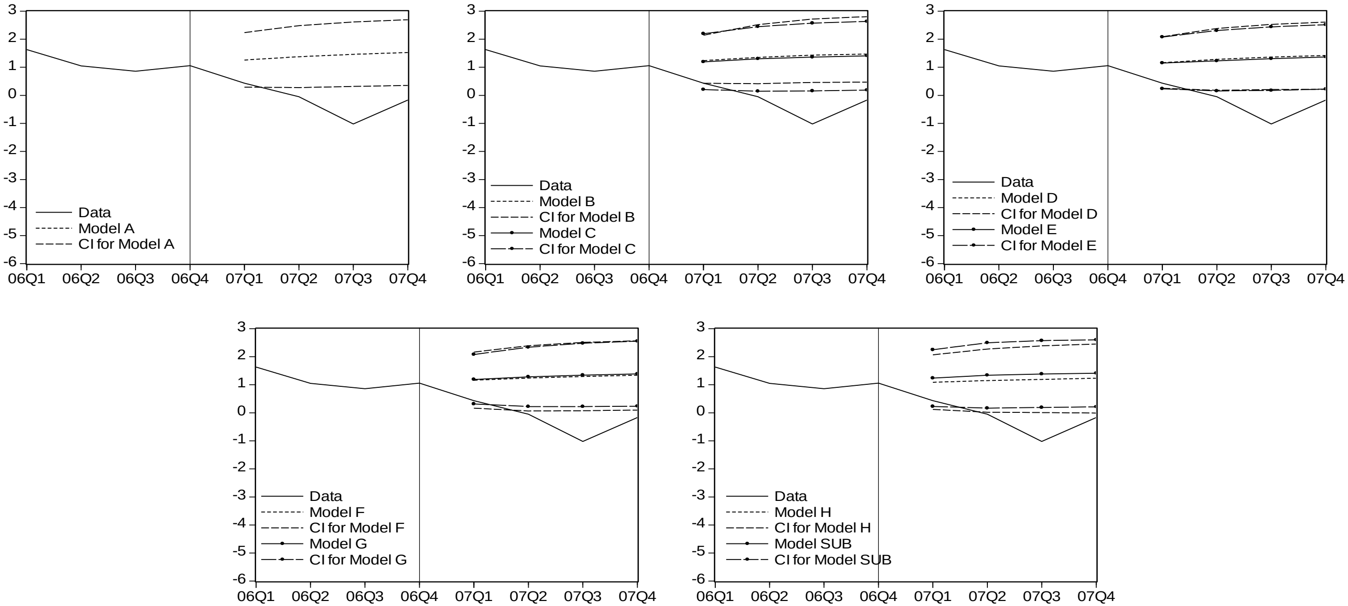
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 4a Simulation-Based Out-of-Sample Forecasts of Housing Returns with 80-Percent Confidence Interval from 2006Q1-2006Q4 based on Information available at 2005Q4



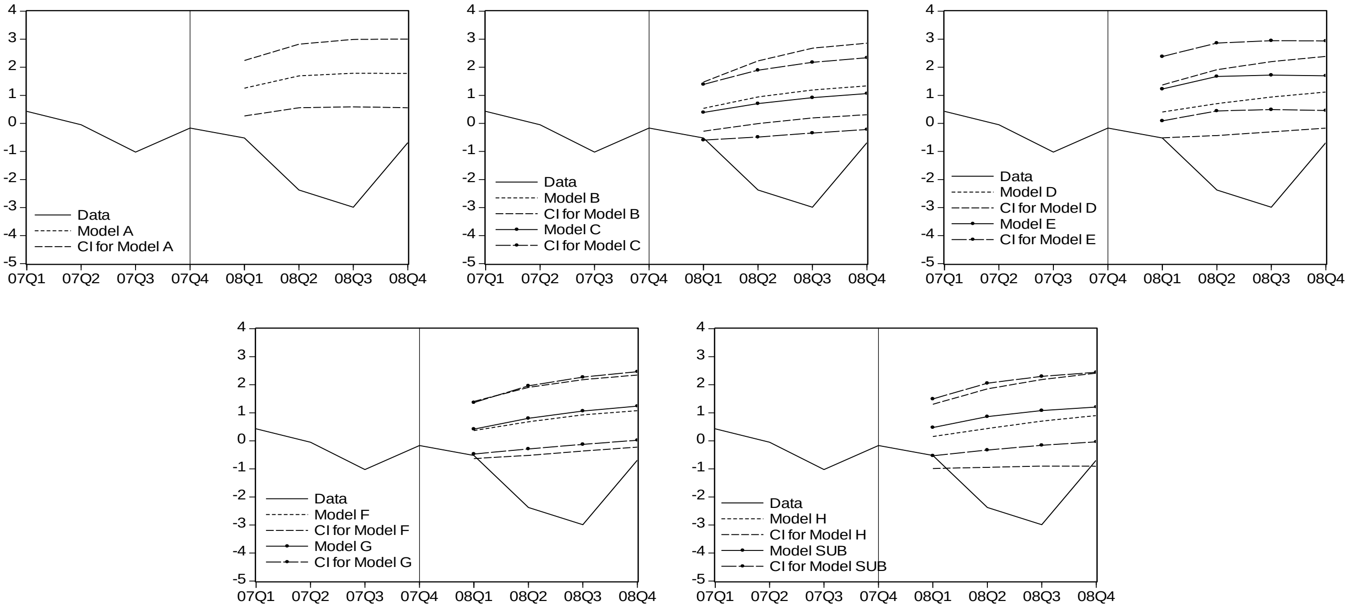
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 4b Simulation-Based Out-of-Sample Forecasts of Housing Returns with 80-Percent Confidence Interval from 2007Q1-2007Q4 based on Information available at 2006Q4



