

# Equity Premium Predictions with Adaptive Macro Indices

## Abstract

Fundamental economic conditions are crucial determinants of equity premia. However, commonly used predictors do not adequately capture the changing nature of economic conditions and hence have limited power in forecasting equity returns. To address the inadequacy, this paper constructs macro indices from large datasets and adaptively chooses optimal indices to predict stock returns. I find that adaptive macro indices explain a substantial fraction of the short-term variation in future stock returns, and have more forecasting power than both the historical average of stock returns and commonly used predictors. The forecasting power exhibits a cyclical pattern, implying the ability of adaptive macro indices in capturing time-varying economic conditions. This finding highlights the importance of using dynamically-measured economic conditions to investigate empirical linkages between equity premium and macroeconomic fundamentals.

# 1 Introduction

Empirical studies increasingly cast doubt on the forecasting power of price-based predictors of equity returns. The return predictability literature has documented numerous problems (see the survey by Spiegel (2008), Campbell and Thompson (2008) and Cochrane (2008)), including small-sample bias with highly persistent predictors (see Lewellen (2004), Campbell and Yogo (2004) among many others),<sup>1</sup> poor out-of-sample performance (see Goyal and Welch (2008)),<sup>2</sup> and unstable forecasting relationships (Pastor and Stambaugh (2001), Lettau and Nieuwerburgh (2008)).<sup>3</sup> One potential reason for their weak performance is that fixed financial predictors fail to capture time-varying economic conditions sufficiently. Nevertheless, fundamental economic forces are crucial determinants of equity premia in the financial markets (Lettau and Ludvigson (2005)).<sup>4</sup> The challenge thus far has been to find a way to capture changing fundamental conditions that affect equity returns.

This paper answers the challenge in two ways. First, I construct macro indices from a large number of economic series as quantitative descriptions of economic conditions. Second, I design an adaptive procedure to choose optimal indices for equity premium predictions. Using this procedure,

---

<sup>1</sup>Examples are Nelson and Kim (1993), Stambaugh (1999), Ang and Bekaert (2007), Ferson, Sarkissian and Simin (2003), and Valkanov (2003). A number of recent papers discuss alternative econometric methods for conducting valid inference, such as Cavanagh, Elliott and Stock (1995), Polk, Thompson and Vuolteenaho (2006), Lewellen (2004), Torous, Valkanov and Yan (2004), Campbell and Yogo (2006), and Eliaz (2005).

<sup>2</sup>Goyal and Welch (2008) have systematically investigated most predictors used in the literature and concludes that, except for equity-issuing activity, current predictors hardly have meaningful and robust predictive power on the equity premium. Additional contributions include Bossaerts and Hillion (1999), and Butler, Grullon and Weston (2005). See also Campbell and Thompson (2008), and Cochrane (2008) for different interpretations of the out-of-sample evidence.

<sup>3</sup>Viceira (1996) and Paye and Timmermann (2006) report evidence that counters the hypothesis of constant coefficient in the forecasting regression. Lettau and Nieuwerburgh (2008) analyze structural shifts in the mean of the dividend-price ratio. Pastor and Stambaugh (2001) use Bayesian methods to estimate structural breaks in the equity premium.

<sup>4</sup>Equity risk premia are closely related to economic conditions. Equity returns seem to be high at business cycle troughs and low at peaks. In line with the pioneering work by Ferson and Merrick (1987), Fama and French (1989), researchers suggest that predictors of excess returns should be correlated with economic conditions. Lettau and Ludvigson (2005) summarize the literature and point out that we should expect to “find evidence from predictive regressions of excess returns on macroeconomic variables over business cycle horizons.”

I find that adaptively selected macro indices are able to predict equity premia. Trading strategies based on these indices significantly and consistently outperform a buy-and-hold strategy benchmark under varying assumptions of transaction costs and risk tolerance. Furthermore, I connect these indices to economic sectors, comparing *ex ante* and *ex post* forecasting. The results provide new evidence that four sectors – interest rates, price indices, housing, and employment – are particularly relevant in predicting the equity premium.

Given the deficiency of financial predictors, previous papers (for example, Lettau and Ludvigson (2001), Piazzesi, Schneider and Tuzel (2007) and Gomes, Kogan and Yogo (2007)) have attempted to explore alternative predictors involving macroeconomic series.<sup>5</sup> Common to all these papers is a focus on a small set of predictors based on theoretical models. From an academic viewpoint, the use of model-based predictors facilitates an understanding of specific aspects of the economic mechanism. From an investor’s viewpoint, however, these predetermined variables may not be enough to capture all information required in decision making.

Taking an investor’s position, this paper examines broad economic conditions by constructing macro indices from a large number of economic series, which is in line with the dynamic factor method by Stock and Watson (2002a, 2002b). These macro indices distinguish themselves for the ability to synthesize multidimensional information inherent in economic conditions. As noted in Ludvigson and Ng (2007), the use of macro indices “eliminates the arbitrary reliance on a small number of exogenous predictors.” Indeed, anecdotal evidence suggests that investors monitor and analyze literally hundreds of data series to make forecasting decision. Since macro indices can consistently cover a broad range of economic conditions, they are more likely to span an investor’s unobservable information set. Applications of macro indices abound in the macroeconomic litera-

---

<sup>5</sup>Lettau and Ludvigson (2001) use “cay”, a measure of the consumption-wealth ratio; Piazzesi, Schneider and Tuzel (2007) employ the housing-expenditure ratio; Gomes, Kogan and Yogo (2009) use the ratio of durables to nondurables consumption. Another example is Campbell and Diebold (2009) who use survey-based measures of expected business conditions.

ture,<sup>6</sup> but are less found in the empirical finance literature except Ludvigson and Ng (2007, 2009) and Moench (2008).<sup>7</sup> This paper contributes to the literature with a new application in real-time equity premium predictions.

Also attempting to address the deficiency of financial predictors, Lettau and Nieuwerburgh (2008) allow for dynamic shifts in the mean of financial ratios (for example, dividend-price ratios) and show that adjusted predictors reconcile some of the controversy regarding return predictability. Their adjustment, however, is impossible in real-time prediction since investors cannot observe future data. This paper instead designs an adaptive prediction process based on macro indices for the real-time equity premium prediction. Adaptability is important since a fixed set of predictors, even using information from a large number of sources, may not accurately reflect investors' learning processes.

This paper is related to the idea of dynamic modeling in Pesaran and Timmermann (1995). However, I focus on dynamic weights of the macro indices constructed from high-dimensional economic variables. In doing so, the adaptive prediction procedure both captures time-varying economic conditions and allows for a time-varying forecasting relationship between equity returns and macro indices.

Applying this procedure to the CRSP value-weighted index over the one-month Treasury bill, I find that adaptively selected macro indices explain a substantial fraction of the short-term variation in future excess returns. Adaptive forecasts conditional on macro indices generate smaller prediction

---

<sup>6</sup>Growing empirical evidence suggests that a large set of macroeconomic variables may in fact be crucial to properly capture the economy's dynamics. For example, Stock and Watson (1999, 2002a) and Forni, Hallin, Lippi and Reichlin (2000) find that macro indices lead to considerable improvements over small scale vector autoregressive models. Bernanke and Boivin (2003), Bernanke, Boivin and Elias (2005), and Giannone, Reichlin and Sala (2004) show that the large information set appears to matter empirically to model monetary policy. Boivin and Giannone (2006) incorporate the dynamic factor approach to estimate dynamic stochastic general equilibrium models.

<sup>7</sup>Ludvigson and Ng (2007) use the dynamic factor method to analyze the empirical risk-return relationship. Ludvigson and Ng (2009) adopt the same method to examine the bond risk premia. Moench (2008) applies the factor method to forecast the bond yield curve.

errors than both unconditional forecasts using the historical average<sup>8</sup> and the conditional forecasts based on many popular predictors (such as financial ratios, traditional business cycle proxies and other individual macroeconomic series).<sup>9</sup> The superior predictive power of adaptive macro indices have persisted over the past three decades, even in the late 1990s when stock return predictability becomes particularly challenging. Moreover, the accuracy of forecasts both in magnitude and in sign exhibits a cyclical pattern: the predictive power of indices (measured by the difference of prediction errors from the benchmark, the prevailing mean) decreases before recessions and lately rebounds during recessions. Timing ability (measured by the percentage of correct forecasting signs) shares the same pattern.

Finally, to test the real-world economic significance of these adaptive predictions, I examine the profitability of trading portfolios under two investment strategies: 100% stock/bond allocation and a utility-based dynamic portfolio allocation. In both cases, investment strategies exploiting the adaptive forecasts yield greater profits when compared with a buy-and-hold strategy in the market portfolio.

The approach in this paper uses time-varying combinations of a large number of variables. It is therefore not immediately clear what specific economic forces contribute to the predictability of returns over time. To overcome this obstacle, I compare several types of forecasts using different information sets. In *ex post* forecasting, I construct indices using the entire sample, and study the performance of each index with fixed weights. The second principal component (F2) stands out with both a relatively good in-sample and simulated out-of-sample performance.<sup>10</sup> In *semi ex post*

---

<sup>8</sup>The unconditional forecast is equal to the regression of excess returns on a constant.

<sup>9</sup>Fama and French (1989) show how dividend-price ratios, term spreads and default spreads forecast stock and bond returns. Since this seminal paper, these three variables have become representatives of business cycle proxies in empirical finance literature. The individual macroeconomic series include the growth rate of industrial production and the consumer price index.

<sup>10</sup>Here I use the same criteria advocated by Goyal and Welch (2008). A good predictor should satisfy three rules: 1) both a significant in-sample and a reasonably good out-of-sample performance over the entire sample period; 2) it should have a general upward trend in the graph of  $\Delta$ SSE, the difference in cumulative prediction errors from the benchmark of historical average; and 3) an upward drift that remains positive over the past several decades.

forecasting, I recursively construct indices using only the historical information available for each month, and I consistently choose the “same” index (principal components with the same order) and study the performance of each index with dynamic loadings. F2 and F5 outperform all other indices. In *ex ante* forecasting, I recursively construct indices using only the historical information available for each month, I sequentially choose indices (principal components with different orders) and I study the performance of adaptive macro indices. Again, F2 and F5 remain the most frequently selected predictors in the adaptive forecasting. Decomposing these indices and projecting them onto each economic sector, I show that the economic fundamentals that contribute most to equity premium prediction are interest rates, price indices, housing and employment.

The paper proceeds as follows. The next section sets up the adaptive prediction process and presents empirical predictability results. Section 3 assesses the economic significance of real-time forecasting. I then move on to explore the connection between *ex ante* and *ex post* forecasting in Section 4. Section 5 concludes the paper.

## **2 Forecasting Equity Returns**

### **A The Adaptive Prediction Process**

In this paper I consider an agnostic investor who has no strong beliefs in any particular model, but who trusts in the time-varying inference drawn from her information set. The evolution of forecasting models over time may reflect an investor’s learning process or the changing nature of the underlying economic conditions. The use of an adaptive prediction procedure explicitly accounts for the continuous uncertainty an investor faces in real time. The advantage of adaptive prediction is that all possible models are constantly re-estimated and re-evaluated in order to reflect an investor’s search for the optimal predictive relation based solely on the historical information.

In related work, Pesaran and Timmermann (1995) develop a recursive method to select optimal predictors over time according to a predetermined in-sample criterion. This paper incorporates their idea of dynamic modeling but with a focus on dynamic weights of the macro indices constructed from high-dimensional economic variables. In doing so, the adaptive prediction procedure both captures time-varying economic conditions and allows for a time-varying forecasting relationship between equity returns and macro indices.

Suppose that at time  $t = N - m$ , an investor constructs an information set including an array of  $K$  macroeconomic variables with  $N - m$  time-series observations. As with Stock and Watson (2002b), the investor uses principal component analysis and decomposes the information set into  $K$  factors,

$$(1) \quad X_t = \Lambda Z_t = \lambda_1 z_{1,t} + \lambda_2 z_{2,t} + \cdots + \lambda_K z_{K,t},$$

where  $X_t$  is the panel of economic variables assumed to cover the investor's information set,  $z_{i,t}$  ( $i = 1, \dots, K$ ) are principal components of the panel data ordered by their ranks in explaining the variance,<sup>11</sup> and  $\lambda_i$  is the loadings on the individual series.

The investor then defines a universe of parsimonious models based on in-sample estimations,<sup>12</sup>

$$(2) \quad r_t = a_i + b_i z_{i,t-1} + \epsilon_t, \quad i = 1, \dots, K,$$

---

<sup>11</sup> $z_1$  explains the most of the variance and it has the biggest eigenvalue.

<sup>12</sup>Since factors  $z_i$  are mutually orthogonal by construction, it is sufficient for an investor to consider only univariate predictive regression. A regression with multiple factors has the same explanatory power as the sum of each individual factor. The explanatory power from different factors does not overlap.

Two stage estimation usually involves the error-in-variable problem. But Bai and Ng (2005) show that the pre-estimation of the factors  $z_t$  does not affect the consistency of the second-stage parameter estimates. Moreover they prove that the least squares estimates from factor-augmented forecasting regressions are  $\sqrt{T}$  consistent and asymptotically normal. Stock and Watson (2002b) provide both theoretical arguments and empirical evidence that factors using estimated principal components are consistent even in the face of temporal instability in the individual time series used to construct the factors. The reason is that such instability may "average out" in the construction of common factors if the instability is sufficiently dissimilar from one to the next.

and chooses the optimal model according to a predefined selection criterion. In other words, the investor treats all models under consideration as equally likely. Choosing a particular model at time  $t$  does not necessary restrict the model choice at subsequent periods. Such a treatment is different from typical dynamic factor models proposed by Stock and Watson (2002a, 2002b), in the sense that this procedure does not restrict the base set of predictors to be the first few principal components. Instead, the investor seek an optimal predictor from all principal components.

Whereas most studies use in-sample model selection criteria based on the predictive regression (2), I introduce an out-of-sample criterion: the *predictive least squares principle*, which is based on a predictor’s forecasting performance in a simulated out-of-sample period, called “Training period.” The model selection period is a moving window with a length of  $m$ -months. In this period, the investor records prediction errors for all macro indices by comparing their forecasts with realized returns.

At the end of the training period  $t = N$ , the investor choose an index with the smallest cumulative prediction errors in previous  $m$ -months and use it to predict the one-month-ahead excess market return at the first point of the trading period,  $t = N + 1$ . Meanwhile, the investor transforms the predicting result into trading strategies that bring profits or losses, depending on the realization of the true market return (Section 3 presents investment strategies).

[Figure 1 about here.]

Figure 1 illustrates this adaptive prediction process. All stages of the adaptive prediction process re-occur on a monthly basis via a rolling window framework. Macro indices are also reconstructed each month via a recursive window starting from the beginning of the sample. For example, the first investment decision made at  $t = N$ , is based on an estimation during the in-sample period of  $[1, N - m]$  and a model selection in the training period of  $[N - m + 1, N]$ . For the second investment decision, the corresponding estimation period is  $[1, N - m + 1]$ ; the training period is  $[N - m + 2, N + 1]$ .



A key step in the adaptive prediction process is to choose the optimal predictor according to a predefined selection criterion. Conventional measures such as  $R^2$ , Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are calculated only from the in-sample information. Thus they have limitations as guides for finding an optimal indicator in forecasting future excess returns. Bossaerts and Hillion (1999) show that the use of these in-sample criteria fails to find sufficient out-of-sample predictability.

In this paper I introduce an out-of-sample model selection criterion, *predictive least squares* (PLS), advocated by Rissanen (1986) and Wei (1992). The PLS principle selects the optimal forecasting model  $M_i$  which minimizes the cumulative squared prediction errors in the training period:

$$(3) \quad M_i : \quad \text{PLS}(z_i) = \sum_{t=N-m}^N \left[ r_t - (\hat{a}_{i,t-1} + \hat{b}_{i,t-1} z_{i,t-1}) \right]^2, \quad i = 1, \dots, K$$

where  $(\hat{a}_{i,t}, \hat{b}_{i,t})$  is estimated from the predictive regression using only historical information up to time  $t$ . The PLS principle is consistent with the main purpose of real-time forecasting: finding a predictor with the least prediction errors. Models selected in this way intuitively tend to have a good performance in the second out-of-sample trading period.<sup>13</sup>

---

<sup>13</sup>Another potential benefit of PLS is its natural link to the information-based criteria. Define  $\epsilon_t(N) = r_t - (\hat{a}_{i,t-1} + \hat{b}_{i,t-1} z_{i,t-1})$  and  $e_t = r_t - (\hat{a}_{i,t-1} + b_{i,t-1} z_{i,t-1})$ , we can decompose PLS into two terms:

$$(4) \quad \sum_{t=N-m}^N e_t^2 = \underbrace{\sum_{t=1}^N \epsilon_t^2(N) - \sum_{t=1}^{N-m} \epsilon_t^2(N-m)}_{\text{goodness of fit}} + \underbrace{\sum_{t=N-m}^N z'_{i,t} \Sigma_{z_i,t} z_{i,t} e_t^2}_{\text{complexity}}.$$

Recall that the in-sample period is from  $t = 1$  to  $N - m$  and the training period is from  $N - m + 1$  until  $N$ . The estimation of  $b_N$  uses only information in the in-sample period and  $\epsilon_t(N)$  is the fitting error of the predictive regression. As with the information-based criteria, the first two terms in the formula (4) can be viewed as the measure of the goodness of fit; the third one as a penalty that reflects the complexity of the model.

