financeUC
Pontificia Universidad Católica de Chile

Working Paper Series

2012-03

Short-Horizon Return Predictability and Oil Prices

Jaime Casassus, Freddy Higuera.

www.finance.uc.cl · Pontificia Universidad Catolica de Chile

Short-Horizon Return Predictability and Oil Prices*

Jaime Casassus
Pontificia Universidad Catolica de Chile

Freddy Higuera
Universidad Catolica del Norte

Revised: November 2012

*We would like to thank Augusto Castillo, Gonzalo Cortazar, Rodrigo Fuentes, Borja Larrain, Chris Telmer, Jose Tessada, Eduardo Walker and seminar participants at UC. Any errors or omissions are the responsibility of the authors. Casassus acknowledges financial support from FONDECYT (grant No. 1110841) and from Grupo Security through FinanceUC. Higuera acknowledges financial support from CONICYT and FinanceUC. Please address any comments to Jaime Casassus, Instituto de Economia, Pontificia Universidad Catolica de Chile, email: jcasassus@uc.cl; Freddy Higuera, Departamento de Ingenieria Industrial, Universidad Catolica del Norte, email: fhiguera@ucn.cl.

Short-Horizon Return Predictability and Oil Prices

Revised: November 2012

Abstract

This paper shows that oil price changes, measured as short-term futures returns, are a strong predictor of excess stock returns at short horizons. Ours is a leading variable for the business cycle and exhibits low persistence which avoids the fictitious long-horizon predictability associated to other predictors used in the literature. We compare our variable with the most popular predictors in a sample period that includes the recent financial crisis. Our results suggest that oil price changes are the only variable with forecasting power for stock returns. This significant predictive ability is robust against the inclusion of other variables and out-of-sample tests. We also study the cross-section of expected stock returns in a conditional CAPM framework based on oil price shocks. Our model displays high statistical significance and a better fit than all the conditional and unconditional models considered including the Fama-French three-factor model. From a practical perspective, ours is a high-frequency, observable variable that has the advantage of being readily available to market-timing investors.

Keywords: Short-horizon predictability, stock returns, business cycle, crude oil, futures prices, conditional CAPM.

JEL Classification: G17, E44, Q43, E32, G12, G14.

1 Introduction

The predictability of stock returns is a controversial topic. Until recently, the prevailing view was that returns could be predicted by the business cycle at long horizons (Cochrane, 2005) and that this evidence was significantly stronger than for short horizons. But in recent years, several studies have questioned the existence of such return predictability. For example, Boudoukh, Richardson, and Whitelaw (2008) show that the dominant findings in this literature are solely the consequence of the high persistence of the predictors. Also recently, Welch and Goyal (2008), who compare the out-of-sample predictive performance of a large number of popular predictors with the prevailing average excess stock return, find that none of these variables predicts stock returns at short horizons better than the historical average return. The current challenge is to propose a variable that predicts equity returns at short horizons and is robust against the new tests suggested in the literature.[1]

This paper shows that unexpected changes in oil prices are a significant predictor of excess stock market returns at short horizons. We measure unexpected oil price shocks by short-term futures returns on crude oil contracts. Our sample period is 1983Q2-2009Q4 and is restricted by the existence of crude oil futures prices. Based on in-sample tests and on the macroeconomic literature (i.e., Hamilton and Herrera, 2004), we use four lags of this variable for the predictive regressions. Our predictive variable has deep macroeconomic roots and allows us to connect the short-horizon predictability of equity returns with the business cycle. Indeed, the existence of negative Granger-causality of oil price shocks on both equity returns and production growth in a trivariate VAR confirms that oil price shocks are a leading variable and are countercyclical. This evidence suggests that increases in oil prices precede recessions and declines in excess stock returns.

We find that at horizons of one to three quarters, the oil price shocks exhibit predictive performance which is both statistically and economically significant. Obtaining significant results in a sample period after the oil crisis in the 1970s is not an easy task since most variables lose their forecasting power within this period (see Welch and Goyal, 2008). In terms

[1] In Welch and Goyal's (2008) words the challenge is still open: "... our article suggests only that the profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power ...".

of global performance, our variable is better than all other variables considered (consumption-wealth ratio, price-dividend ratio, output gap and risk-free rate) with an $\bar{R}^2$ of 6%. This meaningful in-sample result was also detected in out-of-sample tests. Our variable exhibited the best out-of-sample R^2, which was close to 1.2% at one quarter. For longer time horizons, however, no variable showed a significant predictive performance.

Furthermore, oil price shocks have other virtues such as no persistence (they do not produce the pattern reported in Boudoukh, Richardson, and Whitelaw, 2008), they are directly observable by futures changes (unlike variables such as consumption-wealth ratio and product gap, which must be estimated), they have no correlation with the predictive regression's disturbances (do not generate the bias analyzed in Stambaugh, 1999) and they are a high frequency variable available at no cost. All of these characteristics are valued in the practice of portfolio management. To our knowledge, these results position oil price shocks as the best short-term forecasting variable today.

Our study is motivated by two lines of research: the relationship between GDP and oil prices and the linkage between of equity returns and oil price shocks. Because stock return predictability has been detected at business-cycle frequencies, the first strand of literature justifies the relationship between oil prices and the macroeconomy. Intuitively, there are good reasons for believing that oil price leads the business cycle, as nine of ten recessions in the United States since World War II have been preceded by an increase in oil prices (Hamilton, 2008). This has not gone unnoticed by economists and has generated a substantial amount of research, particularly given the fact that oil expenditure represents only 4% of GDP. This literature motivates the main question we address in this paper: Given that oil price shocks precede changes in GDP, do they also have some predictive power for stock market returns? The second line of related research, although less voluminous than the first, provides empirical evidence that past oil shocks have an impact on future equity returns. The following section studies these topics in more detail.

Our paper is also related to recent studies of short-term predictability. For example, Ang and Bekaert (2007) report evidence of predictability, at horizons up to one year, using both the dividend-price ratio and the interest rate. Campbell and Thompson (2008), using a much longer sample, find that the out-of-sample forecasting power of several variables improves significantly when certain restrictions are imposed, and that trading on these predictors can lead to significant welfare benefits when compared to trading on the historical average return. More recently, Cooper and Priestley (2009) propose the output gap as a new forecasting variable for stock market returns. This variable is robust against the tests of Boudoukh, Richardson, and Whitelaw (2008) and Welch and Goyal (2008). Finally, Bakshi, Panayotov, and Skoulakis (2011) find that the Baltic Dry Index, a shipping activity variable that is tied to the business cycle, has predictive ability for a range of stock markets.

While our purpose is to study the effect of oil on the stock market, there is an emerging literature on the financialization of commodities that could extend the short-term predictability of oil to other commodities. Indeed, the emergence of commodities as an asset class has increased the speculative trading in futures markets, specially for those commodities belonging to futures indices, such as the Goldman Sachs Commodity Index and the S&P Commodity Index. Tang and Xiong (2012) find that these indexed commodities have become increasingly correlated with oil after 2004 which makes us believe that these assets would have some predictive power on stock returns too.[2] Unfortunately, the sample period starting from 2004 is not long enough to do a formal study of predictability, so we leave the answer to this question for future research.

The rest of this paper is structured as follows. The next section reviews the relation between oil prices and both the business cycle and excess market returns, and also presents our variable. Section 3 contains the results of in-sample predictability at a quarterly horizon. Section 4 reports the out-of-sample tests. Section 5 discusses the evidence of predictability at longer horizons. Section 6 analyzes the impact of predictability in the cross-section of expected returns. Section 7 concludes.

2 Oil price, the business cycle and stock returns

This section presents the mechanisms that relates the oil prices to the business cycle and excess stock market returns. It also presents our variable.

[2]See Acharya, Lochstoer, and Ramadorai (2011), Basak and Pavlova (2012), Buyuksahin and Robe (2010), Etula (2010), Masters (2008) and Singleton (2012) for more studies about the financialization of commodities and the effect of the speculators on the commodity prices.

2.1 Oil price and the macroeconomy

The study of the relationship between oil prices and the macroeconomy begins with the seminal work of Hamilton (1983). He uses Sims's (1980) six-variable VAR and bivariate VARs with quarterly data for the 1948-1980 period to show that oil price shocks strongly Granger-caused the U.S. GNP growth rate and the unemployment rate. He finds that an increase in the oil price is followed by four successive quarters of lower GNP growth rates. As shown by Ferderer (1996), the common transmission channels of oil price shocks to the real economy are: inflation, terms of trade (Huntington, 2007), the capital utilization rate (Finn, 2000) and increasing returns to scale in production (Aguiar-Conraria and Wen, 2007).

Subsequent studies (Mork, 1989; Lee, Ni, and Ratti, 1995; Hooker, 1996; Hamilton, 2008), note a weakening of this relationship when data from 1980 onwards are included (which coincides with OPEC's loss of control of the oil market).[3] This turned attention towards nonlinear relationships between the variables (Mork, 1989; Ferderer, 1996; Hamilton, 1996; Hamilton, 2003). This evidence required for new explanations to understand the asymmetric impact of oil on the economy. The most commonly cited asymmetric mechanisms are: monetary policy (Ferderer, 1996; Bernanke, Gertler, and Watson, 1997; Hamilton and Herrera, 2004; Balke, Brown, and Yucel, 2002; Leduc and Sill, 2004), imperfect intersectoral mobility of factors (Lee and Ni, 2002; Lilien, 1982; Hamilton, 1988; Davis and Haltiwanger, 2001), investment irreversibility (Bernanke, 1983), wage rigidities (Lee, Ni, and Ratti, 1995), and interest rates (Balke, Brown, and Yucel, 2002). Also, Carruth, Hooker, and Oswald (1998) document a strong and significant relationship between the U.S. rate of unemployment and oil prices.

Recently, Kliesen (2008) adds to the standard regression the variable CFNAI (Chicago Fed National Activity Index), which is the first principal component of 85 monthly indicators of real economic activity, and finds that oil has a significant impact on the U.S. macroeconomic performance. In addition, Cologni and Manera (2009) find a negative influence of oil shocks on GDP growth, although they reject the hypothesis that real GDP growth has no effect on oil prices. Kilian (2008) finds empirical evidence that exogenous oil supply shocks have caused

[3] For example, Hamilton (2008) finds that while in the 1949-1980 period an increase of 10% in the oil price predicts that the GDP growth would be 2.9% slower four quarters later, this figure for the 1949-2005 period is only 0.7%.

significant impacts on the GDP of G-7 countries.[4] Cologni and Manera (2008) observe that oil price shocks affect only the GDP in Italy and in the U.S., albeit temporarily. For almost all of the countries in their sample, oil shocks affect inflation and nominal exchange rates. Gronwald (2008) concludes that only oil shocks that exceed a certain threshold affect the real sector of the economy, while "normal" positive shocks do generate significant nominal impacts. On the contrary, Barsky and Kilian (2004) find inaccuracies in the explanations that support most theories and conclude that "disturbances in the oil market are likely to matter less for U.S. macroeconomic performance than has commonly been thought".

The abovementioned empirical evidence, though not free of debate, largely supports the existence of a significant relationship between oil price shocks and the business cycle in both economic and statistical terms.

2.2 Oil price and the financial market

Contrary to what has occurred with the relationship between oil shocks and the macroeconomy, the linkage between oil and the financial markets has received little attention. Moreover, this reduced literature has not been conclusive about this relationship. For a deeper analysis, Table 1 presents detailed information on the empirical studies reviewed in this section. The table also shows the multiple variables proxying for oil shocks that have been used in the literature.

At a country level, Jones and Kaul (1996) find that oil price shocks produce significant changes in stock returns. The relation can be explained by changes in cash flows and discount rates for United States and Canada, however, they find an overreaction in the United Kingdom and Japan. Driesprong, Jacobsen, and Maat (2008) find that oil price shocks have significant predictive power in developed economies, but not for emerging countries. They find an initial underestimation by agents of the impacts of the shock that is slowly corrected later. Park and Ratti (2008) find evidence that oil price shocks have a significant negative impact on real returns of several net importing countries, unlike what occurs in Norway, a net exporter, where the impact is positive.

[4] This somehow contradicts Kilian (2009) who finds that the impact of oil demand shocks are more significant than supply shocks.

Following Kilian (2009), Kilian and Park (2009) break down oil shocks into three classes: supply, aggregate demand and specific demand (or precautionary demand) shocks. According to their results, these shocks explain 6%, 5% and 11% of the long-run variation of real stock returns, respectively. They do not find a significant response of stock returns to oil supply shocks, however, they do find a positive response to global demand shocks and a negative response to precautionary demand shocks. Sector-level evidence suggests that the mechanism that transmits oil shocks to stock returns is through the demand for industrial products, and not, as widely believed, through the production costs of the firms. Apergis and Miller (2009) criticize the methodology used in Kilian (2009) and Kilian and Park (2009) due to the use of both stationary and non-stationary variables in their VAR specification. They differentiate the I(1) variables and carry out the same breakdown as Kilian (2009), but using only I(0) variables. Instead of including the equity returns in a same VAR, as in Kilian and Park (2009), Apergis and Miller (2009) use a second VAR with the three types of shocks and the stock market returns of each country. Using a sample composed of the G-7 Group and Australia, they conclude that the effects of oil shocks, although statistically significant, produce a minor impact on stock returns.

Some papers report significant intratemporal correlation between oil shocks and portfolio returns. Nandha and Faff (2008) calculate a two-factor model that includes unexpected returns on the world market and the oil market. They find a significant negative impact of oil that is largely symmetrical on all sectors except for mining and oil/gas. On the other hand, Bachmeier (2008) regresses portfolio returns on shocks that are contemporaneous with the oil price. He finds that oil shocks have a significant negative impact in returns.

