What Predicts Stock Returns? – The Role of Expected versus Unexpected Predictors

By

Nina Baranchuk

Yexiao Xu*

School of Management
The University of Texas at Dallas

This version: November 2007

Abstract

Stock returns could be predictable due to time-varying expected returns. In this paper, we model the conditional expected return as an unobservable state variable affected directly by the expected components of predictors. This approach separates the different roles played by expected versus unexpected predictors, while avoiding both the “error-in-variables” problem due to imperfect predictors and bias due to persistent predictors in a typical predictive regression. In addition, the feature of negative correlation between an innovation in the expected return and the unexpected stock return is built into the model from imposing the return identity. At the same time, the unexpected components in predictors are allowed to co-move with the unexpected cash flow component of the stock return. This flexible correlation structure improves the efficiency in estimating the conditional expected returns. More importantly, we are able to study the role of predictors in forming the expected return as well as predicting the cash flow component of returns.

*The address of the corresponding author is: Yexiao Xu, School Of Management, The University of Texas at Dallas, PO Box 688, Richardson, Texas 75080, USA; Email: yexiaoxu@utdallas.edu
What Predicts Stock Returns? – The Role of Expected versus Unexpected Predictors

Abstract

Stock returns could be predictable due to time-varying expected returns. In this paper, we model the conditional expected return as an unobservable state variable affected directly by the expected components of predictors. This approach separates the different roles played by expected versus unexpected predictors, while avoiding both the “error-in-variables” problem due to imperfect predictors and bias due to persistent predictors in a typical predictive regression. In addition, the feature of negative correlation between an innovation in the expected return and the unexpected stock return is built into the model from imposing the return identity. At the same time, the unexpected components in predictors are allowed to co-move with the unexpected cash flow component of the stock return. This flexible correlation structure improves the efficiency in estimating the conditional expected returns. More importantly, we are able to study the role of predictors in forming the expected return as well as predicting the cash flow component of returns.

Key Words: Cash Flow, Expected Return, Kalman Filter, MCMC, Repurchasing Yield.
Introduction

The traditional view of efficient market hypothesis suggests that security prices follow a random walk process with a constant drift, which implies that returns are unpredictable. Recent empirical evidence points to the direction of finding otherwise. The empirical evidence, however, does not necessarily challenge the notion of efficient market hypothesis under the modern asset pricing theory that allows for changing in the investment opportunities (see Merton (1987)). If investment opportunities change over time, especially with business cycle, a long-term investor will deviate from the classical CAPM model in order to hedge such a risk. As a result, the required return on an asset will be different from that predicted by the CAPM. For example, ten years ago, Apple was building personal computers only. Today, a large portion of Apple’s business involves selling iPods. Therefore, there is a significant shift in Apple’s investment opportunities. The uncertainty in the investment opportunities is a risk to investors that should be compensated. The scope of such investment opportunity changes is largely associated with the general economic conditions or the business cycle. If business cycle can be predicted, expected returns should also be predictable due to correlation between business cycle and investment opportunities.

In this sense, predicting returns is equivalent to predicting the expected returns. Conrad and Kaul (1988) have taken this notion directly to the data by modeling the expected return as an unobservable component that evolves according to an autoregressive process. To a large extent, Conrad and Kaul’s approach is a univariate approach that only uses information in return itself. Comparing the autocorrelation pattern of futures returns and the returns on the underlying spot index portfolios, Boudoukh, Richardson, Whitelaw (1994) have concluded that time-varying expected returns can not be an explanation since autocorrelation only exists in the spot index returns. From
a broader perspective, this evidence does not necessarily act against time-varying expected return if there are non-persistent factors that can predict returns. In fact, the widely used predictive regressions uses a multivariate approach that includes the stock return, the dividend yield, and the relative interest rate to predict stock returns. As pointed out by Campbell (1991), a multivariate VAR approach is superior to a univariate approach especially over a long horizon.

There are two issues, however, with predictive multivariate regressions. Since most predictors are persistent and share the price component with the return, return innovation and the innovations to predictors tend to be negatively correlated. The well-known finite sample bias in the estimates of persistent parameters will affect the predictability estimates through the negative correlation. Therefore, estimates have to be adjusted. (see Nelson and Kim, 1993, Stambaugh, 1999, and Lewellen, 2004). Second, according to the rational view, the predictability of predictors comes from predicting the time-varying expected returns. Thus, it implicitly assumes that the expected return is a linear combination of predictors. In practice, however, predictors are noisy estimators of the conditional expected return. In other words, there is an “error-in-variables” issue with predictive regressions if the predictors are predicting the expected returns. As a consequence, the estimates from the predictive regression tend to understate the true predictability the returns.

Like Conrad and Kaul (1988), Pastor and Stambaugh (2006) have proposed a solution to the problem by modeling the expected return as an unobservable component that evolves according to an autoregressive process. At the same time, they model predictors as following an $AR(1)$ process. These predictors are then allowed to affect the expected returns indirectly through correlation between shocks to predictors and shocks to expected returns. Pastor and Stambaugh’s (2006) results suggest that returns are very much predictable and the expected returns depend on all past innovations to
predictors. This approach provides a better link from predictors to future returns.

In this paper, we argue that predictors should be viewed as consisting of both the expected and the unexpected components. Of the two components, we conjecture that the expected component is in fact more likely to be a good estimate for the expected return. The unexpected component may also affect the expected return if, for example, shocks to predictors are related to mispricing. Pastor and Stambaugh’s (2006) approach, however, does not allow the expected component to affect the expected return and thus cannot be used to test our conjecture. Instead, we use Kalman filter to decompose predictors into expected and unexpected components, and allow both components to affect the expected return. Using this approach, it is possible to study the relative importance of each channel for expected returns.

When there is a shock to the expected return, future discount rates tend to increase, which reduces the current stock price for the same cash flow stream. In other words, innovations in expected returns and the unexpected stock returns should be negatively correlated. Imposing this restriction in a statistical model can help improve efficiency in estimating the expected returns. From the return identity, Campbell (1991) decomposes the unexpected return into cash flow news and discount news. Since the expected predictors affect the expected return directly, the innovations to the expected predictors should also be negatively correlated with return innovation. Based on Campbell’s decomposition, we incorporate these negative correlations in our model. The decomposition also helps with interpretation of the correlation estimates.

In addition to the differences in the modeling philosophy, Pastor and Stambaugh (2006) study the importance of the prior information on the negative correlation between innovations in the expected returns and the unexpected return in estimating the expected return using a Baysian approach. Also concurrent is the paper by Rytchkov
(2007) who focuses the role of dividend growth instead of dividend yield in predicting stock returns. He finds improved predictability of expected return by directly allowing innovations in both the expected return and the expected dividend growth to enter the return equation.

We make two contributions to the literature in this paper. First, motivated by Boudoukh, Michaely, Richardson, and Roberts (2006) who document the diminishing role of dividend yield and the increasing importance of share repurchases, we use both dividend yield and the net repurchasing ratio as the predictors in this study. Second, we study how multiple predictors affect returns through both the direct channel that links the expected predictors to the expected returns and the indirect channel from the correlation between the unexpected predictors and the unexpected returns. The rest of the paper is organized as follows. A detailed discussion about our model, estimation strategy, and data construction is given in the next section. Section 2 provides empirical results on different model specifications and comparison with benchmark model. Section 3 concludes.
1 Methodology and Data

1.1 The Model

Denote \( r_t \) and \( \mu_t = E_t(r_{t+1}) \) as the stock return and the expected stock return, respectively, and \( x_t \) and \( z_t = E_t(x_{t+1}) \) as a \( k \)-dimensional vector of predictors and the expected predictors, respectively. We model a stock return in the following way,

\[
\begin{align*}
    r_{t+1} &= \mu_t + \epsilon^r_{t+1}, \\
    x_{t+1} &= z_t + \epsilon^x_{t+1}, \\
    \mu_{t+1} &= \alpha + \beta \mu_t + \gamma' z_t + \eta^\mu_{t+1}, \\
    z_{t+1} &= \psi + \theta z_t + \eta^z_{t+1},
\end{align*}
\]

where \( \theta \) is a \( k \times k \) diagonal matrix. In order to completely separate the role of the expected predictor from the shocks to predictors, we assume that the predictor shock \( \epsilon^x_{t+1} \) and the innovation to the expected predictor \( \eta^x_{t+1} \) are uncorrelated, that is, \( \text{Cov}(\epsilon^x_{t+1}, \eta^x_{t+1}) = 0 \). In addition, it is also reasonable to assume that innovation to the expected return \( \eta^\mu_{t+1} \) and innovation to expected predictor \( \eta^z_{t+1} \) are orthogonal, that is \( \text{Cov}(\eta^\mu_{t+1}, \eta^z_{t+1}) = 0 \).