## B Data and Empirical Implementation

The equity premium is measured by the end-of-month return of the CRSP value-weighted index in excess of the one-month Treasury Bill rate. All variables in this paper are at a monthly frequency covering the period from January 1960 to November 2006, for a total of 563 observations.

The base set of forecasting variables comprises 100 macroeconomic variables. Following Stock and Watson (2002a, 2005), I chose this panel to represent broad sectors of U.S. economic conditions: real output, unemployment and employment, wages, housing, foreign exchange rates, money and credit aggregates, interest rates, inventories and orders, price indices, consumption, and consumer expectation.<sup>14</sup> The series are transformed so as to insure stationarity. In addition, the transformed data are standardized prior to index construction and estimation. All economic series are from the Global Insights Basic Economics Database. A detailed description of the data and its transformation is given in Appendix A.1.

To facilitate the comparison, I adopted ten monthly variables in Goyal and Welch (2008) as the control predictors, including price-dividend ratios, price-earnings ratios, dividend payout, book-to-market ratios, long term government bond returns, inflation measured by Consumer Price Index (all urban consumers), net issuing activity, term spread, default spread and default payout. Their explicit definitions are given in Appendix A.2. The contrast set of forecasting variables does not include cay, the consumption wealth ratios proposed by Lettau and Ludvigson (2001), because the construction of cay uses forward-looking information, which is impossible in real-time forecasting. The base set also excludes the investment to capital ratios proposed by Cochrane (1991), since this variable is not available at the monthly frequency.

---

<sup>14</sup>Strictly speaking, variables in the interest rate sector (bond yields) are not pure macroeconomic variables. But interest rates are important indicators of monetary policy and hence are included here to capture specific information of economic conditions. For a robustness check, I also construct macro indices and redo the experiment without using interest rate variables. I find that the resulting macro indices still have significant predictive power. Detailed results are available from the author on request.

With constructed indices, investors can easily implement the adaptive prediction procedure: estimate the linear prediction models, select the optimal model using the PLS principle and then compute a one-month-ahead forecast. The first in-sample estimation is based on monthly observations over the period 1960:1 to 1974:12. Other estimations occur recursively with the same start. The year 1960 was chosen as the start of the estimation sample since reliable monthly measures for most macroeconomic time series were not available until the late 1950s.<sup>15</sup> Also, a number of studies have suggested the possibility that stock returns may have varied systematically over the business cycles in the early 1960s. The year 1975 was chosen as the start of the out of sample since the oil shock of 1973-1975 probably improved the significance of predictability for many models, as pointed out by Goyal and Welch (2008). Excluding the oil shock recession from the out-of-sample period increases the credibility of a predictive model.

The length of the training period is an ad hoc choice. There is a tradeoff between the length and the quality of the model selection. A long rolling window has more statistical power. But a short rolling window easily captures the dynamic changes of economic conditions and hence an optimal predictor selected under a short window tends to reflect the most recent information. In addition, a short rolling window can alleviate the concern of structural break. Ignoring the structural break can be especially costly when predicting the sign of excess returns as demonstrated by Pesaran and Timmermann (2004), and Pettenuzzo and Timmermann (2011), among others. Motivated by real-time forecasting, this paper adopts a 24-month rolling window under which an investor may switch more frequently from one model to another in response to newly obtained information. For a robustness check, I also conduct the same experiment by choosing an optimal predictor according to their performance in previous 12, 18, 30, 36, 42, and 48 months. The results are not significantly

---

<sup>15</sup>Many macroeconomic series became available after World War II. But the data immediately after the war were volatile due to the unusually intensive re-construction. To obtain reliable macro data, I take a rather conservative stand and commence with the estimation at the beginning of 1960.

different.<sup>16</sup>

## C Empirical Results

### C.1 Real-time Out-of-Sample Performance

Figure 2 illustrates the out-of-sample performance for monthly predictive regressions in real time. Similar to Goyal and Welch (2008), the performance is measured by the difference in the cumulative squared prediction errors between the benchmark (a forecast equal to the historical average excess return measured at each time-point) and the prediction model (a forecast based on a single predictor with the smallest cumulative prediction errors in the previous 24 months). Panel A displays the results using adaptive macro indices; Panel B displays the results using the same adaptive procedure with the application to ten commonly used predictors, summarized in Goyal and Welch (2008).

The graph of cumulative prediction error difference ( $\Delta\text{SSE}$ ) offers special advantages in displaying forecasting performance. First, it provides a continuous evaluation of forecasting performance over the whole out-of-sample period. Its complete records avoid biased judgement based on only single time-point evaluation. In contrast, common practice in the literature conducts out-of-sample tests based solely on the ending point estimate for a fixed sample period, though a good ending cannot guarantee the goodness of the whole process. Second, the time-series pattern of the graph allows for recognizing months with a good, or a bad, performance. An increase in a line indicates better performance of the testing prediction model, whereas a decrease in a line suggests better performance of the benchmark. A good month means that the  $\Delta\text{SSE}$  at that month is included in an upward trend along the line. Finally, the graph is invariant to the choice of the out-of-sample period (though it does affect the in-sample estimation results).

---

<sup>16</sup>Detailed results are available from the author on request.

[Figure 2 about here.]

The first impression on the graph in Panel A is a positive  $\Delta$ SSE line over the entire out-of-sample period, which means that the macro index prediction model generates smaller cumulative prediction errors and hence is superior to the benchmark for the period of January 1977 to December 2005. In addition, the  $\Delta$ SSE line tends to increase in general, suggesting an increasing difference in prediction errors between the benchmark and the prediction model. The  $\Delta$ SSE line edges upward from the beginning of 1977, increases substantially in the early 1980s, but then starts to shake up and down until the early 1990s. During the 1990 recession, the predictive power of the macro indices, as measured by  $\Delta$ SSE, jumps by approximately 70% and then grows steadily until 1998. After the tumultuous period of 1998-1999, the  $\Delta$ SSE line rebounds to a new peak and remains stable until the end of 2005.

Another striking pattern of the out-of-sample performance is its tight correlation with the business cycles. The predictive power of macro indices decreases before the recessions but rebounds quickly during them. As evidenced in the graph, the  $\Delta$ SSE line has three sharp downward drifts in the period 1986:02 - 1986:10 (before the 1987 stock market crash), the period 1989:07 - 1990:02 (before the 1990-91 recession) and the period 1998:02 - 1998:09 (before the “bubble period” of 1999-2001). With each decline, a corresponding upward drift is observed during the recession or market downturns. This phenomenon in turn indicates that macro indices are suitable predictors for excess stock returns since macro indices sufficiently describe economic conditions.

Important information of the real-time forecasting performance is also provided in Table 1, which displays the statistical results using three measures: the out-of-sample R-square statistics ( $R^2_{os}$ ), the difference of the root-mean-squared-error statistics ( $\Delta$ RMSE), and the  $F$ -statistics for the test of equal forecast accuracy (MSE-F). The out-of-sample  $R^2$  statistic is computed as it is in Campbell and Thompson (2008). MSE-F reports the  $F$ -statistics defined by McCracken (2007).

**Table 1****Real-time Out-of-Sample Performance: 1977:01 - 2005:12**

This table presents the statistical results of real-time out-of-sample performance for the monthly predictive regressions. During each out-of-sample month, investors choose a predictor from the base set which generates the smallest cumulative prediction errors in previous 24 months. There are two base sets: A: **Macro Indices** constructed from a panel of 100 economic variables using principal component analysis; and B: the ten monthly Predictors in Goyal and Welch (2008) (**GW Predictors**). The performance uses three measures: the out-of-sample R-square ( $R_{oos}^2$ ), the difference of the root-mean-squared prediction error ( $\Delta RMSE$ ) and the MSE-F statistics (McCracken (2007)) to test for the equal forecast accuracy. The benchmark is the unconditional forecast equal to the historical average. The alternative is the conditional forecast using adaptively selected predictors. One-sided critical values of MSE-F are obtained empirically from bootstrapped distributions. The boldface numbers indicate that the alternative model is superior to the benchmark with at least 90% statistical significance. The formulas for the three statistics are:

$$\begin{aligned}
R_{oos}^2 &= 1 - \sum \hat{u}_{A,\tau}^2 / \sum \hat{u}_{B,\tau}^2 \\
\Delta RMSE &= \sqrt{\sum \hat{u}_{B,\tau}^2} - \sqrt{\sum \hat{u}_{A,\tau}^2} \\
MSE-F &= R \cdot \left( \frac{\sum_r \hat{u}_{A,\tau}^2 - \sum_r \hat{u}_{B,\tau}^2}{\sum_r \hat{u}_{B,\tau}^2} \right)
\end{aligned}$$

where  $\hat{u}_{B,\tau}$  is the forecasting error from the benchmark and  $\hat{u}_{A,\tau}$  is the forecasting error from the alternative adaptive prediction model.  $R$  is the number of out-of-sample observations.

Period	Panel A: Macro Indices			Panel B: GW Predictors		
	$R_{oos}^2$	$\Delta RMSE$	MSE-F	$R_{oos}^2$	$\Delta RMSE$	MSE-F
1977:1 - 2005:12	2.70	0.0604	<b>9.66</b>	-2.26	-0.0500	-12.94
1980:1 - 1989:12	1.73	0.0425	<b>2.11</b>	-0.42	-0.0083	-0.51
1990:1 - 1999:12	1.10	0.0214	1.33	-3.86	-0.0893	-3.01
2000:1 - 2005:12	4.15	0.0924	<b>3.51</b>	-2.26	-0.0500	-8.24

A rejection of the benchmark hypothesis suggests the superiority of the alternative model, rather than the benchmark in the sense that the alternative model generates smaller prediction errors. The benchmark is the unconditional forecast equal to the historical average. The alternative is the conditional forecast using adaptively selected predictors: macro indices in Panel A and financial predictors in Panel B. During each out-of-sample month, investors choose a predictor from the macroeconomic or financial base set, which has the smallest cumulative prediction errors in previous 24-months. All calculations are based on the adaptive prediction process described in Section A.

As shown in Panel A of Table 1, adaptive macro indices robustly predict excess stock returns for

the whole sample of 1977:1 to 2005:12, with a 2.7% out-of-sample  $R^2$  and a 9.66 MSE-F value which is 99% statistically significant. For both subperiods of 1980-1989 and 2001-2005, the  $F$  statistics are both significant above the 90% level: 2.11 for the 1980s and 3.51 for the 2000s. These numbers confirm that the conditional forecasts based on adaptive macro indices have smaller prediction errors and hence are superior to the unconditional forecasts based on the historical average.

The subperiod 1990-1999 has a  $R^2$  value of 1.10, but the  $F$ -statistic (1.33) is not statistically significant, suggesting that adaptive macro indices have a seemingly unimproved forecasting power. However, these numbers only reveal the predictability information at a single point, that is December 1999; it hides the time-series pattern of predictability over the course of the 1990s. As shown in the  $\Delta$ SSE graph in Figure 2, the low  $\Delta$ RMSE number at 1999:12 is due to the sharp drop of predictability during the bubble period 1998-1999. Macro indices do have steadily increasing predictive power from the beginning of 1990 to the beginning of 1998. This example demonstrates that it is important to use the cumulative-prediction-error-difference graph to evaluate out-of-sample performance.

In sharp contrast, the adaptive forecasts of excess returns based on financial/accounting ratios (Panel B) have much higher prediction errors than the benchmark. For either the full out-of-sample period or for three subperiods, the  $R^2_{oos}$  are all negative as are the  $F$ -statistics. These results are consistent with the findings in Goyal and Welch (2008), who show that commonly used predictors in the current literature have a poor out-of-sample performance. Note that a striking difference of this paper from Goyal and Welch (2008) lies in that they examine the forecasting power of fixed predictors using the whole sample information, whereas I study the real-time forecasting using only historical information available up to each prediction time. Supplementing Goyal and Welch's *ex post* results, this paper further shows that those popular predictors also have poor *ex ante* forecasting power even when carefully selected over time.

## C.2 The Sign of Excess Returns Forecasts

The predictive least squares criterion chooses optimal predictors according to the magnitude of their forecasting errors. Investors, however, probably consider the sign of forecasting to be equally important. Leitch and Tanner (1991) found that traditional measures like  $R^2$  fail to pick up predictors that generate equal profits from a trading strategy as a measure of the “sign” criterion. Given investors’ concerns regarding the sign of forecasting, I now examine whether the PLS criterion can choose predictors that also correctly predict the sign of excess returns.

[Figure 3 about here.]

Figure 3 presents the accuracy rate of timing the stock market under the PLS principle, which is measured by the percentage of correct signs of the excess returns predicted by macro indices. On average, the adaptive predictions get the same sign as the realized excess returns for 62 percent of all months in the trading period 1977:01 to 2006:12. Two interesting findings are worth noting. First, predictability from macro indices seems particularly high during the years of recessions and market downturns. In four recent recessions, the forecasting accuracy is locally high with respect to the years before and after recessions; the year of 2001 shows the correct predictions for 10 months with an accuracy rate of 83%, 67% for the year of 1991, 75% for 1980 and 67% for 1982. Three other spikes with the highest accurate forecasting rates occur in 1978 (83%), 1987 (75%) and 1995 (83%). Second, the degree to which the stock returns are predictable seems quite low before each recession.

The accuracy of forecasts both in magnitude and in sign exhibits a strong cyclical pattern. One possible explanation is that the predictability of stock returns may be particularly pronounced in periods of economic “regime switches” where the markets are relatively unsettled and investors are particularly uncertain of which forecasting model to use for trading (Pesaran and Timmermann



(1995)). Another possible explanation of the cyclical pattern is the difficulty investors have in identifying the turning points of business cycles when macro indices carry more noise.

### 3 Economic Significance of Predictability

The ability to predict stock returns ahead of time is the backbone of investment strategies, but predictability, in itself, does not guarantee that an investor can earn profits from a trading strategy that is based on such forecasts.<sup>17</sup> In this section I study the economic significance of the adaptive prediction to see whether an investor can exploit forecasts to make a higher profit than a buy-and-hold strategy in the market portfolio will bring.

#### A Case I: 100% Stock/Bond Allocation

A simple investor considers a hundred-percent switching strategy. According to this strategy, investors should hold stocks during periods when the economic condition indicators suggest that stock returns will outperform returns from holding risk-free bonds (i.e., the predicted excess market return is positive), and otherwise hold one-month Treasury bills. I assume that investors do not allow for the short-selling of assets, nor can they use leverage when selecting their portfolios.

Figure 4 shows the month-to-month cumulative returns for the buy-and-hold market portfolio and the switching portfolios. The switching portfolios are constructed using forecasts from macro indices or from predictors in Goyal and Welch (2008), selected adaptively according to the predictive least squares criterion. This figure confirms that the switching strategy based on

---

<sup>17</sup>Some authors report significant profits using *ex ante* model calibration, see Balduzzi and Lynch (1999, 2000), and Kandel and Stambaugh (1996). Other studies of real-time investment performance, however, document that the use of predictability fails to earn excess profits over the market. Carhart (1997) and Wermers (2000) find that the failure to mutual funds; Barber and Odean (2000) find it in individual investors; Christopherson, Ferson, and Glassman (1998) confirm that for pension funds; and Pirinsky (2001) documents the failure in banks, investment advisors, and insurance companies.

[Figure 4 about here.]

adaptively selected macro indices make great profits in a stable fashion. In contrast, the switching portfolio based on financial predictors earn lower returns than the market portfolio over the entire sample. The performance of the switching portfolios mirror the real-time predictability displayed in Figure 2, suggesting that macro indices have a consistent power in forecasting excess stock market returns, both statistically and economically.