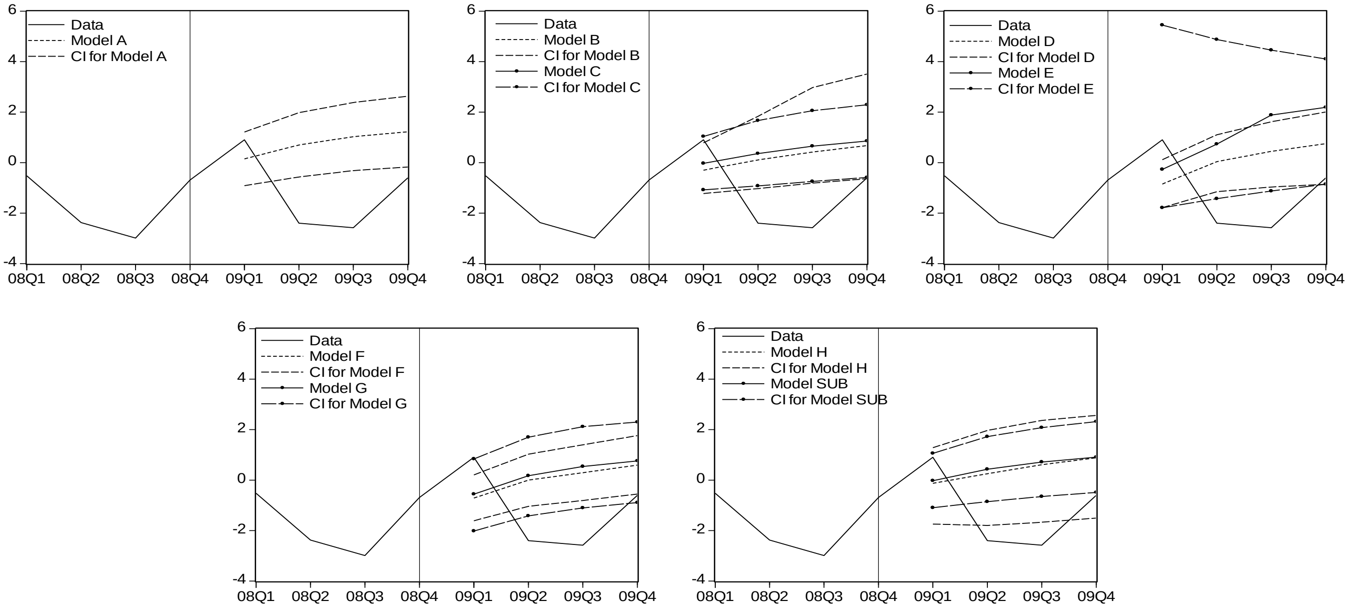
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 4c Simulation-Based Out-of-Sample Forecasts of Housing Returns with 80-Percent Confidence Interval from 2008Q1-2008Q4 based on Information available at 2007Q4



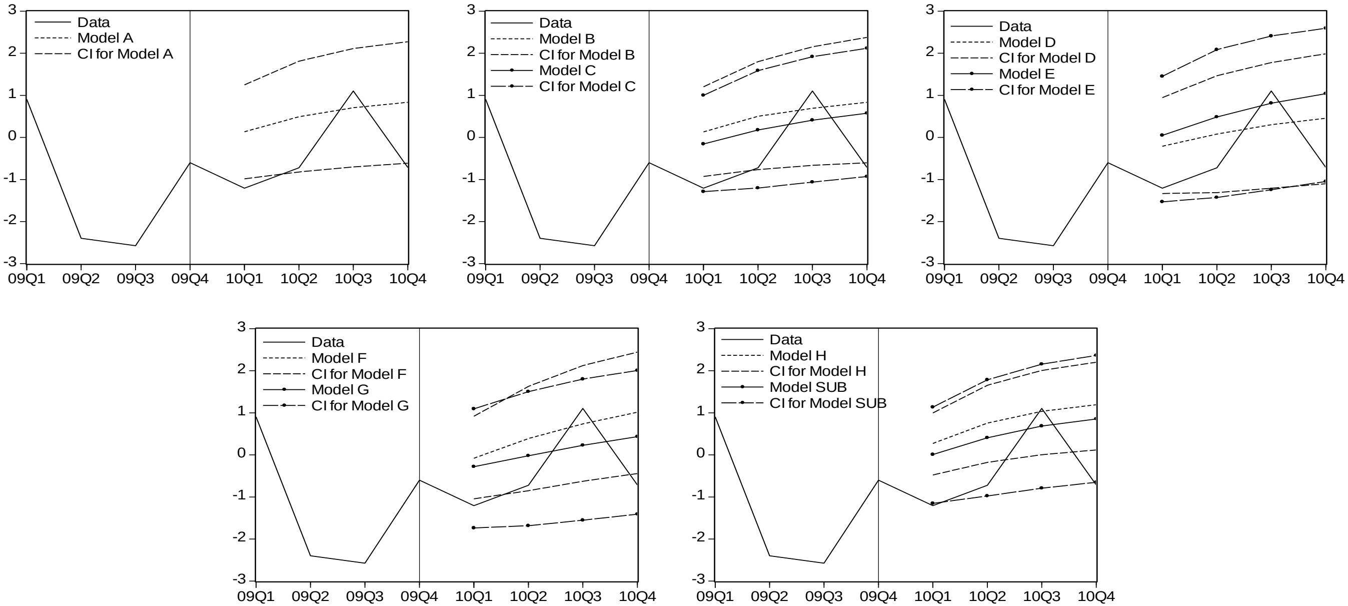
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Univariate benchmark for a single-regime (AR(1))

Figure 4d Simulation-Based Out-of-Sample Forecasts of Housing Returns with 80-Percent Confidence Interval from 2009Q1-2009Q4 based on Information available at 2008Q4