Our model specification (1)–(4) extends Pastor and Stambaugh (2006) by (1) decomposing the predictors into the expected and unexpected components; and (2) allowing the expected predictors directly affect the expected returns through the term \( \gamma z \). In order to see the fundamental difference, we can plug equation (4) into equation (2) as,

\[
x_{t+1} = \theta x_t + (\epsilon^x_{t+1} + \eta^z_{t+1} - \theta \epsilon^x_t).
\]

If we treat the sum of the three random variables in parentheses as an independent residual term, the above equation is in the spirit of Pastor and Stambaugh’s (2006)
specification, where the impact of predictors \( x \) is through correlation between return shock and shocks to predictors. Under our model structure, this correlation can be expressed as,

\[
\text{Cov} [\eta^\mu_{t+1}, (\epsilon^x_{t+1} + \eta^z_t - \theta \epsilon_t^x)] = \text{Cov} (\eta^\mu_{t+1}, \epsilon^x_{t+1}),
\]

which only captures our second channel of how predictors affect the expected return.

It is well-known that estimation of the equation system (1) - (4) runs into identification issues if the error covariance matrix is unrestricted. In our model, we have the following structure on the covariance matrix:

\[
\begin{pmatrix}
\sigma_r^2 & \sigma_{rx} & \sigma_{r\mu} & \sigma_{r\mu} \\
\sigma_{rx} & \sigma_r & \sigma_{x\mu} & 0 \\
\sigma_{r\mu} & \sigma_{x\mu} & \sigma_\mu^2 & 0 \\
\sigma_{r\mu} & 0 & 0 & \sigma_z^2
\end{pmatrix}
\]

(5)

In the above variance covariance matrix, the only unidentifiable parameter is the covariance \( \sigma_{r\mu} \) between return \( r \) and the expected return \( \mu \). While imposing restrictions on \( \sigma_{r\mu} \) does not affect the coefficients \( \beta, \gamma, \) and \( \theta \) in equations (3) and (4), it may affect the rest of the covariance parameters in matrix (5). Thus, in order to have more confidence in our estimates of the matrix (5), we impose only a minimal restriction on \( \sigma_{r\mu} \) and otherwise let this parameter take on any value (see the description of our priors in 1.3). Specifically, we require \( \sigma_{r\mu} < 0 \), reflecting the intuition that stock price innovations are mainly driven by discount rate news rather than cash flow news.

Theoretically, some of the free parameters in the above covariance matrix can be linked to other model parameters. For example, based on the Campbell’s (1991) return decomposition under the no bubble condition, we can link the covariance parameters \( \sigma_{r\mu} \) and \( \sigma_{rz} \) to other parameters. In particular, we can impose a restriction on the error
terms in equations (1), (3), and (4) through the following equation,

\[ r_{t+1} = \mu_t + (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \]  

(6)

Any relation thus derived, however, requires a strong assumption that model (1) - (4) is the true model for the data. In order to be cautious, we use the decomposition (6) only to develop some intuition about the model parameters. Specifically, applying the law of iterative expectation, we have,

\[
\begin{align*}
(E_{t+1} - E_t)r_{t+2} &= \mu_{t+1} - E_t \mu_{t+1} = \eta_{t+1}^\mu \\
(E_{t+1} - E_t)r_{t+3} &= E_{t+1} \mu_{t+2} - E_t \mu_{t+2} = \beta \eta_{t+1}^\mu + \gamma' \eta_{t+1}^z \\
(E_{t+1} - E_t)r_{t+4} &= E_{t+1} \mu_{t+3} - E_t \mu_{t+3} = \beta^2 \eta_{t+1}^\mu + \beta \theta \gamma' \eta_{t+1}^z \\
&\quad \ldots \ldots \\
(E_{t+1} - E_t)r_{t+1+j} &= E_{t+1} \mu_{t+j} - E_t \mu_{t+j} = \beta^{j-1} \eta_{t+1}^\mu + \sum_{l=0}^{j-1} (\beta^{j-1-l} \theta^l) \gamma' \eta_{t+1}^z \\
&= \beta^{j-1} \eta_{t+1}^\mu + (\beta^{j-1} I_k - \theta^{j-1}) (\beta I_k - \theta)^{-1} \gamma' \eta_{t+1}^z,
\end{align*}
\]

where \( I_k \) is a \( k \times k \) identity matrix. Thus, the last term in (6) can be rewritten as

\[
(E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} = \frac{\rho}{1-\rho} \eta_{t+1}^\mu + \frac{\rho^2}{(1-\rho \beta)} (I_k - \rho \theta)^{-1} \gamma' \eta_{t+1}^z
\]

(7)

Using equations (6) and (7), we can rewrite equation (1) as,

\[ r_{t+1} = \mu_t + \epsilon_{t+1}^C - \phi_\mu \eta_{t+1}^\mu - \phi_\eta \eta_{t+1}^z, \]

(8)

where \( \epsilon_{t+1}^C \) represents the cash flow news.

Relationship (8) implies that \( \text{Cov}(r_{t+1}, z_{t+1}) = \text{Cov}(\epsilon_{t+1}^C, z_{t+1}) - \phi_\eta \text{Var}(\eta_{t+1}^z) \). Hence, we expect the covariance term \( \sigma_{r_z} \) to be negative, unless shocks to the expected predictors \( z \) are mostly driven by cash flow news. In other words, a variable can still be
a candidate of predictor even when it has a positive correlation with return shocks. If we additionally assume that $\text{Cov}(\epsilon_{c,t+1}^r, \eta_{\mu,t+1}^\mu) = \text{Cov}(\epsilon_{c,t+1}^c, \eta_{\mu,t+1}^\mu) = 0$, we can identify $\sigma_{r\mu} = \phi_\mu \sigma_\mu$ and $\sigma_{r\zeta} = \phi_\zeta \sigma_\zeta$.

Equation (8) can also be used to understand what determines the correlation between return shocks and predictor shocks.

\[
\text{Cov}(\epsilon_{r,t+1}^\mu, \epsilon_{c,t+1}^c) = \text{Cov}(\epsilon_{c,t+1}^c, \epsilon_{c,t+1}^c) - \phi_\mu \text{Cov}(\eta_{c,t+1}^\mu, \epsilon_{c,t+1}^c). \tag{9}
\]

Thus, the noted covariance (see Stambaugh, 1999) comes from correlation either with the cash flow news in returns or the correlation with innovation to expected return.

In addition to improved data fitting, our specification also provides a new insight to the predictability mechanism. First, as discussed above, our estimates of the covariance terms $\sigma_{x\mu}$ and $\sigma_{r\zeta}$ indicate whether the expected and the unexpected predictors contain information about cash flow news. Second, we can analyze to what extent the autocorrelation of the expected returns is due to persistence in return predictors (Boudoukh, Richardson, and Whitelaw (1994) argue that return autocorrelation is mainly due to market frictions). This can be done by comparing the estimates of the expected return persistence $\beta$ obtained with and without including $\gamma_\zeta$ into the equation (3) for the expected return $\mu$. Third, our estimates of the parameters $\gamma$ and $\sigma_{x\mu}$ inform us about the ability of our predictors to explain the evolution of the expected returns. Note that the last question can be addressed only to a limited extent because the expected returns are not observable and cannot be fully recovered using statistical analysis unless the true model of the data is known with certainty.
1.2 Comparison to a Linear Predictive Regression

Our model relaxes the assumption of a predictive regression that the expected return precisely equals a linear combination of the contemporaneously realized predictors: $\mu_t = b'x_t$. This assumption seems difficult to justify for the following two reasons. First, it seems more reasonable to let the current expected return $\mu_t$ be directly based only on the past (and not simultaneously generated) information $x_{t-1}$. Second, the return expectations formed by the market are likely to include variables that are not observed by a statistician, implying that a statistical model should let expected return differ from a linear combination of the included predictors.

