YALE
INTERNATIONAL CENTER FOR FINANCE

Yale ICF Working Paper No. 02-25
June 16, 2002

MARKET PRICES OF RISK AND RETURN PREDICTABILITY IN A JOINT STOCK-BOND PRICING MODEL

Harry Mamaysky
Yale School of Management

Morgan Stanley
New York

Market Prices of Risk and Return Predictability in a Joint Stock–Bond Pricing Model*

Harry Mamaysky †

Yale School of Management
135 Prospect Street
New Haven, CT 06520-8200

and

Morgan Stanley
1585 Broadway
New York, NY 10036

First version: October 25, 2001
Last revised: June 16, 2002

*This paper represents an extension of results that originally appeared in Mamaysky, 2002, "On the joint pricing of stocks and bonds: Theory and evidence." I would like to thank Monika Piazzesi for her extremely valuable suggestions and advice, as well as Zhiwu Chen, Will Goetzmann, Ken Singleton, Werner Stanzl, Jacob Thomas, and Jiang Wang for helpful discussions and suggestions. Also I would like to thank seminar participants at the Yale School of Management, New York University, the 2002 NBER Asset Pricing Meeting, and the 2002 WFA meeting for valuable comments.

†Email: harry.mamaysky@morganstanley.com, phone (212) 761-1022, fax (212) 761-9589.

Market Prices of Risk and Return Predictability in a Joint Stock–Bond Pricing Model

Abstract

This paper examines the related questions, of the time-series behavior of expected returns and of return predictability, within the framework of the stock–bond pricing model proposed in Mamaysky (2002). The key advantage of the model-based approach adopted in this paper is that the quantities of interest (i.e. expected returns, prices of risk, and R^2's of forecasting regressions of returns on their *true* conditional expectations) are directly observable (once the model has been fitted to the data). Furthemore, the fact that the present model accomodates jointly the pricing of both bonds and stocks allows us to derive estimates of prices of risk and of expected returns that incorporate, by construction, the relevant information from both bond and stock markets. Estimation of the model using U.S. data reveals a rich dynamic structure of prices of risk, some pro- and some countercyclical, and of expected returns. Also, the paper suggests that excess return predictability (as measured by the above mentioned R^2's) for a broad market index is a hump-shaped function of the forecasting horizon, achieving a maximum value of roughly 13.5% at a time horizon of five years.

JEL Classification: G12, G13.

1 Introduction

Expected returns, justifiably, have aroused much interest among both finance scholars and practitioners.[1] After all, many asset pricing models are most conveniently expressed in terms of their implications for expected returns on securities, which makes examination of the latter necessary for testing pricing models. Practitioners, on the other hand, are concerned with knowing expected returns because such information helps to determine whether securities are correctly or incorrectly priced in markets. However, the obvious difficulty with studies involving expected returns is that expected returns are not directly observable.

This inconvenience is solved by conjecturing that future asset returns are a function of today's value of a set of observable economic variables coupled with a noise term assumed to be orthogonal to any information that may be available today. Expected returns are then given by the the above function evaluated using the set of present observables. While this approach is fundamentally sound (indeed what other approach is possible), its implementation suffers from the unfortunate fact that the theoretically motivated explanatory variables, such as market betas and covariances with consumption growth, are themselves difficult to observe, and by the fact that estimated betas and covariances seem to perform very poorly in-sample as determinants of expected security returns. In response to this, selection of the commonly used set of explanatory variables has been guided not by any ex-ante theoretical consideration, but by satisfactory in-sample performance. However, satisfactory in-sample performance does not necessarily imply satisfactory out-of-sample performance. Indeed, whether or not the current set of candidate explanatory variables for expected returns (such as size, value, possibly momentum) withstands the test of time is a question whose answer is not yet known.

This paper proposes and implements a theoretically motivated approach which offers some guidance for determining the importance of a given set of explanatory variables for expected returns. The general idea is to note that if a particular variable is truly important for the determination of expected returns, then this variable must represent a risk that is priced in the economy. That is, owning a security whose return depends on the chosen explanatory variable must entitle that security's owner to some amount of compensation, in the form of expected returns, for bearing the risk for which the chosen explanatory variable proxies. This idea is implemented by first conjecturing a pricing model, based on the idea of no arbitrage, for the relevant set of security prices. A security's dependence on a given explanatory variable is then represented as a loading of that security's return on a pricing factor associated with the risk in question. Within the model's framework, expected returns for a security are then shown to be functions of that security's loading on the risk factors in the economy multiplied by the market prices of risk which these risk factors command. Finally, we assume a tranformation which allows us to move between the physical measure of the data and the risk neutral measure used for pricing. With these ingredients in place, the

[1]Unless otherwise noted, expected returns in this paper refer to expected returns above a riskless interest rate, i.e. expected excess returns. Also, all returns in this paper are in nominal terms.

model is estimated using historical prices of the securities in question, and prices of risk for the conjectured explanatory variables (as well as security expected returns) are then directly observable within the framework of the model.

This approach has been extensively used in the analysis of returns in fixed income markets. Indeed, within the class of affine term structure models (such as Vasicek (1977) or Cox, Ingersoll, and Ross (1985), and their extensions), the above three concepts (risk factors, expected returns, and prices of risk) are all very naturally and conveniently tied together. However, such affine term structure models have to date been used, as the name suggests, only for the study of bond prices. However, Mamaysky (2002a) shows that it is possible to use the technology of affine term structure models to derive an affine security pricing model, capable of jointly pricing bonds and stocks in a unified framework. Within the context of Mamaysky (2002a), it is then possible to set up and estimate a model where a cross section of bond and stock prices jointly depends on a parsimonious set of pricing factors. The end result of this procedure is a set of pricing factors and their associated prices of risk. If it can then be demonstrated that the extracted pricing factors account for a substantial portion of cross sectional return variation for any given set of securities (such as stocks), then using the model's formulas we are able as well to compute an instantaneous expected return for any security of interest.

To make ideas more concrete, let us note that in the present paper a bond is a security which promises its owner a fixed set of nominal cash payments at some future dates. A stock, on the other hand, entitles its owner to a stream of stochastic nominal dividend payments.[2] By choosing convenient processes for the instantaneous interest rate in the economy and for the instantaneous stochastic dividends paid by stocks, Mamaysky (2002a) shows that bond and stock prices can both be expressed as exponential affine functions of the state variables in the economy. Indeed, the bond prices produced by the model are exactly the bond prices of standard affine term structure models (see Duffie and Kan (1996) for a general treatment of affine term structure models). The existence of such closed form solutions for *both* bond and stock prices makes the model amenable for a unified empirical study of prices in both markets. Furthermore, the model allows for quite general interdependence between bond and stock prices: Bond and stocks are allowed to depend on a set of joint pricing factors which possess stationary distributions; Furthermore, stocks are allowed to depend on a set of factors which follow random walks. Together these two sets of factors give the model a great deal of empirical flexibility for matching the historical behavior of bond and stock prices.