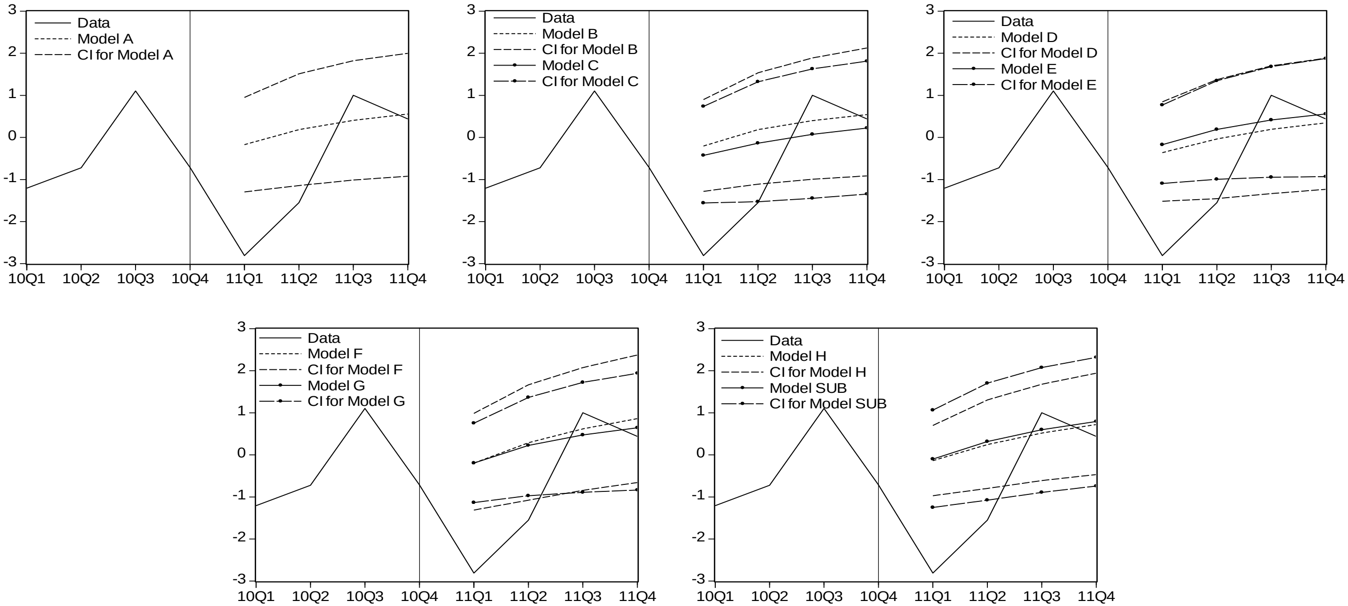
Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 4e Simulation-Based Out-of-Sample Forecasts of Housing Returns with 80-Percent Confidence Interval from 2010Q1-2010Q4 based on Information available at 2009Q4



Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Figure 4f Simulation-Based Out-of-Sample Forecasts of Housing Returns with 80-Percent Confidence Interval from 2011Q1-2011Q4 based on Information available at 2010Q4



Note: Model A: Single-Regime (FFR,SPR,TED,EFP,GDP,SRET,HRET); Model B: Two-Regime (FFR,GDP,SRET,HRET); Model C: Two-Regime (FFR,SPR,SRET,HRET); Model D: Two-Regime (FFR,EFP,SRET,HRET); Model E: Two-Regime (FFR,TED,SRET,HRET); Model F: Two-Regime (EFP,SPR,SRET,HRET); Model G: Two-Regime (EFP,TED,SRET,HRET); Model H: Two-Regime (SPR,TED,SRET,HRET); Model SUB: Univariate benchmark for a single-regime (AR(1))

Table 7a Is the Forecasted Stock Return within The 80% Confidence Interval?

Model	Predicting 2006 based on 1975-2005	Predicting 2007 based on 1975-2006	Predicting 2008 based on 1975-2007	Predicting 2009 based on 1975-2008	Predicting 2010 based on 1975-2009	Predicting 2011 based on 1975-2010
Model A	Yes	Yes	Partly	Partly	Partly	Partly
Model B	Yes	Yes	Partly	Partly	Partly	Partly
Model C	Yes	Yes	Partly	Partly	Partly	Partly
Model D	Yes	Yes	Partly	Partly	Partly	Partly
Model E	Yes	Yes	Partly	Partly	Partly	Partly
Model F	Yes	Yes	Partly	Partly	Partly	Partly
Model G	Yes	Yes	*No*	Partly	Yes	Partly
Model H	Yes	Yes	Partly	Partly	Partly	Partly
USB	Yes	Yes	Partly	Partly	Partly	Partly

Table 7b Is The Forecasted Housing Return within The 80% Confidence Interval?

Model	Predicting 2006 based on 1975-2005	Predicting 2007 based on 1975-2006	Predicting 2008 based on 1975-2007	Predicting 2009 based on 1975-2008	Predicting 2010 based on 1975-2009	Predicting 2011 based on 1975-2010
Model A	Yes	Partly	*No*	Partly	Partly	Partly
Model B	Yes	Partly	*No*	*No*	Partly	Partly
Model C	Yes	Partly	Partly	Partly	Yes	Partly
Model D	Yes	Partly	*No*	Partly	Yes	Partly
Model E	Yes	Partly	*No*	Partly	Yes	Partly
Model F	Yes	Partly	Partly	*No*	Partly	Partly
Model G	Yes	Partly	*No*	Partly	Yes	Partly
Model H	Yes	Partly	Partly	Partly	Partly	Partly
USB	Yes	Partly	Partly	Partly	Partly	Partly

As shown in Figures 3a and 3b, the actual stock return and the predicted paths by using the different models are well within the boundaries of the 80% CIs for all five models. Thus, although the models do not predict what actually happened in 2006 and 2007, their predictions are not that far off. Unfortunately, *with the collapse of the Lehman Brothers, virtually all models are disappointing in the prediction of 2008 returns* (Figure 3c). Among them, Model G (EFP, TED, SRET, HRET) has the worst performance in the sense that its 80-percent CI does not even contain any of the actual quarterly returns in 2008. *As we include data up to 2008Q4 and re-estimate the models, the predictions of 2009 by the models significantly improve*. As shown by Figure 3d, the CI of each model contains at least one quarter of stock return within the CI. Among them, *Model D (FFR, EFP, SRET, HRET), Model E (FFR, TED, SRET, HRET), Model F (EFP, SPR, SRET, HRET) and Model G contain almost the path of the stock return for the whole year*. With the data of 2009 included and model updated, the prediction of 2010 by the models is even better. As Figure 3e shows, the CI of each model contains at least one quarter of stock return. *Interestingly, Model G, which has the worst performance in 2008, becomes the best model in 2010 in the sense that it is the only model in which the CI contains the whole path of the quarterly return of the year*. The prediction of 2011 by the models is similar. As Figure 3f shows, virtually all models contain most of the year return, and all of the models unfortunately "miss" the drop in the stock return in 2011Q3, as no model is able to generate a CI which contains the stock return in 2011Q3.