To obtain the conditions on the model parameters that reduce our model to the predictive regression $\mu_t = b'x_t$, we rewrite our model as follows. From (4), we obtain

$$z_{t+1} = (I_k - \theta L)^{-1}(\psi + \eta_{t+1}^z), \quad (10)$$

where $L$ is the lag operator. Substituting equation (10) into equation (2), we have,

$$(I_k - \theta L)x_{t+1} = \psi + \eta_{t+1}^z + (I_k - \theta L)\epsilon_{t+1}^x. \quad (11)$$

Solving for $z_t$ from equation (2) and substituting it into equation (3), we obtain

$$(1 - \beta L)\mu_{t+1} = \alpha + \gamma'x_{t+1} + \eta_{t+1}^\mu - \gamma'\epsilon_{t+1}^x. \quad (12)$$

Comparing (11) and (12), we obtain the following conditions that guarantee $\mu_t = b'x_t$:

$$\eta_{t+1}^\mu = (1 - \beta L)b'\epsilon_{t+1}^x + [(b - \gamma) - b\beta L]'(I_k - \theta L)^{-1}\eta_{t+1}^z, \quad (13)$$
$$\alpha = [(b - \gamma) - b\beta]'(I_k - \theta)^{-1}\psi. \quad (14)$$

In equation (13), it is reasonable to restrict the last term (which is $[(b - \gamma) - b\beta L]'(I_k - \theta)^{-1}\eta_{t+1}^z$)
\( \theta L^{-1} \) to be a constant. ¹ This implies the following conditions,

\[
\eta_{t+1}^\mu = (1 - \beta L)b'\epsilon_{t+1} + (b - \gamma)\theta \eta_{t+1}^\nu, \tag{15}
\]

\[
b'\beta = (b - \gamma)\theta, \tag{16}
\]

\[
\alpha = (b - \gamma)\psi. \tag{17}
\]

Condition (15) suggests that \( \eta_{t+1}^\mu \) and \( \epsilon_{t+1}^\nu \) should be correlated with each other if the predictive regression is significant. This is the motivation for allowing the unexpected predictor to affect the expected return. When \( \gamma = 0 \), the estimated persistence \( \beta \) of the expected returns should equal the persistence \( \theta \) of the predictors. We know this equality does not hold because predictors are typically very persistent and the expected returns are not. By allowing for \( \gamma \neq 0 \), our model separates persistence in predictors from that in returns.

### 1.3 Estimation Methodology

Our model is given by equations (1) - (4) with covariance structure (5). Using \( k \) predictors, we are interested in estimating the following parameters: \( k + 1 \) long term means \((\bar{\mu}, \bar{z})\), \( k + 1 \) persistence parameters \((\beta, \theta)\) restricted to the interval \((-1, 1)\) to ensure stability, \( k \) slope parameters \( \gamma \), and \( 4k + 2 \) free elements in the covariance matrix 5. Thus, overall, we have \( 7k + 4 \) parameters to estimate. These parameters are difficult to estimate by frequentist methods. The likelihood function is not available directly and even with simulation methods finding the maximum likelihood estimates is a challenge. For this reason we turn to Bayesian methods.

The Bayesian approach views parameters as random, and focuses on the posterior distribution of the parameters of interest. Analytically, the posterior distribution is

¹Otherwise, a persistent predictor (large \( \theta \)) would generate strong autocorrelation in the expected return innovations \( \eta_{t+1}^\mu \). This is almost impossible for security returns.
obtained using Bayes Theorem by combining distributions describing prior beliefs with the likelihood function of the data derived from a statistical model (standard reference texts of the Bayesian approach to statistical inference include Zellner (1971) and Bernardo and Smith (1994)). When the posterior does not fall into the class of the known distribution functions, it can be investigated using Monte Carlo Markov Chain (MCMC) procedures (the details of the MCMC procedures used in this paper are described in the Appendix). The goal of an MCMC procedure is to produce a sample of draws from the posterior distribution for each parameter of interest. The posterior distribution for each parameter is then typically summarized using the mean and the variance of the sample draws.

On all our parameters of interest, we impose normal priors with large variance, insuring that the priors have little effect on the posterior and our conclusions are mainly driven by the data. For the unidentifiable parameter $\sigma_{r\mu}$ that captures the covariance between the realized and the expected returns, we use a non-informative uniform prior, requiring only that this parameter is negative (as discussed in section 1.1). Imposing a uniform prior gives us confidence that our estimates for the remaining elements of the covariance matrix are not driven by our assumptions about the value of an unidentifiable parameter. We offer a detailed description of the estimation procedure in the Appendix.

In our analysis of robustness and search for the best model, we estimate a number of alternative specifications. Specifically, we estimate a model where the factors are assumed to follow and AR(1) process, and also extend our base model by incorporating stochastic volatility for shocks to realized returns. Bayesian approach is particularly useful for estimating a the stochastic volatility model because the model is non-linear and thus difficult to estimate using frequentist methods. Using Bayesian approach to estimate all of the alternative model specifications simplifies model comparison as we
can rule out the concern that differences in model performance are driven by differences in estimation techniques.

1.4 The Data

Most studies on predictability focus on the predictability of dividend yield (or dividend-price ratio) because of the theoretical argument suggesting that this variable should predict returns. A recent study by Boudoukh, Michaely, Richardson, and Roberts (2006) has shown that, with the steady decline in the dividend yield in the past 20 years, repurchasing has played an increasing role in distribution. This is partly due to the implementation of the SEC rule 10b-18 that provides a legal safe harbor for firms repurchasing their shares in 1982. In fact, the total payout yield, which includes both dividend yield and repurchasing yield (common share repurchases to year-end market capitalization) has larger predictability than the pure dividend yield. In this study, we consider both dividend yield and repurchasing yield. While Boudoukh, Michaely, Richardson, and Roberts (2006) construct the repurchasing yield from accounting statement, we are able to back out the repurchasing yield from the following decomposition,

\[
R_{t+1} \equiv \frac{S_{t+1}D_{t+1}}{S_tP_t} + \frac{(S_t - S_{t+1})(P_{t+1} + D_{t+1})}{S_tP_t} + \frac{S_{t+1}P_{t+1}}{S_tP_t}.
\]

(18)

This decomposition is from a representative investor’s perspective since the number of shares changes over time due to either repurchasing or seasonal offering. In other words, the first term represents the dividend yield for a representative shareholder; the second term reflects the net repurchasing yield at the before-dividend price; and the last term is the change in market capitalization, which reflects growth.

Using monthly CRSP composite value-weighted index returns from 1926 to 2005 with \((ret)\) and without \((retx)\) dividends we construct the monthly dividends. In ad-
dition, we can use the total market capitalization to construct the total repurchasing. Similar to Campbell (1991), we use twelve month smoothed dividends and repurchasing to compute the dividend yield and repurchasing yield. If predictability reflects time-varying expected returns, it should be reflected in the low frequency data. In addition, market micro-structure issues existed in high frequency data could easily create predictability. Therefore, we decided to use quarter frequency. Correspondingly, we compound three month returns and compute three month average for dividend yield and repurchasing yield.\(^2\)

\section{The Empirical Evidence}

\subsection{Main Results}

In order to understand the importance of expected versus unexpected predictors, we estimate and compare the following three special cases. In the main model \textbf{Model 1}, we estimate all the identifiable model parameters specified above: \( \alpha, \phi, \beta, \gamma, \theta, \sigma_r, \sigma_x, \sigma_\mu, \sigma_z, \sigma_{rx}, \sigma_{x\mu}, \) and \( \sigma_{rz}. \) In the second case (\textbf{Model 2}), we study the role of unexpected predictors by turning off the expected predictors and thus setting \( \gamma = 0. \) In the last case (\textbf{Model 3}), we focus on the expected predictors by shutting down the correlation between the unexpected predictors and the expected return shock: \( \sigma_{x\mu} = 0. \) To simplify parameter interpretation, instead of reporting \( \alpha \) and \( \phi \) we report the corresponding long term means \( \bar{\mu} \) and \( \bar{z} \) defined as follows: \( \alpha \equiv (1 - \beta)\bar{\mu} + \gamma\bar{z} \) and \( \psi \equiv (1 - \theta)\bar{z} \).

As point estimates for our parameters of interest, we report sample means of the posterior draws. We refer to the estimates as significant at 95\% level if at least 97.5\% of the draws have the same sign, and the interval containing 95\% of the draws as

\(^2\)In all regressions, the return data is multiplied by 100 and the dividend and repurchases data is multiplied by 1000.
the confidence interval. Table 1 reports the estimates for the slope coefficients, and tables 2 through 4 report the estimates for the covariance matrices for each of the three models respectively. The last two rows of Table 1 report statistics on the ability of the model to fit the data. The top of the two rows reports the values of the coefficient of determination $R^2$ calculated as follows. For each draw of the parameters, we calculate $R_i^2$ measured as the correlation between the predicted and the realized returns. The predicted return is obtained using the $i^{th}$ draw of the parameters (predicted returns are obtained using a Kalman Filter procedure as described in the Appendix). Note that, unlike Pastor and Stambaugh, we do not impose any prior on the distribution of $R^2$.³ The reported value is the average of the values calculated for all the retained parameter draws. The second reported measure of model performance, prediction likelihood, is discussed in more detail in section 2.4

Insert Table 1 Approximately Here

Insert Table 2 Approximately Here

From table 1, coefficient $\gamma$ generated by both Models 1 and 3 is positive for both repurchases ($\gamma_1$) and dividends ($\gamma_2$). However, $\gamma_1$ is insignificant for the dividend yield in the complete model. The covariance $\sigma_{x\mu}$ between the expected returns and dividends or repurchases (see Table 2), on the other hand, is significant for both dividends and repurchases, with much stronger effects from the dividends. Thus, the predictability mechanism seems to be very different for dividend yield versus repurchasing yield. The former works largely through the unexpected channel while the latter through the expected channel. The magnitude of the coefficient $\gamma_2$ suggests that an expected increase

³Technically, $R^2$ is not a free parameter. Thus, imposing a prior on $R^2$ is equivalent to altering the priors on the parameters involved in the calculation of $R^2$. However, since $R^2$ also depends on the observed data, imposing a prior on $R^2$ is equivalent to imposing a prior on model parameters that depends on the observed data. This to conflicts with the notion of prior as a belief held before the data was observed.
in repurchases by one unit (approximately 15% of the long term mean) increases the expected quarterly returns by slightly more than 0.1%.