Finally, Huang, Masulis, and Stoll (1996) find that daily oil futures returns have no correlation with stock returns, except for returns on oil companies. Ciner (2001) tests for non-linear Granger-causality of oil futures returns on stock market returns and finds a significant relationship. Using monthly data, Sadorsky (1999) finds that in the U.S. market the price of oil affects the financial market, but that the effect in the other direction is insignificant. He also finds that oil price shocks have an asymmetric effect on industrial production and real stock returns, with positive shocks having a greater impact than negative ones.

On the theoretical side, Wei (2003) builds a general equilibrium model to estimate the impact of an oil price shock on the value of a firm that faces investment irreversibility. His model predicts that an oil shock will have only a small impact. Consequently, he is unable to explain the massive decline in the stock market in 1974 after the oil shock of 1973.

2.3 Stock returns and the business cycle

This section shows that stock market returns vary considerably with the business cycle. This suggest that a variable that anticipates the cycle, such as the oil price shocks, may also do a good job at predicting stock returns.

Figure 1 shows the excess market returns from 1926Q3 to 2009Q4 and the shaded areas represent the NBER (National Bureau of Economic Research) recession periods.[5] The stock returns tend to be negative and grow during recessions, reaching peaks towards the end of each one. In fact, in our sample period the maximum return was reached by the end of the Great Depression of the 1930s. Nevertheless, this commonly cited contracyclical character of stock returns is evident only at the end of recessions; at the beginning of and during recessions, these returns are highly procyclical. For example, the minimum in the sample period was also reached during the Great Depression. In addition, expansions are characterized on average by positive returns, although they are less volatile than those generated during recessions.

NBER's business cycle dates enable the sample to be divided into the four stages: expansion, peak, recession and trough. Table 2 shows the first two conditional sample moments for the excess stock returns. As can be seen in the table, the most frequent stage of the economy is expansion, and during this phase of the cycle the average return is positive (2.9%) and greater than its historical average (2.0%). On the contrary, during the recession stage the average excess return on the market has a similar magnitude as it does during the expansion stage, but with the opposite sign (-3.0%) and almost twice the volatility. Finally, during the peaks (troughs), expectations about the state of the economy are negative (positive) and therefore, excess stock returns are highly negative (positive) in these stages.

[5] For longer sample periods we use the stock returns and the risk-free rate available from Ken French's web page, which contains data from September 1926. Stock returns are from the value-weighted CRSP index and the risk-free rate is from the 3-month Treasury bills from Ibbotson Associates.

2.4 Measuring oil price shocks

Considering the evidence mentioned above, it seems natural to propose oil price shocks as a leading variable with the potential to forecast stock market returns. However, to maximize its predictive power, it is essential to consider only unanticipated changes in the oil price. Although oil spot returns have been a widely used variable in the literature (see Table 1), we do not consider them, because they contain some components that are clearly anticipated by market participants, such as the interest rate and the convenience yield.[6] One way to address this issue is to estimate unexpected oil changes from a model for the spot price dynamics, but this procedure still has disadvantages in that it depends on the model specification and that the information set used by the econometrician to estimate the conditional mean may not coincide with that of the market. Unexpected oil price changes can only be captured with an objective and precise estimate of the expected spot price in the future.

We measure unexpected oil price shocks by short-term futures returns on crude oil. Using Fama and French's (1987) methodology and cointegration tests, Switzer and El-Khoury (2007) show that oil futures prices have significant predictive power for future spot prices. Moreover, Ma (1989) and Kumar (1992) confirm that futures prices, in addition to being unbiased predictors of spot prices, exceed the predictive capacity of a random walk and a wide variety of models. This evidence suggests that unexpected changes in oil prices are correctly captured by our proposed variable.[7] Therefore, we assume that quarterly unexpected oil shocks are proxied by oil futures returns, i.e.,

$$\Delta f(t) \equiv f^1(t) - f^4(t-3) \approx s(t) - \mathbb{E}_{t-3}\left[s(t)\right] \tag{1}$$

where $s(t)$ is the log oil spot price and $f^\tau(t)$ is the log oil futures price of a contract that matures in τ months.

[6] The convenience yield of crude oil is a benefit for immediate ownership of physical units of the commodity attributed to the benefit of protecting regular production from temporary shortages of oil stocks (see for example, Casassus and Collin-Dufresne, 2005).

[7] A direct way to test that futures prices are unbiased estimators of future spot prices is to verify the existence of a risk premium associated to this contract. In our sample, the t-stat of the null hypothesis that the average logarithmic futures return is zero is 1.13. Although this informal test only detects a constant risk premium, the results obtained by Switzer and El-Khoury (2007) suggest that in oil futures this finding is not incompatible with the existence of predictive power for spot prices.

Data on oil futures prices are from NYMEX, which began trading these contracts in March 1983. Therefore, our sample period goes from 1983Q2 to 2009Q4.[8] Figure 1 confirms that recessions are preceded by positive oil shocks, while during recessions these shocks are rapidly reversed. The great variability of the oil shocks is evidence of their predictive potential. Another visual characteristic is the low persistence of the series, which prevents it from being subject to the critiques of existing predictive variables. The optimal number of lags of our variable to be considered in this study was determined using the Akaike Information Criterion (AIC) in Ordinary Least Squares (OLS) regressions of the excess stock market return on lagged oil shocks. In this study we consider four lags of our variable, because this number minimizes the AIC (see Table 3). Interestingly, the number of lags coincides with that obtained in the macroeconomic literature (see Hamilton, 2003 and Hamilton, 2008).

2.5 Oil price, the business cycle and stock returns

To show evidence of the relationship among the oil price, the business cycle and excess stock returns, we study the joint dynamics of these variables using a vector autoregression analysis with four lags. Our proxy for the stock market is the value-weighted CRSP index, from which we obtain the quarterly returns on the market portfolio (R_m). We proxy the risk-free rate (R_f) with the 3-month constant-maturity treasury yields from the Federal Reserve Board of Governors. Following Cooper and Priestley (2009), we use the total Industrial Production Index (IP) from the Fed as a measure of output and a proxy for the business cycle. Table 4 shows the maximum likelihood estimates for the VAR(4) model. As is common in macroeconomic series, industrial production growth rate ($\Delta\%IP$) is the easiest series to predict (its R^2 is 51%) and its own lags have useful information for forecasting its future values. On the other hand, the predictive power of the excess stock return ($R_m - R_f$) on $\Delta\%IP$ is a clear signal that the financial market correctly anticipates future economic growth. Moreover, in our sample the Granger-causality of oil shocks (Δf) on economic growth is also verified, evidence that is in line with the macroeconomic studies mentioned before.

The table also shows that for the excess stock returns, the lack of significance of its own

[8] We use quarterly data to better capture the aggregate impact of oil shocks and make our results more compatible with the macroeconomic evidence. Moreover, this allows us to include in the analysis the consumption-wealth ratio, a variable which is only available at a quarterly frequency (Lettau and Ludvigson, 2001a).

lags suggests the stock market is efficient and that the initial underreaction and later correction documented by Driesprong, Jacobsen, and Maat (2008) is not seen at quarterly frequency.[9] There is also evidence of inverse causality with the industrial production growth rate, which is probably the consequence of an adjustment process of previous expectations about actual economic growth. Furthermore, and as expected, oil shocks demonstrate a significant predictive power for excess stock returns, which will be explored in greater depth in the following sections.

Finally, the results reveal that oil shocks cannot be predicted with any of the lagged variables, which is evidence that our measure for oil shocks effectively captures unanticipated changes in this variable. This result implies that our variable is exogenous to the U.S. economy; however, it may not be so to global aggregate demand. Therefore, the critique of Kilian (2009) regarding the endogeneity of oil price shocks is still valid for our variable.

The following sections analyze and test the predictive power of oil price shocks for stock returns.

3 Short-horizon predictability of stock returns

This section focuses on the in-sample predictability of stock returns at a quarterly horizon. The predictive performance of our variable is evaluated and compared to the performance of the following variables: the risk-free rate (Campbell, 1987), the log dividend-price ratio (Fama and French, 1988), the consumption-wealth ratio (Lettau and Ludvigson, 2001a), and the output gap (Cooper and Priestley, 2009). The log dividend-price ratio ($d - p$) was calculated from the value-weighted CRSP index using the methodology described in Ang and Bekaert (2007). The consumption-wealth ratio (cay) and its individual components are from Martin Lettau's web page and sampled at a quarterly frequency. The output gap (gap) is constructed using the total Industrial Production Index.[10] The output gap is estimated with data from 1948Q1

[9] If the adjustment process of returns after an oil shock is slow, this should be evidenced to some degree in the excess stock return, either positively if the adjustment is gradual or negatively if it is excessive and requires future reversals.

[10] Cooper and Priestley (2009) define four methods for calculating the output gap, but the main one is the quadratic version of gap (based on its greater correlation with procyclical variables), which we also use here. The output gap is calculated from the following regression: $ip(t) = a + b\,t + c\,t^2 + gap(t)$, where $ip(t)$ is the log industrial production, t is a time trend and $gap(t)$ is the error term.

to replicate the series from Cooper and Priestley (2009). The variables $cay(t)$ and $gap(t)$ are assumed to be known at the start of time $t+1$, and therefore can be used to forecast excess stock returns. We omit any complication due to the look-ahead bias in these variables and the normal delay in the publication and subsequent revisions of these and other macroeconomic series.

Table 5 presents the main statistics of the predictive variables and Figure 2 provides graphical evidence. The upper panel of Table 5 shows that our variable (Δf) and $d-p$ exhibit more volatility than the variables R_f, cay and gap. Unlike the other predictors, Δf shows very low persistence; in fact, its first-order serial correlation is as low as that of $R_m - R_f$. The lower panel of the table shows that the intratemporal correlation of our variable with the excess stock return is very low, rejecting Δf as a possible pricing factor. Both the low persistence of our variable and its low correlation with excess stock returns, discard the idea that the forecastability of stock returns from oil returns might be due to a common factor driving both variables. Another important characteristic of our variable is its low correlation with other predictive variables, which suggests that Δf contains business cycle information not captured by the existing predictors. On the contrary, the existing variables show high levels of correlation (in absolute terms) among themselves, revealing the presence of redundant information.

Next we turn our attention to evaluating the predictive performance of our variable and the other variables considered here. We begin with the evidence of in-sample predictability at a quarterly horizon, for which we estimate the following regression:

$$R_m(t) - R_f(t) = \alpha + \beta' X(t-1) + \varepsilon(t) \tag{2}$$

where $X(t-1)$ is a vector of known predictors at $t-1$ and β its associated coefficient vector. It should be emphasized that $X(t-1)$ can include one variable, several variables or several lags of the same variable.

Table 6 shows the OLS results of equation (2). We report asymptotic t-stats and Wald tests that correct for serial correlation and heteroscedasticity using Newey and West (1987).[11] The

[11] Following Newey and West (1994), we chose lag length equal to floor $\left[4 \cdot (T/100)^{2/9}\right]$, where floor $[x]$ denotes the integer part of x.

columns present the results for each predictor variable. The results indicate that our variable has the best in-sample predictive performance with an $\bar{R}^2$ of 6%. The Wald test corresponds to the null hypothesis that all the coefficients in equation (2) are zero, except the constant. This statistic is highly significant for the oil price shocks (p-value of 2%). The annual cumulative impact of our variable is -0.106 and is calculated from the sum of the coefficients corresponding to the four lags of Δf. This impact is also economically significant. To see this, consider a one-time increase in Δf of one standard deviation (19.5% in our sample, see Table 5). This change leads to a 2.1% decrease in expected quarterly excess returns on the value-weighted CRSP index ($-0.106 \times 19.5\% = -2.1\%$), which is equivalent to 29.6% (2.1% / 7.0% = 29.6%) of the historical average annual excess return.

The dynamics of the distributed lags can be explained with aggregate demand shocks (see Kilian and Park, 2009). A positive shock to the global demand for industrial commodities produces both a direct positive impact and an indirect negative one in the financial market. The direct impact manifests as an increase in both the oil price and economic growth with a consequent positive stock return. Increased economic growth pushes the oil price even higher, and thus indirectly affects negatively the future expected economic growth and expected stock returns. The final impact will depend on the relative magnitudes of both impacts.[12] The direct impact is initially stronger, which explains the positive significant effect of the first lag of Δf. Later, the indirect negative impact of the second and third lags begins to gather some strength, although not of a sufficient magnitude to cancel out the initial positive impact (see the sign and low significance of the following two lags). One year after the unexpected aggregate demand shock, the indirect effect becomes dominant; in other words, the high price of oil causes a deceleration in the economy. This is manifested by the negative and significant coefficient of $\Delta f(t-4)$ which is also responsible for the cumulative negative impact reported.

The third column in Table 6 shows the results for the interest rate as a predictor. This variable has the worst predictive performance in our sample, with a $\bar{R}^2$ of -0.01. Contrary to the findings of previous studies (Campbell, 1987), the coefficient that accompanies this variable is positive, although not significant. This poor performance is associated with the low volatility

[12]Note from our VAR analysis in the previous section that the signs of the four lags of oil shocks are the same for both industrial production and excess stock returns (see Table 4). This result gives support to our aggregate demand based explanation.

of this variable in our sample. Its standard deviation is 0.006 (see Table 5), a value well below the 0.032 reported by Ang and Bekaert (2007) for 1935Q2-2001Q4. The last three columns contain the results for the $d-p$, *cay* and *gap* variables. All of these have intuitive signs, although their coefficients are lower in absolute value than those reported in previous studies (see Ang and Bekaert, 2007; Lettau and Ludvigson, 2001a and Cooper and Priestley, 2009; respectively).[13] Moreover, the R^2 statistic for these variables (all about 2%) suggest that they have similarly poor predictive power. Among these variables, *cay* has a better performance being the only one significant at the 5% level.