The main payoff of the present approach, however, comes from the fact that in addition to closed form bond and stock prices, the Mamaysky (2002a) model also produces closed form solutions for the prices of risk associated with the model's pricing factors, and therefore for

[2]All securities in the present model are assumed to be free of default risk. A theoretical development of a joint stock–bond pricing model with default risk is in Mamaysky (2002b). The assumption of no default risk essentially restricts the focus of the present paper either to portfolios of stocks, where default risk is less of an issue, or to extremely high quality credits. If this paper, the focus is on government bond prices and on the returns on large equity portfolios.

bond and stock expected returns. All of these quantities are shown to depend on the model's parameters, as well as on the time series of pricing factors extracted during estimation of the model. Therefore, estimation of the model makes factor prices of risk and security expected returns directly observable! Furthermore, these quantities depend only on the security prices used in the model's estimation, and in particular do not depend on any set of economic variables which must be exogenously chosen by the econometrician.

The question of whether or not a given set of exogenously chosen explanatory variables is truly an important determinant of expected returns can then be naturally addressed in the present framework. To proceed with this analysis we estimate the model using returns on those securities which possess characteristics important for the determination of expected returns in historical data. We then extract a pricing factor associated with each of these security characteristics. Since the model produces a price of risk for each factor, the historical importance of a given factor loading for expected returns is then immediately obvious. For example, if a security has a high loading on a factor at a time when that factor's price of risk is high, then that factor is an important determinant of the instantaneous expected returns of that security.

More precisely, we estimate the model using a time series of government bond prices across a wide span of maturities, as well as using the returns of a broad market index, of a portfolio containing small stocks, and of a portfolio containing stocks with high book-to-market values. The inclustion of these three stock portfolios allows us to extract three stock-specific factors, having to do with broad market exposure, with the size effect, and with the value effect respectively. We then use an estimation approach proposed by Chen and Scott (1993) to extract five pricing factors which best explain the historical behavior of the prices of the securities used in model estimation. Of these factors, two are stationary factors which are joint between bonds and stocks (and which drive prices of risk in the model), and three are the above mentioned stock-specific pricing factors.

By assuming that the moments of the pricing kernel in the model depend on the joint bond–stock factors, we also are able to derive a time series of prices of risk for the above mentioned factors. These prices of risk in turn produce a series of expected returns for the securities used in the model estimation. We then show that the extracted model factors are important determinants of returns across a wide cross-section of stock portfolios, which suggests that knowledge of the prices of risk of the five model factors, together with the knowledge of the loadings of a given security (stock or bond) on these factors, produces an estimate of that's securities instantaneous expected return. An important advantage of the present approach is that the factors which drive prices of risk and expected returns in the model have been derived from an estimation which uses jointly bond and stock price data. Hence, the relevant information for expected returns is naturally extracted from both bond and stock markets.

As an additional benefit of the model, it is shown that security returns can be naturally decomposed into expected returns (known once the model has been estimated) and an orthogonal error term (whose variance depends on model parameters). This decomposition allows for a very natural measure of return predictability to be derived: We simply look at what fraction of the variability of security returns is attributable to the variability of that security's expected return series. This amounts to computing the R^2 from a regression of a security's return on that security's *true* conditional return expectation. Furthermore, this R^2 measure can be derived in closed form as a function of the model's parameters, and of the time horizon of the forecasting regression. It should be noted that such an R^2 can not be derived without recourse to a pricing model because true conditional expectations, as has already been pointed out, are unobservable. Furthermore, let us emphasize that the model produced R^2 measure does not involve actually running any regressions; instead, knowledge of the model parameters and of a given security's loadings on the model factors is sufficient for the computation of the above R^2 for the security in question. This is of great benefit because the R^2's of forecasting regressions are very sensitive to the sample period being used (see Kirby (1997), for example).

Estimation of the model shows that the above mentioned five pricing factors can adequately account for the time series variation of a cross section of government bond prices and of a cross section of stock prices in the U.S. over the last three decades. Interestingly, the extracted model factors are important determinants of a cross section of the returns of a large set of stock portfolios not used in the estimation of the model. This suggests that the prices of risk of the extracted model factors are indeed important determinants of expected returns for a wide cross section of stocks and bonds. It should be noted that the emphasis of the present analysis is on the prices of risk of the stock-specific factors in the model. Careful analysis of bond market prices of risk in an affine framework already exists in the literature (see Duffee (2001) and Dai and Singleton (2001) for example), and for this reason it seems more fruitful for this paper's focus to be on the prices of risk for the stock factors.[3]

The price of risk associated with the market factor is shown to have a rich dynamical behavior, exhibiting substantial time series variability, as well as a pronounced countercyclical behavior.[4] The mean value of this price of risk is approximately 0.40. The size factor price of risk also exhibits a substantial amount of time series variation, but has a long run mean of only 0.08 (despite the fact that this price of risk has assumed values as high as 0.5 historically), suggesting that a security's size is not a significant long run determinant of that security's expected return. Finally, the price of risk associated with the value factor is extremely stable over time, and assumes the high steady state value of approximately 0.53. This suggests that the "valueness" of a given security is an important determinant of that security's long-run expected returns. We also document that expected returns on all

[3]This emphasis in not in name only. The present model is fairly complex (with 2 common and 3 stock-specific factors), and as such certain compromises were necessary for the model's implementation. Such necessary reductions in model complexity were often accomplished at the expense of accuracy for the bond price of risk estimates, while all efforts were made to be very careful about the stock factor price of risk estimates produced by the model. Details about these issues are in the text.

[4]Here countercyclical means that a given variable is low during NBER classified business cycle peaks, and is high during NBER classified business cycle troughs.

three of the stock portfolios used in the model estimation exhibit countercyclical behavior. Finally, we show that is is the prices of risk associated with the stock-specific factors, rather than with the joint bond–stock factors, that are the major determinants of expected returns for equity portfolios. It should be noted that the joint factors, while not having prices of risk which are important for stock expected returns, are nevertheless importnant for stock expected returns because stock expected returns are functions of the values of these joint bond–stock factors!