Table 2 presents the trading results. The mean annual return on the market portfolio is 7.04% over the period of 1977:1 – 2005:12, and it is smaller than the mean return on the switching portfolio based on macro indices, 9.40%. In addition, the standard deviation of the returns on the switching portfolio by macro indices (12.76%) is substantially lower than the standard deviation of returns on the market portfolio (15.38%). Taken together, the higher mean and lower standard deviation of the returns on the switching portfolio by macro indices result in high values of the Sharpe ratio, 0.75, compared to 0.47, which is the Sharpe ratio for the market portfolio. In sharp contrast, the the switching portfolio based on financial predictors generates a Sharpe ratio of 0.43, and is inferior to the performance of the market portfolio.

I also analyze the performance of the switching portfolios over three subperiods: 1980 to 1989, 1990 to 1999, and 2000 to 2005. For all three subperiods, the trading portfolios based on forecasts using macro indices earn a higher average return than the buy-and-hold market portfolio. The difference is especially noticeable in the 2000s: the average excess equity return is -1.05%, while the realized return of the switching portfolio is 3.45%. Correspondingly, the Sharpe ratio of the market is -0.07 and that of the switching portfolio is 0.24. The performance of switching portfolios based on financial predictors is slightly inferior to the benchmark during the 1980s and 1990s, and become even worse in the 2000s.

A switching strategy requires an investor to determine when to move 100% into or out of stocks

**Table 2**  
**Real-time Performance of the Switching Portfolios**

The switching portfolios exploit forecasts of adaptive predictors from two base sets: A. macro indices and B. GW predictors. During each month of the trading period, investors choose the predictor with the smallest cumulative prediction errors in previous 24-months. If the resulting model predicts a positive excess return, investors buy one unit of CRSP stock index; if the resulting model predicts a negative excess return, investors buy one unit of risk-free one-month T-Bill. The returns on the trading portfolio are equal to the realized market returns or the one-month interest rate. The benchmark is the CRSP value-weighted portfolio with a buy-and-hold strategy. The real-time performance uses three measures.  $\mu$  refers to the average excess return of trading portfolios,  $\sigma$  is the standard deviation of realized trading portfolios, and  $SR$  means the Sharpe ratio. Mean returns and standard deviations are annualized percentage data. The model selection criterion is predictive least squares (PLS). The trading period is from January 1977 to December 2005.

Period	Market Portfolio			A: Macro Indices			B: GW Predictors		
	$\mu$	$\sigma$	SR	$\mu$	$\sigma$	SR	$\mu$	$\sigma$	SR
1977:1 - 2005:12	7.04	15.38	0.46	9.40	12.76	<b>0.75</b>	5.30	14.03	0.38
1980:1 - 1989:12	7.29	16.97	0.43	8.55	12.88	<b>0.66</b>	7.23	16.70	0.43
1990:1 - 1999:12	12.75	13.39	0.95	13.61	12.45	<b>1.09</b>	10.47	13.00	0.82
2000:1 - 2005:12	-1.05	16.15	-0.07	3.45	14.13	<b>0.24</b>	-4.47	12.52	-0.34

and bonds. Liquidating entire stock or bond portfolios, however, is very risky. Perhaps the most significant risk of market timing is missing out the stock market’s best performing periods. For example, suppose an investor, believing the market will go down, removes his investment dollars and conservatively places them in bonds. While the money is out of stocks, the market may enjoy its best-performing month. The investor has, therefore, incorrectly timed the market and “missed” those top months. That’s why most mutual fund managers in the real world — especially those striving toward long-term goals — often redistribute their assets among various assets and always hold both stocks and bonds at the same time, a case I study in the next subsection.

## B Case II: Utility-based Stock/Bond Allocation

Suppose an investor has a power utility function over end-of-period wealth at the investment horizon  $T$ . The investor chooses her asset allocation policy between a risky asset (the stock market portfolio) and a risk-free asset (the one-month Treasury Bill) to maximize her expected utility. At each point in time, the investor puts a fraction  $\alpha_t$  of her wealth in the risky asset. Thus the investor's problem is to solve the dynamic portfolio allocation

$$(5) \quad \max_{\{\alpha\}} E_0 \left[ e^{-\rho T} \frac{W_T^{1-\gamma}}{1-\gamma}, \right]$$

subjecting to the dynamic budget constraint:

$$\begin{aligned} dW_t/W_t &= \alpha_t dS_t/S_t + (1 - \alpha_t) dB_t/B_t, \\ dS_t/S_t &= \mu(z_t)dt + \sigma(z_t)dw_t, \\ dB_t/B_t &= r_f dt, \end{aligned}$$

where  $\rho$  is the discount rate,  $\gamma$  is the investor's risk-aversion coefficient. The conditional mean  $\mu$  and the conditional standard deviation  $\sigma$  of stock returns rely on the predictor variable  $z_t$ . The risk-free interest rate  $r_f$  is assumed to be constant over the investment horizon, which is one month.

In solving this dynamic programming problem,<sup>18</sup> the optimal portfolio allocation in the risky asset is

$$(6) \quad \alpha_t = \frac{1}{\gamma} \frac{\mu_t - r_f}{\sigma_t^2}$$

where  $\gamma$  is the local coefficient of relative risk aversion. The proportion invested in the risky asset

---

<sup>18</sup>Cochrane (2008) provides details for solving this dynamic programming problem.

depends on the investor’s risk aversion  $\gamma$  and the conditional distribution of the expected stock returns  $(\mu_t, \sigma_t)$ . In practice, an investor adaptively forecasts the conditional mean and the conditional variance based on the predictive regression. The allocation to the risky asset may rise or fall over time, determined by the investor’s ability to hedge the changes in economic conditions signaled by the optimal macro index, the “market timing” ability.

[Figure 5 about here.]

Figure 5 shows the optimal allocation to risky stocks over time, based on sequentially selected macro indices which reflect the changes of economic conditions over time. A common feature for both the allocation and the predictability is a strong cyclical pattern: reduced investment in risky stocks at the beginning of recession and increased investment at the end of recession. A strong position in stocks coincides with high stock market returns. A weak position in stocks accompanies relatively low stock market returns. The optimal allocation to stocks varies mildly, from -10 to 50 percent. As an additional evidence, the tight correlation of optimal allocations and the stock market returns shows that adaptive macro indices have the ability to capture the changes in economic conditions and this hence improves the investor’s ability of market timing.

Table 3 reports the real-time performance for utility-based portfolios, which exploit the same forecast results based on macro indices as in the case of the switching portfolio. The optimal portfolios have an average annual return of 6.60% over the whole out-of-sample period, which is smaller than the market portfolio, 7.04%. However, the optimal portfolios are much less volatile, with a standard deviation of 10.66%, compared to the market portfolio’s standard deviation of 15.38%. Overall the lower mean and even lower standard deviation of the returns on the optimal portfolio result in a slightly higher value of the Sharpe ratio, 0.62, compared to that of the market portfolio, 0.46. The utility-based portfolio do not perform as well as the market portfolio in the

**Table 3****Real-time Performance of the Utility-Based Optimal Portfolios**

The Utility-based optimal portfolios are constructed by allocating  $\alpha_t$  percent of the wealth to risky stocks and  $(1 - \alpha_t)$  to the one-month Treasury Bill at the beginning of each month. The allocation is calculated as  $\alpha_t = \frac{1}{\gamma} \frac{E_t(r^e)}{\sigma_t^2}$ , where expected excess returns  $E_t(r^e)$  come from the adaptive forecasts based on macro indices, and the variance of excess returns  $\sigma_t^2$  is estimated by the residual variance in predictive regressions:  $r_{t+1}^e = a_i + b_i z_{i,t} + \epsilon_{t+1}$ . The coefficient of risk aversion is set as unity in this table,  $\gamma = 1$ . The performance of trading portfolios over varying subperiods is measured by the average excess return ( $\mu$ ), the standard deviation ( $\sigma$ ) and the Sharpe ratio (SR). “Market Portfolio” refers to the CRSP value-weighted index. “Market Index” refers to utility-based optimal portfolios exploiting forecasts based on macro indices. Mean excess returns and the standard deviation are annualized data.

Period	Market Portfolio			Macro Index		
	$\mu$	$\sigma$	SR	$\mu$	$\sigma$	SR
1977:1 - 2005:12	7.04	15.38	0.46	6.60	10.66	<b>0.62</b>
1980:1 - 1989:12	7.29	16.97	0.43	6.92	12.46	<b>0.56</b>
1990:1 - 1999:12	12.75	13.39	<b>0.95</b>	7.36	11.01	0.67
2000:1 - 2005:12	-1.05	16.15	-0.07	4.52	8.94	<b>0.51</b>

1990s, with a much lower value of the Sharpe ratio (0.67 vs 0.95), but it noticeably outperforms the market in the early 2000s, with a significant mean return of 4.52%, a mild volatility of 8.94% and a Sharpe ratio of 0.51, which is much higher than the Sharpe ratio of the market portfolio, -0.07.

## C The Effects of Transaction Costs

For any investment strategy, two elements are important for investors: transaction costs and taxes. Transaction costs clearly impose a drag on the performance of investment strategies. Taxes also erode trading profits but are difficult to deal with, partly because they are investor- and investment-specific, and will therefore not be considered in this paper. For trading costs, I make a simplifying assumption that they are constant through time and symmetric with respect to whether the investor is buying or selling assets. I further assume that trading costs are simply proportional to the value of the trade, letting  $c_1$  and  $c_2$  be the percentage trading costs on stock shares and bonds,

respectively.<sup>19</sup> Besides zero trading costs, I consider two additions scenarios: “Low” transaction costs with 25 basis points on trading in one unit share of stock ( $c_1 = 0.0025$ ) and 5 basis points in one unit of bond ( $c_2 = 0.0005$ ), and “High” transaction costs with 50 basis points on trading in one unit share of stock ( $c_1 = 0.005$ ) and 5 basis points in one unit of bond ( $c_2 = 0.0005$ ).

The buy-and-hold strategy is a relatively passive investment strategy and it hence incurs low transaction costs. Compared with this benchmark, an investment strategy based on adaptive predictions is likely to incur considerably higher transaction costs and may not be profitable when taking into account transaction costs.<sup>20</sup> This subsection explores the performance of switching portfolios and utility-based portfolios under varying scenarios of trading costs and risk tolerance.

Table 5 presents the trading results. For each panel, I report the mean and the standard deviation of excess returns, and their Sharpe ratio. I also list the percentage of the total months when trading portfolios have negative returns and I show the final amount of wealth obtained from trading portfolios, assuming that investors start off with \$100 at the beginning of 1977 and reinvest portfolio income every month until the end of 2005. In the case of a buy-and-hold market portfolio, only the dividends are reinvested on a monthly basis. In contrast, the switching portfolios may reallocate funds between stocks and bonds, depending on whether a change in the sign of the excess return is predicted. The utility-based portfolios always have positions in both stocks and bonds but they adjust optimal allocations according to the forecasts of adaptive macro indices. A transaction cost is incurred only when investors switch or adjust positions.

First I consider the results for switching portfolios. Under all three transaction cost scenarios, the switching portfolios based on conditional forecasts by adaptively selected macro indices perform better than a buy-and-hold market portfolio: they have higher mean returns, higher Sharpe ratio and more final wealth. With zero cost the end-of-period funds of the switching portfolios are ap-

---

<sup>19</sup>This paper doesn’t consider other types of trading costs like commission fees and bid-ask spreads.

<sup>20</sup>As Vanguard founder Jack Bogle said, the strength of indexing (buy-and-hold strategy) stems primarily from its inherent cost advantage.

**Table 4****The Effect of Transaction Costs**

The columns headed Zero, Low, and High refer to the portfolio performance under three transaction cost scenarios: “Zero”, “Low” (25 basis points on trading in one unit share of stock and 5 basis points in one unit of the one month T-bill), and “High” (50 basis points on trading in one unit share of stock and 5 basis points in one unit of the one month T-bill). The benchmark is a buy-and-hold strategy applied to the market portfolio. Panel A shows the performance of switching portfolios based on conditional forecasts of adaptive macro indices or conditional forecasts of monthly predictors used in Goyal and Welch (2008). Panel B shows the performance of utility-based portfolios with different levels of risk-aversion ( $\gamma$ ). Negative performance is the percentage of the total months when portfolios have negative returns. Investors start off with \$100 at the beginning of 1977 and reinvest the portfolio income every month until the end of 2005. The model selection criterion is the predictive least squares (PLS). The rolling window for model selection has a length of 24 months. The trading period is from January 1977 to December 2005. Mean returns and the standard deviations are in percentage form and are annualized data.

Panel A: Switching Portfolios							
Transaction Costs	buy and hold	Macro Indices			GW Predictors		
		Zero	Low	High	Zero	Low	High
Mean return ( $\mu$ )	7.04	9.40	8.92	8.53	5.30	5.03	4.81
Std. Deviation ( $\sigma$ )	15.38	12.76	12.77	12.78	14.03	14.04	14.06
Sharpe Ratio (SR)	0.46	0.74	0.70	0.67	0.38	0.36	0.34
Negative Performance(%)	39.94	39.08	43.68	43.68	40.23	44.25	44.54
Final Wealth (\$)	541	<b>1195</b>	<b>1042</b>	<b>929</b>	347	321	300

Panel B: Utility-Based Optimal Portfolios							
Transaction Costs	buy and hold	Macro Indices ( $\gamma = 1$ )			Macro Index ( $\gamma = 3$ )		
		Zero	Low	High	Zero	Low	High
Mean return ( $\mu$ )	7.04	6.60	5.13	3.91	2.17	1.69	0.49
Std. Deviation ( $\sigma$ )	15.38	10.66	10.65	10.66	3.50	3.50	3.52
Sharpe Ratio (SR)	0.46	0.62	0.48	0.37	0.62	0.48	0.14
Negative Performance (%)	39.94	38.51	43.68	47.13	39.66	45.11	54.02
Final Wealth (\$)	541	<b>572</b>	374	263	188	163	114



proximately twice as large as the end-of-period funds of the market portfolio. In sharp contrast, the switching portfolio based on forecasts by financial/accounting ratios and individual macroeconomic series performs rather poorly with lower mean returns, lower Sharpe ratios and less final wealth.

Unlike the switching strategy, the utility-based investment strategy holds more conservative positions in risky assets; hence portfolios exploiting it earn lower returns and also have lower volatility. Using the same macro indices' forecasts as in the switching strategy, the utility-based optimal portfolio has only slightly higher values of the Sharpe ratio (0.62) than the buy-and-hold market portfolio (0.46) under the zero transaction cost scenario. The slight difference is also reflected in the end-of-period wealth accrued to the investment strategies: \$572 for the optimal portfolio and \$541 for the market portfolio.

In sum, active portfolio strategies provide incremental benefits to investors, but they seem sensitive to the level of transaction costs and investors' risk tolerance. The benefits will not be substantial to style investors if the expenses are significant or if investors have a high tolerance of risk.

## 4 Forecasting: *Ex Post* vs *Ex Ante*

Adaptive macro indices have shown both statistical and economic significance in forecasting stock market excess returns. A related interesting question is whether the predictability is due to adaptability or due to the dynamic weights of macro indices. The adaptive nature of the prediction procedure leads to the variation in both the indices selected and their loadings, which complicates the interpretation. To overcome this obstacle, I compare several types of forecasts using different information sets. Beginning with *ex post* forecasting, I construct indices using the entire sample and examine the performance of each index with fixed weights. Then I recursively construct indices using only the historical information available for each month. In this *semi ex post* forecasting, I

consistently choose the “same” index (principal components with the same order in explaining the variation of the panel data) and study the performance of each index with dynamic loadings. Finally, I revisit *ex ante* forecasting adopted in this paper, where I sequentially choose indices (principal components with different orders) rather than use the fixed index as in *ex post* forecasting. If there exists a consistent connection from *ex post* to *ex ante* forecasting, we may understand the important economic forces that contribute to equity premium predictions.

## A *Ex Post* Forecasting

I first construct fixed weight indices using the entire sample and study *ex post* forecasting performance for each fixed index.

[Figure 6 about here.]

Among all indices, I entertain both in-sample and out-of-sample (oos) tests to select the best predictor.<sup>21</sup> The second principal component (F2) stands out according to the criteria proposed by Goyal and Welch (2008): it has 1) both a significant in-sample and a reasonable good out-of-sample performance; 2) decreasing cumulative prediction errors compared with the benchmark (the prevailing mean); and 3) smaller cumulative prediction errors over the past several decades. Compared with the performance of adaptive macro indices in Figure 2, the fixed index F2 has weaker forecasting power, especially in the early 1980s. This evidence reinforces the importance of adaptability in real-time predictions.

[Figure 7 about here.]