Insert Table 3 Approximately Here

Insert Table 4 Approximately Here

When shutting down the unexpected channel by imposing the restriction \( \sigma_{xp} = 0 \) as in Model 3, the expected dividend yield becomes significant in affecting the expected return, which make perfect sense. In fact, shutting down the correlation between the unexpected dividend yield and the expected return – the main channel for dividend – increases the volatility for the expected dividends, which increases the role of expected dividends (see Table 4). In contrast, there is not much impact on the repurchasing yield coefficients since correlation is not the main channel here. This restriction on correlation, however, reduces the model’s \( R^2 \) by approximately 1%.

Comparing the results for Model 1 and Model 2, we can infer that restricting \( \gamma \) to zero has an even larger impact on the \( R^2 \), reducing it by more than 2.7%. This suggests that the large part of the predictability comes from the expected repurchasing yield. Surprisingly, the covariance between the unexpected repurchasing yield and the expected return has also been reduced. (see Table 3).

Eliminating either the expected or the unexpected channel will affect the estimated persistence \( \beta \) of the expected returns. While the estimate of \( \beta \) in Model 2 is slightly larger than that in Model 3, these estimates are 12% to 15% larger than the persistence parameter in the complete model. This implies that our predictors (dividends and repurchases) account for a reasonable amount of the autocorrelation in the expected returns.
It is also interesting to see that the estimated long term mean of quarterly stock returns tends to be low from Model 2 (2.7%), and to be high for Model 3 (4.3%). The complete model (Model 1) produces a more reasonable estimate of 3.8%. This provides further evidence for the importance of allowing both the expected and the unexpected predictors. Finally, since the correlation between the unexpected repurchasing yield and the return shock is almost zero, it indicates that repurchasing co-move positively with the cash flow news. (see equation (9))

Focusing on equation (3), we can use the estimates of Models 1 - 3 to evaluate the explanatory power of the included variables (last period expected returns, expected dividends, and expected repurchases). We calculate this explanatory power as follows. First, we generate the expected returns using a Kalman filter Smoothing procedure (De Jong and Shepard, 1995) and find their total variance. Next, we estimate the explanatory power as \( \frac{(\text{total variance} - \text{error variance})}{\text{total variance}} \), taking the estimate of the error variance from Tables 2 - 4. We find that all models produce similar results: Model 2 produces the explanatory power of \( \frac{(39.914 - 16.4263)}{39.914} \approx 58.8\% \), Model 1 produces \( \frac{(34.961 - 19.728)}{34.961} \approx 43.6\% \), and Model 3 produces \( \frac{(44.069 - 19.728)}{44.069} \approx 55.2\% \). All of our explanatory power estimates are at the high end of those found in Pastor and Stambaugh,\(^4\) which range from 3% to 63%. If we take into account the significant covariance estimates produced by Model 1 for the unexpected dividends and repurchases and the expected returns, similar to Pastor and Stambaugh we find that the explanatory power increases to \( \frac{(34.961 - 14.634)}{34.961} \approx 58.1\% \). This explanatory power calculation, however, uses information that is not available before time \( t \) and thus cannot be used to predict the expected returns, which may be viewed as conflicting with the notion of explanatory power. Note also that, similar to

\(^4\)Our methodology differs from that of Pastor and Stambaugh. Their model specification does not include expected predictors as explanatory variables for the expected returns. Thus, they first use their model to infer the expected returns, and then regress the inferred expected returns on the observed and unexpected predictors, omitting the last period expected returns.
estimates of covariance matrices, the estimates of the expected returns $\mu_t$ are directly affected by the assumed values of the unidentifiable parameter $\sigma_{r\mu}$, suggesting some caution in interpreting this set of results.

### 2.2 Stochastic Volatility

There are overwhelming evidence suggesting time varying volatilities for stock returns. Allowing persistent conditional volatilities not only improves the efficiency of estimates, but also can incorporate kurtosis in return distribution. Therefore, we augment the base model by letting return volatility change over time. Specifically, return volatility is assumed to follow an $AR(1)$ process, resulting in the following model specification (referred to as STV):

\[
\begin{align*}
    r_{t+1} &= \mu_t + \epsilon_{t+1}^{h_{t+1}/2} e_{t+1}^r, \\
    x_{t+1} &= z_t + \epsilon_{t+1}^x, \\
    \mu_{t+1} &= \alpha + \beta \mu_t + \gamma' z_t + \eta_{t+1}^\mu, \\
    z_{t+1} &= \psi + \theta z_t + \eta_{t+1}^z, \\
    h_{t+1} &= \lambda_0 + \lambda_1 h_t + \nu_{t+1}. 
\end{align*}
\]

(19)

We retain the assumptions on the error structure made for our base model and also assume that the shocks to return volatility are independent of the observed data. Thus, we have the following variance-covariance matrix:

\[
Var \begin{bmatrix}
\epsilon_{t+1}^r \\
\epsilon_{t+1}^x \\
\eta_{t+1}^\mu \\
\eta_{t+1}^z \\
\nu_{t+1}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & -\phi_\mu \sigma_{\eta}^\mu & -\phi_z \sigma_{\eta}^z & 0 \\
0 & \sigma_\epsilon & \sigma_\epsilon \sigma_{\eta}^\mu & 0 & 0 \\
-\phi_\mu \sigma_{\eta}^\mu & \sigma_{\epsilon}^\mu & \sigma_{\eta}^\mu & 0 & 0 \\
-\phi_z \sigma_{\eta}^z & 0 & 0 & \sigma_{\eta}^z & 0 \\
0 & 0 & 0 & 0 & \sigma_{\nu}
\end{bmatrix}.
\]

(20)
The estimation results for the STV model are reported in Tables 5 and 6 (the details of the estimation procedure are given in the Appendix). The results show that adding stochastic volatility improves the model fit (the model $R^2$ increases from 3.5% for Model 1 to 4.0% for the STV model). The mean estimate of the expected return has decreased from 3.8% per quarter to 3.65% per quarter. The expected repurchasing yield plays an even more significant role in affecting the expected return. At the same time the persistent parameter estimate $\beta$ has gone down by 20%. This result suggests that less persistent expected return does not necessarily mean low predictability. Similarly dividend yield works through the indirect channel with much stronger effect. (see Table 5) Other estimated coefficients are robustness when comparing the implications produced by Model 1.

Insert Table 5 Approximately Here

As before, focusing on equation (3), we can use the estimates of Model STV to evaluate the explanatory power of the included variables (last period expected returns, expected dividends, and expected repurchases). This model generates expected returns with total variance of 18.04. Using the estimate of the error variance from Table 6, we find that the explanatory power is approximately 30.3%, which is noticeably lower than that produces by Models 1 – 3. If we take into account the significant covariance estimate for the unexpected dividends and the expected returns, the explanatory power increases to 82%.

Insert Table 6 Approximately Here
2.3 Benchmark Model

As the Benchmark Model, we estimate the main model specification used in Pastor and Stambaugh. Their model can be summarized as follows:

\[
\begin{align*}
    r_{t+1} &= \mu_t + \epsilon^r_{t+1}, \\
    x_{t+1} &= \psi + Ax_t + \epsilon^x_{t+1}, \\
    \mu_{t+1} &= \alpha + \beta \mu_t + \eta^\mu_{t+1},
\end{align*}
\]

(21)

with the following covariance matrix:

\[
\operatorname{Var} \left( \begin{bmatrix} \epsilon^r_{t+1} \\ \epsilon^x_{t+1} \\ \eta^\mu_{t+1} \end{bmatrix} \right) = \begin{bmatrix} \sigma^r_\epsilon & 0 & -\phi_\mu \sigma^\mu_\eta \\ 0 & \sigma^x_\epsilon & \sigma^{x\mu}_{\epsilon\eta} \\ -\phi_\mu \sigma^\mu_\eta & \sigma^{x\mu}_{\epsilon\eta} & \sigma^\mu_\eta \end{bmatrix}. \quad (22)
\]

The main difference between the above model and our main model (1) – (4) is the specification of the time series process for the predictors \(x\). The above model assumes that \(x\) follows an AR(1) process and only the unexpected component has an effect on the expected returns. Our main model, on the other hand, separates \(x\) into the predicted and the unexpected components and allows both to affect the expected returns.