Finally, a diagnostic test is derived which shows that the present model does a good job of accounting for the amount of data-observed predictability of expected excess returns for actual excess returns. This diagnostic suggests that the model assumed factor dynamics are sufficiently rish so as to properly account for the time series behavior of prices of risk of the stock-specific factors extracted during model estimation. As has already been mentioned, this predictability is measured as the model implied value of the R^2 over a certain forecasting horizon of a regression of excess returns on their true conditional expectations. It is proven in the paper that this R^2 must be a hump-shaped function of the forecasting horizon, starting at zero for short time periods, and again going to zero at long forecasting horizons. The parameter estimates obtained in estimation of the model suggest that for a broad market index the maximal value of the model forecasting R^2 is equal to 13.5% at a time horizon of five years.

The remainder of the paper proceeds as follows. Section 2 develops the model. Section 3 discusses the data sources, as well as the estimation procedure for the model. Section 4 presents the results of model estimation, as well as a series of diagnostic tests meant to assess the validity of the model's implementation. Section 5 concludes. All derivations are in the Appendix.

2 A Joint Model for Stocks and Bonds

The model presented in this section is a special case of the model developed in Mamaysky (2002a). To conserve space, we will keep the development as free of technical details as possible, and simply refer the interested reader to Mamaysky (2002a). We assume the usual regularity conditions on the economy in question (such as the existence of a filtered probability space and market completeness), and in particular we assume the existence of an equvalent martingale measure $\mathcal{Q}$, under which all discounted gains processes are martingales. Under certain regularity conditions (see Harrison and Pliska (1981), Dybvig and Huang (1988), or Duffie (2001) for a textbook treatment), the existence of $\mathcal{Q}$ is equivalent to the absence of arbitrage opportunities in the economy. Going forward let us refer to the actual (or physical) probability measure as $\mathcal{P}$.

In the present setting, we assume that bonds and stocks are default-free. Hence any promised payment is guaranteed to be made in the model. A joint model for pricing bonds and stocks in a setting with default risk is given in Mamaysky (2002b).

2.1 Bonds

Bonds are the standard bonds of affine term structure models.[5] As such, bonds are assumed to pay a one dollar dividend at some point in the future. Under $\mathcal{Q}$, the fact that discounted gains processes are martingales is equivalent to the price of a bond which pays \$1 at time T being given by

$$P_T(t) = \mathbb{E}_t^{\mathcal{Q}} \left[\exp\left(-\int_t^T r(u)du \right) \right]. \tag{1}$$

Here $r(t)$ is the short term interest rate, or simply the short rate. In order to solve for the expectation in (1), we need to specify the behavior of this short rate. In particular, we assume that the short rate $r(t)$ is given by

$$r(t) = r_0 + r_Y' Y(t) \tag{2}$$

where $r_0 \in \mathbb{R}$ and $r_Y \in R^N$ are constant, and $Y(t)$ is an N-dimnesional vector of state variables. We will often write $r_{Y,n}$ to indicate the n^{th} element in the vector r_Y. As is usually done in term structure models, we will assume that these Y-type factors are stationary. The expectation in (1) can be solved for quite a general Y process (within the affine class), as shown in Duffie and Kan (1996), but we will have no need for such generality in this paper. Instead we assume that each of the N Y-tpye factors follows an Ornstein-Uhlenbeck process with dynamics under $\mathcal{Q}$ given by (for $n = 1, \ldots, N$)

$$dY_n(t) = \tilde{K}_{Y,n}(\tilde{\Theta}_n - Y_n(t))dt + \sigma_{Y,n} d\tilde{W}_n(t), \tag{3}$$

where $\tilde{K}_{Y,n}$, $\tilde{\Theta}_n$, and $\sigma_{Y,n}$ are all constants, and where $\tilde{W}_n(t)$ is a standard Brownian motion under $\mathcal{Q}$, and is independent of all other Brownian motions in the economy. Stationarity is insured by requiring that $\tilde{K}_{Y,n} > 0$ for all n. We will write $\tilde{\Theta}$ for the vector given by

$$\tilde{\Theta} = \begin{bmatrix} \tilde{\Theta}_1 \\ \vdots \\ \tilde{\Theta}_N \end{bmatrix}.$$

Also, we will write $\tilde{K}_Y$ for the $N \times N$ matrix whose diagonal elements are given by $\tilde{K}_{Y,n}$. Given these assumptions for the dynamics of the Y-type factors, we have specified a multi-dimensional Vasicek-type model (the original one-dimensional version of this model is in Vasicek (1977)).

The solution for bond prices in such a setting are widely known (see, for example, Mamaysky (2002a)), and are given by

$$P_T(t) = \exp\Big(A_T(t) - B_T(t)' Y(t) \Big), \tag{4}$$

[5]To date, the most general treatment of affine term structure models is in Duffie and Kan (1996).

where the A and B functions are given by

$$A_T(t) = -(r_0 + \tilde{\Theta}' r_Y)(T-t) + \tilde{\Theta}' B_T(t) + \frac{1}{2}\sum_{n=1}^{N} \sigma_{Y,n}^2 \int_t^T [B_T(u)]_n^2 du, \tag{5}$$

and where

$$[B_T(t)]_n = \frac{r_{Y,n}}{\tilde{K}_{Y,n}} \left(1 - e^{-\tilde{K}_{Y,n}(T-t)}\right). \tag{6}$$

We will write $[\cdot]_n$ to indicate the b^{th} element of a vector, and $[\cdot]_{nm}$ to indicate the element in the n^{th} row and m^{th} column of a matrix.[6]

Spot rates, or yields on zero-coupon bonds, are given by

$$-\frac{\log P_T(t)}{T-t} = \frac{1}{T-t}\Big(-A_T(t) + B_T(t)' Y(t)\Big).$$

It is easy to check that the yield on an infinitely lived bond, i.e. as $T-t$ goes to infinity, is given by

$$y_\infty = r_0 + \tilde{\Theta}' r_Y - \frac{1}{2}\sum_{n=1}^{N}\left(\frac{\sigma_{Y,n}\, r_{Y,n}}{\tilde{K}_{Y,n}}\right)^2, \tag{7}$$

and is simply a function of the model's parameters, not dependent on the current values of the Y-type factors.