The previous results suggest that the oil shocks have a significant in-sample forecasting power for short horizons. We now see whether our variable is robust to the inclusion of the other predictors considered here. To evaluate this, we estimate the following extended predictive regression:

$$R_m(t) - R_f(t) = \alpha + \sum_{j=1}^{4} \beta_j \Delta f(t-j) + \theta' Z(t-1) + \epsilon(t) \tag{3}$$

where $Z(t-1)$ is a vector of predictor variables and θ its associated coefficients vector. Lack of robustness in the predictive power of our variable should be reflected in changes in sign and/or loss of significance in the coefficients that accompany the lags of oil shocks.

The results of the estimation of equation (3) are presented in Table 7. The last row contains the p-value of an asymptotic Wald test for the combined null hypothesis that all the coefficients associated with the lags of our variable are zero. The columns show the results of including each of the other variables in the predictive regression of Δf, while the inclusion of all of them is considered in the last column. First, given the low correlation of our variable with the others, the forecasting power of our variable remains intact. According to the Wald test for the coefficients of Δf, in all of the estimates these coefficients keep their joint significance. In addition, their signs and individual significance remain the same, and they are roughly the same size. Second, consistent with previous results, the greatest increase in predictive power is reached when our variable is used in combination with *cay*, obtaining an $\bar{R}^2$ of 8%. Third, contrary to what occurs with the oil shock coefficients, the results of the last column provide

[13] According to Stambaugh (1999), the estimation of the β coefficient in a predictive regression of $d-p$ is biased. This bias further reinforces our conclusion regarding this variable's lack of forecasting power.

evidence of great instability in the predictive power of the other variables. None of these are significant. The coefficients of R_f and *gap* experience changes in sign and those of $d - p$ and *cay* vary dramatically in size. Of course, this evidence is consistent with the high correlation between these variables reported in Table 5.

In summary, this section demonstrated that our variable has significant and robust short-horizon in-sample forecasting power for stock returns. Out-of-sample predictability, also at a quarterly horizon, is considered in the next section.

4 Out-of-sample predictability of stock returns

In-sample predictive performance is essential for establishing the existence of predictability. However, in order for a predictor to be used by an investor, it must also demonstrate a good out-of-sample performance. That is, a predictive variable must be able to forecast excess returns reasonably well with information available at the time of the forecast, which is not guaranteed by the in-sample tests of equations (2) and (3), as the coefficients are estimated using the full sample.

Welch and Goyal (2008) conclude that it is very difficult to find variables with short-horizon out-of-sample forecasting power that outperforms the average excess return in a recent sample period. In fact, when considering the sub-period from 1965 to 2005, they only find variables that outperform the prevailing historical average return at a five-year horizon. Although the out-of-sample predictive performance can be increased by imposing certain restrictions, as shown in Campbell and Thompson (2008), here we choose to keep the simplicity and linearity of the predictive model. We also test the other variables out-of-sample performance despite their poor in-sample predictive power.

In order to contribute to this discussion, we compare forecasts from nested linear models to determine whether each variable has predictive content for stock returns. The prevailing historical average of excess stock returns is used as a benchmark. Therefore, we define the following benchmark and competing models:

$$\text{benchmark:} \qquad R_m(t+1) - R_f(t+1) = \alpha_1 + u_1(t+1) \tag{4}$$

$$\text{competing:} \qquad R_m(t+1) - R_f(t+1) = \alpha_2 + \beta' X(t) + u_2(t+1) \tag{5}$$

where the coefficients α_1, α_2 and β are estimated recursively. The sample size T is divided into in-sample and out-of-sample portions. Q is defined as the minimum number of observations used for estimating the coefficients and $P \equiv T - Q$ denotes the maximum number of one-step-ahead predictions. Thus, forecasts of $R_m(t+1) - R_f(t+1)$, $t = Q, \ldots, T-1$, are generated recursively using the two linear models in equations (4) and (5), where all coefficients are re-estimated with new observations as forecasting moves forward through time.

Our assessment of out-of-sample predictability involves three metrics that are presented with more detail in Appendix A. The first approach is the forecast encompassing test of Clark and McCracken (2001) that verifies whether the competing forecasts incorporate any useful information absent in the benchmark forecasts. Their statistic ENC-NEW tests the null hypothesis that the benchmark model encompasses the competing model. This test suffers from a look-ahead bias, since ENC-NEW depends on a parameter that is estimated with the full sample. The second metric we use is the MSE-F test developed by McCracken (2007) that compares the predictive accuracy between nested models measured as the mean squared forecasting errors (MSE). Although this test does not have the best small-sample properties, it enables testing of out-of-sample predictive power under the same conditions that an investor faces in reality. The final measure of out-of-sample forecasting performance is the out-of-sample R^2, R^2_{OS}, proposed by Campbell and Thompson (2008). Following Welch and Goyal (2008), we use bootstrap in order to address size distortions in the t-stat for long horizons (Ang and Bekaert, 2007) and to make the correct inference from nested out-of-sample predictability tests (Clark and McCracken, 2005). We also consider covariance matrices of coefficients corrected for heteroscedasticity and autocorrelation that arise from the use of distributed lags and overlap of returns (Newey and West, 1987).

In general, when performing out-of-sample tests in a small sample, there is a trade-off with the number of in-sample observations that is hard to resolve. On the one hand, the objective is to use a relatively large in-sample proportion of the sample (Q/T), so that the out-of-sample forecasts are done with estimates which are as similar as possible to those obtained with the full sample. But at the same time, as suggested by the results of Inoue and Kilian (2005), the out-of-sample proportion (P/T) must be large enough to prevent significant differences in power between the in-sample and out-of-sample tests. Thus, to achieve a reasonable level of power without producing excessive forecasting errors at the beginning of the out-of-sample sub-period, the optimal choice should be around $\pi = P/Q = 1$. However, to make our test more rigorous, we choose 1997Q4, which is when the Asian crisis hit the U.S. economy, as the starting point for the out-of-sample sub-period. That is, given that our adjusted (for lags) sample spans the period 1984Q2-2009Q4, our choice implies the following sample portions: $Q = 54$, $P = 49$ and $\pi = 0.91$.

The results of the out-of-sample tests are presented in Table 8. All of the tests coincide in that our variable is the only one with out-of-sample forecasting power for the excess stock returns at a 10% significance level.[14] Although the *cay* variable is marginally significant according to the ENC-NEW test (its bootstrapped p-value is 10.3%), as explained above, the only test that measures forecasting power under the effective conditions faced by a potential investor is the MSE-F test. Our variable (Δf) has the highest and most significant R^2_{OS} among all of the variables considered; however, the size of this statistic is only 1.2%. The table also shows that given the wide differences between the bootstrapped and asymptotic critical values, controlling for considerations of small sample and differences in the relative out-of-sample portion (i.e., π) is essential for a reliable inference, especially when highly persistent predictor variables are used. Also, as a consequence of the close relationship between the MSE-F and the R^2_{OS} statistics shown in Appendix A, the inference using the bootstrap method produces the same results for both tests.

Finally, the low R^2_{OS} for all the predictors is evidence that out-of-sample forecasting of stock returns has become an increasingly difficult challenge in recent times (one of the main points emphasized by Welch and Goyal, 2008). The forecasting ability of a predictor variable depends exclusively on its capacity to successfully summarize the conditioning information used by the market participants, which has become increasingly complex.

[14] Our results could be distorted by a possible look-ahead bias in the *cay* and *gap* variables, since for the sake of simplicity, both were calculated using the full sample of observations. Strictly speaking, in an out-of-sample test, forecasting of the excess stock returns implies the estimation of the coefficients with data only up to the prevailing quarter.

5 Long-horizon predictability of stock returns

This section examines the in-sample predictability of stock returns at longer horizons. The evidence presented below is based on the following long-horizon regression:

$$R_m^h(t+h) - R_f^h(t+h) = \alpha_h + \beta_h' X(t) + \varepsilon^h(t+h) \quad (6)$$

where $R_i^h(t+h) = \prod_{j=1}^{h} \left(1 + R_i^1(t+j)\right) - 1$ is the h-period return for asset $i = m, f$ and $R_i^1(t+j)$ is the respective one-period return from time $t+j-1$ to $t+j$.

The evidence of long-run predictability has been the subject of a great deal of criticism. According to Boudoukh, Richardson, and Whitelaw (2008) long-horizon regressions are misleading for persistent predictors.[15] That is, long-horizon regressions associated with the variables R_f, $d-p$, *cay* and *gap* cannot show anything different to what was shown in Section 3 due to their high persistence (see Table 5). On the other hand, given the almost null persistence of our variable, the Δf shocks are absolutely short-lived. Therefore, at the most our variable could have forecasting power for stock returns over a one-year horizon.

Following Kilian (1999), we adapt the bootstrap algorithm described in Appendix A to support the inference from the long-horizon regressions presented here. In addition, to evaluate the impact of the findings in Boudoukh, Richardson, and Whitelaw (2008) regarding persistent predictors, we report both expected regression coefficient and the R^2 statistic at the hth horizon conditional on their one-period counterpart, under the null of nonpredictability. The expected coefficients are given by the following equations:

$$\mathbb{E}\left[\hat{\beta}_h | \hat{\beta}_1 = \hat{\beta}_1^*\right] = \left(1 + \frac{\rho\left(1-\rho^{h-1}\right)}{1-\rho}\right)\hat{\beta}_1^* \quad (7)$$

$$\mathbb{E}\left[R_h^2 | R_1^2 = R_1^{2*}\right] = \frac{\left(1 + \frac{\rho\left(1-\rho^{h-1}\right)}{1-\rho}\right)^2}{h} R_1^{2*} \quad (8)$$

where $\hat{\beta}_1^*$ is the actual estimate of the regression coefficient in equation (6) for $h = 1$, ρ is the

[15] Other problems documented for long-horizon regressions are: (1) serial correlation in residuals induced by the overlap of observations; (2) inefficient use of the data that provide spurious forecasts about the dynamics of variables, especially for non-exogenous predictors (Campbell, 1991); and (3) by aggregating returns, long-horizon regressions invalidate the inference from standard asymptotic methods (Valkanov, 2003).

first-order serial correlation coefficient of the predictor variable and R_1^{2*} is the actual estimate of the R^2 statistic for $h = 1$, also from equation (6).

Our analysis covers return horizons up to five years (h = 20). For each horizon we consider the same number of observations and the same sample period 1984Q1-2004Q4 (i.e., 84 observations for each horizon). Table 9 presents the OLS estimates of equation (6) for $h = 3, 4, 8, 12, 16, 20$. We report bootstrapped p-values for t-stats and Wald tests that correct for serial correlation and heteroscedasticity using Newey and West (1987). Because of overlapping observations, we increase lag length for the Newey-West estimator by $h - 1$.[16] The table shows that our variable has significant forecasting power for stock returns up to an horizon of three quarters. For $h = 3$ the signs of the coefficients associated with Δf lags are the same as those for $h = 1$, the $\bar{R}^2$ is 7% and the bootstrapped p-value for the test that all coefficients for oil lags are zero is 0.02. At longer horizons, as expected, there is no evidence of predictability with our variable.

Regarding the other variables, Table 9 shows that at the 10% significance level, none of the variables demonstrates forecasting power for stock returns at long horizons. The *cay* variable is marginally significant at a five-year horizon; however, given the absence of predictive ability at other horizons, it is very likely that this result can be explained by the look-ahead bias. The variable R_f has so little forecasting power that despite its high persistence, it does not exhibit the pattern predicted by Boudoukh, Richardson, and Whitelaw (2008). Instead, its regression coefficient changes sign through horizons and its $\bar{R}^2$ is always negative. Finally, the other persistent predictor variables ($d-p$, *cay* and *gap*) effectively follow the pattern predicted by Boudoukh, Richardson, and Whitelaw (2008). That is, both regression coefficients and R^2 are always (in absolute value) growing with the horizon. In fact, both the regression coefficients and the R^2s look pretty much like those predicted by equations (7) and (8). These results reinforce the skepticism with which the evidence of predictability with highly persistent variables should always be viewed.

[16] This is, we use floor $\left[4 \cdot (T/100)^{2/9}\right] + (h - 1)$.

6 Implications for the cross-section of expected returns

In this section we study the cross-sectional implications of our variable. Despite the clear relationship between predictability, time-varying risk premium and conditional asset pricing models, there is an important methodological asymmetry between studies on predictability and those on conditional asset pricing. Papers on predictability have only sought to prove that the equity risk premium is variable and do not consider cross-sectional tests. On the other hand, studies on conditional asset pricing only present evidence of in-sample predictability and generally lack the empirical rigor of the predictability studies.

In our opinion, if a variable exhibits significant forecasting power for stock market returns, it should also be tested as a conditioning variable to explain the cross-section of expected stock returns. For this reason and for greater robustness, we offer empirical evidence in the cross-section of expected returns through CAPM and CCAPM models conditioned on our variable.[17] In addition, the empirical performance of our conditional model is compared to the performance of the same models conditioned on the other predictor variables considered in the previous sections, as well as the following unconditional models: CAPM (Sharpe, 1964), CCAPM (Breeden and Litzenberger, 1978) and the three-factor FF model (Fama and French, 1993).