The estimation results for the Benchmark model are reported in Tables 7 and 8 (the details of the estimation procedure are in the Appendix). In this table, we use notation \(\bar{x}\) to denote the long-term values for dividends and repurchases (similar to the definition of \(\bar{z}\) used for Models 1 – 3).

Insert Table 7 Approximately Here

Insert Table 8 Approximately Here
The Benchmark model produces the lowest $R^2$ among all of our alternative model specifications. While the persistence coefficients $\beta$ and $\theta$ are similar to those obtained from Model 2, the estimates of the long-term values for dividends and expected returns are lower.\footnote{Strictly speaking, the long-term values for dividends are not comparable between the two model specifications because in the benchmark model it represents the long term value of the dividends themselves, while in the main model it represents the long-term value of the expected dividends. It seems reasonable to expect, however, that the two should be similar.} Unlike the model specifications analyzed in the earlier sections, the benchmark model produces significantly positive correlation estimates between the expected returns and the unexpected components for both dividends and repurchases.

As before, we can use the estimates to evaluate the explanatory power of the included variables for the expected returns. This model generates expected returns with a fairly large total variance of 104.9. Using the estimate of the error variance from Table 8, we find that the explanatory power of the expected predictors is approximately 64.8%, which is slightly larger than that generated by Models 1 – 3.

### 2.4 Model Comparison

We use the parameter estimates for Models 1 - 3, STV, and Benchmark to generate predicted values for returns using the Kalman filter procedure described in the Appendix. Using each draw of the parameters generated by the Markov Chain, we calculate $\hat{r}_{t+1} = E(r_{t+1} | r_t)$ for $t = 1, .., T$ (see the Appendix for details on our Markov Chain estimation procedure). We access the quality of each draw of the predicted return $\hat{r} = (\hat{r}_1, .., \hat{r}_T)$ by calculating $R^2$; the mean value of $R^2$ for each of the three models is reported in Table 1. The largest average value of 0.040 is generated by the stochastic volatility version of our base model. Additionally, Models 2 and 3 as well as the Benchmark generated significantly lower values of $R^2$ than the base model, Model 1.
We also compare the five models in terms of out of sample prediction. To generate out of sample predictions, we use the following procedure. For $t = 177, \ldots, 216$, we use the first $t$ observations of $r_t$ and $x_t$ for parameter estimation. For each draw of the parameters, we evaluated the expected return $\hat{r}_{t+1} \equiv \mu_t$. Given the predicted return, we calculated a natural log of the likelihood of the observed $r_{t+1}$, given our model assumption that $r_{t+1} \sim N(\mu_t, \sigma_r)$. The average value of the log-likelihood across all parameter draws is reported in Tables 1, 5, and 7 for Models 1-3, STV, and Benchmark correspondingly. The ranking of the models is the same as suggested by comparing $R^2$: the highest likelihood value of -165.465 is achieved by the stochastic volatility version of our base model. The remaining models produce considerably lower likelihood values, with the smallest value produced by the Benchmark Model.

3 Concluding Comments

One of the challenges to the efficient market hypothesis has been return predictability. The traditional view on the evidence is either denying the evidence with the help of statistical methods or attributing the phenomenon to market frictions. The modern view takes a more positive approach by arguing that the risk premium is time-varying due to changing either in investment opportunities or in investors risk tolerance. If this is indeed the case, there are return predictors, for example the dividend yield. These predictors should predict return because price should summarize all future changes in the expected return or cash flow distribution. From empirical perspective, however, we tend to find weak evidence coupled with many estimation issues. For example, predictive regressions use imperfect predictors and thus suffer from "error-in-variables" problem. The downward bias in the persistence of predictors tends to overstate the role of predictors due to possible contemporaneous correlation between return shocks.
and predictors.

In this study, we propose a structured model for the time-varying expected returns that not only avoids the potential bias issues but also allows us to investigate the different roles played by the expected versus unexpected predictors. Understanding the difference is important in affecting the asset allocation decisions of investors since the expected predictors will directly alter the expected return over time while the unexpected predictors will affect shocks, thus the volatility, of the expected return.

In the empirical study, we simultaneously examine two predictors—dividend yield and repurchasing yield. Since a representative investor cares about the total distribution and there is an increasing trend in repurchasing by corporations after relaxing restrictions on repurchasing, both predictors might be important in predicting returns but with different roles. Our results suggest that dividend yield affects the expected return through the indirect channel while repurchasing yield acts on the expected return directly.
Appendix

Estimation

Models 1 – 3

Models 1 – 3 are all special cases described by equations (1) – (4). For notational convenience, we express these equations in matrix form,

\[ \begin{align*}
    y_{t+1} &= c_t + G\epsilon_{t+1}, \\
    c_{t+1} &= a + Bc_t + H\epsilon_{t+1},
\end{align*} \tag{23, 24} \]

where \( \epsilon_t \) are i.i.d. standard Gaussian shocks, \( G \) and \( H \) are defined by \( \text{Chol}(\Sigma) \equiv (G', H')' \) with \( \Sigma \) denoting the covariance matrix (5), and

\[ y_t = \begin{bmatrix} r_t \\ x_t \end{bmatrix}, \quad c_t = \begin{bmatrix} \mu_t \\ z_t \end{bmatrix}, \quad a = \begin{bmatrix} \alpha \\ \psi \end{bmatrix}, \quad B = \begin{bmatrix} \beta & \gamma \\ 0 & \theta \end{bmatrix}. \]

We also use the following notation: \( \Sigma = (\sigma_{ij})_{i,j=1...k+2} \) and \( B = (b_{ij})_{i,j=1...k+1} \).

We are interested in the posterior distributions of the following parameters: \( a, (b_{ij}), \) and \( (\sigma_{ij}) \). According to Bayes theorem, the posterior for each parameter is proportional to the product of the corresponding prior and the likelihood function. We impose the following priors on the parameters of interest:

\[ \begin{align*}
    a &\sim N(0, s^2_a) \\
    b_{ij} &\sim N(0, s^2_{b_{ij}}) \\
    \sigma_{ij} &\sim N(0, s^2_{\sigma_{ij}}), \ i \neq j \\
    \sigma_{ii} &\sim T N_{[0,\infty)}(1, s^2_{\sigma_{ii}}).
\end{align*} \]

In our estimations, we let the priors be fairly diffuse, and thus uninformative, by setting
$s_a^2 = s_{b_{ij}}^2 = s_{\sigma_{ij}} = 100$. As part of our robustness analysis, we also investigated the results obtained by increasing the prior variance to 1000 or 10000. We find that the increases in prior variance do not affect the estimates, suggesting that the estimates are determined mainly on the data and not driven by the priors.

The likelihood function for the data can be calculated using a Kalman filter procedure (25) described in the next section. This procedure in particular implies that $y_{t+1} | y_t \sim N(f_t, \Omega_t)$, where $f_t$ and $\Omega_t$ are functions of the model parameters obtained using a recursive procedure described in the next section. The lack of closed form expressions for $f_t$ and $\Omega_t$ implies that the shape of the likelihood function $\ell = \prod_{t=1}^T N(f_t, \Omega_t)$ is difficult to describe cannot be reduced to a standard distribution.

Investigating the posterior distributions (=likelihood×priors) for the parameters of interest is simplified if the parameter set is enlarged by including the latent parameters $c_t$. These latent parameters can be generated using the Kalman filter with a simulation smoother. We use the procedure developed in De Jong and Shepard (1995) and briefly summarized for our model in the following section (see equation (27)). Conditional on $c_t$ and $B$, model (23) - (24) becomes a standard multivariate Gaussian regression, simplifying the sampling of the covariance matrix $\Sigma$. If we did not have any restrictions on the shape of this matrix, we could use a conjugate prior of Inverse Wishart, which would result in an Inverse Wishart conditional posterior distribution for $\Sigma$. Unfortunately, the restrictions imposed in equation (5) imply that the posterior distribution for $\Sigma$ is of an unknown shape. Conditional on $c_t$ and $B$, however, it can be efficiently investigated using Metropolis-Hastings (MH) algorithm (a detailed description of the algorithm can be found in Chib and Greenberg (1995)). Conditioning on $c_t$ and $B$ additionally produces a normal conditional posterior distribution for $a$. We find, however, that sampling efficiency is dramatically improved if both $a$ and $B$ are sampled together in one block. Thus, we also rely on MH algorithm to sample the remaining
parameters of interest $\mathbf{a}$ and $\mathbf{B}$. To summarize, the estimation procedure consists of the following steps.