As shown in Table 7b and Figure 4, the prediction of housing return is worse than that of the stock return. Figure 4a shows that every model generates a CI that contains the path of the quarterly housing return of 2006, yet as early as 2007, Figure 4b shows that the CIs generated by our models fail to contain at least one quarter of housing return. It should be noted that the same set of models successfully contains the whole year path of stock return of the same year (2007). With 2007 data included and models updated, the failure of 2008 is in a sense unexpected. Both Table 7b and Figure 4c show that the CIs generated by more than half of our models - namely, Model A (linear VAR with all 7 variables), Model B (FFR, GDP, SRET, HRET), Model D (FFR, EFP, SRET, HRET), Model E (FFR, TED, SRET, HRET), and Model G (EFP, TED, SRET, HRET) - fail to contain any quarterly housing return in 2008. The 4 remaining models all miss at least one quarterly return of housing of the year. Thus, *while our models do not perform well in 2008 to predict the stock return, the prediction of housing return in the same year is much worse*. *With the information of 2008 included and models updated again, the prediction of 2009 by the models is improved but perhaps still disappointing*. Recall that for the year 2009, the CIs generated by all of our models all contain some quarterly return of stock, which suggests that the enlargement of the sample with model updating might improve the stock return forecasting. In the case of housing, Figure 4d shows that Model B continues to fail to contain any quarterly return of housing in 2009. While Models A, D, E, and G show some improvement, the CI generated by Model F (EFP, SPR, SRET, HRET) fails to contain any quarterly housing return. Furthermore, if 2008 and 2009 are disappointing years of housing return forecasting, 2010 is a year with positive surprises. As shown by Table 7b and Figure 4e, all of the models successfully contain some quarterly return of housing. Moreover, Models C (FFR, SPR, SRET, HRET), D, E and G are able to generate CIs that contain the whole year of housing return. For the case of stock return, only Model G contains the whole year of stock return in the CI. In this sense, the prediction of housing returns in 2010 by the models is a success. Unfortunately, Figure 4f shows that all of the models fail to contain the drop in housing return in 2011Q1, although they contain some quarterly return in the later part of the same year. This is comparable to the performance of the stock return prediction of the same year (2011).

In summary, *it seems that the OSF of asset returns is particularly difficult during this period*. In the case of stock return, Table 7a suggests that since 2008, most models will miss at least one quarter of stock return, and while Model G "fails" in 2008, it becomes very successful in 2010. This confirms the intuition that the regular incorporation of new data and re-estimations of the model lead to better forecasting. In the case of the housing return, most of our models, namely, Models A, B, D, E, F, and G, all experience at least one "missing year" in either 2008 or 2009, i.e. a year in which the CI generated by the models does not contain any quarterly return of the year. At the same time, when we reestimate the model with data up to the end of 2009, Models C, D, E, and G successfully capture the year 2010. Therefore, *Model G (EFP, TED, SRET, HRET) seems to be the "best learner" in the sense that while it made mistakes in 2008 or 2009, when it is re-estimated with the data up to the end of 2009, it successfully captures the movements in both stock and housing returns in 2010*. Note that according to Tables 7a and 7b, neither the USB for stock return nor that for housing return enjoys a "perfect" year (i.e. the asset return movement over the whole year within the 80% CI) after 2006, which suggests that *while the USB may be classified in the same EPPC as other models during the out-of-sample period* (2006 and afterwards) *as a whole, it may not "learn" as much and as fast as the other models which incorporate other macroeconomic and financial variables*.

4.4 Some Robustness Checks

Thus far, our analysis is based on the use of data from 1975 to 2005 as the in-sample, and the periods afterwards as the out-of-sample, and then we progressively update the in-sample. *As a robustness check, we also reestimate our models with the period from 1975 to 2006 as the in-sample*. Table 8 provides a summary and the details can be found in the Appendix. Note that Table 8 is analogously constructed to Table 6 in order to facilitate a comparison. A few observations are in order. Table 8a shows that the best ISF performance comes from Model C (FFR, SPR, SRET, HRET) for stock return and Model E (FFR, TED, SRET, HRET) for housing return, which is the same as the results in Table 6a. While the details of the model classifications in Table 8b slightly differ from Table 6b, some principal findings sustain. First, *most, if not all, of the models are equally good at predicting stock return during the in-sample* *period.* Second, *in terms of predicting housing return, Model B (FFR, GDP, SRET, HRET) and the USB are often the worst*. Again, this supports the idea that the *GDP is not very helpful even for the in-sample prediction of housing return, and information about the future housing returns are reflected in economic variables other than the current housing return*. Table 8c shows that the principal result of Table 6c also sustains, namely, *Model H (SPR, TED, SRET, HRET) outperforms the other models in terms of the OSF of stock return*, and *Model D (FFR, EFP, SRET, HRET) outperforms the other models in terms of the OSF of housing return*. Table 8d, which provides the model classification in terms of the out-of-sample forecasts, is also similar to Table 6d. More specifically, *the model classifications in terms of the out-of-sample 4-quarter ahead forecasts of stock return are identical*. In terms of the counterpart of housing return forecasting, Model A (linear VAR with all 7 variables), Model B which contains FFR and GDP and Model E which contains FFR and TED are inferior to Models C, D, F, and G. A noticeable difference is that now Model H which contains SPR and TED, and USB are also in the second class. Thus, it seems that *while the choice of choosing 2005 as the end of the in-sample might not be the consensus among researchers, the results are not as sensitive as one may think*. In addition, our results based on the recursive approach of allowing models to "update and re-estimate" in terms of the OSF in the previous section has also been presented.