2.2 Stocks

Stocks in the present model entitle their owners to two types of dividend payments. First, the owner of stock i receives an instantaneous dividend of $D_i(t)$ per unit time, implying that the cumulative dividend paid by the stock through time t' is given by

$$\int_0^{t'} D_i(u) du.$$

Furthermore, the stock entitles its owner to a terminal dividend $\bar{D}_i(T)$ paid at some time T in the future. After the terminal dividend $\bar{D}_i(T)$ is paid, the owner of the stock is entitled to no future dividend payments. Let us write $S_{i,T}(t)$ for the price of a stock with this dividend process. Then the fact that discounted gains processes for stocks must be martingales under $\mathcal{Q}$ is equivalent to the following formula for the price $S_{i,T}(t)$ of stock i:

$$S_{i,T}(t) = \mathbb{E}_t^{\mathcal{Q}}\left[\int_t^T e^{-\int_t^u r(s)ds} D_i(u) du + e^{-\int_t^T r(s)ds} \bar{D}_i(T)\right]. \tag{8}$$

[6]Given $B_T(t)$ in (6), it is to check that

$$\int_t^T [B_T(u)]_n^2 du = \frac{r_{Y,n}^2}{\tilde{K}_{Y,n}^2}\left[\tau - \frac{2}{\tilde{K}_{Y,n}}\Big(1 - \exp(-\tilde{K}_{Y,n}\tau)\Big) + \frac{1}{2\tilde{K}_{Y,n}}\Big(1 - \exp(-2\tilde{K}_{Y,n}\tau)\Big)\right],$$

where $\tau \equiv T - t$.

While this representation for the stock price is completely standard, finding a closed form solution for $S_{i,T}(t)$ may not always be possible. In particular, the choice of the dividend process which the stock pays out is crucial in order to be able to solve the above expectation in closed form. While a closed form solution for the stock price in (8) is not, strictly speaking, necessary because the integral may be computed numerically, the existence of such a closed form solution clearly makes implementation of the model susbtantially simpler.

With these considerations in mind, Mamaysky (2002a) shows that a particular choice of intermediate and terminal dividend processes leads to a very simple solution for the stock price. The crucial feature of the dividend process which leads to closed form solutions is that the intermediate dividend turns out to equal the stock price multiplied by an affine dividend yield. Before proceeding to specify the actual dividend process, we note that the only factors in the economy (i.e. the Y-type factors of the previous section) are stationary by assumption. Since ample empirical evidence points to the possibility that stock prices contain a random-walk component, a model with only stationary factors will most likely be unable to provide an adequate match to actual stock prices. Because of this we introduce into the economy another set of factors, which we will call the Z-type factors. We will assume that the $Z(t)$ is an M-dimensional vector which gives the time t values of the Z-type factors, and that the dynamics of $Z(t)$ are given by

$$dZ(t) = \tilde{\mu} dt - \tilde{K}_Z Y(t) dt + \Sigma_Z d\tilde{W}(t), \tag{9}$$

where $\tilde{\mu} \in \mathbb{R}^M$, $\tilde{K}_Z \in \mathbb{R}^{M \times N}$, and $\Sigma_Z \in \mathbb{R}^{M \times (N+M)}$. The $N+M$-dimensional process $\tilde{W}(t)$ is assumed to be vector of standard, independent Brownian motions, the first N of which are the Brownian motions from (3). Also we assume that Σ_Z only has elements along its right diagonal (i.e. locations $\{M, N+M\}$, $\{M-1, N+M-1\}$, and so on). Note that we have assumed that the Z-type factors have constant volatilities, an assumption which can be relaxed (as is done in Mamaysky (2002a)), but one which suffices for the purposes of this paper.

With the Y and Z-type factors defined, we assume that the intermediate dividend $D_i(t)$ which is paid by stock i is given by the following formula:

$$D_i(t) = \left(\delta_{i,0} + \delta_{i,Y}' Y(t)\right) \exp\left(a_i(\delta_{i,0}, \delta_{i,Y}, C_i)\, t - B_i(\delta_{i,Y}, C_i)' Y(t) - C_i' Z(t)\right), \tag{10}$$

where $\delta_{i,0} \in \mathbb{R}$, $\delta_{i,Y} \in \mathbb{R}^N$, $C_i \in \mathbb{R}^M$, and where $a_i : \delta_{i,0}, \delta_{i,Y}, C_i \to \mathbb{R}$ and $B_i : \delta_{i,Y}, C_i \to \mathbb{R}^N$ are functions depending on the values of $\{\delta_{i,0}, \delta_{i,Y}, C_i\}$. Indeed we assume that a_i is given by

$$a_i = r_0 - \delta_{i,0} + \tilde{\Theta}' \tilde{K}_Y' B_i + \tilde{\mu}' C_i - \frac{1}{2} \sum_{n=1}^{N} \sigma_{Yn}^2 [B_i]_n^2 - \frac{1}{2} \sum_{m=1}^{M} \sigma_{Z,m}^2 [C_i]_m^2, \tag{11}$$

and that B_i is given by

$$B_i = \left(\tilde{K}_Y'\right)^{-1} \left(r_Y - \delta_{i,Y} - \tilde{K}_Z' C_i\right). \tag{12}$$

Going forward, we will suppress the dependence of a_i and B_i on $\{\delta_{i,0}, \delta_{i,Y}, C_i\}$, though these two quantities are assumed to satisfy (11) and (12) throughout the paper. We furthermore assume that the terminal dividend $\bar{D}_i(T)$ is given by

$$\bar{D}_i(T) = \exp\Big(a_i\, t - B_i'Y(t) - C_i'Z(t)\Big). \tag{13}$$

This is the same as the exponential in (10), but without the affine multiplicative term in front.

Before we proceed, let us comment briefly on this choice of dividend process. In order to have a workable model for stock prices, we must specify a dividend process so that the expectation in (8) may be computed. An important feature of the chosen dividend process is that is have enough empirical flexibility to be able to provide an adequate description for the data. On the other hand, in order to maintain the tractability of the model, we cannot choose an arbitrary dividend process.

It is the hope that the dividend process specified in (10) and (13) provides us with a good compromise with regard to these two objectives. First off, we do have a great deal of empirical flexbiility in writing down the dividend process. The quantities $\{\delta_{i,0}, \delta_{i,Y}, C_i\}$ can all be specified exogenously (subject to some parameter restrictions which will be given later). Hence the dividend can load arbitrarily on the Z-type factors in the economy, and also can load arbitrarily on the Y-type loadings, as long as this dependence occurs outside the exponential in (10). Compromises for tractability take two forms. First, we assume that the instantaneous dividend is given by the product of an exponential affine function and an affine function, and that the terminal dividend is given by the exponential inside the instantaneous dividend. Second, we must restrict the dependence of the exponential on the Y-type factors.[7] Together, these two restrictions lead to a very simple solution for the expectation in (8).