The standard equilibrium condition for any asset pricing model is given by:

$$1 = \mathbb{E}\left[M(t+1) \cdot (1 + R_i(t+1))\,|\Omega(t)\right] \tag{9}$$

where $M(t+1)$ is the stochastic discount factor or pricing kernel and $\Omega(t)$ is the agent's information set at time t. We assume that the pricing kernel is exponentially affine on the pricing factor:

$$M(t+1) = \exp\left(b^0(t) + b^1(t)F(t+1)\right) \tag{10}$$

[17] Multiple conditioning variables have been used in the literature. For example, Lettau and Ludvigson (2001b) propose CAPM and CCAPM models conditioned on the *cay* variable. Lustig and Van Nieuwerburgh (2005) derive a conditional CCAPM on the housing collateral ratio (ratio of housing wealth to human wealth). Santos and Veronesi (2006) present a conditional CAPM on the labor income to consumption ratio (fraction of consumption funded by labor income). In these studies the evidence of predictability is only an intermediate step that enables the authors to fulfill the first requisite needed for the existence of a conditional asset pricing model, or in other words, that the equity risk premium is variable. Then, given that the predictability condition is satisfied, there should be a conditional model that correctly values the cross-section of returns using the proposed predictor.

where $b^i(t) = b^i_0 + b^i_1 z_1(t) + \ldots + b^i_L z_L(t);\ i = 0, 1$ and $F(t+1)$ is the risk factor. We use the aggregate market return, $R_m(t+1)$, as the risk factor for the CAPM models, and the real per-capita consumption growth rate, $\Delta c(t+1)$, for the CCAPM models. We allow for more than one conditioning variable, in particular, we use $L = 4$ for our models whereas $L = 1$ for conditional models on other variables. The exponentially affine form in equation (10), as opposed to the standard linear assumption, is to keep the pricing kernel positive and to avoid the large negative values implied in linear models (see Nagel and Singleton, 2011).

As noted by Lewellen, Nagel, and Shanken (2010), conditional asset pricing models present serious problems in pricing a risk-free portfolio. Although they report high R^2 in the cross-section, this is typically achieved at the expense of estimated intercepts that are substantially greater than their theoretical values (i.e., the risk free rate). To address this, we follow Nagel and Singleton (2011) and include the risk-free portfolio as an extra asset in our study. We consider the orthogonal relationship between the risk-free rate and the pricing kernel for the following extra moment condition:[18]

$$0 = \mathbb{E}\left[M(t+1) - \frac{1}{1 + R_f(t+1)}\right] \tag{11}$$

In addition, we use the beta representation of equation (9) to produce the unconditional expressions for the expected risk-free return and the expected return of asset i given by:

$$\mathbb{E}\left[R_f(t+1)\right] = \frac{1}{\mathbb{E}\left[M(t+1)\right]} - 1 \tag{12}$$

$$\mathbb{E}\left[R_i(t+1)\right] = \mathbb{E}\left[R_f(t+1)\right] - \mathbb{E}\left[1 + R_f(t+1)\right] \mathbb{C}ov\left[M(t+1), R_i(t+1)\right] \tag{13}$$

where in equation (12) we are assuming that the risk-free asset is unconditionally orthogonal to $M(t+1)$ (i.e., the risk-free asset is a zero-beta asset). Equations (12) and (13) can be used to assess the fit of an estimated model to the cross-section of average returns.

For the cross-sectional tests we use the standard 25 portfolio of Fama and French (1993) ordered by size and book-to-market, in addition to the risk-free portfolio, therefore, $N = 26$. These series as well as the SMB (small minus big) and HML (high minus low) factors

[18] The grounds for this moment condition is that the risk-free rate is known at time t, therefore, $\frac{1}{1+R_f(t+1)} = \mathbb{E}\left[M(t+1)|\Omega(t)\right]$. By taking the unconditional mean on both sides yields equation (11).

from the FF three-factor model are available from Ken French's web page. The real per-capita consumption series was constructed using data from the Bureau of Economic Analysis. Specifically, we construct our quarterly series from nominal consumption of nondurables and services, seasonally adjusted, per capita (NIPA Table 7.1). Real consumption was calculated by deflating the nominal series by the PCE (personal consumption expenditures) price index, 2005=100 (NIPA Table 2.3.4).

Tables 10 to 12 present the results of the estimation by the Generalized Method of Moments (GMM) for the moment conditions in equations (9) and (11).[19] We report the asymptotic t-stats, and the Wald and J_T tests based on the covariance matrices of pricing errors corrected for heteroscedasticity and serial correlation using the Newey-West estimator. We use the root of mean square errors (RMSE) to measure the fit of an estimated model to the cross-section of average returns. Figures 3 to 15 plot the fitted expected returns for the 26 portfolios against their realized average returns.[20]

Table 10 and Figures 3 to 5 show the results for the unconditional CAPM, CCAPM and FF three-factor models. The J_T tests do not reject any conditional or unconditional model, so our inference is based solely on the t-stats and Wald tests. The latter tests the hypothesis that all coefficients except the constant are zero. According to the Wald test, no model is significant at the 5% level; nevertheless, the FF three-factor model presents the best cross-sectional adjustment ($RMSE = 0.55\%$).

Table 11 and Figures 6 to 10 present the results for the conditional CAPM models. A Wald test for the null hypothesis that conditional CCAPM model does not improve the adjustment relative to the unconditional CCAPM model is included in the p-value Wald *CAPM* row (i.e., it tests whether the additional coefficients in the conditional CCAPM model are zero). The table shows that the CAPM conditional on Δf exhibits by far the best forecasting power for the cross-section of expected returns ($RMSE = 0.34\%$); moreover, in accordance with both Wald tests, it is the only significant one at the 5% level. The effect of the risk factor $R_m(t)$ is

[19] We use the identity weighting matrix for all estimates, based on the following reasons. First, we do not have theoretical arguments for giving more or less importance to a particular portfolio. Second, the number of moment conditions ($N = 26$) is large relative to our sample size ($T = 103$), so this choice avoids dealing with estimates that depend on unstable and near singular error covariance matrices.

[20] The 25 portfolios, sorted by size and book-to-market ratio, are labeled with two digits. The first digit refers to the size quintile (1 indicating the smallest firms, 5 the largest), and the second digit refers to book-to-market quintile (1 indicating the portfolio with the lowest book-to-market ratio, 5 with the highest).

significant only through its interaction with $\Delta f(t-4)$. This implies that a positive oil price shock accompanied by a subsequent decline in market return ($\Delta f(t-4) \cdot R_m(t) < 0$) causes a drop in future portfolio returns. This effect strongly suggests that our variable also has forecasting power for returns on more disaggregated stock portfolios. Finally, CAPM models conditioned on $d-p$ and gap slightly outperform the FF three-factor model, but none of these models is statistically significant.

Table 12 and Figures 11 to 15 display the outcomes for conditional CCAPM models. Based on these outcomes, we note that the CCAPM conditional on Δf is the only one that outperforms the FF three-factor model; however, it is not significant at the 5% level. Also, the CCAPM model conditioned on R_f is statistically significant but is outperformed by several conditional and unconditional models ($RMSE = 0.60\%$).

In summary, the conditional CAPM on our variable has significant predictive power for the cross-section of expected stock returns and has a better fit than all unconditional and conditional models considered here.

7 Conclusions

Although the predictability of stock market excess returns has been associated with the business cycle, the evidence that supports this relationship is far from being conclusive. Moreover, when the sample is extended to include the period of the subprime crisis, none of the popular predictors exhibit forecasting power for market excess returns. In this paper, we show that such a relationship exists.

We find that unexpected oil price changes, a non-persistent variable with deep macroeconomic roots, have significant forecasting power for stock returns at short horizons. Our variable, proxied by futures returns on crude oil, shows statistically and economically significant predictive power for stock returns at horizons from one to three quarters. Its predictive power outperforms those of the risk-free rate, the dividend-price ratio, the consumption-wealth ratio and the output gap, with quarterly $\bar{R}^2$ between 6% and 7%. This result is robust against the inclusion of other variables and out-of-sample tests. However, at longer horizons, none of the variables displays significant forecasting ability. Our results also validate the recent findings of Boudoukh, Richardson, and Whitelaw (2008) that unstable results in previous studies in this literature are due to the high persistence of the predictors used.

Our variable also shows a significant forecasting power for the cross-section of expected returns. We build a conditional CAPM model on oil price shocks, which shows high statistical significance and better adjustment than all conditional and unconditional models considered, including the Fama and French (1993) three-factor model.

From a practical perspective, unlike variables based on macroeconomic series, such as the consumption-wealth ratio and the output gap, our variable can be directly observed and is available on a daily basis at no cost. These characteristics make use of our variable by potential investors highly feasible.

Finally, an open question motivated by the emerging literature on the financialization of commodities is how well is the forecasting ability of other commodities on stock returns. The correlation of indexed commodities and oil has increased dramatically since 2004 because of the speculative trading in futures markets, suggesting that these commodities would also have some predictive power. We leave this topic for future research.

References

Acharya, Viral V., Lars A. Lochstoer, and Tarun Ramadorai, 2011, Limits to arbitrage and hedging: Evidence from commodity markets, NBER Working Paper No. 16875.

Aguiar-Conraria, Luis, and Yi Wen, 2007, Understanding the large negative impact of oil shocks, *Journal of Money, Credit and Banking* 39, pp. 925–944.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

Apergis, Nicholas, and Stephen M. Miller, 2009, Do structural oil-market shocks affect stock prices?, *Energy Economics* 31, 569 – 575.

Bachmeier, Lance, 2008, Monetary policy and the transmission of oil shocks, *Journal of Macroeconomics* 30, 1738 – 1755.

Bakshi, Gurdip, George Panayotov, and Georgios Skoulakis, 2011, The baltic dry index as a predictor of global stock returns, commodity returns, and global economic activity, Working Paper, University of Maryland.

Balke, Nathan S., Stephen P.A. Brown, and Mine K. Yucel, 2002, Oil price shocks and the u.s. economy: Where does the asymmetry originate?, *Energy Joumal* 23, 27–52.

Barsky, Robert B., and Lutz Kilian, 2004, Oil and the macroeconomy since the 1970s, *Journal of Economic Perspectives* 18, 115–134.

Basak, Suleyman, and Anna Pavlova, 2012, A model of financialization of commodities, Working paper, London Business School.

Berkowitz, Jeremy, and Lutz Kilian, 2000, Recent developments in bootstrapping time series, *Econometric Reviews* 19, 1–48.

Bernanke, Ben S., 1983, Irreversibility, uncertainty, and cyclical investment, *The Quarterly Journal of Economics* 98, pp. 85–106.

———, Mark Gertler, and Mark W. Watson, 1997, Systematic monetary policy and the effects of oil price shocks, *Brookings Papers on Economic Activity* 1997, 91–157.

Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw, 2008, The myth of long-horizon predictability, *Review of Financial Studies* 21, 1577–1605.

Breeden, Douglas T., and Robert H. Litzenberger, 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* 51, 621–651.

Buyuksahin, Bahattin, and Michel A. Robe, 2010, Speculators, commodities and cross-market linkages, Working paper, American University.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373 – 399.

———, 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.

———, and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Carruth, Alan A., Mark A. Hooker, and Andrew J. Oswald, 1998, Unemployment equilibria and input prices: Theory and evidence from the united states, *Review of Economics and Statistics* 80, 621–628.

Casassus, Jaime, and Pierre Collin-Dufresne, 2005, Stochastic convenience yield implied from commodity futures and interest rates, *Journal of Finance* 60, 2283–2332.

Ciner, Cetin, 2001, Energy shocks and financial markets: Nonlinear linkages, *Studies in Nonlinear Dynamics and Econometrics* 5, 203–212.

Clark, Todd E., and Michael W. McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85 – 110.

——— , 2005, Evaluating direct multistep forecasts, *Econometric Reviews* 24, 369–404.

Cochrane, John H., 2005, *Asset Pricing* (Princeton University Press).

Cologni, Alessandro, and Matteo Manera, 2008, Oil prices, inflation and interest rates in a structural cointegrated var model for the g-7 countries, *Energy Economics* 30, 856 – 888.

——— , 2009, The asymmetric effects of oil shocks on output growth: A markov-switching analysis for the g-7 countries, *Economic Modelling* 26, 1 – 29.

Cooper, Ilan, and Richard Priestley, 2009, Time-varying risk premiums and the output gap, *Review of Financial Studies* 22, 2801–2833.

Davis, Steven J., and John Haltiwanger, 2001, Sectoral job creation and destruction responses to oil price changes, *Journal of Monetary Economics* 48, 465 – 512.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, pp. 253–263.

Driesprong, Gerben, Ben Jacobsen, and Benjamin Maat, 2008, Striking oil: Another puzzle?, *Journal of Financial Economics* 89, 307 – 327.

Etula, Erkko, 2010, Broker-dealer risk appetite and commodity returns, Working paper, Federal Reserve Bank of New York.

Fama, Eugene F., and Kenneth R. French, 1987, Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage, *Journal of Business* 60, 55–73.

——— , 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

——— , 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Ferderer, J. Peter, 1996, Oil price volatility and the macroeconomy, *Journal of Macroeconomics* 18, 1–26.

Finn, Mary G., 2000, Perfect competition and the effects of energy price increases on economic activity, *Journal of Money, Credit and Banking* 32, 400–416.

Gronwald, Marc, 2008, Large oil shocks and the us economy: Infrequent incidents with large effects, *Energy Journal* 29, 151–171.

Hamilton, James D., 1983, Oil and the macroeconomy since World War II, *Journal of Political Economy* 91, 228–248.

———, 1988, A neoclassical model of unemployment and the business cycle, *Journal of Political Economy* 96, 593–617.

———, 1996, This is what happened to the oil price-macroeconomy relationship, *Journal of Monetary Economics* 38, 215 – 220.

———, 2003, What is an oil shock?, *Journal of Econometrics* 113, 363–398.

———, 2008, oil and the macroeconomy, in Steven N. Durlauf, and Lawrence E. Blume, ed.: *The New Palgrave Dictionary of Economics* (Palgrave Macmillan: Basingstoke).

———, and Ana Maria Herrera, 2004, Comment: Oil shocks and aggregate macroeconomic behavior: The role of monetary policy, *Journal of Money, Credit and Banking* 36, 265–286.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold, 1998, Tests for forecast encompassing, *Journal of Business & Economic Statistics* 16, pp. 254–259.

Hooker, Mark A., 1996, What happened to the oil price-macroeconomy relationship?, *Journal of Monetary Economics* 38, 195 – 213.

Huang, Roger D., Ronald W. Masulis, and Hans R. Stoll, 1996, Energy shocks and financial markets, *Journal of Futures Markets* 16, 1–27.