Table 8a Summary of In-sample Forecasting Performances (4-Quarter Ahead Forecasts) (1975Q2-2006Q4)

		Stock Return		Housing Return	
		SLC	**ALC**	**SLC**	**ALC**
Model A	Single-regime model (FFR, SPR, TED, EFP, GDP, SRET, HRET)	56.1661	5.5860	0.7463	0.6412
Model B	Two-regime model (FFR, GDP, SRET, HRET)	56.3073	5.5655	0.8174	0.7001
Model C	Two-regime model (FFR, SPR, SRET, HRET)	55.6801	5.5256	0.7232	0.6599
Model D	Two-regime model (FFR, EFP, SRET, HRET)	56.6837	5.5849	0.7252	0.6521
Model E	Two-regime model (FFR, TED, SRET, HRET)	56.7498	5.5731	0.6995	0.6216
Model F	Two-regime model (EFP, SPR, SRET, HRET)	58.9283	5.6407	0.7190	0.6447
Model G	Two-regime model (EFP, TED, SRET, HRET)	58.2390	5.6850	0.7622	0.6450
Model H	Two-regime model (SPR, TED, SRET, HRET)	57.7033	5.7300	0.7727	0.6579
USB	Uni-variate, Single-regime Benchmark	56.4531	5.6016	0.8218	0.6636

Note: SLC: Square Loss Criteria; ALC: Absolute Loss Criteria

Table 8b Summary of EPPC (In-sample 4-Quarter Ahead Forecasts)

		Class 1	Class 2
Stock Return	SLC	All models	/
	ALC	A, B, C, D, E, F, G, USB	H
Housing Return	SLC	A, C, D, E, F, G, H	B, USB
	ALC	A, C, D, E, F, G, H, USB	B

Note: SLC: Square Loss Criterion; ALC: Absolute Loss Criterion. Our convention is that if $i > j$, then any model in Class i has less predictive power than any model in Class j.

Table 8c Summary of Out-of-Sample Forecasting Performances (4-Quarter Ahead Forecasts) (2007Q4-2012Q1)

		Stock Return		Housing Return	
		SLC	ALC	SLC	ALC
Model A	Single-regime model (FFR, SPR, TED, EFP, GDP, SRET, HRET)	152.4294	11.0135	5.3034	1.9487
Model B	Two-regime model (FFR, GDP, SRET, HRET)	143.7629	10.5411	5.2328	1.9505
Model C	Two-regime model (FFR, SPR, SRET, HRET)	147.2156	10.8114	3.9551	1.6780
Model D	Two-regime model (FFR, EFP, SRET, HRET)	153.8418	11.0492	3.9000	1.6292
Model E	Two-regime model (FFR, TED, SRET, HRET)	139.7478	10.2042	5.5364	2.0180
Model F	Two-regime model (EFP, SPR, SRET, HRET)	161.8230	11.1754	4.1199	1.7119
Model G	Two-regime model (EFP, TED, SRET, HRET)	153.6449	10.3014	4.3944	1.7674
Model H	Two-regime model (SPR, TED, SRET, HRET)	132.4850	9.9965	4.6400	1.8598
USB	Univariate benchmark for a single-regime	137.1296	10.0236	5.0374	1.9263

Note: SLC: Square Loss Criteria; ALC: Absolute Loss Criteria.

Table 8d Summary of EPPC (Out-Of-Sample 4-Quarter Ahead Forecasts)

		Class 1	Class 2
Stock Return	SLC	A, B, C, E, H, USB	D, F, G
	ALC	B, E, G, H, USB	A, C, D, F
Housing Return	SLC	C, D, F, G	A, B, E, H, USB
	ALC	C, D, F, G	A, B, E, H, USB

Note: SLC: Square Loss Criterion; ALC: Absolute Loss Criterion. Our convention is that if $i > j$, then any model in Class i has less predictive power than any model in Class j.

5. Concluding Remarks

Dramatic movements in asset prices often occupy the media headlines and carry implications in real economic activities, even though there are changes in political personnel. Yet there are competing and sometimes even conflicting explanations in the media and even the academic circles. In this paper, some explanations for these events are brought forth for formal testing. Note that while an econometric model comparison is not equivalent to an explanation comparison, the exercise in this paper may provide information for future model development. This is especially true if models contain certain variables that consistently outperform other models which do not contain those variables. *From a policy point of view, identifying the empirically more relevant model not only satisfies intellectual curiosity, but may also assist governments to make more appropriate policy decisions*. Thus, some stylized facts about the asset return dynamics at the aggregate level are established in this paper. In particular, this paper separates the data into the pre-crisis period 1975-2005 (in-sample) and the period afterwards (2006-2011), and examines whether an estimated model based on the pre-crisis period can reasonably forecast for the crisis period (OSF), with the effect of other variables such as GDP growth and monetary policy taken into consideration.

Our first contribution is *to demonstrate how to categorize competing models into different EPPCs. We find that it can shed light on the EMH debate, as well as the structural-break discussion on the asset markets*. During the in-sample period, all of the models have the same predictive power as the USB, i.e. AR(q) on the stock return. This is consistent with the EMH that all information about future stock returns has been reflected in the current period stock return and hence additional variables do not improve the forecasting. For housing return, the situation is different. Models D, E, F, and H consistently outperform the USB, which suggests the *existence of cross-market information flow and the value of multivariate modeling*. For the out-of-sample period as a whole, the picture changes. *No model has superior predictive power than the USB on either the stock return or housing return forecast*. Whether we use the SLC or the ALC , Models A, B and E are shown to have lower predictive power than the USB in the housing return forecast. Thus, these results seem to be consistent with the notion that *there are some "structural changes" in the determinants of asset return dynamics since 2006*, *even when regime-switching behaviors are explicitly modeled*. In particular, since both Models A and B are the only models that contain GDP as a variable in the dynamical system, the evidence then points to the possibility that *GDP has lost its power in predicting the housing return after the crisis period, perhaps due to the deep recession*. Obviously, this does not mean that "economic fundamentals" such as GDP are not important. One interpretation is that the *information contained in the GDP has already been reflected in other financial variables that we include in the econometric model, such as the TED spread*. This is consistent with the theoretical result in Telmer and Zin (2002) that with incomplete markets, asset returns can predict (future) asset prices well.