It is shown in Mamaysky (2002a), that given the dividend process in (10) and (13), the stock price (which is equal to the expectation in (8) is given by

$$S_i(t) = \exp\Big(a_i\, t - B_i'Y(t) - C_i'Z(t)\Big). \tag{14}$$

Note that a_i and B_i are assumed to satisfy equations (11) and (12) respectively. We note as well that the stock price does not depend on the date of the terminal dividend $\bar{D}_i(T)$. This is why we write $S_i(t)$ in (14), instead of $S_{i,T}(t)$, and we will continue to neglect the dependence of the stock price on T for the remainder of the paper.

Given the stock price in (14) and the instantaneous dividend process $D_i(t)$ in (10), we see that the instantanoues dividend is in fact equal to

$$D_i(t) = \Big(\delta_{i,0} + \delta_{i,Y}'Y(t)\Big)\, S_i(t).$$

[7]The reason for doing so is to accomodate the no-arbitrage restrictions which are placed on these coefficients for any security whose price is exponential affine in the model's factors (which we will soon see to be the case for the stock price). The loadings inside the exponential on time and on the Y-type factors must be restricted to allow for the stock plus dividend process to have a return equal to the risk-free rate under the risk neutral measure $\mathcal{Q}$. The derivation of the stock price in Mamaysky (2002a) makes these points obvious.

Note that this relationship is a result, rather than an assumption: We start with a dividend process of the form in (10), solve for the stock price, and then it turns out that the exogenously specified instantaneous dividend and the endogenously determined stock price satisfy the above relationship. Let us define $\delta_i(t)$ as

$$\delta_i(t) \equiv \delta_{i,0} + \delta_{i,Y}' Y(t). \tag{15}$$

Going forward, we will refer to $\delta_i(t)$ as the dividend yield on stock i.

The Transversality Condition

Our intuition for the way in which stock prices are determined tells us that as the terminal dividend becomes ever farther away in the future, the value of a given stock should be mostly determined by the value of the interemediate dividend stream. This intuition is made concrete in the present setting by imposing a transversality condition on the present value (under the risk neutral measure) of the terminal dividend: We require that as the date of the terminal dividend goes to infinity, the value of the terminal dividend (and hence its contribution to the stock price) should go to zero. This means that we require that the following limit exists, and that it equal zero:

$$\lim_{T\to\infty} \mathbb{E}_T^Q \left[e^{-\int_t^T r(u)du} \bar{D}_i(T) \right] = 0. \tag{16}$$

Since from (13) and (14), we see that $S_i(t) = \bar{D}_i(t)$, the tranversality condition in (16) also implies that the present value of a share of stock given infinitely far off in the future is zero.

It is shown in Mamaysky (2002a), that for the model we have specified in this paper, the transversality condition in (16) holds as long as the following parameter restriction is satisfied

$$\delta_{i,0} + \delta_{i,Y}' \tilde{\Theta} + \sum_{n=1}^{N} \frac{[\delta_{i,Y}]_n \sigma_{Y,n}^2}{[\tilde{K}_Y]_{nn}^2} \left(\frac{1}{2}[\delta_{i,Y}]_n - r_{Y,n} + [C_i' \tilde{K}_Z]_n \right) > 0. \tag{17}$$

Going forward, we will assume that this restriction is indeed satisfied.

It is furthermore shown in Mamaysky (2002a) that imposing the transversality condition allows us to define an infinitely lived stock (i.e. one which pays $D_i(t)$ forever, with no terminal dividend) whose price will be equal to $S_i(t)$, the price of any finitely lived stock. However, this point is of no practical interest for the questions at hand, and so going forward in this paper we wiil assume that all stocks are finitely lived, though with a terminal dividend which is potentially very far off in the future. Also, as has been shown in (14), the date of the terminal dividend does not affect the stock price.

2.3 Total Returns Processes

While it is possible to estimate the present model using data on stock prices and on the associated dividend processes, doing so is often unnecessary for the questions at hand. For example, to look at the risk premia implied by the present model, separate data on stock prices and dividends is unncecessary. Instead, it is often more convenient to work with a total returns process associated with a given stock. The total returns process is the value of a portfolio which starts off holding one share of the stock, and then reinvests all dividends back into the stock itself.

If we write $s_i(t)$ for the value of a total returns process associated with stock i, then the "full reinvestment of dividends" condition allows us to write the dynamics of $s_i(t)$ as written as

$$\frac{ds_i(t)}{s_i(t)} = \frac{dS_i(t)}{S_i(t)} + \delta_i(t). \tag{18}$$

Details are provided in Mamaysky (2002a). We would now like to construct a convenient solution for $s_i(t)$, just as we have done for $S_i(t)$. Mamaysky (2002a) shows how this may be done, and we will follow that construction.

Let us select a set of M stocks. For these stocks, referenced by $m = 1, \ldots, M$, let us define the following two matrixes:

$$C \equiv \left[C_1 \, C_2 \, \cdots \, C_M \right],$$

and

$$\delta_Y \equiv \left[\delta_{1,Y} \, \delta_{2,Y} \, \cdots \, \delta_{M,Y} \right].$$

We see therefore that $C \in \mathbb{R}^{M \times M}$ and that $\delta_Y \in \mathbb{R}^{N \times M}$. Also we assume that the stocks are chosen so that C is invertible (i.e. the M stocks' loadings on the Z-type factors are not colinear). Let us now define a new set of state varuiables, which we will call $z(t)$. These will be M-dimensional, and we will assume that $z(0) = Z(0)$. The z's differ from the Z's because their dynamics are given by

$$dz(t) = \tilde{\mu} dt - \tilde{k}_Z Y(t) dt + \Sigma_Z d\tilde{W}(t), \tag{19}$$

where $\tilde{k}_Z \in \mathbb{R}^{M \times N}$ and is given by

$$\tilde{k}_Z = \tilde{K}_Z + (C')^{-1} \delta'_Y. \tag{20}$$

Mamaysky (2002a) then shows that for any of the above M stocks, the total returns process $s_m(t)$ may be solved for explicitly, and that it is given by

$$s_m(t) = \exp\Big((a_m + \delta_{m,0})\, t - B'_m Y(t) - C'_m z(t) \Big). \tag{21}$$

There are two differences between $s_m(t)$ and the stock price $S_m(t)$. First, the loading of $s_m(t)$ on time is given by $a_m + \delta_{m,0}$, while for the stock price it is given by a_m. Second, the stock loads on the $Z(t)$'s, whereas the total returns process loads on the $z(t)$'s. We now make the following two observations. First, given the form of a_m in (11), we see that $a_m + \delta_{m,0}$ does not depend on the behavior of the dividend yield. Let us define $a_m^{TR} \equiv a_m + \delta_{m,0}$. Second, given the form for B_m in (12) and the definition of $\tilde{k}_Z$ in (20), we see that B_m can be written as

$$B_m = \left(\tilde{K}'_Y\right)^{-1}\left(r_Y - \tilde{k}'_Z C_m\right). \tag{22}$$