Huntington, Hillard G., 2007, Oil shocks and real u.s. income, *Energy Joumal* 28, 31–46.

Inoue, Atsushi, and Lutz Kilian, 2005, In-sample or out-of-sample tests of predictability: Which one should we use?, *Econometric Reviews* 23, 371–402.

Jones, Charles, and Gautam Kaul, 1996, Oil and the stock markets, *Journal of Finance* 51, 463–491.

Kilian, Lutz, 1998, Small-sample confidence intervals for impulse response functions, *The Review of Economics and Statistics* 80, pp. 218–230.

———, 1999, Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions?, *Journal of Applied Econometrics* 14, pp. 491–510.

———, 2008, A comparison of the effects of exogenous oil supply shocks on output and inflation in the g7 countries, *Journal of the European Economic Association* 6, 78–121.

———, 2009, Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, *American Economic Review* 99, 1053–1069.

———, and Cheolbeom Park, 2009, The impact of oil price shocks on the u.s. stock market, *International Economic Review* 50, 1267–1287.

Kliesen, Kevin L., 2008, Oil and the u.s. macroeconomy: An update and a simple forecasting exercise, *Federal Reserve Bank of St. Louis Review* 90, 505–516.

Kumar, Manmohan S., 1992, The forecasting accuracy of crude oil futures prices, *Staff Papers - International Monetary Fund* 39, pp. 432–461.

Leduc, Sylvain, and Keith Sill, 2004, A quantitative analysis of oil-price shocks, systematic monetary policy, and economic downturns, *Journal of Monetary Economics* 51, 781–808.

Lee, Kiseok, and Shawn Ni, 2002, On the dynamic effects of oil price shocks: A study using industry level data, *Journal of Monetary Economics* 49, 823–852.

——— , and Ronald A. Ratti, 1995, Oil shocks and the macroeconomy: The role of price variability, *Energy Journal* 16, 39–56.

Lettau, Martin, and Sydney C. Ludvigson, 2001a, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–849.

——— , 2001b, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175 – 194.

Lilien, David M., 1982, Sectoral shifts and cyclical unemployment, *Journal of Political Economy* 90, pp. 777–793.

Lustig, Hanno N., and Stijn G. Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective., *Journal of Finance* 60, 1167–1219.

Ma, Cindy W., 1989, Forecasting efficiency of energy futures prices, *Journal of Futures Markets* 9, 393–419.

Masters, Michael W., 2008, Testimony before committee on homeland security and governmental affairs of the united states senate, .

McCracken, Michael W., 2007, Asymptotics for out of sample tests of granger causality, *Journal of Econometrics* 140, 719 – 752.

Mork, Knut Anton, 1989, Oil and the macroeconomy when prices go up and down: An extension of hamilton's results, *Journal of Political Economy* 97, 740–744.

Nagel, Stefan, and Kenneth J. Singleton, 2011, Estimation and evaluation of conditional asset pricing models, *The Journal of Finance* 66, 873–909.

Nandha, Mohan, and Robert Faff, 2008, Does oil move equity prices? a global view, *Energy Economics* 30, 986–997.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, pp. 703–708.

——— , 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, pp. 631–653.

Park, Jungwook, and Ronald A. Ratti, 2008, Oil price shocks and stock markets in the u.s. and 13 european countries, *Energy Economics* 30, 2587 – 2608.

Sadorsky, Perry, 1999, Oil price shocks and stock market activity, *Energy Economics* 21, 449–469.

Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Sims, Christopher A., 1980, Macroeconomics and reality, *Econometrica* 48, pp. 1–48.

Singleton, Kenneth J., 2012, Investor flows and the 2008 boom/bust in oil prices, Working paper, Stanford University.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Stine, Robert A., 1987, Estimating properties of autoregressive forecasts, *Journal of the American Statistical Association* 82, pp. 1072–1078.

Switzer, Lorne N., and Mario El-Khoury, 2007, Extreme volatility, speculative efficiency, and the hedging effectiveness of the oil futures markets, *Journal of Futures Markets* 27, 61–84.

Tang, Ke, and Wei Xiong, 2012, Index investment and the financialization of commodities, Forthcoming *Financial Analysts Journal*.

Valkanov, Rossen, 2003, Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics* 68, 201–232.

Wei, Chao, 2003, Energy, the stock market, and the putty-clay investment model, *American Economic Review* 93, 311–323.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Appendix

A Tests for out-of-sample predictability

This appendix presents the three metrics we use to test the out-of-sample performance of the predictors. First, we estimate the forecast errors for the benchmark and competing models in equations (4) and (5) as:

$$\text{benchmark:} \qquad \hat{u}_1(t+1) = [R_m(t+1) - R_f(t+1)] - \hat{\alpha}_1(t) \tag{A1}$$

$$\text{competing:} \qquad \hat{u}_2(t+1) = [R_m(t+1) - R_f(t+1)] - \hat{\alpha}_2(t) - \hat{\beta}(t)'X(t) \tag{A2}$$

for $t = Q, \ldots, T-1$ and the coefficients $\hat{\alpha}_1(t)$, $\hat{\alpha}_2(t)$ and $\hat{\beta}(t)$ are estimated with data through period $1, \ldots, t$. Then, one-step-ahead forecasts from the competing model can be compared to forecasts from the benchmark model (that is, a restricted version of the competing model) by using statistics based on the time series $\hat{u}_1(t+1)$ and $\hat{u}_2(t+1)$.

The first test we use for out-of-sample predictability is the forecast encompassing test of Clark and McCracken (2001). To clarify how it works, we follow Harvey, Leybourne, and Newbold (1998) and specify a regression of the excess stock return on a weighted average of forecasted values from the benchmark and competing models:

$$R_m(t+1) - R_f(t+1) = (1-\lambda)\,[\alpha_1] + \lambda\,[\alpha_2 + \beta' X(t)] + \nu(t+1) \tag{A3}$$

where $0 \leq \lambda \leq 1$ and $\nu(t+1)$ is a error term. Substituting both forecasts from equations (4) and (5) yields:

$$u_1(t+1) = \lambda\,[u_1(t+1) - u_2(t+1)] + \nu(t+1) \tag{A4}$$

Then, as λ is also the coefficient of the regression model in equation (A4):

$$\lambda = \frac{\mathbb{C}ov\,[u_1(t+1), u_1(t+1) - u_2(t+1)]}{\mathbb{V}ar\,[u_1(t+1) - u_2(t+1)]} \tag{A5}$$

Thus, the combined forecast will have a smaller expected squared error than the benchmark model forecast unless the covariance between $u_1(t+1)$ and $u_1(t+1) - u_2(t+1)$ is zero (i.e., $\lambda = 0$). This way, Clark and McCracken (2001) tests the null hypothesis that $\lambda \leq 0$ and is given by:

$$\text{ENC-NEW} = P\,\frac{\sum_{t=Q}^{T-1}\left(\hat{u}_1(t+1)^2 - \hat{u}_1(t+1)\hat{u}_2(t+1)\right)}{\sum_{t=Q}^{T-1}\hat{u}_2(t+1)^2} \tag{A6}$$

Under the null hypothesis that the benchmark model encompasses the competing model, the covariance between series $u_1(t+1)$ and $u_1(t+1) - u_2(t+1)$ will be less than or equal to zero. Under the alternative that the competing model contains added information, the covariance should be positive. Hence, the encompassing test presented above is one-sided. Clark and McCracken (2001) demonstrate that the limiting distribution of ENC-NEW is not normal when the forecasts are nested under the null, but they provide asymptotic critical values for this statistic.

The second test used is the one developed by McCracken (2007). This test, unlike the one proposed by Diebold and Mariano (1995) in the context of non-nested models, allows for comparison of predictive accuracy between nested models. In particular, we use it to test for equality of the mean squared forecasting errors (MSE) from the benchmark and competing models, which is given by:

$$\text{MSE-F} = P\frac{\sum_{t=Q}^{T-1}\left(\hat{u}_1(t+1)^2 - \hat{u}_2(t+1)^2\right)}{\sum_{t=Q}^{T-1}\hat{u}_2(t+1)^2} = P\left[\frac{\text{MSE}_1 - \text{MSE}_2}{\text{MSE}_2}\right] \tag{A7}$$

where $\text{MSE}_j = \sum_{t=Q}^{T-1}\hat{u}_j(t+1)^2/P;\ j = 1, 2$. Based upon the value of this statistic the null of equal MSE is either rejected or not rejected. McCracken (2007) shows that when the two models are nested the alternative is one-sided, rather than two-sided. Moreover, since the asymptotic distribution of MSE-F under the null is nonstandard, tables of asymptotically valid critical values are provided in McCracken (2007).

Clark and McCracken (2001) use simulations to examine the small-sample properties of the ENC-NEW and MSE-F tests. They report that although both tests have good sample size properties, the ENC-NEW test is clearly the more powerful out-of-sample test of predictive ability. While this evidence indicates that the inference from the ENC-NEW test is more reliable, Welch and Goyal (2008) highlight an important problem of encompassing tests in general. The ENC-NEW test uses the entire out-of-sample test to estimate the parameter λ, but an investor trying to use a combined forecast to predict $R_m(t+1) - R_f(t+1)$ will only have the information available up to t to calculate the combination coefficient λ.

The final measure of out-of-sample forecasting performance is the out-of-sample R^2, R^2_{OS}. This statistic is the analog to the in-sample R^2 and in terms of our notation is computed as:

$$R^2_{OS} = 1 - \frac{\sum_{t=Q}^{T-1}\hat{u}_2(t+1)^2}{\sum_{t=Q}^{T-1}\hat{u}_1(t+1)^2} = \frac{\text{MSE}_1 - \text{MSE}_2}{\text{MSE}_1} = \frac{\text{MSE-F}}{P}\left(\frac{\text{MSE}_2}{\text{MSE}_1}\right) \tag{A8}$$

As can be seen from equation (A8), if R^2_{OS} is positive then the competing model has a lower MSE than the benchmark model. Also, as shown in the last equality, the R^2_{OS} is not a statistic that provides new information with respect to the other tests, since it is merely a scaled-up version of the MSE-F statistic.[21] That is, predictor variables with greater MSE-F will also exhibit greater R^2_{OS}. Then, this could also be considered a test of equal MSE, assuming that it has an asymptotic distribution.[22]

As mentioned above, in the context of one-step ahead forecasts, Clark and McCracken (2001) and McCracken (2007) provide asymptotic critical values for the ENC-NEW and MSE-F statistics, respectively. These critical values depend on two parameters: $\pi = P/Q$ and $K_2 - 1$, the number of variables included in $X(t)$. Because the tables with the critical values for these nonstandard tests do not contain the particular value of π chosen by us, we follow Clark and McCracken (2005) and obtain these values with an inference technique based on bootstrapping. In addition, based on the bootstrapped time series, we obtain the empirical distribution of the R^2_{OS} statistic and critical values for the tests. In particular, we use a parametric bootstrap (Berkowitz and Kilian, 2000) and our algorithm has five steps, which we briefly describe below:[23]

1. We estimate a bivariate VAR for the excess stock return, $R_m(t) - R_f(t)$, and the variable $X(t)$ under the null hypothesis of nonpredictability. The model is estimated with OLS and using the full sample. The excess return is modeled according to equation (4) and for the variable $X(t)$ the optimal number of lags of $R_m(t) - R_f(t)$ and $X(t)$ were chosen with the AIC criterion.

[21] The MSE-F, as derived originally by McCracken (2007), is designed to be used with any loss function. Here we use only one particular case.

[22] Just as the in-sample R^2 has an adjusted counterpart for degrees of freedom, $\bar{R}^2$, Welch and Goyal (2008) use a version of the R^2_{OS} adjusted for degrees of freedom. However, since the forecasting errors are not part of the OLS estimation, and therefore there is no loss of degrees of freedom in its calculation, we consider it inappropriate to adjust this statistic.

[23] For more details on this methodology see Clark and McCracken (2005).

2. The coefficients of the VAR were adjusted for the small-sample bias using Kilian's (1998) procedure with 10,000 bootstrap draws.

3. We bootstrapped 999 time series for the excess stock return and the variable $X(t)$ by drawing from the rescaled sample residuals with replacement (Berkowitz and Kilian, 2000), using the adjusted VAR coefficients and initial observations selected by sampling from actual data (Stine, 1987).

4. Each artificial bivariate time series is used to estimate the benchmark and competing models (equations (4) and (5)) in a recursive way. Forecasting errors are calculated according to equations (A1) and (A2) and using the sample portions described above. The ENC-NEW, MSE-F and R^2_{OS} statistics were calculated based on these estimated forecasting errors with equations (A6), (A7) and (A8).

5. For each statistic, critical values are simply computed as percentiles of the corresponding empirical distribution. The p-values are calculated using the standard method.

Table 1: **Literature on financial markets and oil prices**

The table provides information on studies on financial markets and oil prices. The variables used are: Δf: log returns on oil futures prices; Δs: log returns on nominal spot prices; Δs_r: log returns on real spot prices; Sop: oil prices scaled by volatility, unexpected changes in oil prices (Lee, Ni, and Ratti, 1995); $Nopi$: net oil price increases (Hamilton, 1996; Hamilton, 2003); Vol: rolling volatility of oil price changes; $x^+ = max(0, x)$; $x^- = min(0, x)$.