1. Sample (MH step; likelihood of the data (26) is calculated using Kalman Filter (25))

   $$\mathbf{a}, (b_{ij})|\mathbf{y}, \mathbf{c}, (\sigma_{ij})$$

2. Sample (using Simulation Smoother (27))

   $$\{\mathbf{c}_t\}|\mathbf{y}_t, \mathbf{a}, (b_{ij}), (\sigma_{ij})$$

3. Sample (MH step; the conditional likelihood of the data is Gaussian)

   $$(\sigma_{ij})|\mathbf{y}, \mathbf{c}, \mathbf{a}, (b_{ij})$$

4. Goto 1

After a period of convergence, the above procedure generates draws from the posterior distributions of the parameters of interest. In all our estimations, we discarded the first 15000 draws and retained the remaining 25000 draws which showed little persistence. All the reported statistics is based on these last 25000 draws for each parameter.

**Kalman Filter and Simulation Smoother**

We use Kalman Filter to evaluate the likelihood function for the data and calculate predicted values for future returns. We use simulation smoother of De Jong and Shephard (1995) to generate the unobserved expected returns and predictors $\mathbf{c}_t$. For our model (23) – (24), the procedures can be described as follows.
Kalman Filter: for \( t = 0, \ldots, T - 1 \), run

\[
e_{t+1} = y_{t+1} - c_t, \quad D_t = P_t + GG', \quad K_t = (BP_t + HG')D_t^{-1}
\]

\[
c_{t+1} = a + Bc_t + K_t e_{t+1}, \quad P_{t+1} = BP_t(B - K_t)' + H(H - K_tG)',
\]

where \( c_0 = a(I - B)^{-1} \) and \( P_0 = HH' \). The above Kalman Filter suggests that the likelihood of the data \( f(y_1, \ldots, y_T) \) can be calculated as follows:

\[
f(y_1, \ldots, y_T) = f(y_1) \prod_{t=1}^{T} f(y_t|y_1, \ldots, y_{t-1}) = \prod_{t=1}^{T} N(c_{t-1}, D_{t-1}).
\]

Simulation Smoother: first run the Kalman Filter (25) and save the quantities \( D_t, K_t, \) and \( e_t \). Next, for \( t = T - 1, \ldots, 0 \), run

\[
N_t = D_t^{-1} + K_t'U_tK_t, \quad C_t = \Gamma - \Gamma N_t \Gamma, \quad u_t = N(0, C_t), \quad V_t = \Gamma(N_t - K_t'U_tB),
\]

\[
r_{t-1} = D_t^{-1}e_{t+1} + L_t'r_t - V_{t-1}'C_{t-1}^{-1}u_t, \quad U_{t-1} = D_t^{-1} + L_t'U_tL_t' + V_t'C_t^{-1}V_t,
\]

\[
c_t = y_{t+1} - \Gamma n_t - u_t,
\]

where \( r_T = 0, U_T = 0, \) and \( \Gamma = GG' \).

Given model (23) – (24), the unbiased predictor for the \((t + 1)\)-period value of \( y \) is given by \( c_t \). Using Kalman Filter (25), \( c_t \) can be computed as:

\[
c_t = a + Bc_{t-1} + K_{t-1}(y_t - c_{t-1})
\]

\[
= \bar{c} + (B - K_{t-1})(c_{t-1} - \bar{c}) + K_{t-1}(y_t - \bar{c})
\]

\[
= \bar{c} + \sum_{\tau=1}^{t}(B - K_{\tau-1})^{t-\tau}K_{\tau-1}(y_\tau - \bar{c}),
\]

where \( \bar{c} = (I - B)^{-1}a \) is the long term mean of \( c_t \). Equation (28) suggests that the predicted return (the first element of vector \( c_t \)) is a weighted average of past returns and predictors, where the weights depend on the structure of the model.
Stochastic Volatility Model

In the stochastic volatility model, we are interested in estimating the following parameters: $\alpha$, $\psi$, $\beta$, $\gamma$, $\theta$, $\lambda = (\lambda_1, \lambda_2)$, and the covariance matrix $\Sigma = (\sigma_{ij})$ given by (20). As before, we enlarge the parameter space with the latent variables, which in this model include the expected returns $\mu_t$, the expected predictors $z_t$, and the return volatility $h_t$ (on log scale). The estimation method relies on the procedure developed in Omori, Chib, Shephard, and Nakajima (2004). First, we introduce the following notation: $d_t \equiv \text{sign}(r_t)$ and

$$ r^*_{t+1} = \ln \left[ (r_{t+1} - \mu_t)^2 \right]. $$

Then, the first equation of the stochastic volatility model (19) implies that

$$ r^* = h_t + \xi_t, $$

where $\xi_t = \ln((\epsilon^2_{t+1})^2)$. Conditional on $d_t$, the multivariate density of $(\xi_t, u_t)$, where $u_t \equiv (\epsilon_t^x, \eta_t^x, \eta_t^z, \nu_t)$ can be approximated as follows:

$$ g(\xi_t, u_t|d_t) = f(\xi_t)f(u_t|\xi_t, d_t) $$

$$ = \sum_{i=1}^{10} p_i f_N \left( \xi_t|m_i, v_i^2 \right) f_n \left( u_t|d_t, \Sigma_{\xi u} \exp(m_i/2)[a_i + b_i(\xi_t - m_i)], \Sigma_u - \Sigma'_{\xi u} \Sigma_{\xi u} \right), $$

where $\Sigma_{\xi u} = (\sigma_{12}, \ldots, \sigma_{1(2k+3)})$, is the covariance between $r_t$ and $u_t$, $\Sigma_u$ is the covariance matrix for $u_t$, and $p_i$, $m_i$, $v_i$, $a_i$ and $b_i$ are constants that are chosen to minimize the approximation error as described in Omori, Chib, Shephard, and Nakajima (see their Table 1). Then $(\xi_t, u_t)$ can be sampled by first identifying the state $s_t \in 1, \ldots, 10$, where $s_t = i$ with probability $p_i$, and next sampling $(\xi_t, u_t)$ from the $i^{th}$ component of the above mixture of multivariate normal distributions. Conditional on the state $\{s_t\}$, the system of equations (19) where the first equation is substituted with (29) becomes a
system of Gaussian regressions which can be estimated using Kalman Filter, as we
illustrated above for the main model (1)–(4). Letting $\phi = (\alpha, \psi, \beta, \gamma, \theta)$, the final
estimation procedure can be described as follows.

1. Sample (as described in Omori, Chib, Shephard, and Nakajima)
   $\{s_t\} | r^*, x, \{d_t\}, \{\mu_t\}, \{z_t\}, \{h_t\}, \phi, \lambda, (\sigma_{ij})$

2. Sample (MH step that uses Kalman filter (25))
   $\phi | r, x, \{d_t\}, \{s_t\}, \{h_t\}, \lambda, (\sigma_{ij})$

3. Sample (using simulation smoother (27))
   $\{(\mu_t, z_t)\} | r, x, \{d_t\}, \{s_t\}, \{h_t\}, \phi, \lambda, (\sigma_{ij})$

4. Sample (MH step that uses Kalman filter (25))
   $\lambda | r^*, x, \{d_t\}, \{s_t\}, \{\mu_t\}, \{z_t\}, \phi, (\sigma_{ij})$

5. Sample (using simulation smoother (27))
   $\{h_t\} | r^*, x, \{d_t\}, \{s_t\}, \{\mu_t\}, \{z_t\}, \phi, \lambda, (\sigma_{ij})$

6. Sample (MH step; the conditional likelihood of the data is Gaussian)
   $(\sigma_{ij}) | r, x, \{d_t\}, \{s_t\}, \{\mu_t\}, \{z_t\}, \phi, \lambda$

7. Goto 1

As before, we discard the first 15000 draws and retain the remaining 25000 draws,
which we use for all of the reported statistics for each parameter.
Benchmark Model

In the benchmark model, we are interested in estimating the following parameters: $A$, $\alpha$, $\beta$, $\gamma$, and the covariance matrix $\Sigma = (\sigma_{ij})$ given by (22). As before, we enlarge the parameter space with the latent variables, which in this model include only the expected returns $\mu_t$. The estimation procedure uses the method developed in (__) for estimating ARMA type models and consists of the following steps.