Our second contribution shows that *all of the models do not have the same capacity to "learn and update"*. Note that the results based on the EPPC clearly treat the model performance during the out-of-sample period as a whole. However, *it is possible that some models make more mistakes in the beginning of the financial crisis, and later significantly improve with the new data supplied*. To investigate such a possibility, we conduct OSF based on a simulation approach. We find that the USB of either the stock return or the housing return has never enjoyed a perfect year (i.e. a year that the asset return is totally inclusive within the CI) since 2006. On the other hand, after revising with data up to the end of 2009, Model G, which includes both the EFP and TED spread, can include the actual asset returns of 2010 in the CIs. In other words, *while univariate models may be as good as any model in predicting asset returns in the "long run", they may learn "slower" than multivariate models in the "short-run".*[41] This also suggests that *macroeconomic models, which emphasize the role of imperfect capital markets that non-financial firms as well as banks face, may be promising in enhancing the understanding of stock and housing returns*.[42]

In addition, we also methodologically demonstrate a few things in terms of multivariate modeling. First, the widely used *linear VAR model (i.e. single regime) with 7 variables can actually underperform as opposed to its regime-switching counterparts with only 4 variables most of the time*. Perhaps more importantly, we demonstrate that we can combine the multi-model comparison test of Mariano and Preve (2012) with the MCS procedure developed by Hansen et al. (2011) to classify models into different EPPCs. We also demonstrate that *such an EPPC approach can be complemented by the simulation-based method* used by Sargent, Williams and Zha (2006) and others to further distinguish the empirical performance of models within the same EPPC. Future research may want to further explore along these lines.

This research can clearly be extended in several directions. First, it would be interesting to apply the current econometric framework to other economies, such as the European Union countries. The current framework can also be applied to more disaggregate data. Also, it would be interesting to incorporate higher frequency variables and study the (potential) price discovery processes among different asset markets. Needless to say, one can re-examine the two asset markets with Bayesian econometrics. Finally, one could develop a theoretical framework that mimics the stylized facts found in this paper, which will further enhance our understanding of the interactions between the real economy and the asset markets.

---

[41] This may also provide support to the model-averaging approach in forecasting. Clearly, this is beyond the scope of the current paper. Among others, see Clark and McCracken (2010), Clements and Hendry (2004), Hansen (2007), Inoue and Kilian (2006), among others.

[42] Among others, see Gertler and Kiyotaki (2010), Davis (2010), Andres and Arce (2012), Jin et al. (2012).

Acknowledgement

An earlier version of the paper was circulated as "Would some models please give me some hints?". The current version is revised after receiving helping comments from (in alphabetical order) Gianni Amisano, Wing Hong Chan, Marcelle Chauvet, Andrew Filrado, Xu Han, Ivan Jaccard, Yuichiro Kawaguchi, Kyung-Hwan Kim, Fred Kwan, Rose Lai, Giovanni Lombardo, Thomas Lubik, Joe Ng, Liang Peng, Daniel Preve, John Quigley, Tim Riddiough, Douglas Rolph, Anna Samarina, Jim Shilling, Edward Tang, Peng Wang, an anonymous referee, seminar participants at the Asian Econometric Society meeting, AsRES-AREUEA meeting, Bank of International Settlements, City University of Hong Kong, European Central Bank, European Real Estate Society meeting, NUS, and WEAI. Chen thanks the National Science Council (Taiwan) for financial support. The work in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No. CityU 144709]. We are grateful for all of their support. The usual disclaimer applies.

The full version of this paper, which contains all the appendices, can be downloaded from https://ideas.repec.org/p/dpr/wpaper/0932.html

References

Andres, J. and Arce, O. (2012), Banking competition, housing prices and macroeconomic stability, *Economic Journal*, 122, 1346-372.

Austen-Smith, D. and Banks, J. (2000), Positive political theory i: collective preference, Ann Arbor: University of Michigan Press.

Bank for International Settlements (2003), 73rd Annual Report, Basle.

Bernanke, B. and Gertler, M. (1995), Inside the black box: The credit channel of monetary policy transmission, *Journal of Economic Perspectives* 9, 27-48.

Bernanke, B. and Gertler, M. (2001), Should central banks respond to movements in asset prices? *American Economic Review*, 91, 253-257.

Bernanke, B., Gertler, M. and Gilchrist, S. (1999), The financial accelerator in a quantitative business cycle framework, in John Taylor and Michael Woodford (ed.), Handbook of Macroeconomics, Vol. 1, New York: Elsevier.

Bertay, A.C. and Gong, D. (2014), Securitization, financial stability and macroeconomy: evidence from an international panel, mimeo.

Campbell, J.Y. and Cocco, J.F. (2007), How do house prices affect consumption? Evidence from micro data, *Journal of Monetary Economics*, 54, 591-621.

Case, K.E. and Shiller, R.J. (1990), Forecasting prices and excess returns in the housing market, *Real Estate Economics*, 18, 253-73.

Case, K.E., Quigley, J. and Shiller, R.J. (2005), Comparing wealth effects: the stock market versus the housing market, *Advances in Macroeconomics*, 5, 1235-1235.

Cetorelli, N. and Goldberg, L. (2011), Global banks and international shock transmission: Evidence from the crisis, *IMF Economic Review*, 59, 41-76.

Cetorelli, N. and Goldberg, L. (2012), Banking globalization and monetary transmission, *Journal of Finance*, 67, 1811-1843.