---

<sup>21</sup>Goyal and Welch (2008) believe that the out-of-sample performance is an important diagnostic tool for predictability. However, Inoue and Kilian (2004), Campbell and Thompson (2008) and Cochrane (2008) argue that out-of-sample tests provide lower power than in-sample tests. These debates are ongoing, but researchers tend to agree that in-sample tests alone cannot identify a reliable predictive model. Out-of-sample tests should be a necessary supplement for in-sample performance, and a more reliable experiment should explore whether, when they are conditional on observed in-sample significance, out-of-sample diagnostics are still reasonably powerful.

Figure 6 shows that the index F2 has significant in-sample predictive power with an  $R^2$  value of 4.77%. The pseudo out-of-sample predictive performance is poor in the early 1980s, but it recovers after the recessions, further improves in the 1990s, and again drops during the bubble period and the 2001 recession. The out-of-sample  $R^2$  over the past three decades is 0.87%, which still outperforms the historical average at a 95% significance. To understand its economic meaning, I run a univariate regression of F2 on all the individual variables in the panel and report the marginal  $R^2$  in Figure 7. Clearly, F2 constructed from the whole sample mostly relates to sectors like interest rates and price indices. If projecting F2 onto each sector instead of the individual series, the price index sector explains 86.90 percent, whereas the interest rate sector explains 30.88 percent of the total variance in F2.

## B *Semi Ex Post* Forecasting

[Figure 8 about here.]

[Figure 9 about here.]

Relaxing the fixed-weight restriction, I now reconstruct indices using only the historical information and focus on the ex post performance of indices with dynamic loadings. Using the same criteria as in the fixed weight case, F2 and F5 surpass all other indices. Figure 8 presents the in-sample and pseudo out-of-sample performance for selected indices in Panel A and for popular predictors such as financial ratios and business cycles proxies in Panel B. F2 with dynamic loadings shares a similar forecasting pattern to F2 with fixed loadings, but the former differs with regard to higher and relatively stable predictive power. F5 predicts well until the early 1980s but gradually loses power, although it still outperforms the historical average over the past three decades. In addition, the performance of F2 and F5 is superior to that of financial ratios and business cycles proxies. These

**Table 5**  
**Pseudo Out-of-Sample Performance: 1977:1 - 2005:12**

This table presents pseudo out-of-sample performance of selected predictors. The performance uses four measures: in-sample and out-of-sample R-square ( $R_{in}^2$  and  $R_{oos}^2$ ), the difference of the root-mean-squared prediction error ( $\Delta RMSE$ ) and the MSE-F statistic (McCracken, 2004) to test equal forecast accuracy. The benchmark is the unconditional forecast equal to the historical average of excess market returns. The alternative is the conditional forecast using selected predictors listed in the first column. See Table 1 for formulas of the statistics. Macro indices are recursively constructed at each month of the trading period. Other financial predictors are from Goyal and Welch (2008), including dividend-price ratios, earnings-price ratios, book-to-market ratios, net issuing activity, term spread, default spread, default return and long-term bond return.

Predictor	$R_{in}^2(\%)$	$R_{oos}^2(\%)$	$\Delta RMSE$	MSE-F
Panel A: Significant Macro Indices				
F2	4.97	2.58	0.0581	9.85
F5	0.20	1.73	0.0389	6.54
Panel B: Selected Insignificant Macro Indices				
F1	0.90	-0.07	-0.0015	-0.25
F3	0.82	-0.29	-0.0060	-1.09
Panel C: Other Insignificant Predictors				
Dividend-Price Ratio	0.26	-0.87	-0.0194	-3.20
Earnings-Price Ratio	0.24	-1.31	-0.0293	-4.82
Book-to-Market Ratio	0.03	-1.36	-0.0304	-4.99
Net Issuing Activity	0.69	0.43	0.0096	1.60
Term Spread	0.87	-0.37	-0.0082	-1.36
Default Spread	0.72	0.16	0.0035	0.59
Default Return	0.05	-0.37	-0.0083	-1.37
Long-term Bond Return	1.02	-0.63	-0.0141	-2.33

results confirm the findings of poor out-of-sample performance for the well-documented predictors in Goyal and Welch (2008). Results of formal statistical tests are reported in Table 5.

Throughout the out-of-sample period, F2 and F5 by definition are the second and fifth principal components in explaining the total variance of the panel data. However, both their loadings and economic interpretations have evolved over time, in contrast to a sharp identity under the fixed weight scenario. Figure 9 illustrates the dynamic loadings of F2 and F5 for the period of 1975:1 to 2006:11.<sup>22</sup> The changing identities make the interpretation complex.

[Figure 10 about here.]

[Figure 11 about here.]

Yet behind the good performance of F2 and F5, do some economic intuitions exist? To gain insight into this question, I project the key indices onto each economic sector at each month to see which sectors relate the most to F2 and F5 over time. Table 6 reports the average, the minimum and the maximum values for the adjusted  $R^2$  across economic sectors. In spite of the varying meanings of each single index, the essential economic sectors that F2 and F5 remix still include interest rates and price indices, plus housing and employment.

## C Ex Ante Forecasting

The ex ante performance coincides with the ex post forecasting in the sense that F2 and F5 are also the most frequently selected predictors in the adaptive forecasting process. Figure 10 shows the top five macro indices most frequently selected over time and through different rolling windows. The value of the  $y$ -axis indicates whether a variable was included ( $y = 1$ ) or excluded ( $y = 0$ ) at

---

<sup>22</sup>The loadings are in the absolute form, which makes it easy to compare the relative importance of each component. The sign of the loadings may also be of interest, but the scale is sufficient here since the purpose is to illustrate the absolute correlation between components and the index.

**Table 6****Economic Meanings of the Best Performing Macro Indices**

Macro indices are recursively constructed during the period of 1975:1 to 2006:11. The best performing indices, F2 and F5, are selected according to their in-sample and pseudo out-of-sample performance reported in Table 5 and Figure 8. During each month I run the multiple regression of index F2 and F5 onto all variables in each economic sector. This table reports the average, the minimum and the maximum values for the adjusted  $R^2$  for twelve economic sectors that span the panel data. A detailed description of economic sectors are listed in Appendix A.1. The boldface blue numbers emphasize the sectors that are most closely correlated with indices F2 and F5.

Economic Sector	F2			F5		
	$\bar{R}_{mean}^2$	$\bar{R}_{min}^2$	$\bar{R}_{max}^2$	$\bar{R}_{mean}^2$	$\bar{R}_{min}^2$	$\bar{R}_{max}^2$
Real output	5.85	1.84	20.63	13.88	4.96	32.43
Unemployment	1.05	-1.25	2.23	15.31	5.04	35.39
Employment	<b>28.33</b>	21.38	53.13	<b>40.88</b>	4.97	75.17
Wages	15.82	5.78	19.68	4.60	0.62	27.11
Housing	19.50	8.41	43.89	<b>32.87</b>	3.47	47.51
Exchange Rates	5.65	0.99	11.30	4.69	-0.54	9.60
Money & Credit Aggregate	2.95	-2.06	6.24	6.76	-0.67	71.48
Interest Rates	<b>42.93</b>	11.78	56.93	5.78	0.42	43.41
Inventory and Orders	12.49	6.75	47.52	<b>22.44</b>	3.06	31.53
Price Indices	<b>72.35</b>	60.02	86.91	9.66	1.07	18.51
Consumer Confidence	1.76	0.32	10.42	1.80	0.00	4.33
Consumption	8.37	6.88	12.47	4.80	0.83	16.43

each time point for the “one month ahead” prediction of excess returns. Inclusion percentages are reported in parentheses. It is possible that a predictor occasionally gets included by chance but not because its predictive power. In such cases, however, it is unlikely that the predictor under consideration would be included in the adaptive process for long in subsequent periods. In contrast, when a forecasting variable is selected in a large proportion of the time periods and on a continuous basis, then it is reasonable to expect this predictor to be an important factor in generating the observed predictability of stock returns.

[Figure 12 about here.]

Under a 24-month selection window, F2 becomes an optimal predictor for one-third of the time in the past three years. If choosing an optimal predictor based on its previous 120-month performance, the included frequency of F2 increases to 50.57% which means F2 does well for half of the time over the course of three decades. After the second principal component, the index F5 is the second most frequently selected predictor. The included frequency for F5 is 6.96% using a 24-month rolling window; it increases to 23.19% with a ten year selection window. Since indices here are also recursively reconstructed in the same way as they were in *ex post* forecasting based on indices with dynamic loadings, it is natural to conclude that the real-time predictability also benefits from joint economic forces such as interest rates, price indices, housing and employment.

## D Summary

The increasing predictive power of indices comes at the cost of diminishing clarity in economic identity. Real-time prediction using adaptively selected macro indices has strong and robust performance, but obscure economic interpretations. The *ex post* prediction using fixed weight indices has a relatively poor performance but a sharp economic identity. The more adaptable the procedure is, the more predictive power it has.

In the best effort to understand the meaning of adaptive macro indices, I sequentially project each selected macro index onto the twelve economic categories and record economic categories that are most correlated with the selected predictor. Suppose at month  $t$ , the  $j$ th principal component is selected as the best predictor of future excess stock returns, I run the regression of  $F_t^j$  on all individual series in sector  $Z_t^i$ , where  $i$  denotes one of the twelve economic categories (real output, unemployment, employment, wages, housing, exchange rates, money and credit aggregates, interest

rates, inventory and orders, price indices, consumption and consumer expectation),

$$(7) \quad F_t^j = c' Z_t^i + \varepsilon_t^j.$$

If the selected macro index  $F^j$  contains the information most from sector  $i$ , for example the interest rates sector, then we expect to believe that interest rates play an important role in forecasting excess stock returns at that period. Figure 6 illustrates the dynamic meaning of adaptive macro indices. At each month, two solid blue squares indicate economic categories that is most correlated with the selected macro index, where the correlation is measured by the R-square value.

[Figure 13 about here.]

The major impression on Figure 6 is that economic forces which affect stock returns varies over time. In times of prosperity attention seems to focus on the problems that can accompany strong growth, such as inflation. Optimal predictive macro indices in expansion periods mainly capture information of economic categories such as consumer and producer price indices, and interest rates, which usually serve as measures of inflation. However, in other times like the early 2000, attention shifts to other indicators. Top of the list is the payrolls data which plot the number of jobs created or lost, a timely and sensitive indicator to the U.S. economy.

Accumulating over time, particular categories – interest rates, price indices, employment, and housing – contribute most to the predictability of the equity premium, in the sense that they significantly explain the meaning of optimal predictors.

## 5 Concluding Remarks

Real-time equity premium predictability is always a challenging task. We certainly do not claim that adaptive macro indices will outperform the historical average over every possible out-of-sample



period. Nevertheless, we find that the simple adaptive prediction model based on macro indices performs surprisingly well, certainly much more consistently than numerous individual predictive regression models from the literature. By recursively choosing the optimal index under the predictive least squares criterion, the conditional forecasts outperform both the unconditional forecasts (historical average) and forecasts conditional on prominent variables from the literature. It is robust — the ex ante forecasting performs well over the past three decades. More importantly, investment strategies exploiting real-time forecasts are able to earn excess profits over the market portfolio with moderate transaction costs.

Our evidence highlights the importance of using dynamically-measured economic conditions to investigate the real-time stock return prediction. The major predictive power comes from the dynamic weights of macro indices. By selecting the optimal combination of macroeconomic variables in the immediate past period, the concern of structural instability in the data-generating process gets alleviated. Such concern is usually difficult to handle with a single predictive regression model.

Macro indices explored in this paper provide a fresh opportunity to investigate the determinants of asset risk. The findings on the time-series behavior of excess returns can be linked to the large body of literature on cross-sectional asset pricing. According to the ICAPM of Merton (1973), innovations in state variables – only those state variables that are capable of predicting the expected returns over time – are likely to describe stochastic investment opportunities and hence command risk premia. Macro indices F2 and F5 are the most frequently selected optimal predictors in real-time forecasting; thus, innovations to them could be potential risk factors that explain the cross-sectional pattern of expected returns. No matter what these macro indices turn out to be, whether priced factors or merely good conditioning variables, the cross-sectional investigation takes us one step closer to understanding the main challenge of financial markets: what are the fundamental macroeconomic forces that drive risk premia in the stock market?

# Appendix

## A Data Description

### A.1 Macroeconomic Series

Following Stock and Watson (2002, 2005), I choose series to capture broad U.S. economic conditions. This panel data comprises 100 economic series, a subset of Stock and Watson’s 132 dataset excluding stock market data and various author-calculated spreads of interest rates. I also extends the panel to include the most recent available observations, a sample period from 1960:01 to 2006:11. The format is: series number, short name of each series, its mnemonic (the series label used in the source database), the transformation applied to the series, and a brief data description. All series come from the Global Insights Basic Economics Database. In the transformation column,  $\ln$  denotes logarithm,  $\Delta \ln$  and  $\Delta^2 \ln$  denote the first and second differences of the logarithm,  $lv$  and  $\Delta v$ , denote the level and the difference of the series.

Table 7: Macro Economic Series

	Short Name	Mnemonic	Tran	Description
<b>Real Output</b>				
1	IP: total	ips10	$\Delta \ln$	Industrial Production Index- Total Index
2	IP: products	ips11	$\Delta \ln$	Industrial Production Index - Products, Total
3	IP: final prod	ips299	$\Delta \ln$	Industrial Production Index - Final Products
4	IP: cons gds	ips12	$\Delta \ln$	Industrial Production Index - Consumer Goods
5	IP: cons dble	ips13	$\Delta \ln$	Industrial Production Index - Durable Consumer Goods
6	IP: cons nondble	ips18	$\Delta \ln$	Industrial Production Index - Nondurable Consumer Goods
7	IP: bus eqpt	ips25	$\Delta \ln$	Industrial Production Index - Business Equipment
8	IP: matls	ips32	$\Delta \ln$	Industrial Production Index - Materials
9	IP: dble matls	ips34	$\Delta \ln$	Industrial Production Index - Durable Goods Materials
10	IP: nondble matls	ips38	$\Delta \ln$	Industrial Production Index - Nondurable Goods Materials
11	IP: mfg	ips43	$\Delta \ln$	Industrial Production Index - Manufacturing (Sic)
12	IP: res util	ips307	$\Delta \ln$	Industrial Production Index - Residential Utilities
13	IP: fuels	ips306	$\Delta \ln$	Industrial Production Index - Fuels
14	NAPM prodn	pmp	$lv$	Napm Production Index (Percent)
<b>Unemployment</b>				
15	Help wanted indx	lhel	$\Delta v$	Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)
16	Help wanted/emp	lhelx	$\Delta v$	Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf
17	Emp CPS total	lhemp	$\Delta \ln$	Civilian Labor Force: Employed, Total (Thous.,Sa)
18	Emp CPS nonag	lhempnag	$\Delta \ln$	Civilian Labor Force: Employed, Nonagric.Industries (Thous.,Sa)
19	U: all	lhur	$\Delta v$	Unemployment Rate: All Workers, 16 Years Over (%Sa)
20	U: mean duration	lhu680	$\Delta v$	Unemploy.By Duration: Average(Mean)Duration In Weeks (Sa)
21	U ; 5 wks	lhu5	$\Delta \ln$	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)
22	U 5-14 wks	lhu14	$\Delta \ln$	Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)
23	U 15+ wks	lhu15	$\Delta \ln$	Unemploy.By Duration: Persons Unempl.15 Wks + (Thous.,Sa)

24	U 15-26 wks	lhu26	Δln	Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)
25	U 27+ wks	lhu27	Δln	Unemploy.By Duration: Persons Unempl.27 Wks + (Thous,Sa)

#### Employment

26	Emp: total	ces002	Δln	Employees On Nonfarm Payrolls: Total Private
27	Emp: gds prod	ces003	Δln	Employees On Nonfarm Payrolls - Goods-Producing
28	Emp: mining	ces006	Δln	Employees On Nonfarm Payrolls - Mining
29	Emp: const	ces011	Δln	Employees On Nonfarm Payrolls - Construction
30	Emp: mfg	ces015	Δln	Employees On Nonfarm Payrolls - Manufacturing
31	Emp: dble gds	ces017	Δln	Employees On Nonfarm Payrolls - Durable Goods
32	Emp: nondbles	ces033	Δln	Employees On Nonfarm Payrolls - Nondurable Goods
33	Emp: services	ces046	Δln	Employees On Nonfarm Payrolls - Service-Providing
34	Emp: TTU	ces048	Δln	Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities
35	Emp: wholesale	ces049	Δln	Employees On Nonfarm Payrolls - Wholesale Trade
36	Emp: retail	ces053	Δln	Employees On Nonfarm Payrolls - Retail Trade
37	Emp: FIRE	ces088	Δln	Employees On Nonfarm Payrolls - Financial Activities
38	Emp: Govt	ces140	Δln	Employees On Nonfarm Payrolls - Government
39	Avg hrs	ces151	lv	Avg Weekly Hrs of Workers On Private Nonfarm Payrolls - Goods-Producing
40	Overtime: mfg	ces155	lv	Avg Weekly Hrs of Workers On Private Nonfarm Payrolls-Mfg Overtime
41	NAPM empl	pmemp	lv	Napm Employment Index (Percent)

#### Wages

42	AHE: goods	ces275	Δln	Avg Hourly Earnings of Workers On Private Nonfarm Payrolls-Goods-Producing
43	AHE: const	ces277	Δln	Avg Hourly Earnings of Workers On Private Nonfarm Payrolls-Construction
44	AHE: mfg	ces278	Δln	Avg Hourly Earnings of Workers On Private Nonfarm Payrolls-Manufacturing

#### Housing

45	Starts: nonfarm	hsfr	ln	Housing Starts: Nonfarm(1947-58);Total Farm and Nonfarm(1959-)(Thous.,Saar)
46	Starts: NE	hsne	ln	Housing Starts:Northeast (Thous.U.)S.A.
47	Starts: MW	hsmw	ln	Housing Starts:Midwest(Thous.U.)S.A.
48	Starts: South	hssou	ln	Housing Starts:South (Thous.U.)S.A.
49	Starts: West	hswst	ln	Housing Starts:West (Thous.U.)S.A.
50	BP: total	hsbr	ln	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)
51	BP: NE	hsbne	ln	Houses Authorized By Build. Permits:Northeast(Thou.U.)S.A
52	BP: MW	hsbmw	ln	Houses Authorized By Build. Permits:Midwest(Thou.U.)S.A.
53	BP: South	hsbsou	ln	Houses Authorized By Build. Permits:South(Thou.U.)S.A.
54	BP: West	hsbwst	ln	Houses Authorized By Build. Permits:West(Thou.U.)S.A.