Paper	Subject	Frequency	Data		
			Stock market	Oil	
				Series	Variable
Huang, Masulis, and Stoll (1996)	Joint dynamics	Daily	S&P 500, industrial portfolios, oil companies	Heating and crude oil futures	Δf
Jones and Kaul (1996)	Market efficiency	Quarterly	Real country indexes	PPI oil and related products	Δs
Sadorsky (1999)	Joint dynamics	Monthly	S&P 500/CPI	PPI fuels/CPI	Δs_r, Δs_r^+, Δs_r^-, Sop, Sop^+, Sop^-
Ciner (2001)	Joint dynamics	Daily	S&P 500	Heating and crude oil futures	Δf
Driesprong, Jacobsen, and Maat (2008)	Predictability	Monthly	Country indexes, sector indexes	Brent, WTI, Dubai, Arab Light, Brent futures, Oil futures	Δs, Δf
Park and Ratti (2008)	Joint dynamics	Monthly	Real country indexes	Brent/PPI all commodities	Δs_r, Sop, $Nopi$, Δs_r^+, Δs_r^-, Sop^+, Sop^-, Vol
Kilian and Park (2009)	Oil demand and supply shocks	Monthly	Real CRSP value weighted, industry portfolios	EIA refiner acquisition cost/CPI	Δs_r
Apergis and Miller (2009)	Oil demand and supply shocks	Monthly	Real country indexes	EIA Refiner Acquisition Cost/CPI	Δs_r

Table 2: **Stock market returns and the business cycle**

Conditional excess stock returns for the sample period 1926Q3 to 2009Q4.

Classification of states based on NBER Business Cycle Dates.

State	Frequency	Average excess return	Standard deviation of excess returns
Expansion	252	2.9	9.1
Peak	15	-5.5	8.1
Recession	52	-3.0	17.3
Trough	15	12.3	10.0
Total	334	2.0	11.3

Table 3: **Optimal lags of oil price shocks**

OLS regressions of excess stock returns on lags of oil price shocks, 1983Q2-2009Q4.

All estimations use full sample and include a constant.

Lags of oil price shocks	AIC
0	-4.884
1	-4.891
2	-4.867
3	-4.846
4	-4.893
5	-4.870
6	-4.844
7	-4.824
8	-4.803

Table 4: **VAR estimation results, 1983Q2-2009Q4**

Maximum likelihood estimates of the VAR(4) model for the rate of growth of industrial production ($\Delta\%IP(t)$), excess stock market returns ($R_m(t) - R_f(t)$) and log returns on crude oil futures ($\Delta f(t)$). Asymptotic t-stat in parentheses. Granger causality was tested using the asymptotic Wald test. The *All* row at the bottom of the table refers to all coefficients except the constant.

	$\Delta\%IP(t)$	$R_m(t) - R_f(t)$	$\Delta f(t)$
Constant	0.000	0.011	0.020
	(0.45)	(1.20)	(0.88)
$\Delta\%IP(t-1)$	0.276	0.956	-0.205
	(3.04)	(1.25)	(-0.11)
$\Delta\%IP(t-2)$	-0.027	-0.852	0.404
	(-0.26)	(-0.99)	(0.19)
$\Delta\%IP(t-3)$	-0.113	-1.215	0.460
	(-1.10)	(-1.40)	(0.22)
$\Delta\%IP(t-4)$	0.347	2.327	-1.194
	(3.70)	(2.93)	(-0.62)
$R_m(t-1) - R_f(t-1)$	0.061	0.038	0.269
	(5.59)	(0.41)	(1.19)
$R_m(t-2) - R_f(t-2)$	0.030	-0.069	0.191
	(2.52)	(-0.68)	(0.77)
$R_m(t-3) - R_f(t-3)$	0.005	-0.038	0.268
	(0.42)	(-0.37)	(1.07)
$R_m(t-4) - R_f(t-4)$	0.026	0.019	-0.087
	(2.12)	(0.19)	(-0.35)
$\Delta f(t-1)$	0.010	0.065	0.035
	(2.09)	(1.59)	(0.35)
$\Delta f(t-2)$	-0.006	-0.044	-0.183
	(-1.22)	(-1.04)	(-1.79)
$\Delta f(t-3)$	-0.012	-0.004	0.027
	(-2.35)	(-0.10)	(0.26)
$\Delta f(t-4)$	-0.005	-0.119	-0.057
	(-1.00)	(-2.74)	(-0.54)
R^2	0.51	0.18	0.06
$\bar{R}^2$	0.44	0.07	-0.06
p-value Granger causality test			
All	0.00	0.03	0.85
$\Delta\%IP$	0.00	0.04	0.98
$R_m - R_f$	0.00	0.94	0.53
Δf	0.01	0.04	0.49

Table 5: **Statistics for 1983Q2-2009Q4**

Autocorrelation is the first-order serial correlation.

	$R_m - R_f$	Δf	R_f	$d - p$	cay	gap
Average	0.016	0.021	0.012	-3.781	0.005	-0.014
Standard deviation	0.087	0.195	0.006	0.386	0.019	0.063
Autocorrelation	0.031	0.031	0.972	0.974	0.910	0.977

	Correlation matrix					
	$R_m - R_f$	Δf	R_f	$d - p$	cay	gap
$R_m - R_f$	1.000					
Δf	-0.039	1.000				
R_f	0.006	-0.035	1.000			
$d - p$	0.139	-0.101	0.524	1.000		
cay	-0.108	-0.100	0.348	0.472	1.000	
gap	-0.121	0.111	-0.210	-0.847	-0.601	1.000

Table 6: **Predictive regressions for excess stock returns, 1983Q2-2009Q4**

OLS regressions for excess stock returns on the predictor variables in the first row. All tests are based on covariance matrices of coefficients corrected for heteroscedasticity and serial correlation using Newey and West (1987). Lag length in the Newey-West estimator is floor $\left[4 \cdot (T/100)^{2/9}\right]$, where floor$[x]$ denotes the integer part of x (Newey and West, 1994). Asymptotic t-stat in parentheses. At the bottom of the table, the p-value is for asymptotic Wald test and *All* refers to all coefficients except the constant.

	Δf	R_f	$d-p$	cay	gap
Constant	0.019	0.013	0.154	0.014	0.015
	(2.59)	(0.55)	(2.04)	(1.52)	(1.66)
$\Delta f(t-1)$	0.068				
	(1.95)				
$\Delta f(t-2)$	-0.035				
	(-0.53)				
$\Delta f(t-3)$	-0.024				
	(-0.48)				
$\Delta f(t-4)$	-0.115				
	(-2.68)				
$R_f(t)$		0.403			
		(0.25)			
$d(t-1)-p(t-1)$			0.036		
			(1.75)		
$cay(t-1)$				0.729	
				(2.09)	
$gap(t-1)$					-0.219
					(-1.58)
R^2	0.10	0.00	0.02	0.02	0.02
$\bar{R}^2$	0.06	-0.01	0.01	0.01	0.02
p-value Wald *All*	0.02	0.80	0.08	0.04	0.11

Table 7: **Predictive regressions: Additional controls, 1983Q2-2009Q4**

OLS regressions for excess stock returns on the predictor variables in the first row. All tests are based on covariance matrices of coefficients corrected for heteroscedasticity and serial correlation using Newey and West (1987). Lag length in the Newey-West estimator is floor $\left[4 \cdot (T/100)^{2/9}\right]$, where floor$[x]$ denotes the integer part of x (Newey and West, 1994). Asymptotic t-stat in parentheses. At the bottom of the table, the p-value is for asymptotic Wald test, *All* refers to all coefficients except the constant and Δf refers to the coefficients associated with the four lags of that variable.

	Δf & R_f	Δf & $d-p$	Δf & cay	Δf & gap	Δf & All
Constant	0.012	0.140	0.015	0.017	0.269
	(0.57)	(2.02)	(1.81)	(2.01)	(1.03)
$\Delta f(t-1)$	0.067	0.075	0.075	0.076	0.078
	(1.96)	(2.13)	(2.07)	(2.20)	(1.87)
$\Delta f(t-2)$	-0.036	-0.026	-0.029	-0.025	-0.028
	(-0.53)	(-0.41)	(-0.47)	(-0.38)	(-0.47)
$\Delta f(t-3)$	-0.026	-0.017	-0.023	-0.013	-0.026
	(-0.49)	(-0.33)	(-0.45)	(-0.26)	(-0.53)
$\Delta f(t-4)$	-0.116	-0.108	-0.114	-0.105	-0.117
	(-2.72)	(-2.50)	(-2.67)	(-2.33)	(-2.54)
$R_f(t)$	0.650				-1.708
	(0.45)				(-0.64)
$d(t-1)-p(t-1)$		0.032			0.061
		(1.68)			(1.01)
$cay(t-1)$			0.784		0.985
			(2.00)		(1.78)
$gap(t-1)$				-0.172	0.303
				(-1.50)	(0.93)
R^2	0.10	0.11	0.12	0.11	0.14
$\bar{R}^2$	0.05	0.07	0.08	0.06	0.06
p-value Wald *All*	0.02	0.00	0.00	0.00	0.00
p-value Wald Δf	0.01	0.01	0.01	0.02	0.02

Table 8: **Out-of-sample predictability tests, 1983Q2-2009Q4**

Out-of-sample tests of stock return predictability. Each column reports the results using the predictor from the first row. Out-of-sample period is from 1997Q4 to 2009Q4. ENC-NEW, MSE-F and R^2_{OS} statistics are described in equations (A6), (A7) and (A8). Asymptotic critical values for ENC-NEW test are from Table 1 in Clark and McCracken (2001) using $\pi = 1.0$. Asymptotic critical values for MSE-F test are from Table 4 in McCracken (2007) using $\pi = 1.0$. Bootstrapped p-values and critical values are based on the methodology of Clark and McCracken (2005).

	Δf	R_f	$d-p$	cay	gap
ENC-NEW					
Sample value	2.218	-0.555	0.202	1.630	0.063
0.10 Asymptotic critical value	2.169	0.984	0.984	0.984	0.984
0.05 Asymptotic critical value	3.007	1.584	1.584	1.584	1.584
Bootstrapped p-value	0.077	0.791	0.278	0.103	0.505
0.10 Bootstrapped critical value	1.939	1.325	1.325	1.653	2.163
0.05 Bootstrapped critical value	2.970	2.033	2.276	2.559	3.294
MSE-F					
Sample value	0.603	-1.287	-0.034	0.261	-0.750
0.10 Asymptotic critical value	0.545	0.751	0.751	0.751	0.751
0.05 Asymptotic critical value	1.809	1.548	1.548	1.548	1.548
Bootstrapped p-value	0.081	0.655	0.172	0.197	0.463
0.10 Bootstrapped critical value	0.443	0.789	0.431	1.335	1.816
0.05 Bootstrapped critical value	1.420	1.860	1.268	2.237	3.006
R^2_{OS}					
Sample value	0.012	-0.027	-0.001	0.005	-0.016
Bootstrapped p-value	0.081	0.655	0.172	0.197	0.463
0.10 Bootstrapped critical value	0.009	0.016	0.009	0.027	0.036
0.05 Bootstrapped critical value	0.028	0.037	0.025	0.044	0.058

Table 9: **Long-horizon predictability, 1983Q2-2009Q4**

OLS regressions of $R_m^h(t+h) - R_f^h(t+h)$ on the predictor variables using the same number of observations. All tests are based on covariance matrices of coefficients corrected for heteroscedasticity and serial correlation (Newey and West, 1987). Lag length in the Newey-West estimator is floor $\left[4 \cdot (T/100)^{2/9}\right] + (h-1)$, where floor$[x]$ denotes the integer part of x. Bootstrapped p-values for t-stats in parentheses. $\mathbb{E}[\hat{\beta}_h|\hat{\beta}_1 = \hat{\beta}_1^*]$ and $\mathbb{E}[R_h^2|R_1^2 = R_1^{2*}]$ are described in equations (7) and (8), respectively. p-value Wald *All* is the bootstrapped p-value for the asymptotic Wald test that all coefficients except the constant are zero.

	Forecast horizon in quarters (h)					
	$h=3$	$h=4$	$h=8$	$h=12$	$h=16$	$h=20$
$\Delta f(t)$	0.153	0.019	-0.019	0.105	0.112	-0.143
	(0.01)	(0.44)	(0.45)	(0.33)	(0.35)	(0.28)
$\Delta f(t-1)$	-0.051	-0.057	-0.057	0.029	-0.147	-0.323
	(0.27)	(0.31)	(0.39)	(0.47)	(0.30)	(0.18)
$\Delta f(t-2)$	-0.081	-0.142	-0.128	-0.105	-0.051	-0.380
	(0.24)	(0.18)	(0.28)	(0.35)	(0.42)	(0.14)
$\Delta f(t-3)$	-0.208	-0.146	-0.154	-0.119	-0.138	-0.432
	(0.05)	(0.14)	(0.21)	(0.31)	(0.30)	(0.08)
R^2	0.12	0.05	0.02	0.01	0.01	0.04
$\bar{R}^2$	0.07	0.00	-0.03	-0.04	-0.04	0.00
p-value Wald *All*	0.02	0.58	0.82	0.75	0.71	0.45
$R_f(t+1)$	0.682	0.926	1.129	-0.084	-3.804	5.304
	(0.39)	(0.38)	(0.38)	(0.57)	(0.47)	(0.34)
$\mathbb{E}[\hat{\beta}_h\|\hat{\beta}_1 = \hat{\beta}_1^*]$	-0.125	-0.165	-0.317	-0.456	-0.585	-0.702
R^2	0.00	0.00	0.00	0.00	0.00	0.00
$\mathbb{E}\left[R_h^2\|R_1^2 = R_1^{2*}\right]$	0.00	0.00	0.00	0.00	0.00	0.00
$\bar{R}^2$	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
p-value Wald *All*	0.87	0.87	0.90	0.99	0.78	0.85
$d(t) - p(t)$	0.083	0.111	0.220	0.324	0.440	0.672
	(0.13)	(0.14)	(0.18)	(0.18)	(0.18)	(0.14)
$\mathbb{E}[\hat{\beta}_h\|\hat{\beta}_1 = \hat{\beta}_1^*]$	0.078	0.102	0.190	0.266	0.331	0.388
R^2	0.06	0.08	0.13	0.14	0.17	0.25
$\mathbb{E}\left[R_h^2\|R_1^2 = R_1^{2*}\right]$	0.05	0.06	0.10	0.14	0.16	0.17
$\bar{R}^2$	0.04	0.06	0.12	0.13	0.16	0.24
p-value Wald *All*	0.28	0.29	0.36	0.36	0.36	0.28
$cay(t)$	2.501	3.436	7.493	12.588	18.904	25.065
	(0.26)	(0.28)	(0.30)	(0.36)	(0.25)	(0.10)
$\mathbb{E}[\hat{\beta}_h\|\hat{\beta}_1 = \hat{\beta}_1^*]$	2.867	3.662	6.212	7.986	9.222	10.083
R^2	0.07	0.10	0.21	0.30	0.44	0.49
$\mathbb{E}\left[R_h^2\|R_1^2 = R_1^{2*}\right]$	0.09	0.11	0.16	0.17	0.17	0.17
$\bar{R}^2$	0.06	0.09	0.20	0.30	0.43	0.48
p-value Wald *All*	0.27	0.30	0.32	0.39	0.26	0.11
$gap(t)$	-0.444	-0.578	-1.548	-2.635	-4.161	-5.938
	(0.60)	(0.62)	(0.54)	(0.45)	(0.33)	(0.25)
$\mathbb{E}[\hat{\beta}_h\|\hat{\beta}_1 = \hat{\beta}_1^*]$	-0.345	-0.454	-0.862	-1.229	-1.558	-1.853
R^2	0.04	0.05	0.14	0.21	0.34	0.43
$\mathbb{E}\left[R_h^2\|R_1^2 = R_1^{2*}\right]$	0.02	0.03	0.05	0.07	0.08	0.09
$\bar{R}^2$	0.02	0.03	0.13	0.20	0.33	0.43
p-value Wald *All*	0.64	0.66	0.57	0.47	0.34	0.26