Therefore, neither a_m^{TR} nor B_i depends on any of the dividend yield paramters or on the parameters of the $Z(t)$ process directly. Instead, the dependence goes through the $\tilde{k}_Z$ matrix, which can be estimated directly, without needing to know $\tilde{K}_Z$ or δ_Y separately. Also notice that the loadings of the total returns process of stock on the $Z(t)$ is exactly the same as the loadings of the associated total returns process on the $z(t)$'s. With this, we can write the total returns process as

$$s_m(t) = \exp\left(a_m^{TR}\,t - B'_m Y(t) - C'_m z(t)\right). \tag{23}$$

Conveniently, therefore, the total returns process on a stock also has a closed form solution, and is also exponential affine in the model's state varuiables (once a change of variables is performed on the Z-type factors). We note that in the empirical implementation we will be working with total returns processes, and hence will be using equation (23).

We make one additional point. The parameter condition in (17) which guarantees that the transversality condition in (16) holds is stated in terms of the parameters of the stock price process. For stocks $m = 1, \ldots, M$ this condtion can be restated in terms of the parameters of the total returns process as follows

$$\delta_{m,0} + \delta'_{m,Y}\tilde{\Theta} + \sum_{n=1}^{N}\frac{[\delta_{m,Y}]_n\sigma_{Y,n}^2}{[\tilde{K}_Y]_{nn}^2}\left(-\frac{1}{2}[\delta_{m,Y}]_n - r_{Y,n} + [C'_i\tilde{k}_Z]_n\right) > 0. \tag{24}$$

Portfolios Versus Stocks

So far, we have been refering to $S_i(t)$ as the price of a single stock with a dividend process parameterized by $\{\delta_{i,0}, \delta_{i,Y}, C_i\}$. However, in the empirical implementation of the present model we will be working with portfolios of stocks, rather than with stock individually. It is shown in Mamaysky (2002a) that when factor innovations are Gaussian (and only when factor innovations are Gaussian), as is the case here, the total return on a portfolio of stocks is exactly equal to the total return on a single stock, whose dividend specification $\{\delta_{i,0}, \delta_{i,Y}, C_i\}$ is the average of the dividend specifications of each of the stocks in the portfolio. Therefore, in this paper, we are justified in treating the portfolios used in the estimation as if they were individual stocks. The one caveat with this is that the portfolio weights must remain constant in order for the argument in Mamaysky (2002a) to go through. This will likely not be the case for the portfolios used in this paper, but we will proceed with parameterizing a portfolio as a single stock despite this fact. See Mamaysky (2002a) for a derivation of the total returns process for a portfolio of stocks in a more general setting than the one used in this paper.

2.4 The Pricing Kernel

The final step in specifying the model is to write down a pricing kernel, which allows us to change between the physical measure $\mathcal{P}$ and the equivalent martingale measure $\mathcal{Q}$. The development presented in this section is from Mamaysky (2002a), and the interested reader should refer there for more details.

Let us write $m(t)$ for the pricing kernel in the model. Let us write the evolution of the pricing kernel as

$$\frac{dm(t)}{m(t)} = -r(t)dt - \Lambda(t)'dW(t), \tag{25}$$

where $r(t)$ is the short rate process, $W(t)$ is a standard $N+M$-dimensional Brownian motion under $\mathcal{P}$, and where $\Lambda(t)$ is a $N+M$-dimensional process which is called the *price of risk*. This name is justified because of the following result. Let us write the evolution of stock and bonds prices as follows

$$\begin{aligned} \frac{dP_T(t)}{P_T(t)} &= \mu_T(t)dt + \sigma_T(t)'dW(t), \\ \frac{dS_i(t)}{S_i(t)} &= \mu_i(t)dt + \sigma_i(t)'dW(t). \end{aligned}$$

An application of Ito's lemma to bond and stock prices in (4) and (14) shows that $\sigma_T(t)' = -B_T(t)'\Sigma_Y$ and that $\sigma_i(t)' = -B_i'\Sigma_Y - C_i'\Sigma_Z$.[8] Also, from standard results (see, for example, Mamaysky (2002a)) we have that

$$\mu_T(t) - r(t) = \Lambda(t)'\sigma_T(t), \tag{26}$$

and that

$$\mu_i(t) + \delta_i(t) - r(t) = \Lambda(t)'\sigma_i(t). \tag{27}$$

Therefore the time-varying vector $\Lambda(t)$ determines how much excess return each security is entitled to by virtue of its loadings on each of the sources of uncertainty in the economy (i.e. the $N+M$-dimensional vector of Brownian motions $W(t)$). Thus the name: price of risk.

It is important to note that one of the main advantages of the present model is that it gives us risk premia processes for both stocks and bonds. Therefore we can use data from both bond and stock markets to study the behavior of risk premia in the economy. Furthermore we have a unified framework for seeing how prices of risk translate into risk premia across different asset classes. For example, given the present specification, we have $N+M$ price of risk processes in the economy (one for each Brownian motion). Bonds can load only on the first N prices of risk (i.e. those associated with the Y-type factors), whereas stocks can load on all $N+M$ prices of risk. Therefore, an empirical estimation of the model should reveal which prices of risk are joint between stock and bond markets, which prices of risk are specific only to stocks, and how the behaviors of these prices of risk differ from one another. This analysis is pursued in Section 4.

[8] Here, Σ_Y is a $N \times (N+M)$ matrix with element $[\Sigma_Z]_{nn}$ equal to $\sigma_{Y,n}$ for $n = 1, ..., N$ and where all other elements are zero.