#### Exchange Rates

55	Ex rate: avg	exrus	Δln	United States;Effective Exchange Rate(Merm)(Index No.)
56	Ex rate: Switz	exrsw	Δln	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.\$ )
57	Ex rate: Japan	exrjan	Δln	Foreign Exchange Rate: Japan (Yen Per U.S.\$ )
58	Ex rate: UK	exruk	Δln	Foreign Exchange Rate: United Kingdom (Cents Per Pound)
59	EX rate: Canada	exrcan	Δln	Foreign Exchange Rate: Canada (Canadian \$ Per U.S.\$ )

#### Money and Credit Aggregates

60	M1	fm1	Δ <sup>2</sup> ln	Money Stock: M1(Curr,Trav.Cks,Dem Dep,Other Ck'able Dep)(Bil\$,Sa)
61	M2	fm2	Δ <sup>2</sup> ln	Money Stock:M2(M1+O'nite Rps,Euro\$,G/P&B/D Mmmfs&Sav&Sm Time Dep)(Bil\$,Sa)
62	M3	fm3	Δ <sup>2</sup> ln	Money Stock: M3(M2+Lg Time Dep,Term Rp's&Inst Only Mmmfs)(Bil\$,Sa)
63	MB	fmfba	Δ <sup>2</sup> ln	Monetary Base, Adj For Reserve Requirement Changes(Mil\$,Sa)
64	Reserves tot	fmrra	Δ <sup>2</sup> ln	Depository Inst Reserves:Total, Adj For Reserve Req Chgs(Mil\$,Sa)
65	Reserves nonbor	fmrnba	Δ <sup>2</sup> ln	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil\$,Sa)
66	Cons credit	ccinrv	Δ <sup>2</sup> ln	Consumer Credit Outstanding -Nonrevolving(G19)

#### Interest Rate

67	Fed Funds	fyff	Δlv	Interest Rate: Federal Funds (Effective) (% Per Annum,Nsa)
68	3 mo T-bill	fygm3	Δlv	Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.(% Per Ann,Nsa)
69	6 mo T-bill	fygm6	Δlv	Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.(% Per Ann,Nsa)
70	1 yr T-bond	fygt1	Δlv	Interest Rate: U.S.Treasury Const Maturities,1-Yr.(% Per Ann,Nsa)
71	3 yr T-bond	fygt3	Δlv	Interest Rate: U.S.Treasury Const Maturities,3-Yr.(% Per Ann,Nsa)
72	5 yr T-bond	fygt5	Δlv	Interest Rate: U.S.Treasury Const Maturities,5-Yr.(% Per Ann,Nsa)
73	10 yr T-bond	fygt10	Δlv	Interest Rate: U.S.Treasury Const Maturities,10-Yr.(% Per Ann,Nsa)
74	Aaa bond	fyaaac	Δlv	Bond Yield: Moody's Aaa Corporate (% Per Annum)
75	Baa bond	fybaac	Δlv	Bond Yield: Moody's Baa Corporate (% Per Annum)

#### Inventories and Orders

76	PMI	pmi	lv	Purchasing Managers' Index (Sa)
77	NAPM new ordrs	pmno	lv	Napm New Orders Index (Percent)

78	NAPM vendor del	pmdel	lv	Napm Vendor Deliveries Index (Percent)
79	NAPM Invent	pmnv	lv	Napm Inventories Index (Percent)
80	Orders:coms gds	mocmq	$\Delta$ ln	New orders -consumer goods & materials, 1996 dollars (BCI)
81	Orders:cap gds	msondq	$\Delta$ ln	New orders, nondefense capital goods,in 1996 dollars(BCI)
<b>Price indices</b>				
82	PPI: fin gds	pwfsa	$\Delta$ ln	Producer Price Index: Finished Goods (82=100,Sa)
83	PPI: cons gds	pwfcsa	$\Delta$ ln	Producer Price Index: Finished Consumer Goods (82=100,Sa)
84	PPI: int mat'ls	pwmsa	$\Delta$ ln	Producer Price Index:Intermed Mat.Supplies & Components(82=100,Sa)
85	PPI: crude mat'ls	pwcmsa	$\Delta$ ln	Producer Price Index: Crude Materials (82=100,Sa)
86	NAPM com price	pmcp	lv	Napm Commodity Prices Index (Percent)
87	CPI-U: all	punew	$\Delta$ ln	Cpi-U: All Items (82-84=100,Sa)
88	CPI-U: apparel	pu83	$\Delta$ ln	Cpi-U: Apparel & Upkeep (82-84=100,Sa)
89	CPI-U: transp	pu84	$\Delta$ ln	Cpi-U: Transportation (82-84=100,Sa)
90	CPI-U: comm.	puc	$\Delta$ ln	Cpi-U: Commodities (82-84=100,Sa)
91	CPI-U: dbles	pucd	$\Delta$ ln	Cpi-U: Durables (82-84=100,Sa)
92	CPI-U: services	pus	$\Delta$ ln	Cpi-U: Services (82-84=100,Sa)
93	CPI-U: ex food	puxf	$\Delta$ ln	Cpi-U: All Items Less Food (82-84=100,Sa)
94	CPI-U: ex shelter	puxhs	$\Delta$ ln	Cpi-U: All Items Less Shelter (82-84=100,Sa)
95	CPI-U: ex med	puxm	$\Delta$ ln	Cpi-U: All Items Less Midical Care (82-84=100,Sa)
<b>Misellaneous</b>				
96	Consumer expect	hhsntn	$\Delta$ lv	U. Of Mich. Index Of Consumer Expectations(Bcd-83)
<b>Consumption</b>				
97	PCE defl	gmde	$\Delta$ ln	Pce, Impl Pr Defl:Pce (1987=100)
98	PCE defl: dlbes	gmdd	$\Delta$ ln	Pce, Impl Pr Defl:Pce; Durables (1987=100)
99	PCE defl: nondble	gmddn	$\Delta$ ln	Pce, Impl Pr Defl:Pce; Nondurables (1996=100)
100	PCE defl: service	gmdds	$\Delta$ ln	Pce, Impl Pr Defl:Pce;Services (1987=100)

## A.1 Predictors in Goyal and Welch (2008)

In the monthly prediction, Goyal and Welch (2008) uses the following predictors: the dividend price ratio, the earnings price ratio, the dividend-earnings ratio, the book-to-market ratio, the long-term government bond return, net equity expansion, inflation measured by the first difference of the logarithm in Consumer Price Index (All Urban Consumers), the term spread, the default yield spread, and the default return spread. For a detailed description and data sources, please see the original paper. In this paper, I use these data for a sample period of January 1960 to December 2005.

## B Real-time Availability of Macro Data

The availability of economic data in real time is an important concern. The data obtained by economists may not be historically available to investors. For example, the Consumer Price

**Table 8****Example of Data Revision: CPI-U**

The Consumer Price Index for All Urban Consumers (CPI-U) is one variable in the panel data listed in Appendix A.1. Both the unadjusted and the adjusted data come from the archives of economic news release in the Bureau of Labor Statistics. The adjusted data are also the same ones used in this paper from the Global Insight Basic Economic Database. The difference is the adjusted number minus the unadjusted one. The noise ( $0.1 * \sigma_{i,s} * e$ ) is generated according to the formula (8).

Time	Unadjusted (Original Release)	Adjusted (Final Release)	difference	Noise
Nov 2006	201.5	201.7	-0.2	-0.34
Oct 2006	201.8	201.7	0.1	0.22
Sep 2006	202.9	202.7	0.2	0.23
Aug 2006	203.9	203.7	0.2	-0.06

Index for All Urban Consumers (CPI-U) is subject to revision for up to five years after its original release. Each year, the previous five years of data are seasonally adjusted using annually-computed seasonal factors; data from January 2002 through December 2006 were replaced in January 2007. As another example, all unadjusted Producer Price Indices (PPI) are routinely subject to a single revision, four months after the original publication,<sup>23</sup> to reflect late reports and corrections by respondents. Despite various rules, revisions of macroeconomic data usually have two common features: 1) indices are considered final once revised; and 2) revisions are usually minor, especially for highly aggregated data.

[Figure 14 about here.]

<sup>23</sup>The Bureau of Labor Statistics does not use the term “preliminary” to describe the originally released numbers, because preliminary usually describes data that are based on a small information sample, typically subject to large revisions. When Consumer Price Indices and Producer Price Indices are first released, they are typically based on a substantial portion of the total number of returns that eventually will be received from respondents. Hence, subsequent revisions are normally minor, especially at the more highly aggregated grouping levels. “First published” and “originally released” are more appropriate terms than “preliminary”. Changes in previously published data caused by a processing error are so indicated in a subsequent news release or detailed report; such occurrences are rare.

To check the robustness of the predictability results, I add noise to the data used throughout this paper and conduct the same experiment. Specifically, the new data come from the final released observations plus one-tenth of the standard deviation in each year multiplied by a random number generated from the standard normal distribution,

$$(8) \quad x_{i,t}^{new} = x_{i,t} + \text{noise} = x_{i,t} + 0.1 \cdot \sigma_{i,s} \cdot e, \quad e \sim N(0, 1)$$

where  $\sigma_{i,s}$  is the standard deviation of the variable  $x_i$  in the year  $s$ . The seemingly arbitrary number “0.1” actually effectively captures the magnitude of revision. I collect the historical releases of the Consumer Price Index (CPI-U) for the period of August to November 2006. Table 8 lists the originally released numbers, the adjusted numbers (identical to those used in the paper), their difference and the noise generated using formula (8). The noise might not have the same sign as the difference, but they at least have the same magnitude. Hence the noise-added data capture the revision in the real world to some extent.

Figure 12 presents the performance for the noise-added data. The performance is based on Monte Carlo Simulations with 100 iterations based on the revision formula (8). They are slightly different from the performance using the adjusted data (used in this paper), but generally the performance remains robust. Indeed, Stock and Watson (2002) provide both theoretical arguments and empirical evidence that factors estimated by the principal component analysis are consistent even in the face of temporal instability in the individual time series used to construct the factors. The reason for this is because the construction of indices may “average out” such instability from one series to the other.

## References

- [1] Ang, Andrew and Geert Bekaert, 2007, “Stock Return Predictability: Is It There?,” *Review of Financial Studies* 20, 651 – 707.
- [2] Avramov, Doron, 2002, “Stock Return Predictability and Model Uncertainty,” *Journal of Financial Economics*, Vol. 64, No. 3, 423 – 458.
- [3] Bai, Jushan, and Serena Ng, 2002, “Determining the Number of Factors in Approximate Factor Models,” *Econometrica* 70 (1), 191 – 221.
- [4] Bai, Jushan, and Serena Ng, 2006, “Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions,” *Econometrica* 74(4), 1133 – 1150.
- [5] Balduzzi, Pierluigi and Anthony W. Lynch, 1999, “Transaction Costs and Predictability: Some Utility Cost Calculations,” *Journal of Financial Economics* 54, 47 – 78.
- [6] Balduzzi, Pierluigi and Anthony W. Lynch, 2000, “Predictability and Transaction Costs: The Impact on Rebalancing Rules and Behavior,” *Journal of Finance* 55, 2285 – 2310.
- [7] Barber, Brad, and Terrance Odean, 2000, “Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors,” *Journal of Finance* 55, 773 – 806.
- [8] Bauer Rob, Jeroen Derwall, and Roderick Molenaar, 2004, “The Real-time Predictability of the Size and Value Premium in Japan,” *Pacific-Basin Finance Journal* 12, 503 – 523.
- [9] Bernanke, Ben, and Jean Boivin, 2003, Monetary Policy in a Data-Rich Environment, *Journal of Monetary Economics* 50(3), 525 – 546.
- [10] Bernanke, Ben, Jean Boivin and Piotr S. Eliaz, 2005, “Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach,” *The Quarterly Journal of Economics* 120(1), 387 – 422.
- [11] Boivin, Jean, and Marc Giannoni, 2006, “DSGE Models in a Data-Rich Environment,” Unpublished paper, Columbia University.
- [12] Bossaerts, Peter, and Pierre Hillion, 1999, “Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?” *Review of Financial Studies* 12, 405 – 428.
- [13] Brennan, Michael. J., and Yihong Xia, 2004, “Persistence, Predictability, and Portfolio Planning,” Working Paper, UCLA.
- [14] Butler, Alexander W., Gustavo Grullon, and James P. Weston, 2005, “Can Managers Forecast Aggregate Market Returns?” *Journal of Finance* 60, 963 – 986,

- [15] Campbell, S. D. and F. X. Diebold, 2009, "Stock Returns and Expected Business Conditions: Half a Century of Direct Evidence," *Journal of Business & Economic Statistics* 27, 266-278.
- [16] Campbell, John Y., and Samuel Thompson, 2007, "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?" Forthcoming, *Review of Financial Studies*.
- [17] Campbell, John Y. and Luis M. Viceira, 1999, "Consumption and Portfolio Decisions When Expected Returns are Time Varying", *Quarterly Journal of Economics* 114, 433 – 495.
- [18] Campbell, John Y., and Luis M. Viceira, 2002, "Strategic Asset Allocation: Portfolio Choice for Long-Term Investors," Oxford University Press, Oxford, UK.
- [19] Campbell, John Y., and Motohiro Yogo, 2006, "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics* 81, 27 – 60.
- [20] Carhart, Mark, 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57 – 82.
- [21] Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, "Inference in Models with Nearly Integrated Regressors," *Econometric Theory* 11(5), 1131 – 1147.
- [22] Christopherson, Jon A., Wayne E. Ferson, and Debra A. Glassman. 1998, "Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance," *Review of Financial Studies* 11, 111 – 142.
- [23] Cochrane, John H., 1991, "Production-based Asset Pricing and the Link between Stock Returns and Macroeconomic Fluctuations," *Journal of Finance* 46, 209 – 238.
- [24] Cochrane, John H., 1999, "Portfolio Advice for a Multifactor World," *Economic Perspectives* 23 (3), Federal Reserve Bank of Chicago, 59 – 78.
- [25] Cochrane, John H., 2005, *Asset Pricing*, Princeton University Press, Princeton, Revised Edition.
- [26] Cochrane, John H., 2007a, "Financial Markets and the Real Economy," in John H. Cochrane, ed., *Financial Markets and the Real Economy*, Volume 18 of the International Library of Critical Writings in Financial Economics, London: Edward Elgar, p. xi-lxix.
- [27] Cochrane, John H., 2007b, "The Dog That Did Not Bark: A Defense of Return Predictability," Forthcoming, *Review of Financial Studies*.
- [28] Cochrane, John H., 2007c, "Portfolio Theory", a new chapter in the new revision of *Asset Pricing*.