Chang, K.L., Chen, N.K. and 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., Chen, N.K. and 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., Chen, N.K. and 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.

Chauvet, M. and Potter, S. (2013), Forecasting output, G. Elliot and A. Timmenmann et., Handbook of Economic Forecasting, volume 2, part A, 141-194.

Chen, N.K., Cheng, H.L. and Mao, C.S. (2014), Identifying and forecasting house prices: A macroeconomic perspective, *Quantitative Finance*, 14, 2105-2120.

Chen, N.K. and Leung, C.K.Y. (2008), Asset price spillover, collateral and crises: with an application to property market policy, *Journal of Real Estate Finance and Economics*, 37, 351-385.

Christiano, L., Motto, R. and Rostagno, M. (2007), Financial factors in business cycles, Northwestern University, mimeo.

Clapp, J.M. and Giaccotto, C. (1994), The Influence of economic variables on local house price dynamics, *Journal of Urban Economics*, 36, 161-83.

Clark, T. (2011), Real-time density forecasts from Bayesian vector autoregressions with Stochastic Volatility, *Journal of Business and Economic Statistics*, 29, 327-341.

Clark, T. and McCracken, M. (2010), Averaging forecasts from VARS with uncertain instabilities, *Journal of Applied Econometrics*, 25, 5-29.

Clements, M. and Hendry, D. (2004), Pooling of forecasts, *Econometrics Journal*, 7, 1-31.

Cochrane, J. (2001), Asset Pricing, Princeton: Princeton University Press.

Collin-Dufresne, P. (2004), Notes on term structure, mimeo.

Cooley, T. and Dwyer, M. (1998). Business cycle analysis without much theory: a look at structural VARs. *Journal of Econometrics*, 83(1-2), 57-88.

Davis, S. (2010), The adverse feedback loop and the real effects of financial sector uncertainty, Vanderbilt University, mimeo.

De Walque, G., Pierrard O. and Rouabah, A. (2010), Financial (in)stability, supervision and liquidity injections: A dynamic general equilibrium approach, *Economic Journal*, 120(549), 1234-1261.

Darrat, A. and Glascock, J. (1989), Real estate returns, money and fiscal deficits: Is the real estate market efficient? *Journal of Real Estate Finance and Economics*, 2, 197-208.

Diebold, F. and Mariano, R. (1995), Comparing predictive accuracy, *Journal of Business & Economic Statistics*, 13, 253-263.

Emiris, M. (2006), The term structure of interest rates in a DSGE model, National Bank of Belgium, mimeo.

Estrella, A. (2005), Why does the yield curve predict output and inflation? *Economic Journal*, 115, 722-44.

Estrella, A. and Trubin, M.R. (2006), The yield curve as a leading indicator: Some practical issues, *Current Issues in Economics and Finance*, 12, Federal Reserve Bank of New York.

Fama, E. (1970), Efficient capital markets: A review of theory and empirical work, *Journal of Finance*, 25(2), 383-417.

Ferreira, F., Gyourko, J. and Tracy, J. (2010), Housing busts and household mobility, *Journal of Urban Economics*, 68(1), 34--45.

Francq, C. and Zakoian, J.M. (2001), Stationary of multivariate Markov-switching ARMA models, *Journal of Econometrics*, 10, 339-364.

Gertler, M. and Lown, C. (1999), The information in the high-yield bond spread for the business cycle: Evidence and some implications, *Oxford Review of Economic Policy*, 15, 132-150.

Gertler, M. and Kiyotaki, N. (2010), Financial intermediation and credit policy in business cycle analysis, in B. M. Friedman and M. Woodford ed., Handbook of Monetary Economics, Vol. 3, 547-599.

Goodhart, C. (2001), What weight should be given to asset prices in the measurement of inflation? *Economic Journal*, 111, 335-356.

Goodhart, C. and Hofmann, B. (2007), House prices and the macroeconomy: implications for banking and price stability, Oxford: Oxford University Press.

Green, R. (2002), Stock prices and house prices in California: New evidence of a wealth effect? *Regional Science and Urban Economics*, 32, 775-783.

De Graeve, F. (2007), The external finance premium and the macroeconomy: US Post-WWII evidence, working paper.

Grishchenko, O.V. and Rossi, M. (2012), The role of heterogeneity in asset pricing: The effect of a clustering approach, *Journal of Business and Economic Statistics*, 30, 297-311.

Hamilton, J. (1994), Time series analysis, Princeton: Princeton University Press.

Hansen, B. (2007), Least Squares Model Averaging, *Econometrica*, 75 (4), 1175-1189.

Hansen, P., Lunde, A. and Nason, J. (2011), The model confidence set, *Econometrica*, 79(2), 453-497.

Hempel, C.G. (1942), The function of general laws in history, *Journal of Philosophy*, 39(2), 35-48.

Himmelberg, C., Mayer, C. and Sinai, T. (2005), Assessing high house prices: Bubbles, fundamentals and misperceptions, *Journal of Economic Perspective*, 19, 67-92.

Hott, C. and Monnin, P. (2008), Fundamental real estate prices: An empirical estimation with international data, J*ournal of Real Estate Finance and Economics*, 36, 427-450.

Huang, H., and Tang, Y. (2012), Residential land use regulation and the US housing price cycle between 2000 and 2009, *Journal of Urban Economics*, 71(1), 93--99.

Huang, M. (2014), Bubble-like housing boom--bust cycles: Evidence from the predictive power of households' expectations, *Quarterly Review of Economics and Finance*, 54, 2-16

Ibbotson, R.G. and Siegel, L.B. (1984), Real estate returns: A comparison with other investments, Jo*urnal of American Real Estate and Urban Economics Association*, 12, 219-242.

In't-Veld, J., Raciborski, R., Ratto, M. and Roeger, W. (2011), The recent boom--bust cycle: The relative contribution of capital flows, credit supply and asset bubbles. *European Economic Review*, 55(3), 386--406.

Inoue, A. and Kilian, L. (2006), On the selection of forecasting models, *Journal of Econometrics*, 130, 273-306.