1. Sample (MH step that uses Kalman filter (25))

   $\alpha, \beta, \gamma | r, x, \psi, A, (\sigma_{ij})$

2. Sample (using simulation smoother (27))

   $\{\mu_t\} | r, x, \alpha, \beta, \gamma, \psi, A, (\sigma_{ij})$

3. Sample (MH step; the proposal density is Gaussian as described in Chib and Greenberg (1994))

   $A | r, x, \alpha, \beta, \gamma, (\sigma_{ij})$

4. Sample (Gibbs step as described in Chib and Greenberg (1994))

   $\psi | r, x, \alpha, \beta, \gamma, A, (\sigma_{ij})$

5. Sample (MH step; the conditional likelihood of the data is Gaussian)

   $(\sigma_{ij}) | y, \{\mu_t\} \alpha, \beta, \gamma, \psi, A$

6. Goto 1

As before, we discard the first 15000 draws and retain the remaining 25000 draws, which we use for all of the reported statistics for each parameter.
References


Table 1: Parameter Estimates for Models 1 - 3

This table shows the mean values and the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,

\[ r_{t+1} = \mu_t + \epsilon_{t+1}, \]
\[ x_{t+1} = z_t + \epsilon_{t+1}, \]
\[ \mu_{t+1} = \alpha + \beta \mu_t + \gamma' z_t + \eta_{t+1}^\mu, \]
\[ z_{t+1} = \psi + \theta z_t + \eta_{t+1}^z, \]

where \( x \) contains the dividend yield (\( dp \)) and the repurchasing yield (\( fp \)).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{\mu} )</td>
<td>3.797</td>
<td>2.673</td>
<td>4.307</td>
</tr>
<tr>
<td></td>
<td>(1.930 5.602)</td>
<td>(1.619 3.731)</td>
<td>(2.281 6.547)</td>
</tr>
<tr>
<td>( \bar{z}_1 ) (dividends)</td>
<td>16.637</td>
<td>16.919</td>
<td>16.269</td>
</tr>
<tr>
<td></td>
<td>(15.264 18.035)</td>
<td>(15.593 18.240)</td>
<td>(14.742 17.738)</td>
</tr>
<tr>
<td>( \bar{z}_2 ) (repurchases)</td>
<td>-6.615</td>
<td>-6.586</td>
<td>-6.668</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.514</td>
<td>0.596</td>
<td>0.578</td>
</tr>
<tr>
<td></td>
<td>(0.366 0.662)</td>
<td>(0.461 0.708)</td>
<td>(0.456 0.695)</td>
</tr>
<tr>
<td>( \gamma_1 ) (dividends)</td>
<td>0.101</td>
<td>0.131</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>(-0.037 0.262)</td>
<td>(0.000 0.274)</td>
<td>(0.000 0.274)</td>
</tr>
<tr>
<td>( \gamma_2 ) (repurchases)</td>
<td>0.111</td>
<td>0.110</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>(0.011 0.237)</td>
<td>(0.003 0.220)</td>
<td>(0.003 0.220)</td>
</tr>
<tr>
<td>( \theta_1 ) (dividends)</td>
<td>0.994</td>
<td>0.995</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>(0.985 0.999)</td>
<td>(0.987 0.999)</td>
<td>(0.981 0.999)</td>
</tr>
<tr>
<td>( \theta_2 ) (repurchases)</td>
<td>0.898</td>
<td>0.896</td>
<td>0.896</td>
</tr>
<tr>
<td></td>
<td>(0.840 0.956)</td>
<td>(0.841 0.955)</td>
<td>(0.838 0.954)</td>
</tr>
<tr>
<td>( R^2 ) (in-sample)</td>
<td>0.035</td>
<td>0.008</td>
<td>0.024</td>
</tr>
<tr>
<td>predict. lik. (out of sample)</td>
<td>-237.520</td>
<td>-246.750</td>
<td>-230.660</td>
</tr>
</tbody>
</table>

32
Table 2: Error Covariance Matrix for Model 1
This table shows the mean values and the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,

\[
\begin{align*}
\tau_{t+1} &= \mu_t + \epsilon_{t+1}, \\
\tau_{t+1} &= \mu_t + \epsilon_{t+1}, \\
\mu_{t+1} &= \alpha + \beta \mu_t + \gamma' z_t + \eta_{t+1} \\
z_{t+1} &= \psi + \theta z_t + \eta_{t+1},
\end{align*}
\]

where \( x \) contains the dividend yield (\( dp \)) and the repurchasing yield (\( fp \)).

\[
\begin{array}{|c|ccc|ccc|}
\hline
 & \epsilon_t & \epsilon_x.dp & \epsilon_x.fp & \eta_t^\mu & \eta_t^z.dp & \eta_t^z.fp \\
\hline
\epsilon_t & 22.283 & -1.620 & -0.018 & -9.974 & -0.780 & -0.659 \\
(18.611, 25.393) & (-1.929, -1.452) & (-0.024, -0.013) & (-10.717, -9.490) & (-1.020, -0.516) & (-0.794, -0.548) \\
\epsilon_x.dp & -1.620 & 0.149 & -0.001 & 0.832 & 0 & 0 \\
(-1.929, -1.452) & (0.131, 0.199) & (-0.001, -0.001) & (0.797, 0.883) & & \\
\epsilon_x.fp & -0.018 & -0.001 & 0.001 & 0.012 & 0 & 0 \\
(-0.024, -0.001) & (-0.001, 0.001) & (0.001, 0.001) & (0.007, 0.018) & & \\
\eta_t^\mu & -9.974 & 0.832 & 0.012 & 19.591 & 0 & 0 \\
(-10.717, -9.490) & (0.797, 0.883) & (0.007, 0.018) & (12.704, 24.946) & & \\
\eta_t^z.dp & -0.780 & 0 & 0 & 0.408 & 0 & 0 \\
(-1.020, -0.516) & & & (0.333, 0.523) & & \\
\eta_t^z.fp & -0.659 & 0 & 0 & 0 & 0 & 5.674 \\
(-0.794, -0.548) & & & (5.143, 6.288) & & \\
\hline
\end{array}
\]
Table 3: Error Covariance Matrix for Model 2
This table shows the mean values the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,
\[
\begin{align*}
    r_{t+1} &= \mu_t + \epsilon_{t+1}, \\
    x_{t+1} &= z_t + \epsilon_{x_{t+1}}, \\
    \mu_{t+1} &= \alpha + \beta \mu_t + \gamma' z_t + \eta_{\mu_{t+1}}, \\
    z_{t+1} &= \psi + \theta z_t + \eta_{z_{t+1}},
\end{align*}
\]
where \(x\) contains the dividend yield (\(dp\)) and the repurchasing yield (\(fp\)).

<table>
<thead>
<tr>
<th></th>
<th>(\epsilon^r)</th>
<th>(\epsilon^{x,dp})</th>
<th>(\epsilon^{x,fp})</th>
<th>(\eta^{\mu})</th>
<th>(\eta^{x,dp})</th>
<th>(\eta^{x,fp})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\epsilon^r)</td>
<td>23.799</td>
<td>-1.680</td>
<td>-0.007</td>
<td>-10.487</td>
<td>-0.809</td>
<td>-0.402</td>
</tr>
<tr>
<td></td>
<td>(20.729 26.145)</td>
<td>(-1.879 -1.479)</td>
<td>(-0.024 0.002)</td>
<td>(-10.982 -9.973)</td>
<td>(-0.982 -0.468)</td>
<td>(-0.623 -0.288)</td>
</tr>
<tr>
<td>(\epsilon^{x,dp})</td>
<td>-1.680</td>
<td>0.157</td>
<td>0.001</td>
<td>1.050</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(-1.879 -1.479)</td>
<td>(0.138 0.203)</td>
<td>(0.001 0.001)</td>
<td>(0.856 1.224)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\epsilon^{x,fp})</td>
<td>-0.007</td>
<td>0.001</td>
<td>0.001</td>
<td>0.009</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(-0.024 0.002)</td>
<td>(0.001 0.001)</td>
<td>(0.001 0.001)</td>
<td>(0.004 0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\eta^{\mu})</td>
<td>-10.487</td>
<td>1.050</td>
<td>0.009</td>
<td>16.426</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(-10.982 -9.973)</td>
<td>(0.856 1.224)</td>
<td>(0.004 0.017)</td>
<td>(12.664 23.494)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\eta^{x,dp})</td>
<td>-0.809</td>
<td>0</td>
<td>0</td>
<td>0.364</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(-0.982 -0.468)</td>
<td>(0.279 0.555)</td>
<td>(0.279 0.555)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\eta^{x,fp})</td>
<td>-0.402</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.728</td>
</tr>
<tr>
<td></td>
<td>(-0.623 -0.288)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.860 6.631)</td>
</tr>
</tbody>
</table>
Table 4: Error Covariance Matrix for Model 3
This table shows the mean values the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,

\begin{align*}
\tau_{t+1} &= \mu_t + \epsilon_t^{r_{t+1}}, \\
x_{t+1} &= z_t + \epsilon_t^{x_{t+1}}, \\
\mu_{t+1} &= \alpha + \beta \mu_t + \gamma \zeta_t + \eta_{\mu_{t+1}}, \\
\zeta_{t+1} &= \psi + \theta z_t + \eta_{\zeta_{t+1}}.
\end{align*}

where \( x \) contains the dividend yield (\( dp \)) and the repurchasing yield (\( fp \)).