- [29] Cooper, Michael and Huseyin Gulen, 2002, "Is Time-Series Based Predictability Evident in Real Time?" *Journal of Business* 79, 1263 – 1292.
- [30] Elias, Piotr, 2005, "Optimal Median Unbiased Estimation of Coefficients on Highly Persistent Regressors," Unpublished paper, Princeton University.
- [31] Emanuel Monch, 2006, "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach," working paper, Humboldt University.
- [32] Fama, Eugene F. and Kenneth R. French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* 25, 23 – 49.
- [33] Ferson, Wayne E. and Merrick John Jr., 1987, "Non-stationarity and stage-of-the-business-cycle effects in consumption-based asset pricing relations," *Journal of Financial Economics* 18(1), 127 – 146.
- [34] Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, "Spurious regressions in financial economics?" *Journal of Finance* 58, 1393 – 1413.
- [35] Forni, Mario, Marc Hallin, Marco Lippi and Lucrezia Reichlin, 2005, "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association* 100, 830 – 840.
- [36] Gomes, João F., Leonid Kogan, and Motohiro Yogo, 2007, "Durability of Output and Expected Stock Returns," working paper, University of Pennsylvania.
- [37] Goyal, Amit and Ivo Welch, 2007, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," Forthcoming, *Review of Financial Studies*.
- [38] Guo Hui, 2006, "On the Out-of-Sample Predictability of Stock Market Returns." *Journal of Business* 79, 645 – 670.
- [39] Hansen, Lars P., 2007, "Belief, Doubts and Learning: Valuing Macroeconomic Risk," Richard T. Ely Lecture at the ASSA/AEA Conference at Chicago.
- [40] Inoue, A., and L. Kilian, 2004, "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews* 23(4), 371 – 402.
- [41] Kandel, Shmuel, and Robert F. Stambaugh, 1996, "On the Predictability of Stock Returns: An Asset-Allocation Perspective," *Journal of Finance* 51, 385 – 424.
- [42] Leitch, Gordon, and Ernest J. Tanner, 1991, "Economic Forecast Evaluation: Profits versus the Conventional Error Measures," *American Economic Review* 81, 580 – 590.
- [43] Lettau, Martin, and Sydney Ludvigson, 2001, "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance* 56, 815 – 849.

- [44] Lettau, Martin, and Sydney Ludvigson, 2005, "Measuring and Modeling Variation in the Risk-Return Tradeoff," Forthcoming in the *Handbook of Financial Econometrics*, edited by Yacine Ait-Shalia and Lars-Peter Hansen.
- [45] Lettau, Martin, and Stijn Van Nieuwerburgh, 2007, "Reconciling the Return Predictability Evidence," Forthcoming, *Review of Financial Studies*.
- [46] Lewellen, John W., 2004, "Predicting Returns With Financial Ratios," *Journal of Financial Economics* 74, 209 – 235.
- [47] Ludvigson, Sydney, and Serena Ng, 2007a, "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics* 83, 171 – 222.
- [48] Ludvigson, Sydney, and Serena Ng, 2007b, "Macro Factors in Bond Risk Premia," Working paper, New York University.
- [49] McCracken, Michael W., 2007, "Asymptotics for Out-of-Sample Tests of Causality," *Journal of Econometrics* 140, 719 – 752.
- [50] Nelson, Charles R., and Myung J. Kim, 1993, "Predictable stock returns: The role of small sample bias," *Journal of Finance* 48, 641 – 662.
- [51] Pstor, Lubos and Robert F. Stambaugh, 2001, "The Equity Premium and Structural Breaks," *Journal of Finance* 56, 1207 – 1239.
- [52] Paye, Bradley S., and Allan Timmermann, 2006, "Instability of Return Prediction Models," *Journal of Empirical Finance* 13(3), 274 – 315.
- [53] Pesaran, Hashem M., and Allan Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance* 50, 1201 – 1228.
- [54] Pesaran, Hashem M., and Allan Timmermann, 2004, "How costly is it to ignore breaks when forecasting the direction of a time series?" *International Journal of Forecasting* 20(3), 411-425.
- [55] Pettenuzzo, Davide and Allan Timmermann, 2011, "Predictability of Stock Returns and Asset Allocation under Structural Breaks." *Journal of Econometrics* 164, 60-78.
- [56] Piazzesi, Monika, Martin Schneider and Selale Tuzel, 2007, "Housing, Consumption, and Asset Pricing," *Journal of Financial Economics* 83, 531-569.
- [57] Pirinsky, C. 2001, "Are Financial Institutions Better Investors?" Unpublished paper, Ohio State University.
- [58] Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2006, "Cross-sectional forecasts of the equity risk premium," *Journal of Financial Economics* 81, 101 – 141.

- [59] Rapach, D. E. and M. E. Wohar (2006), “In-sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining,” *Journal of Empirical Finance* 13, 2381–247.
- [60] Rapach, D. E., J. K. Strauss, and G. Zhou (2010), “Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy,” *Review of Financial Studies* 23, 821 – 862.
- [61] Rissanen, Jorma, 1986, “A Predictive Least-Squares Principle,” *Journal of Mathematical control & Information* 3, 211 – 222.
- [62] Sargent, Thomas J., 1999, *The Conquest of American Inflation*, Princeton University Press, Princeton.
- [63] Stambaugh, Robert F., 1999, “Predictive Regressions,” *Journal of Financial Economics* 54, 375 – 421.
- [64] Stock, James H., and Mark W. Watson, 1999, “Forecasting Inflation,” *Journal of Monetary Economics* 44, 293 – 335.
- [65] Stock, James H., and Mark W. Watson, 2002a, “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business & Economic Statistics* 20(2), 147 – 162.
- [66] Stock, James H., and Mark W. Watson, 2002b, “Forecasting Using Principal Components From a Large Number of Predictors,” *Journal of the American Statistical Association* 97, 1167 – 1179.
- [67] Stock, James H., and Mark W. Watson, 2005, “Implications of Dynamic Factor Models for VAR Analysis,” NBER Working Paper No. 11467.
- [68] Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, “On Predicting Stock Returns with Nearly Integrated Explanatory Variables,” *Journal of Business* 77(4), 937 – 966.
- [69] Valkanov, Rossen, 2003, “Long-horizon regressions: Theoretical results and applications,” *Journal of Financial Economics* 68, 201 – 232.
- [70] Viceira, Luis, 1996, “Testing for Structural Change in the Predictability of Asset Returns,” Unpublished paper, Harvard University.
- [71] Wei, C. Z., 1992, “On Predictive Least Squares Principles,” *The Annals of Statistics* 20(1), 1 – 42.
- [72] Wermers, R., 2000, “Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses,” *Journal of Finance* 55, 1655 – 1695.

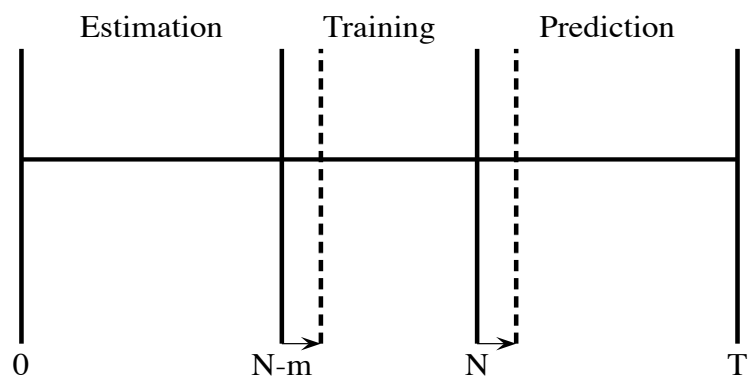
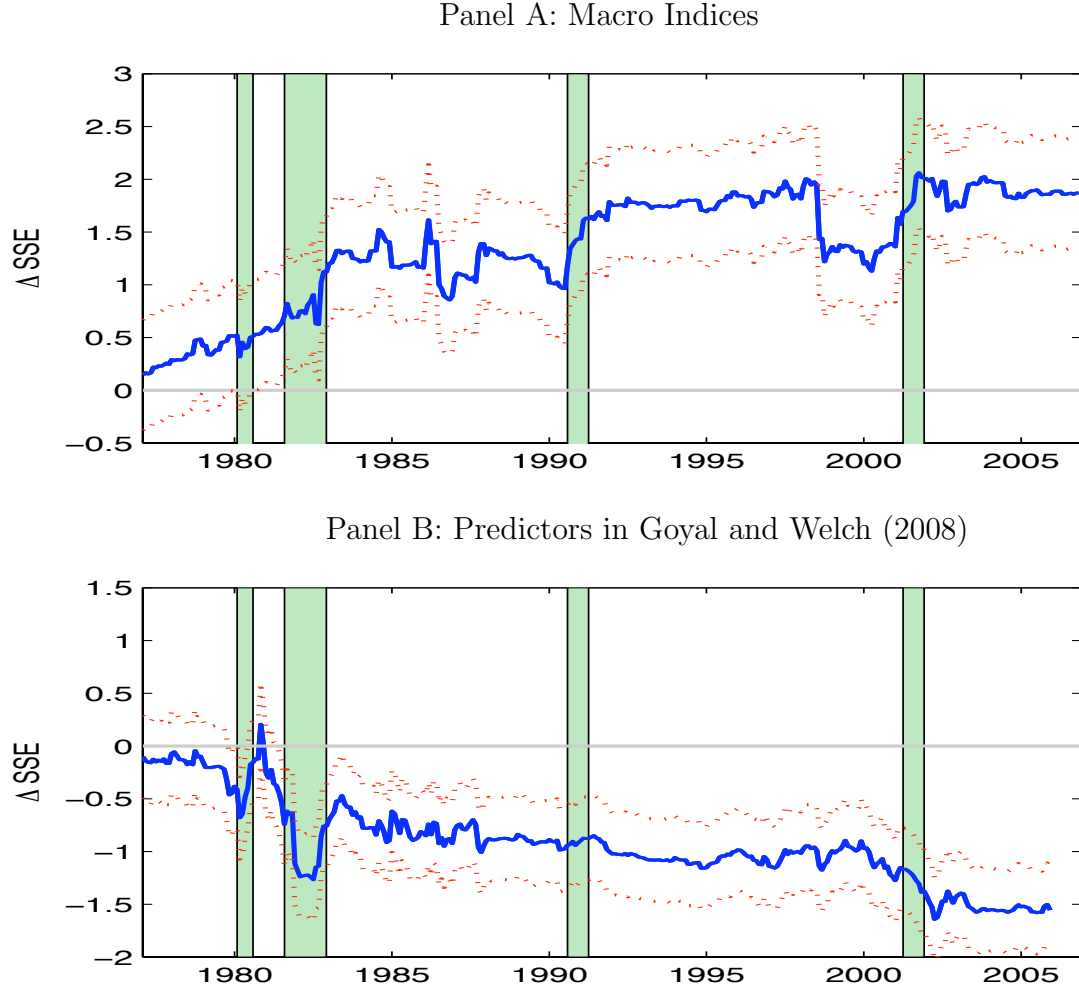


Figure 1: **Graphical Illustration of Real-time Forecasting**



**Figure 2: Out-of-Sample Performance of Monthly Prediction Models.** The performance is measured by the cumulative squared prediction errors of the benchmark ( $u_{B,t}^2$ ) minus those of the alternative ( $u_{A,t}^2$ ),  $\Delta SSE = \sum_t (u_{B,t}^2 - u_{A,t}^2)$ . The benchmark is the unconditional forecast equal to the historical average. The alternative is the conditional forecast using adaptively selected predictors from two base sets: Panel A. macro indices constructed from a panel of 100 economic variables using principal component analysis; and Panel B. 10 monthly predictors in Goyal and Welch (2008). The forecasting horizon is one month. The prediction period is from January 1977 to December 2005. Dotted red lines denote  $\Delta SSE$  minus/plus one standard deviation. Shading denotes months of recessions identified by NBER.

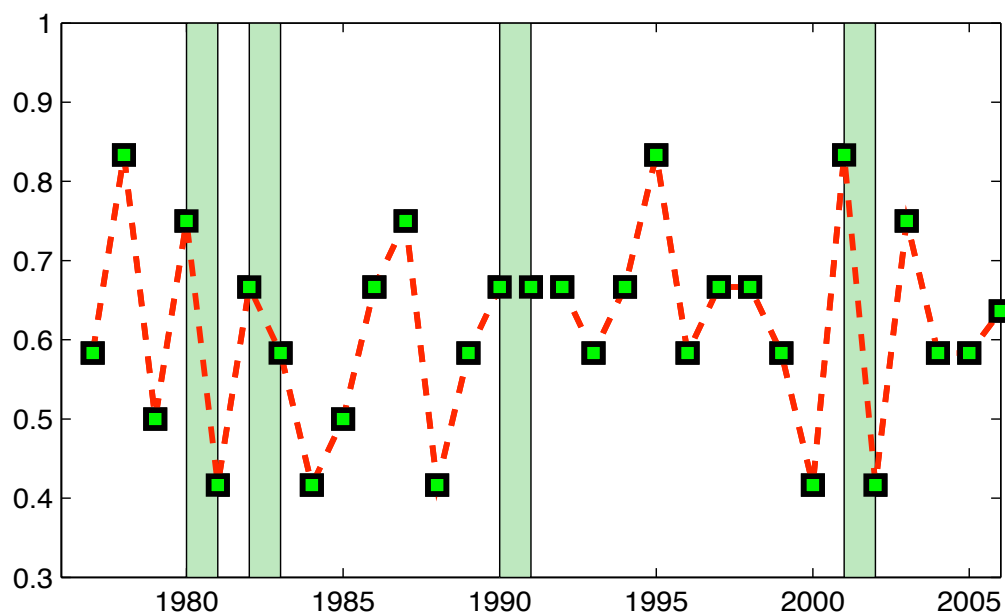


Figure 3: **Accuracy Rate of Excess Return Forecasts**, measured by the proportion of correct signs predicted by macro indices in each year of the trading period: 1977:1 - 2006:11. For example, an 83% accuracy rate means that excess return forecasts get the same sign as the realized returns for ten months within a year. Shading denotes years of recessions identified by NBER.

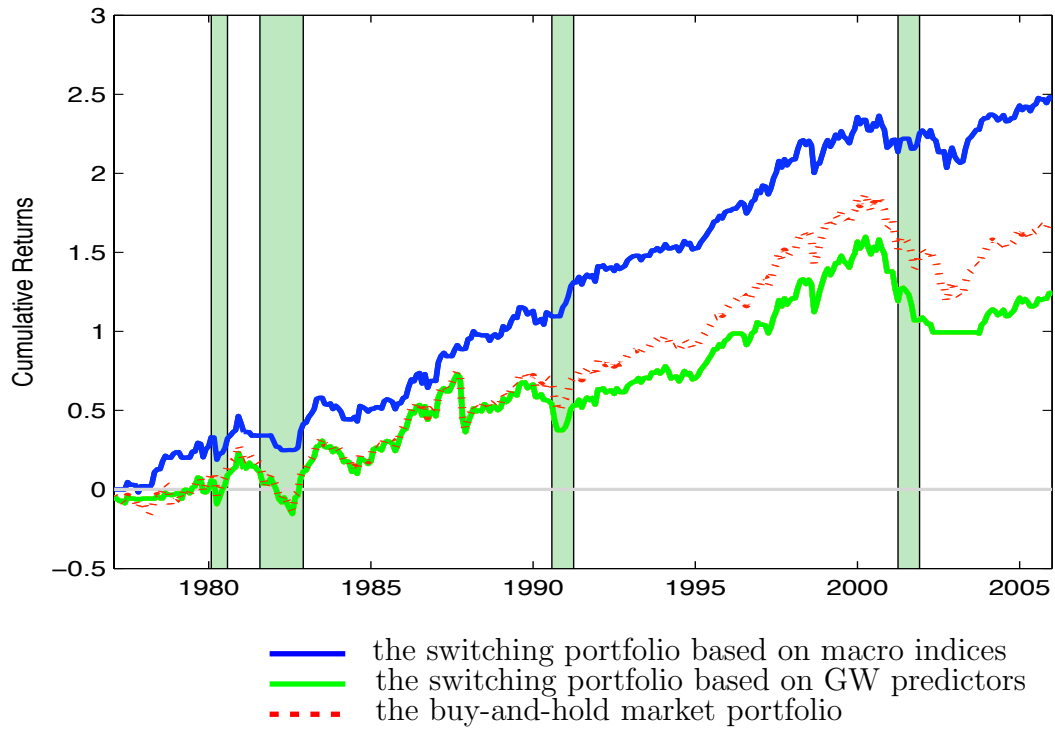


Figure 4: **Cumulative Month-to-Month Excess Returns.** The benchmark market portfolio is the CRSP value-weighted index. All numbers are raw returns in excess of the one-month Treasury bill rate at that month. The trading portfolios have a 100% position in stocks if the conditional forecast of future excess return is positive, and a 100% position in bonds if the forecast is negative. Shading denotes months of recessions designated by NBER.