Specification of the Price of Risk Process

It remains for us to specify the behavior of the price of risk process $\Lambda(t)$. It is shown in Mamaysky (2002a) that the "essentially affine" price of risk process proposed in Duffee (2001) (see also Dai and Singleton (2001)) works as well in the context of a joint bond-stock pricing model. We will assume that the $N+M$-dimensinoal process $\Lambda(t)$ is given by

$$\Lambda(t) \equiv \lambda_0 + \lambda_Y Y(t), \tag{28}$$

where $\lambda_0 \in \mathbb{R}^{N+M}$ and $\lambda_Y \in \mathbb{R}^{N+M} \times N$. That this is a valid price of risk process in the context of a model having factors with Gaussian innovations is shown in Dai and Singleton (2001). It is shown in Mamaysky (2002a) that with this price of risk process, factor dynamics under $\mathcal{P}$ may be written as follows:[9]

$$dY(t) = K_Y\Big(\Theta - Y(t)\Big)dt + \Sigma_Y dW(t), \tag{29}$$

$$dz(t) = \mu dt - k_Z Y(t)dt + \Sigma_Z dW(t), \tag{30}$$

where $K_Y \in \mathbb{R}^{N\times N}$, $\Theta \in \mathbb{R}^N$, $\mu \in \mathbb{R}^M$, $k_Z \in \mathbb{R}^{M\times N}$, where Σ_Y and Σ_Z have already been defined, and where

$$\begin{aligned} K_Y &= \tilde{K}_Y - \Sigma_Y \lambda_Y, \\ \Theta &= K_Y^{-1}\Big(\tilde{K}_Y\tilde{\Theta} + \Sigma_Y\lambda_0\Big), \\ k_Z &= \tilde{k}_Z - \Sigma_Z\lambda_Y, \\ \mu &= \tilde{\mu} + \Sigma_Z\lambda_0. \end{aligned}$$

Given factor dynamics under both measures, it is straightforward to check that λ_0 and λ_Y satisfy the following:

$$\lambda_0 = \begin{bmatrix} \Sigma_Y \\ \Sigma_Z \end{bmatrix}^{-1} \begin{pmatrix} K_Y\Theta - \tilde{K}_Y\tilde{\Theta} \\ \mu - \tilde{\mu} \end{pmatrix}, \tag{31}$$

$$\lambda_Y = \begin{bmatrix} \Sigma_Y \\ \Sigma_Z \end{bmatrix}^{-1} \begin{pmatrix} \tilde{K}_Y - K_Y \\ \tilde{k}_Z - k_Z \end{pmatrix}. \tag{32}$$

Therefore estimation of the model (discussed in Section 3 below) will allow us to compute the price of risk process in (28). With this we will be able to analyze the commonalities and differences in the behaviors of the risk premia of across stock and bond markets in a unified, theoretically coherent framework.

So far we have discussed how to obtain the price of risk process using the parameters of the total returns process under both measures. We now note that the $\Lambda(t)$ process implied

[9] This follows from the fact that the $N+M$-dimensional $\mathcal{Q}$ Brownian motion $\tilde{W}(t)$ has the following decomposition $\tilde{W}(t) = \int_0^t \Lambda(u)du + W(t)$, where $W(t)$ is our standard $N+M$-dimensional Brownian motion under $\mathcal{P}$.

by (31) and (32) is the same one that we would have obtained had we used stock price and dividend data, rather than total returns data. To see this observe that changing measures from $\mathcal{Q}$ to $\mathcal{P}$ yields the following relationship between $\tilde{K}_Z$ and K_Z

$$K_Z = \tilde{K}_Z - \Sigma_Z \Lambda_Y.$$

Hence $\tilde{K}_Z - K_Z = \tilde{k}_Z - k_Z$, and the estimate of λ_Y in (32) does not change regardless of our choice of stock price versus total returns data. Also note that this data choice does not affect $\sigma_i(t)$. Hence the equity risk premium $\mu_i(t) + \delta_i(t) - r(t)$ in (27) can be recovered using only total returns data.

2.5 Forecasting Returns

In equations (26) and (27), we derived instantaneous risk premia on securities in the model. It is also relatively straightforward to compute expected conditional returns over longer time horizons. From (23) we see that log total returns on stocks are affine functions of the state variables. Hence we can express the continuously compounded return on a stock over a given time horizon τ as follows

$$R_j(t,\tau) \equiv \log \frac{s_j(t+\tau)}{s_j(t)} = a_j^{TR} \times \tau - B_j'\Big(Y(t+\tau) - Y(t)\Big) - C_j'\Big(z(t+\tau) - z(t)\Big). \quad (33)$$

We can therefore write

$$\mathbb{E}_t[R_j(t,\tau)] = a_j^{TR} \times \tau - B_j'\Big(\mathbb{E}_t[Y(t+\tau)] - Y(t)\Big) - C_j'\Big(\mathbb{E}_t[z(t+\tau)] - z(t)\Big). \quad (34)$$

These conditional expectations are rather straightforward to compute. This is done in Proposition D.2 in the Appendix. Note that since we are interested here in making statements about the predictability of returns, all moments are being computed under $\mathcal{P}$, the physical measure.

Another quantity of interest is the log return from rolling over investments at the short rate over a given time period. This quantity is given by

$$R_r(t,\tau) \equiv \int_t^{t+\tau} r(u)du. \quad (35)$$

Under the factor dynamics assumed in this paper, it is possible to compute $\mathbb{E}_t[R_r(t,\tau)]$ in closed form. This is done in Proposition D.1 in the Appendix.

Given equations (33) and (34), we see that it is possible to write the log return of an investment in any stocks, or of a rolled over investment growing at the short rate, as follows

$$R_j(t,\tau) = \mathbb{E}_t[R_j(t,\tau)] + \epsilon_j(t+\tau),$$

where the error term is, by definition, orthogonal to the conditional expectation.

This decomposition allows us to compute, in closed form, the R^2 of a forecasting regression of the actual return on its expectation, as a function of the forecasting time horizon τ. This quantity is given by

$$R^2(\tau) \equiv \frac{\mathrm{Var}(\mathbb{E}_t[R_j(t,\tau)])}{\mathrm{Var}(R_j(t,\tau))} = \frac{\mathrm{Var}(\mathbb{E}_t[R_j(t,\tau)])}{\mathrm{Var}(\mathbb{E}_t[R_j(t,\tau)]) + \mathrm{Var}(\epsilon_j(t+\tau))}, \tag{36}$$

since the error term is orthogonal to the conditional expectation. Also note that we are computing unconditional variances. This calculation is possible because from (34), and the factor dynamics, we see that $\mathbb{E}_t[R(t,\tau)]$ is stationary, and hence has a long-run variance. In a specialization of the model, we will later see that this R^2 can be computed in closed form. The R^2's for forecasting regressions of stock returns, stock returns in excess of the short rate, and of rolled over investments at the short rate are given in Propositions D.1 and D.2 in the Appendix.

This analysis is extremely useful for the following reason: The conditional expectation used in the forecasting regression above is derived within the framework of the model. Hence any strategy for estimating the present model will yield an estimate for the predictive power of the true conditional expectation of a given return, were such an expectation actually known. This exercise is by construction free of any data snooping bias (though not free from problems of model misspecification). On the other hand, any purely empirical attempt to calculate predictive power must necessarily rely on actual forecasting variables. Since the forecasting variables must be found empirically, it is always possible to data mine and find "overly predictive" variables. To the extent that the present model provides an adequate description of the data, the measure of R^2 derived above is likely to be more accurate (and lower) than a purely empirical measure.