Jin, Y., Leung, C.K.Y. and Zeng, Z. (2012), Real estate, the external finance premium and business investment: A quantitative dynamic general equilibrium analysis, *Real Estate Economics*, 40(1), 167-195.

Kan, K., Kwong, S.K.S. and Leung, C.K.Y. (2004), The dynamics and volatility of commercial and residential property prices: Theory and evidence, *Journal of Regional Science*, 44(1), 95-123.

Kapetanios, G., Pagan, A. and Scott, A. (2007), Making a match: Combining theory and evidence in policy-oriented macroeconomic modeling, J*ournal of Econometrics*, 136, 565-594.

Khandani, A.E., Lo, A.W. and Merton, R.C. (2013), Systemic risk and the refinancing ratchet effect, *Journal of Financial Economics*, 108(1), 29--45.

King, R., Plosser, C. and Rebelo, S. (2002), Production, growth and business cycles: Technical appendix, *Computational Economics*, 20, 87--116.

Kiyotaki, N. and Moore, J. (1997), Credit cycles, *Journal of Political Economy*, 105, 211-248.

Kouwenberg, R. and Zwinkels, R. (2014), Forecasting the US housing market, *International Journal of Forecasting*, 30, 415-425.

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

LeRoy, S. (1989), Efficient capital markets and Martingales, *Journal of Economic Literature*, 27(4), 1583-1621.

Lettau, M. and Ludvigson, S.C. (2005), Expected returns and expected dividend growth, *Journal of Financial Economics*,76, 583-626.

Leung, C. (2004), Macroeconomics and housing: a review of the literature, *Journal of Housing Economics*, 13, 249-267.

Leung, C.K.Y. (2007), Equilibrium correlations of asset price and return, *Journal of Real Estate Finance and Economics*, 34, 233-256.

Leung, C.K.Y. (2014), Error correction dynamics of house price: An equilibrium benchmark, *Journal of Housing Economics*, 25, 75-95.

Lubik, T., and Schorfheide, F. (2004), Testing for indeterminacy: An application to U.S. monetary policy, *American Economic Review*, 94(1), 190-217.

Lustig, H. and Van Nieuwerburgh, S. (2005), Housing collateral, consumption insurance and risk premia: An empirical perspective, *Journal of Finance*, 60, 1167--1219.

Lutz, B., Molloy, R. and Shan, H. (2011), The housing crisis and state and local government tax revenue: Five channels. *Regional Science and Urban Economics*, 41(4), 306--319.

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

Maheu, J., McCurdy, T. and Song, Y. (2012), Components of Bull and Bear markets: Bull corrections and Bear rallies, *Journal of Business and Economic Statistics*, 30, 391-403.

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

Mehra, R. and Prescott, E. (2003), The equity premium in retrospect, NBER Working Paper 9525.

Menzly, L., Santos, T. and Veronesi, P. (2004), Understanding predictability, *Journal of Political Economy*, 112, 1-47.

Mishkin, F. (2001), The transmission mechanism and the role of asset prices in monetary policy, NBER Working Paper 8617.

Nichols, J.B., Oliner, S.D. and Mulhall, M.R. (2013), Swings in commercial and residential land prices in the United States. *Journal of Urban Economics*, 73(1), 57-76.

Pagan, A. and Robinson, T. (2014), Methods for assessing the impact of financial effects on business cycles in macroeconometric models, *Journal of Macroeconomics*, 41, 94-106.

Phillips, P. (2014), On confidence intervals for autoregressive roots and predictive regression, *Econometrica*, 82(3), 1177-1195.

Piazzesi, M., Schneider, M. and Tuzel, S. (2007), Housing, consumption, and asset pricing, *Journal of Financial Economics*, 83, 531-569.

Plosser, C. and Rouwenhorst, K.G. (1994), International term structures and real economic growth, *Journal of Monetary Economics*, 33, 133-155.

Quan, D. and Titman, S. (1997). Commercial real estate prices and stock market returns: An international analysis, *Financial Analysts Journal*, 53(3), 21-34.

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

Rigobon, R. and Sack, B. (2003). Measuring the reaction of monetary policy to the stock market, *Quarterly Journal of Economics*, 118, 639--669.

Rosenberg, J.V. and Maurer, S. (2008), Signal or noise? Implications of the term premium for recession forecasting, *FRBNY Economic Policy Review*, 14(3), 1-11.

Sargent, T., Williams, N. and Zha, T. (2006), Shocks and government beliefs: The rise and fall of American inflation, *American Economic Review*, 96, 1193-1224.

Shiller, R. (2007), Understanding recent trends in house prices and home ownership. Retrieved from http://www.kc.frb.org/publicat/sympos/2007/pdf/2007.09.27.shiller.pdf.

Shiller, R. (2008), The Subprime Solution, Princeton: Princeton University Press.

Sims, C. (1980a), Comparison of interwar and postwar business cycles: Monetarism reconsidered, *American Economic Review*, 70(2), 250-257.

Sims, C. (1980b), Macroeconomics and reality, *Econometrica*, 48, 1-48.

Sims, C. and Zha, T. (2006), Were there regime switches in U.S. monetary policy? *American Economic Review*, 96, 54-81.

Sutton, G.D. (2002), Explaining changes in house prices, *BIS Quarterly Review*, September, 46-55.

Telmer, C. and Zin, S. (2002), Prices as factors: Approximate aggregation with incomplete markets, *Journal of Economic Dynamics and Control*, 26, 1127-1157.

Torous, W., Valkanov, R. and Yan, S. (2004), On predicting stock returns with nearly integrated explanatory variables, *Journal of Business*, 77, 937-966.

Tsatsaronis, K. and Zhu, H. (2004), What drives housing price dynamics: Cross-country evidence, *BIS Quarterly Review*, March, 65-78.

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

Yoshida, J. (2008), Technology shocks and asset price dynamics: The role of housing in general equilibrium, University of Tokyo, mimeo.

Zivot, E. (2004), Notes on forecasting, University of Washington at Seattle, mimeo.