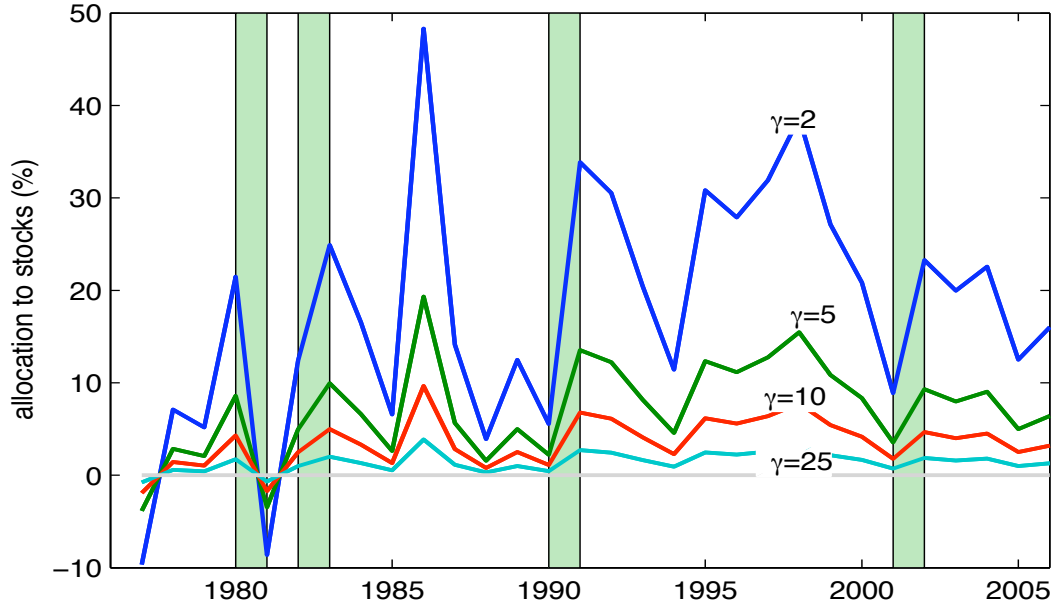


Figure 5: **Optimal Portfolio Allocation Over Time.** The allocation to risky stocks is  $\alpha_t = \frac{1}{\gamma} \frac{E_t(r^e)}{\sigma_t^2}$ . Expected excess returns come from the adaptive forecasts based on macro indices,  $r_{t+1}^e = \hat{a}_i + \hat{b}_i z_{i,t}$ . The variance of excess returns is estimated by the variance of predictive regression residuals.  $\gamma$  is the coefficient of risk aversion. This figure shows the annual allocation, which is calculated as the average of monthly allocations. The sample period is from January 1977 to November 2006. Shading denotes recessions identified by NBER.



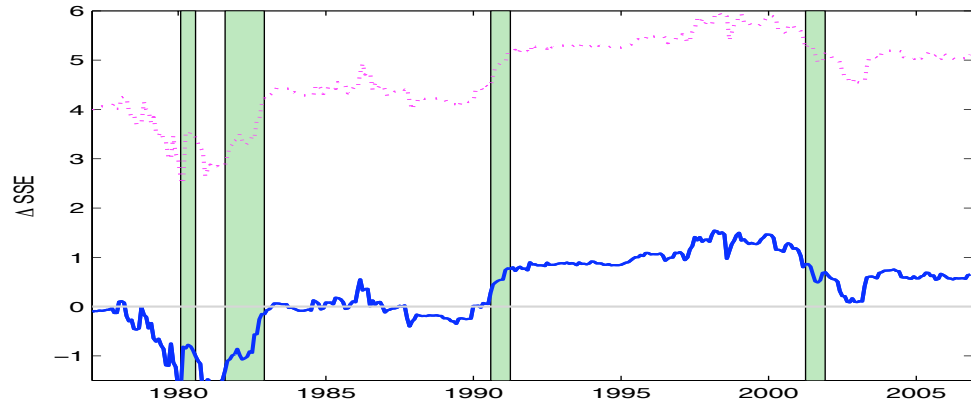


Figure 6: **Pseudo Out-of-sample Performance of F2.** The index F2 is constructed using the full sample of 1960:1-2006:11. The solid blue line (—) is the difference of cumulative prediction errors from the benchmark (the historical mean). The dotted red line (- -) is the difference of cumulative fitting errors from the benchmark. Shading denotes months of recessions designated by NBER.

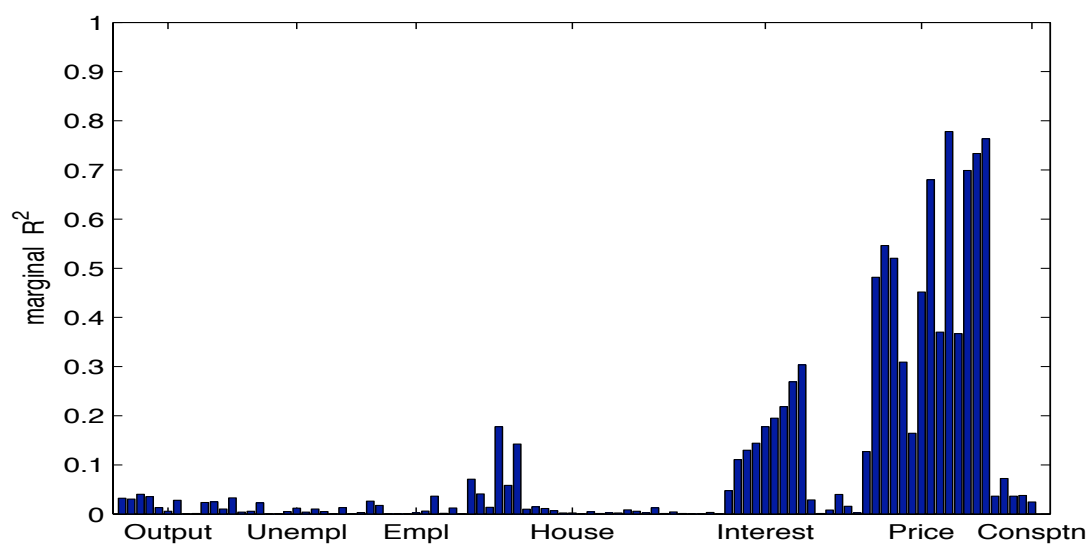


Figure 7: **Marginal  $R^2$  of F2**, calculated from the univariate regression of F2 on 100 individual variables in the panel data. The index F2 is constructed using the full sample of 1960:1-2006:11.

Panel A: Selected Macro Indices

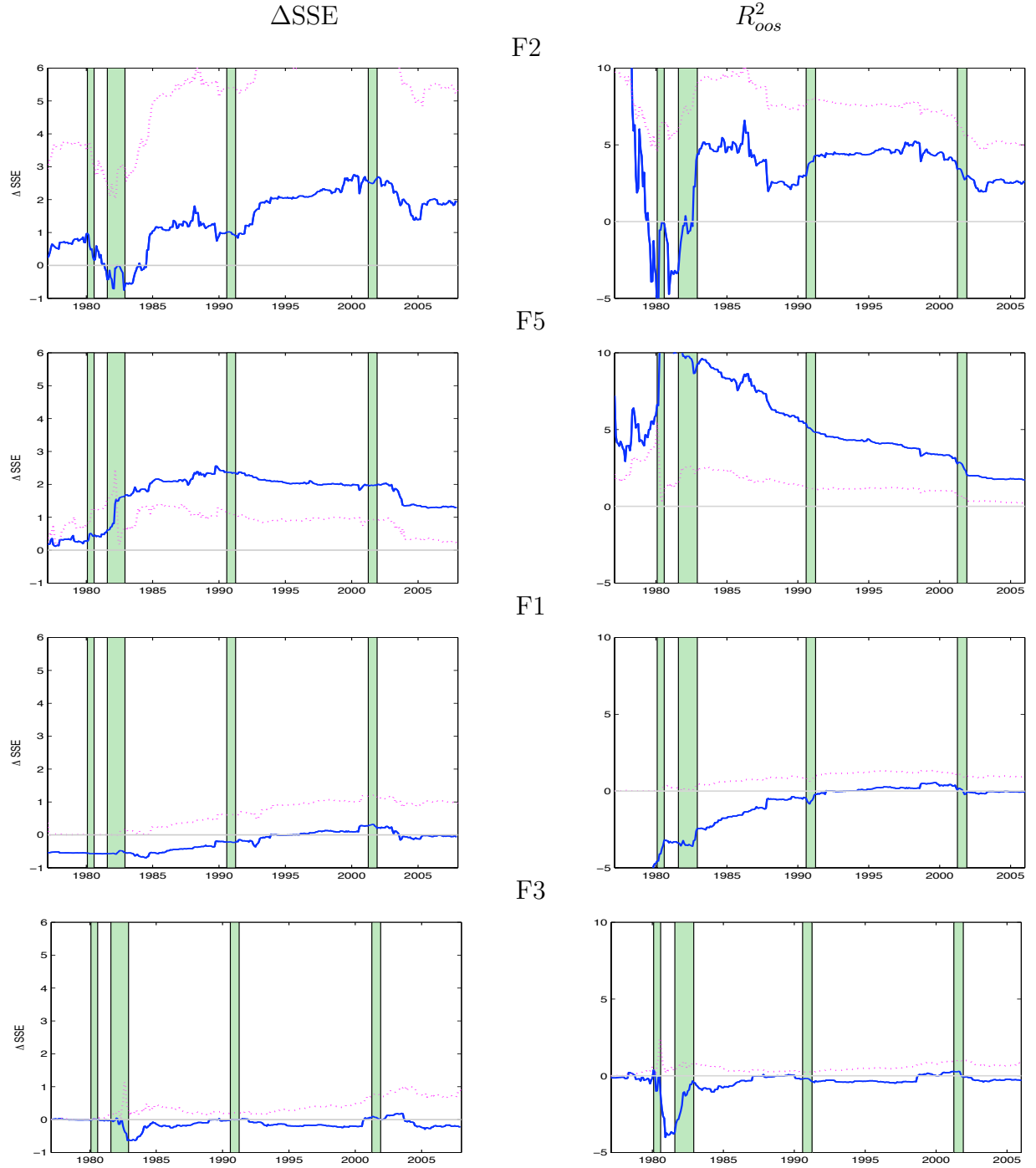


Figure 8: **In-sample (—) and Pseudo Out-of-sample (—) Performance of *Semi Ex Post* Forecasting.**  $\Delta SSE$  measures the difference in cumulative squared prediction errors from the benchmark (the historical average) to the conditional forecasts based on macro indices. The out-of-sample  $R^2_{oos}$  statistic is computed as it is in Campbell and Thompson (2008). Shading denotes months of recessions identified by NBER.

Panel B: Selected Popular Predictors

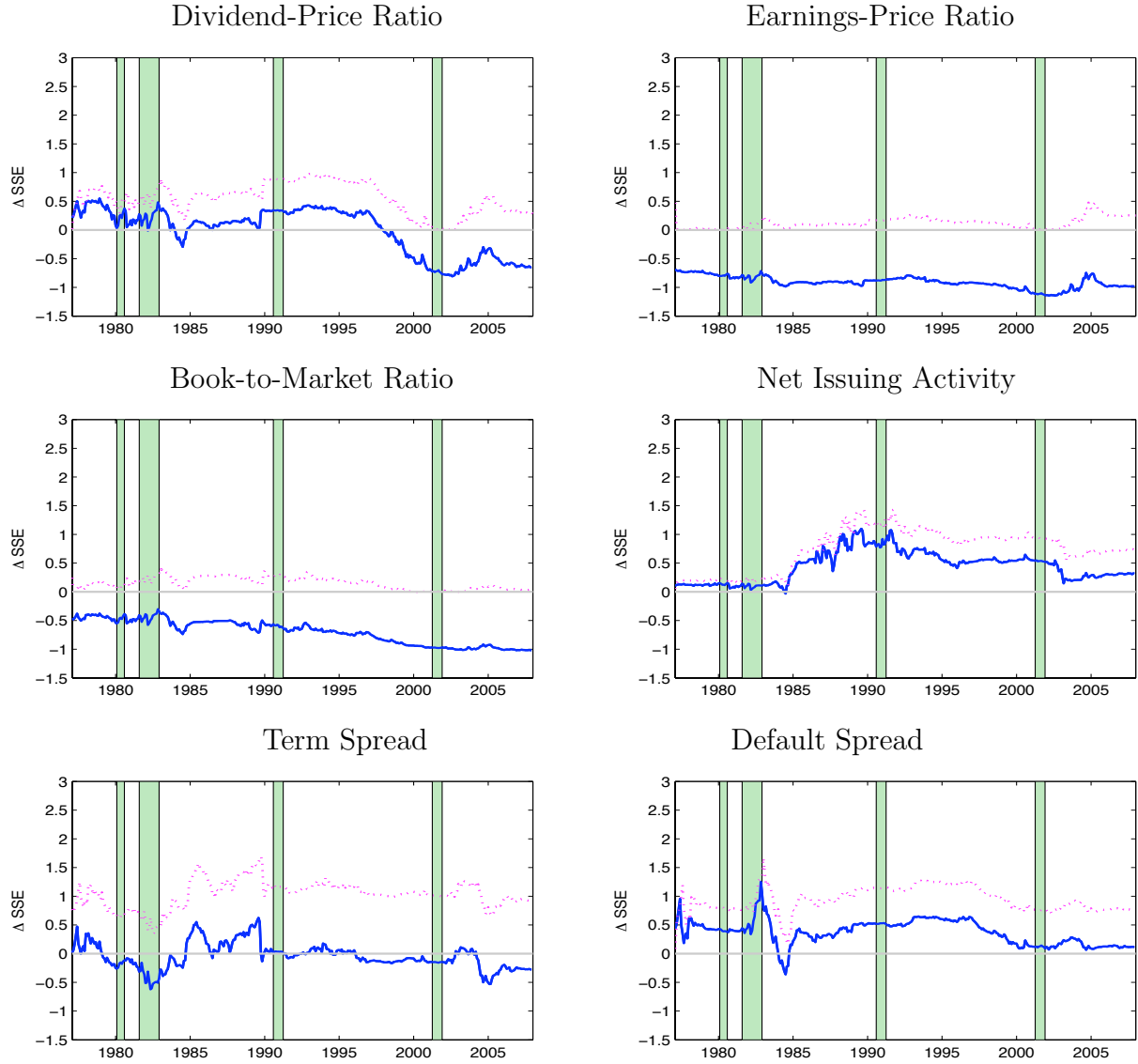


Figure 8 (Cont'd): **In-sample** (---) and **Pseudo Out-of-sample** (—) Performance of *Semi Ex Post* Forecasting.  $\Delta SSE$  measures the difference in cumulative squared prediction errors from the benchmark (the historical average) to the conditional forecasts based on selected predictors summarized in Goyal and Welch (2008). Shading denotes months of recessions identified by NBER.