Table 10: **Unconditional asset pricing models, 1983Q2-2009Q4**

GMM estimates of pricing kernel coefficients for the unconditional models. The models are estimated using returns on the Fama and French's (1993) portfolios and a risk-free portfolio ($N = 26$). The identity-weighting matrix is used in all estimates. All tests are based on covariance matrices of errors corrected for heteroscedasticity and serial correlation (Newey and West, 1987). Lag length in the Newey-West estimator is floor $\left[4 \cdot (T/100)^{2/9}\right]$, where floor$[x]$ denotes the integer part of x. Asymptotic t-stat in parentheses. At the bottom of the table, we report the p-value for J_T test of the null that all pricing errors are zero. p-value Wald *All* is the p-value for the asymptotic Wald test that all coefficients except the constant are zero. *RMSE* is the root of mean square errors and measures the fit of the estimated model to the cross-section of average returns.

	CAPM	CCAPM	FF three-factor
Constant	-0.027	0.506	0.022
	(-0.66)	(1.31)	(0.60)
$R_m(t)$	-0.139		
	(-0.11)		
$\Delta c(t)$		-121.603	
		(-1.21)	
$R_m(t) - R_f(t)$			-2.058
			(-1.25)
$SMB(t)$			0.259
			(0.09)
$HML(t)$			-3.986
			(-1.80)
p-value J_T	0.44	0.43	0.32
p-value Wald *All*	0.91	0.23	0.28
RMSE	0.78%	0.74%	0.55%

Table 11: **Conditional CAPM models, 1983Q2-2009Q4**

GMM estimates of pricing kernel coefficients for the conditional CAPM models on variables are in the first row. The models are estimated using returns on Fama and French's (1993) 25 portfolios and the risk-free asset ($N = 26$). The identity-weighting matrix is used in all estimates. All tests are based on covariance matrices of errors corrected for heteroscedasticity and serial correlation. Lag length in the Newey-West estimator is floor $\left[4 \cdot (T/100)^{2/9}\right]$, where floor$[x]$ denotes the integer part of x. Asymptotic t-stat in parentheses. At the bottom of table, we report the p-value for J_T test of the null that all pricing errors are zero. The p-values presented are for asymptotic Wald tests. *All* means all coefficients except the constant. *CAPM* means the coefficients that are not in the unconditional CAPM model. *RMSE* is the root of mean square errors and measures the adjustment of an estimated model to the cross-section of average returns.

	Δf	R_f	$d-p$	cay	gap
Constant	-1.262	0.266	1.962	-0.413	0.007
	(-2.61)	(0.41)	(0.38)	(-0.75)	(0.09)
$\Delta f(t-1)$	-2.744				
	(-2.18)				
$\Delta f(t-2)$	-3.239				
	(-2.68)				
$\Delta f(t-3)$	-1.387				
	(-0.46)				
$\Delta f(t-4)$	6.932				
	(3.07)				
$R_f(t)$		-39.419			
		(-0.66)			
$d(t-1)-p(t-1)$			0.520		
			(0.38)		
$cay(t-1)$				42.908	
				(1.13)	
$gap(t-1)$					0.283
					(0.03)
$R_m(t)$	1.157	-7.826	-59.513	1.602	-1.634
	(0.59)	(-1.42)	(-1.96)	(0.44)	(-0.96)
$\Delta f(t-1) \cdot R_m(t)$	-20.573				
	(-1.21)				
$\Delta f(t-2) \cdot R_m(t)$	-17.374				
	(-1.63)				
$\Delta f(t-3) \cdot R_m(t)$	11.929				
	(0.54)				
$\Delta f(t-4) \cdot R_m(t)$	-63.751				
	(-2.92)				
$R_f(t) \cdot R_m(t)$		785.090			
		(1.42)			
$(d(t-1)-p(t-1)) \cdot R_m(t)$			-15.186		
			(-1.90)		
$cay(t-1) \cdot R_m(t)$				-251.228	
				(-0.95)	
$gap(t-1) \cdot R_m(t)$					89.443
					(1.79)
p-value J_T	0.07	0.32	0.33	0.33	0.35
p-value Wald *All*	0.00	0.46	0.19	0.45	0.25
p-value Wald *CAPM*	0.00	0.33	0.15	0.52	0.16
RMSE	0.34%	0.67%	0.53%	0.72%	0.53%

Table 12: **Conditional CCAPM models, 1983Q2-2009Q4**

GMM estimates of pricing kernel coefficients for the conditional CCAPM models on variables are in the first row. The models are estimated using returns on Fama and French's (1993) 25 portfolios and the risk-free asset ($N = 26$). The identity-weighting matrix is used in all estimates. All tests are based on covariance matrices of errors corrected for heteroscedasticity and serial correlation. Lag length in the Newey-West estimator is floor $\left[4 \cdot (T/100)^{2/9}\right]$, where floor$[x]$ denotes the integer part of x. Asymptotic t-stat in parentheses. At the bottom of table, we report the p-value for J_T test of the null that all pricing errors are zero. The p-values presented are for asymptotic Wald tests. *All* means all coefficients except the constant. *CCAPM* means the coefficients that are not in the unconditional CCAPM model. *RMSE* is the root of mean square errors and measures the adjustment of an estimated model to the cross-section of average returns.

	Δf	R_f	$d-p$	cay	gap
Constant	0.811	1.033	1.877	0.375	0.485
	(2.25)	(1.48)	(0.44)	(0.86)	(1.50)
$\Delta f(t-1)$	-0.483				
	(-0.38)				
$\Delta f(t-2)$	-1.858				
	(-2.15)				
$\Delta f(t-3)$	0.384				
	(0.29)				
$\Delta f(t-4)$	-0.407				
	(-0.17)				
$R_f(t)$		-186.224			
		(-2.18)			
$d(t-1)-p(t-1)$			0.364		
			(0.32)		
$cay(t-1)$				19.948	
				(0.76)	
$gap(t-1)$					-0.253
					(-0.04)
$\Delta c(t)$	-308.738	-320.549	-390.098	-160.464	-144.812
	(-2.77)	(-1.83)	(-0.41)	(-1.57)	(-1.51)
$\Delta f(t-1)\cdot\Delta c(t)$	116.051				
	(0.35)				
$\Delta f(t-2)\cdot\Delta c(t)$	-134.842				
	(-0.59)				
$\Delta f(t-3)\cdot\Delta c(t)$	343.002				
	(1.09)				
$\Delta f(t-4)\cdot\Delta c(t)$	651.304				
	(1.46)				
$R_f(t)\cdot\Delta c(t)$		34600.413			
		(2.88)			
$(d(t-1)-p(t-1))\cdot\Delta c(t)$			-68.620		
			(-0.27)		
$cay(t-1)\cdot\Delta c(t)$				1269.334	
				(0.34)	
$gap(t-1)\cdot\Delta c(t)$					1302.922
					(1.05)
p-value J_T	0.08	0.34	0.33	0.32	0.34
p-value Wald *All*	0.10	0.03	0.44	0.33	0.47
p-value Wald *CCAPM*	0.24	0.01	0.93	0.74	0.54
RMSE	0.48%	0.60%	0.73%	0.69%	0.68%

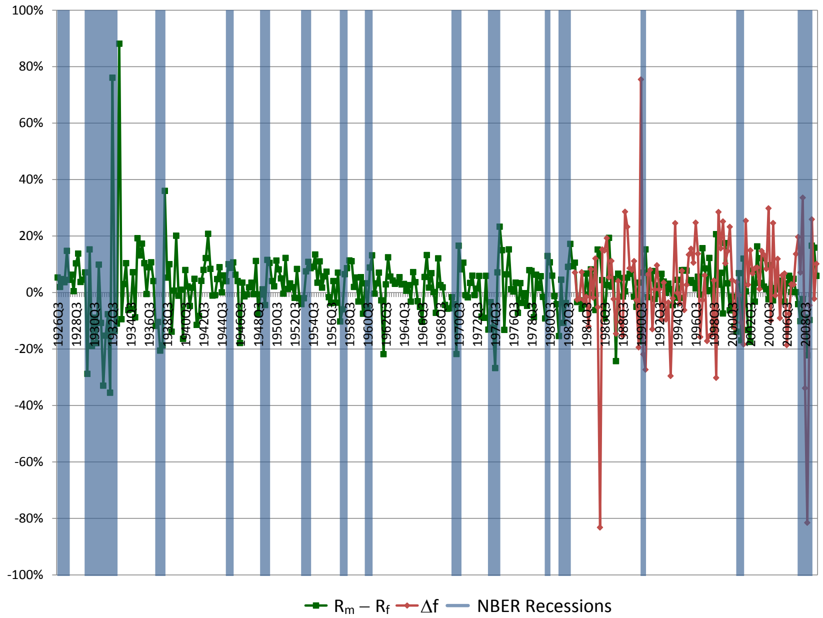

Figure 1: Excess stock market returns, oil shocks and the business cycle, 1926Q3-2009Q4.

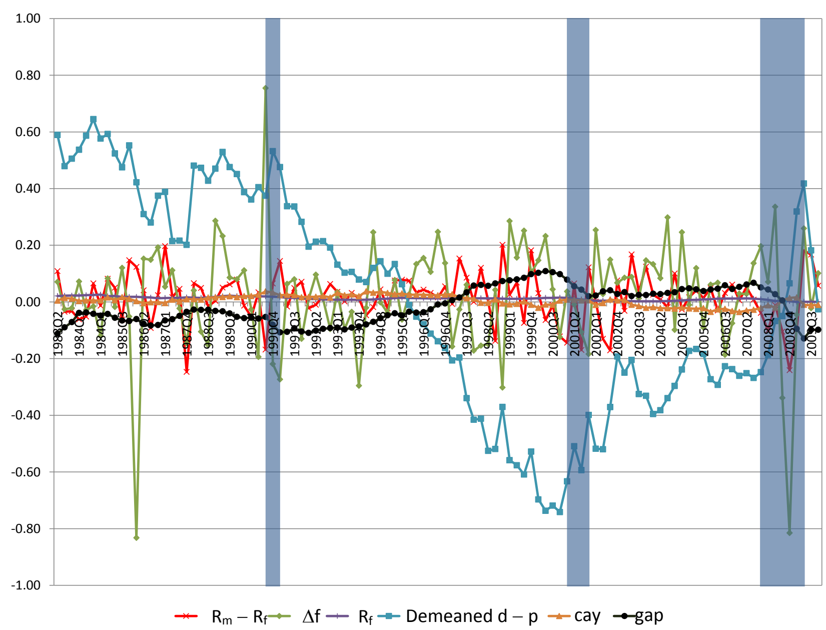

Figure 2: Excess stock market returns and predictor variables, 1983Q2-2009Q4.

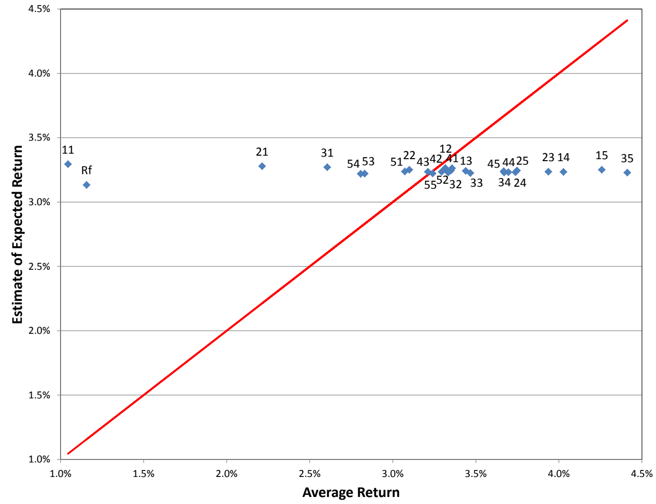

Figure 3: Realized vs. expected returns by CAPM.