<table>
<thead>
<tr>
<th></th>
<th>( \epsilon^r )</th>
<th>( \epsilon^{x.dp} )</th>
<th>( \epsilon^{x.fp} )</th>
<th>( \eta^\mu )</th>
<th>( \eta^{z.dp} )</th>
<th>( \eta^{z.fp} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon^r )</td>
<td>26.785</td>
<td>-1.824</td>
<td>0.010</td>
<td>-11.565</td>
<td>-0.571</td>
<td>-0.458</td>
</tr>
<tr>
<td>( \epsilon^{x.dp} )</td>
<td>(23.509 30.494)</td>
<td>(-2.101 -1.466)</td>
<td>(-0.017 0.029)</td>
<td>(-12.602 -10.372)</td>
<td>(-0.709 -0.474)</td>
<td>(-0.607 -0.327)</td>
</tr>
<tr>
<td>( \epsilon^{x.fp} )</td>
<td>-1.824</td>
<td>0.217</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \eta^\mu )</td>
<td>(-2.101 -1.466)</td>
<td>(0.183 0.277)</td>
<td>(0.001 0.001)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \eta^{z.dp} )</td>
<td>0.010</td>
<td>0.001</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \eta^{z.fp} )</td>
<td>(0.001 0.002)</td>
<td>(0.001 0.002)</td>
<td>(0.001 0.002)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(35)
Table 5: Parameter Estimates for STV Model
This table shows the mean values and the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,

\[
\begin{align*}
    r_{t+1} &= \mu_t + \epsilon_{t+1}^r, \\
    x_{t+1} &= z_t + \epsilon_{t+1}^x, \\
    \mu_{t+1} &= \alpha + \beta \mu_t + \gamma' z_t + \eta^\mu_{t+1}, \\
    z_{t+1} &= \psi + \theta z_t + \eta^z_{t+1}, \\
    h_{t+1} &= \lambda_0 + \lambda_1 h_t + \nu_{t+1},
\end{align*}
\]

where \( x \) contains the dividend yield (\( dp \)) and the repurchasing yield (\( fp \)).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{\mu} )</td>
<td>3.647</td>
<td>(1.211)</td>
<td>(5.895)</td>
</tr>
<tr>
<td>( \bar{z}_1 ) (dividends)</td>
<td>16.242</td>
<td>(14.564)</td>
<td>(17.738)</td>
</tr>
<tr>
<td>( \bar{z}_2 ) (repurchases)</td>
<td>-6.370</td>
<td>(-9.958)</td>
<td>(-3.876)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.403</td>
<td>(0.085)</td>
<td>(0.655)</td>
</tr>
<tr>
<td>( \gamma_1 ) (dividends)</td>
<td>0.043</td>
<td>(-0.196)</td>
<td>(0.247)</td>
</tr>
<tr>
<td>( \gamma_2 ) (repurchases)</td>
<td>0.160</td>
<td>(0.002)</td>
<td>(0.350)</td>
</tr>
<tr>
<td>( \theta_1 ) (dividends)</td>
<td>0.992</td>
<td>(0.981)</td>
<td>(0.999)</td>
</tr>
<tr>
<td>( \theta_2 ) (repurchases)</td>
<td>0.899</td>
<td>(0.833)</td>
<td>(0.963)</td>
</tr>
<tr>
<td>( \lambda_1 ) (volatility)</td>
<td>1.698</td>
<td>(1.062)</td>
<td>(2.325)</td>
</tr>
<tr>
<td>( \lambda_2 ) (volatility)</td>
<td>0.975</td>
<td>(0.925)</td>
<td>(0.998)</td>
</tr>
<tr>
<td>( R^2 ) (in-sample)</td>
<td>0.040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>predict. lik. (out of sample)</td>
<td>-165.465</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6: Error Covariance Matrix for STV Model

This table shows the mean values the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,

\[ r_{t+1} = \mu_t + e^{h_{t+1}/2} \epsilon_{t+1}, \]
\[ x_{t+1} = \gamma + \epsilon_{t+1}, \]
\[ \mu_{t+1} = \alpha + \beta_1 \mu_t + \gamma' z_t + \eta_{t+1}, \]
\[ z_{t+1} = \psi + \theta z_t + \eta_{t+1}^z, \]
\[ h_{t+1} = \lambda_0 + \lambda_1 h_t + \nu_{t+1}. \]

where \( x \) contains the dividend yield (\( dp \)) and the repurchasing yield (\( fp \)).

\[
\begin{array}{cccccccc}
\epsilon^r & \epsilon^x.jp & \epsilon^x.dp & \eta^\mu & \eta^x.dp & \eta^x.fp & \nu \\
1 & 0 & 0 & -0.085 & -0.064 & -0.060 & 0 \\
0 & 0.220 & 0 & (0.143 0.303) & (0.961 1.895) & (0.418 0.021) & 0 \\
0 & 0 & 0.194 & (0.040 0.393) & (0.040 0.393) & (0.040 0.393) & 0 \\
\eta^\mu & -0.085 & 1.434 & -0.155 & 12.581 & 0 & 0 & 0 \\
(-0.125 -0.047) & (0.961 1.895) & (-0.418 0.021) & (8.022 18.429) & 0 & 0 & 0 \\
\eta^x.dp & -0.064 & 0 & 0 & 0 & 0.617 & 0 & 0 \\
(-0.099 -0.023) & 0 & 0 & 0 & 0 & 0 & 0 \\
\eta^x.fp & -0.060 & 0 & 0 & 0 & 0 & 6.044 & 0 \\
(-0.079 -0.039) & 0 & 0 & 0 & 0 & 0 & (4.897 7.313) \\
\nu & 0 & 0 & 0 & 0 & 0 & 0.277 & (0.076 0.658) \\
\end{array}
\]
Table 7: **Parameter Estimates for the Benchmark Model**

This table shows the mean values and the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,

\[
\begin{align*}
    r_{t+1} &= \mu_t + \epsilon_{t+1}, \\
    x_{t+1} &= \psi + Ax_t + \epsilon_{t+1}, \\
    \mu_{t+1} &= \alpha + \beta \mu_t + \eta_{t+1},
\end{align*}
\]

where $x$ contains the dividend yield ($dp$) and the repurchasing yield ($fp$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{\mu} )</td>
<td>2.766</td>
<td>(1.148 4.402)</td>
</tr>
<tr>
<td>( \bar{x}_1 ) (dividends)</td>
<td>6.599</td>
<td>(0.333 12.758)</td>
</tr>
<tr>
<td>( \bar{x}_2 ) (repurchases)</td>
<td>-6.430</td>
<td>(-9.958 -3.926)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.547</td>
<td>(0.373 0.695)</td>
</tr>
<tr>
<td>( \theta_1 ) (dividends)</td>
<td>0.983</td>
<td>(0.951 0.998)</td>
</tr>
<tr>
<td>( \theta_2 ) (repurchases)</td>
<td>0.866</td>
<td>(0.777 0.967)</td>
</tr>
<tr>
<td>( R^2 ) (in-sample)</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>predict. lik. (out of sample)</td>
<td>-273.596</td>
<td></td>
</tr>
</tbody>
</table>
Table 8: Error Covariance Matrix for the Benchmark Model

This table shows the mean values the 95% intervals (in parentheses) of the posterior distributions for the estimated parameters from the following model,

\[
\begin{align*}
    r_{t+1} &= \mu_t + \epsilon_{t+1}, \\
    x_{t+1} &= \psi + Ax_t + \epsilon_{x, t+1}, \\
    \mu_{t+1} &= \alpha + \beta \mu_t + \eta_{t+1},
\end{align*}
\]

where \( x \) contains the dividend yield (\( dp \)) and the repurchasing yield (\( fp \)).

<table>
<thead>
<tr>
<th></th>
<th>( \epsilon^r )</th>
<th>( \epsilon^{x, dp} )</th>
<th>( \epsilon^{x, fp} )</th>
<th>( \eta^\mu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon^r )</td>
<td>17.566</td>
<td>0</td>
<td>0</td>
<td>-8.872</td>
</tr>
<tr>
<td></td>
<td>(8.172, 29.220)</td>
<td>(-12.643, -4.982)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \epsilon^{x, dp} )</td>
<td>0</td>
<td>1.007</td>
<td>0</td>
<td>1.412</td>
</tr>
<tr>
<td></td>
<td>(0.732, 1.301)</td>
<td>(0.233, 2.100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \epsilon^{x, fp} )</td>
<td>0</td>
<td>0</td>
<td>5.982</td>
<td>2.717</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4.932, 7.171)</td>
<td>(0.547, 4.783)</td>
</tr>
<tr>
<td>( \eta^\mu )</td>
<td>-8.872</td>
<td>1.412</td>
<td>2.717</td>
<td>36.958</td>
</tr>
<tr>
<td></td>
<td>(-12.643, -4.982)</td>
<td>(0.233, 2.100)</td>
<td>(0.547, 4.783)</td>
<td>(21.996, 46.414)</td>
</tr>
</tbody>
</table>