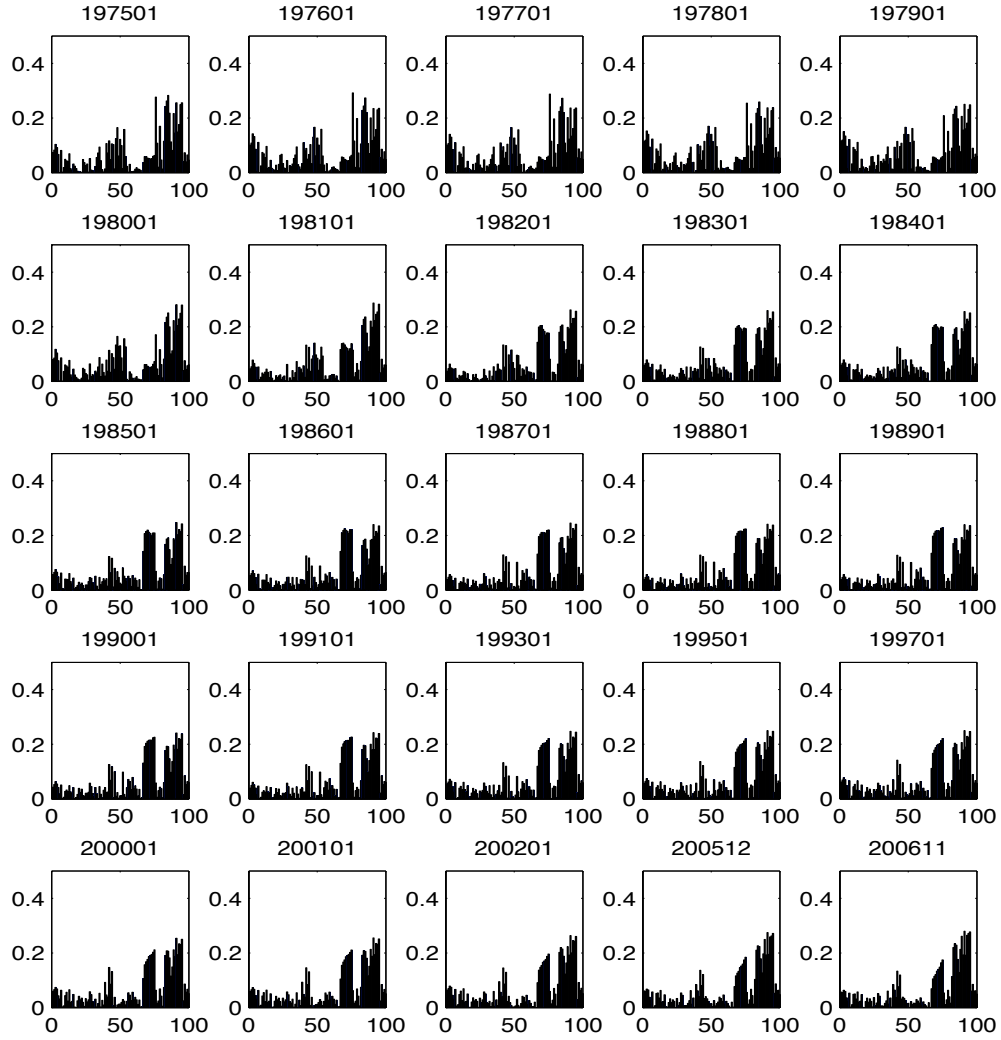


Figure 9: **Dynamic Loadings for Macro Index F2**. Index F2, as the second principal component of the panel data, is recursively constructed at each month using historical information. The vertical axis reports the absolute loadings of index F2 on each individual series. Along the horizontal axis, economic variables are located as Real Output (1-14), Unemployment (15-25), Employment (26-41), **Wages** (42-44), Housing (45-54), Exchange Rates (55-59), Money & Credit Aggregates (60-66), **Interest Rates** (67-75), Inventory & Orders (76-81), **Price Indices** (82-95), Confidence Index (96), and Consumption (97-100).

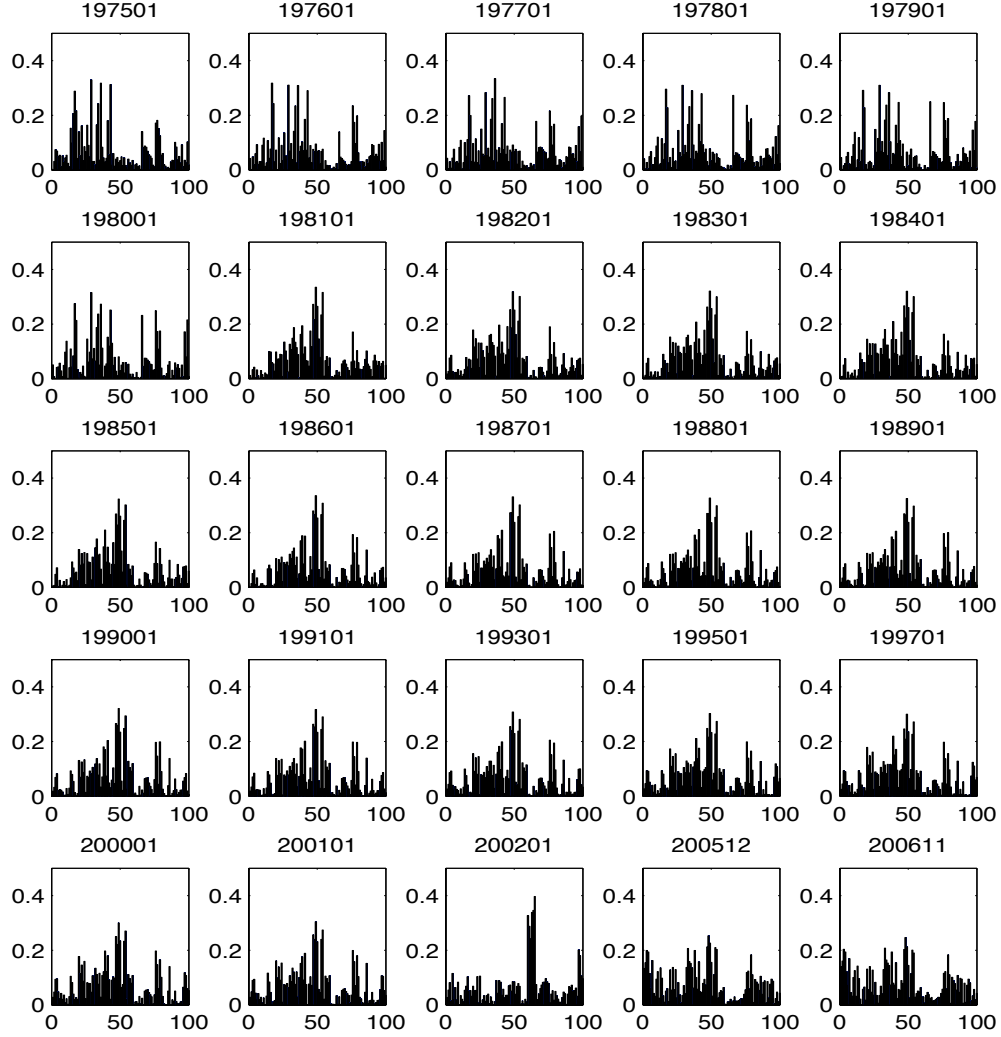


Figure 9 (Cont'd): **Dynamic Loadings for Macro Index F5**. Index F5, as the fifth principal component of the panel data, is recursively constructed at each month using historical information. The vertical axis reports the absolute loadings of index F5 on each individual series. Along the horizontal axis, economic variables are located as Real Output (1-14), Unemployment (15-25), **Employment** (26-41), Wages (42-44), **Housing** (45-54), Exchange Rates (55-59), Money & Credit Aggregates (60-66), Interest Rates (67-75), Inventory & Orders (76-81), Price Indices (82-95), Confidence Index (96), and Consumption (97-100).

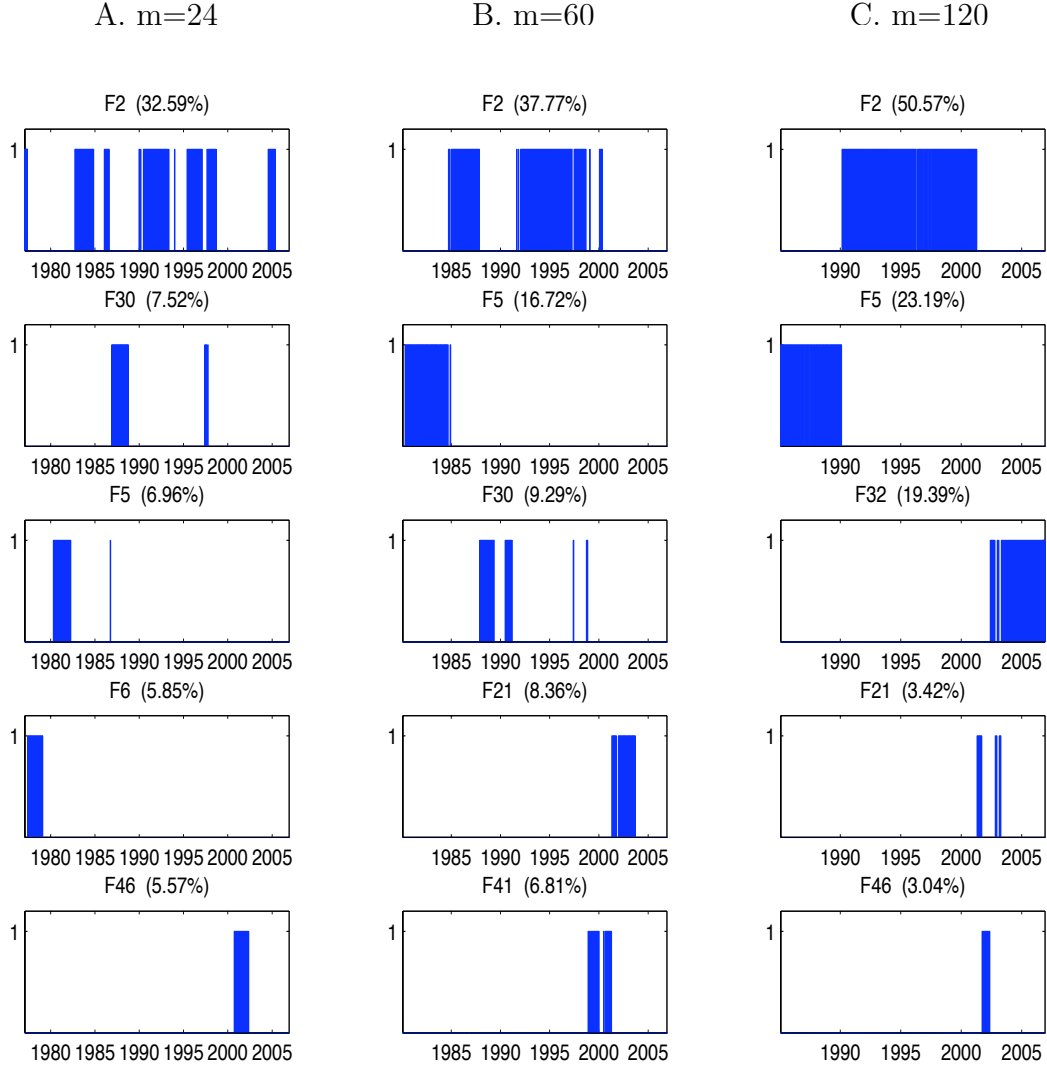


Figure 10: **Top Five Most Frequently Selected Macro Indices** under varying lengths of model selection windows,  $m = 24, 60, 120$ . Index  $F_i$  denotes the  $i$ -th principal component of the panel data, recursively constructed at each month. The value of  $y$ -axis indicates whether the index is selected as an optimal predictor over time, unity for inclusion ( $y = 1$ ) and zero otherwise ( $y = 0$ ). The values in the brackets report the overall including percentage. The model selection criterion is predictive least squares.

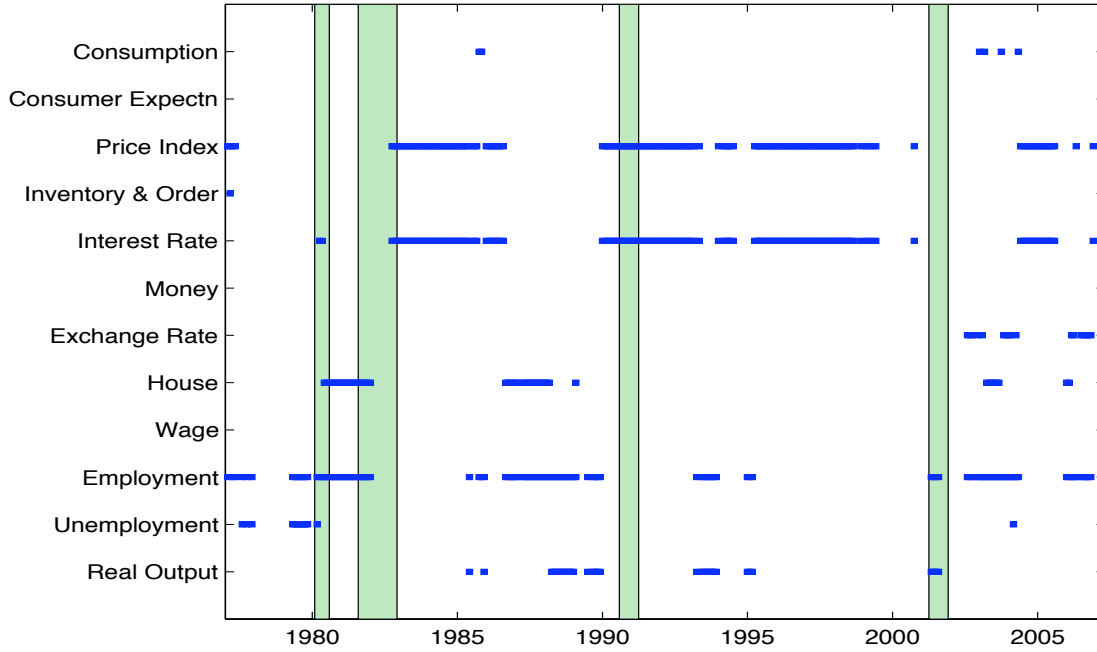
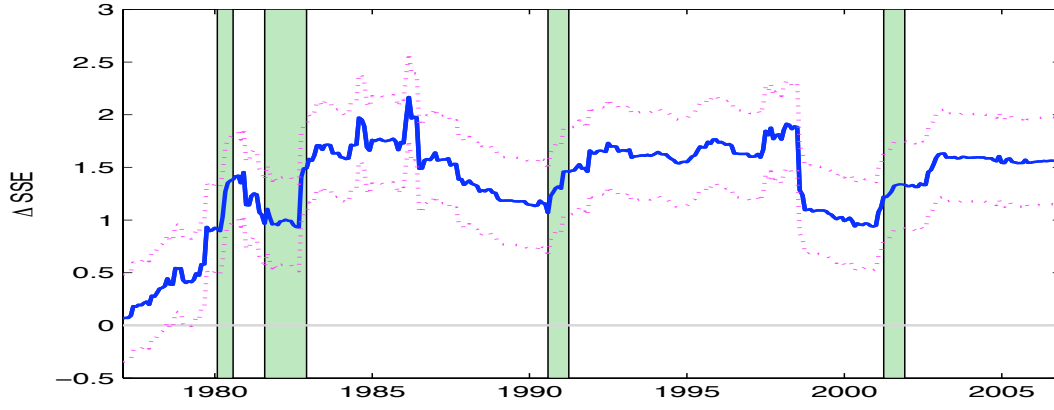


Figure 11: **Dynamic Meanings of Adaptive Macro Indices.** Solid blue squares at each month indicate top two economic categories that are most correlated with the macro index selected as the predictor of future excess stock returns. The correlation is measured by the adjusted R-square value from the regression of macro index  $F_t^j$  on individual series in each sector  $Z_t^i$ , where  $F^j$  is the  $j$ th principal component of the panel data, and  $Z^i$  is a vector of economic series in sector  $i$ , one of the twelve economic categories. The prediction period is from January 1977 to November 2006. Shading denotes months of recessions identified by NBER.



A: Cumulative Prediction Errors



B: Cumulative Month-to-Month Returns

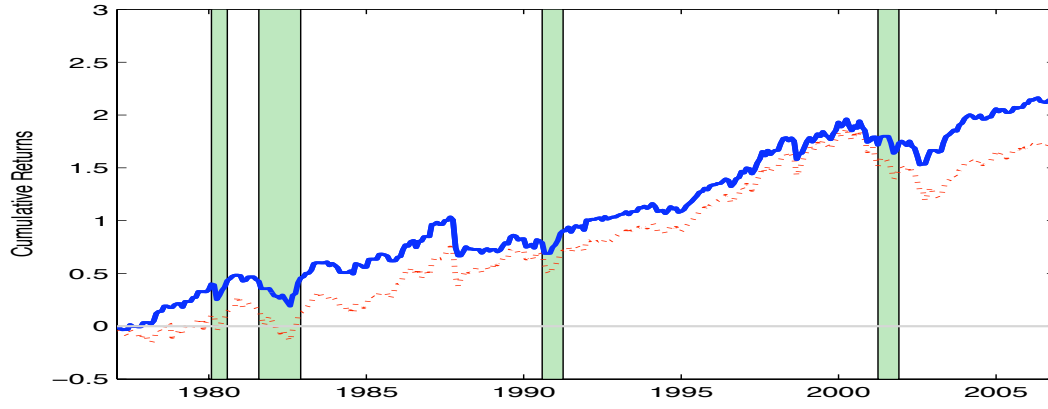


Figure 12: **Real-time Performance of the Noise-Added Data.** The performance is based on Monte Carlo Simulations with 100 iterations based on the revision formula:  $x_{i,t}^{new} = x_{i,t} + \text{noise} = x_{i,t} + 0.1 \cdot \sigma_{i,s} \cdot e$ ,  $e \sim N(0,1)$ . Shading denotes months of recessions identified by NBER. The sample period is from January 1977 to November 2006.