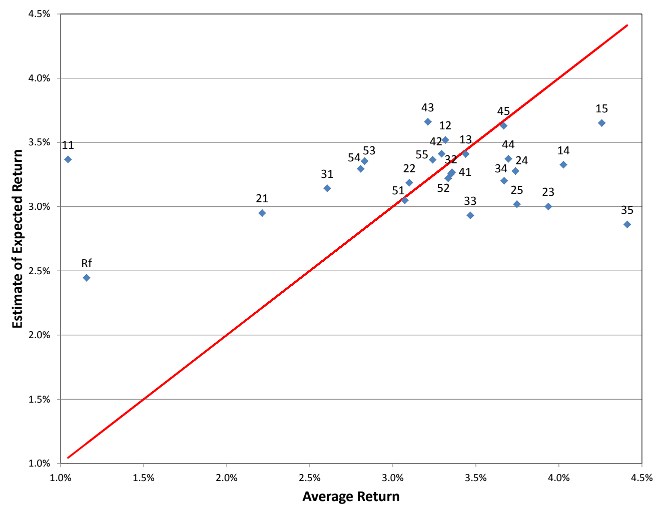

Figure 4: Realized vs. expected returns by CCAPM.

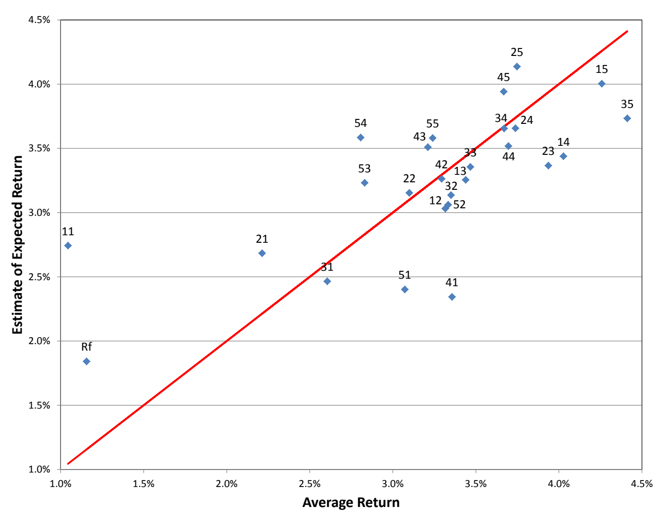

Figure 5: Realized vs. expected returns by FF three-factor.

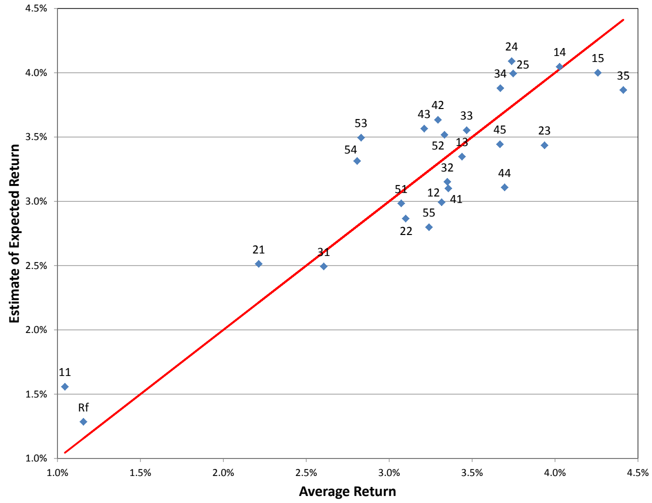

Figure 6: Realized vs. expected returns by CAPM conditioned on Δf.

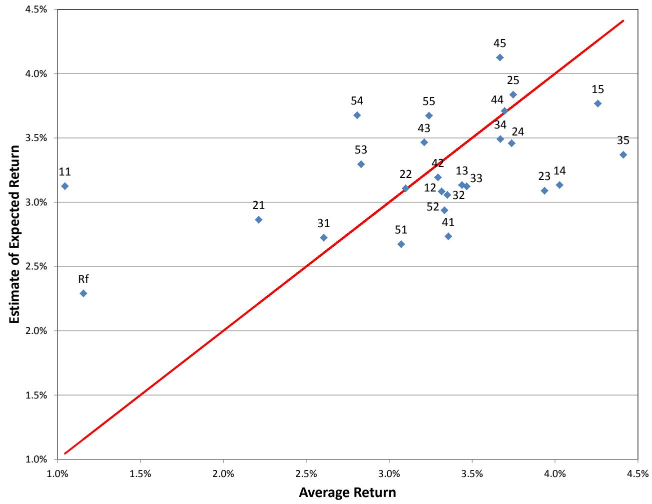

Figure 7: Realized vs. expected returns by CAPM conditioned on R_f.

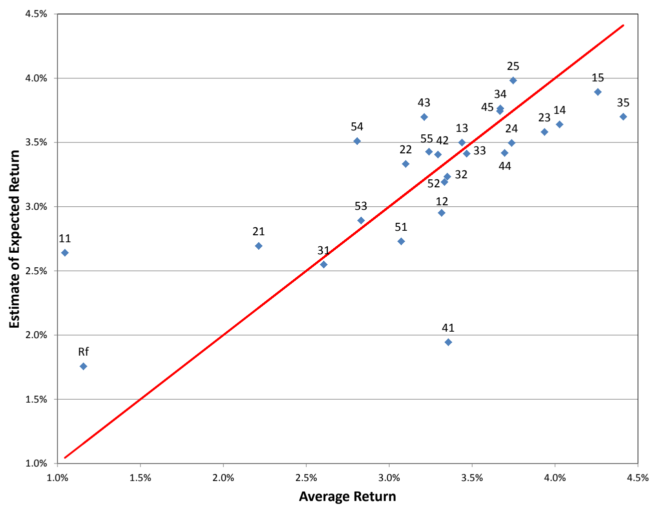

Figure 8: Realized vs. expected returns by CAPM conditioned on $d - p$.

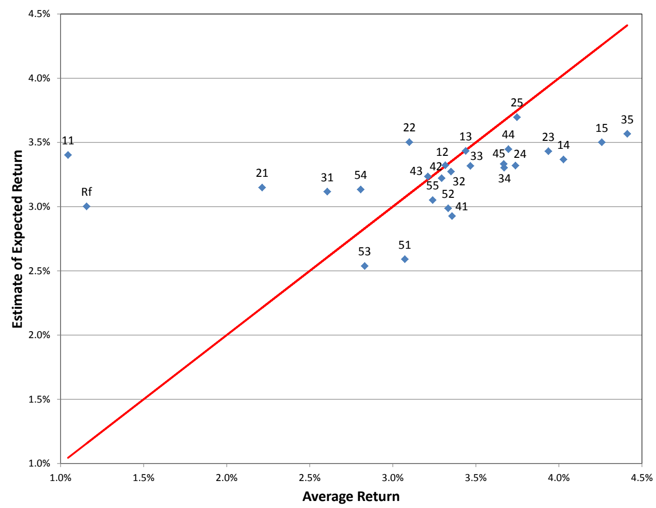

Figure 9: Realized vs. expected returns by CAPM conditioned on *cay*.

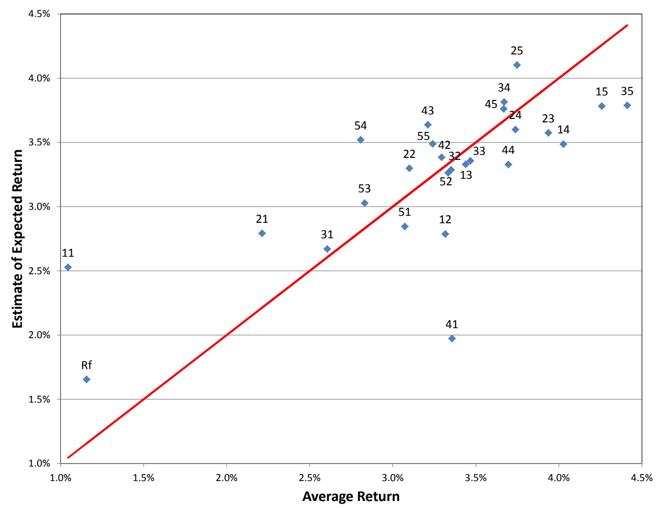

Figure 10: Realized vs. expected returns by CAPM conditioned on *gap*.

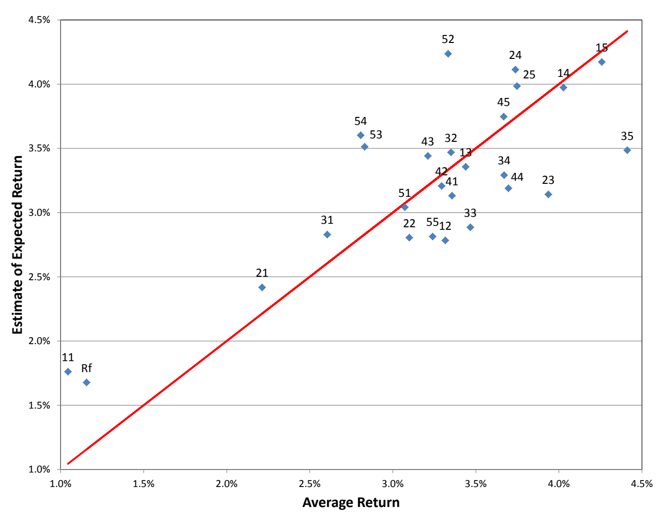

Figure 11: Realized vs. expected returns by CCAPM conditioned on Δf.

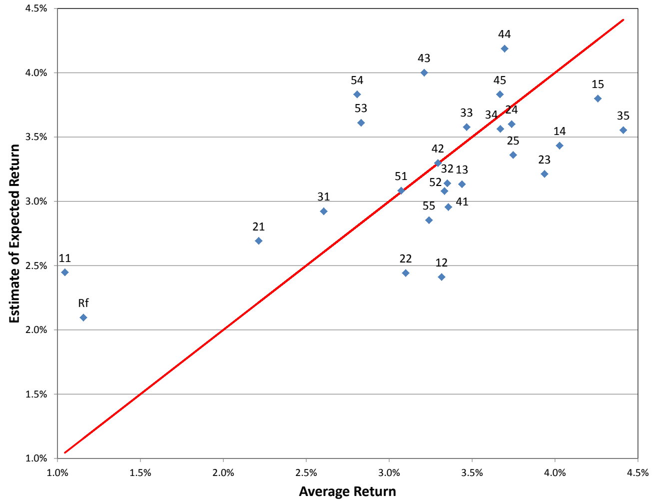

Figure 12: Realized vs. expected returns by CCAPM conditioned on R_f.

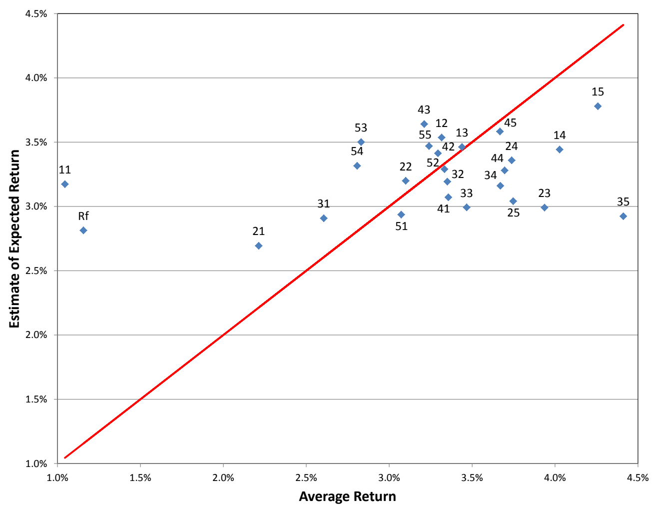

Figure 13: Realized vs. expected returns by CCAPM conditioned on $d - p$.

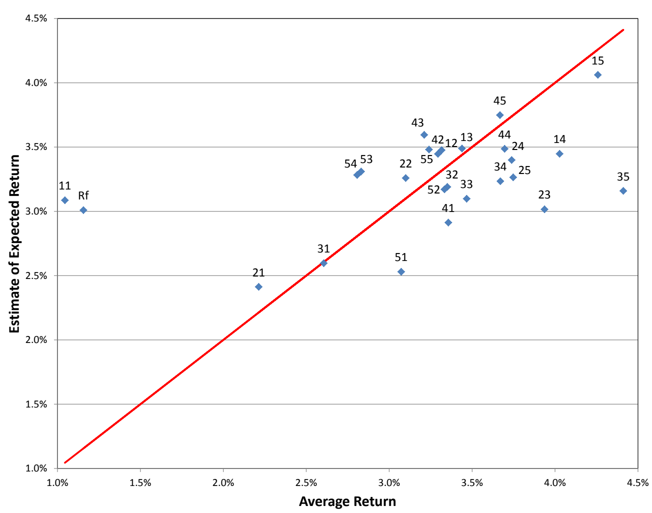

Figure 14: Realized vs. expected returns by CCAPM conditioned on cay.

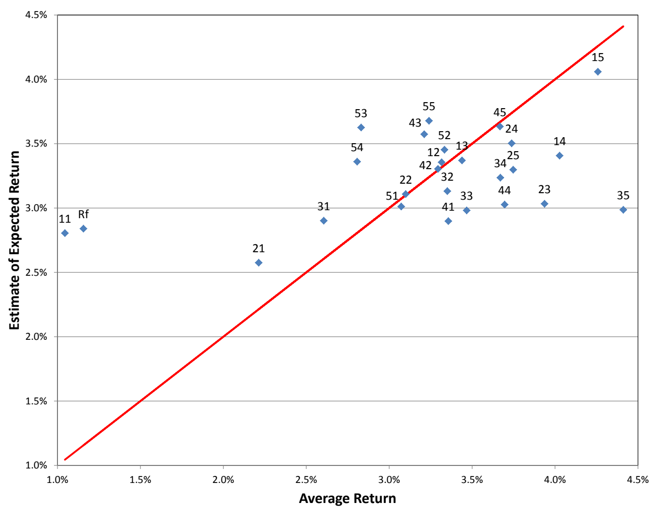

Figure 15: Realized vs. expected returns by CCAPM conditioned on *gap*.