3 Model Estimation

This section provides details of the estimation of the model set out in the previous section.

3.1 Data

All data are in nominal terms and at a monthly frequency. The time period of the sample is from February 1973 through December 2000. The sample therefore contains 335 months of data. This choice of time period deserves some justification. The data used for this study actually goes back to June 1952 (some goes further back than that). However, the present model, when estimated from June 1952 to December 2000, does an extremely poor job of pricing long term government bonds in the time period prior to 1973. Furthermore, the correlation between the dividend yield and the short rate switches signs around the early seventies. For example, in the time period from June 1952 to January 1973, the correlation between the average annual dividend yield of the CRSP value weighted index and the one year governemtn spot rate (from the Fama-Bliss zeros file, see below) is -0.7743. In the time period from February 1973 to June 2000, this correlation is 0.6040. Since a critical component of the present model is the relationship between dividend yields and interest rates, it is important to estimate the model over a time period during which this relationship does not undergo regime shifts, as apparently hapenned sometime around the early seventies. For these two reasons, the time period for this study does not include data prior to 1973.

Data on government bonds is obtained from two sources in CRSP. First, we use the CRSP fixed term indexes file. This file provides monthly data on the seven government bonds whose maturities are closest to 1, 2, 5, 7, 10, 20, and 30 years. The data runs from 1952 to 2000. In some months, there may not be information on all seven bonds in the sample (for example, the bond with maturity closest to 20 years may also be the bond with maturity closest to 30 years). Also we use the CRSP Fama-Bliss zero coupon bond prices. This series runs from 1952-2000, and gives in every month the prices for zeros of maturities equal to 1, 2, 3, 4, and 5 years. Details on how these zero prices are extracted from coupon bond prices are available from CRSP.

Stock data is taken from CRSP and from Ken French's website

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

at Dartmouth. From CRSP we use the value-weighted market index from 1952–2000. From French's website we have data on five sets of decile portfolios, where the sorts are based on the following equity characteristics: size, book to market, dividend to price ratio, earnings to price, and cashflow to price. Descriptions of how these portfolios are formed are available on the above website. We use monthly return data on the 3^{rd} decile size portfolio (small stock) and on the 8^{th} decile book to market (high book to market, or value) portfolio.[10] All stock data are total returns.

3.2 Model Specification

We will assume that there are two Y-type factors in the economy. For specifying the bond portion of the model we follow the estimation strategy of Chen and Scott (1993) who assume that certain bond prices are exactly in agreement with the model, and that certain other ones are observed with error. The advantage of this strategy (see also the discussion in Duffie and Singleton (1997)) is its simplicity and the fact that the assumption that certain bonds are priced exactly according to the model renders the Y-type factors directly observable.

We assume that the two and five year zeros from the Fama-Bliss file are priced exactly by the model, which, together with equation (4), implies that

$$\log \begin{bmatrix} \text{2 Year Zero}(t) \\ \text{5 Year Zero}(t) \end{bmatrix} = \begin{bmatrix} A_{t+2}(t) \\ A_{t+5}(t) \end{bmatrix} - \begin{bmatrix} B_{t+2}(t)' \\ B_{t+5}(t)' \end{bmatrix} Y(t). \tag{37}$$

[10] We do not use the extreme size or value portfolios (i.e. the first and tenth deciles respectively) in order to exclude the small, distressed, and/or thinly traded companies which these portfolios contain.

For the bond prices which are obeved with error, we use the CRSP fixed term index bonds closest in maturity to ten and twenty years, as well as the one, three, and four year zeros from the Fama-Bliss files. The i^{th} error $\epsilon_i(t)$ is given by

$$\epsilon_i(t) = \text{Price of Bond}(t) - \text{Model Price for Bond}(t).$$

We will write $\epsilon(t)$ for the vector of the stacked pricing errors. We assume that the pricing errors associated with these bonds are independently (cross-setionally and over time) and Normally distributed in each month of the sample, with means of zero and standard deviations given by $\sigma_{\epsilon,1}, \ldots, \sigma_{\epsilon,5}$ respectively.

We assume that there are three Z-type factors in the economy. Among the stock portfolios, we assume that the CRSP value weighted index, the 3^{rd} size decile portfolio, and the 8^{th} book to market decile portfolio are all priced exactly by the model. From equation (23), this means that

$$\log \begin{bmatrix} \text{CRSP VW Index}(t) \\ 3^{rd}\text{Decile Size}(t) \\ 8^{th}\text{Decile Book-to-Market}(t) \end{bmatrix} = \begin{bmatrix} a_1^{TR} \\ a_2^{TR} \\ a_3^{TR} \end{bmatrix} t - \begin{bmatrix} B_1' \\ B_2' \\ B_3' \end{bmatrix} Y(t) - \begin{bmatrix} 1 & 0 & 0 \\ c_1 & 1 & 0 \\ c_2 & c_3 & 1 \end{bmatrix} Z(t). \tag{38}$$

Notice that the left hand side in the above equation is a vector of the total returns processes associated with the three equity portfolios. Recall that a_i^{TR} and B_i are not free parameters, but must satisfy the pricing equations of Section 2. The loadings of the total returns processes on the Z-type factors are left unrestricted by the model (and arise from the fact the associated dividend processes of these portfolios may depend on the Z-type factors). The exact nature of this dependence, therefore, must be specified by the econometrician. We choose the loadings in equation (38) because this is the most restrictive specification of the loadings of total returns processes on the Z-type factors which allows for all innovations in the Z's to be independent (which has been the maintained assumption in the development of the model in Section 2).

With the loadings on the Z's given in (38), the role that each of the Z-type factors plays in the economy is easily delineated. For

$$Z(t) = \begin{bmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \end{bmatrix},$$

we see that $Z_1(t)$ is the portion of total returns of the CRSP value weighted index which is not due to its dependence on the Y-type factors. Hence we will refer to $Z_1(t)$ as the *market* factor. $Z_2(t)$ is the portion of total returns on the small stock portfolio which is not due to the Y-type factors or to the market factor. We will refer to $Z_2(t)$ as the *size* factor. Finally, $Z_3(t)$ is the portion of the total returns on the value (high book to market) portfolio which is not due to the Y-type factors, or to either of the market or the small stock Z-type factors. We will therefore refer to $Z_3(t)$ as the *value* factor.

3.3 Estimation of the Model

The present model is estimated in a single step. One estimation strategy would have been to first estimate the parameters of the term-structure portion of the model, and then to use these parameter estimates and the extracted Y-type factors as given in estimation of the stock portion of the model. This approach, while computationally far less intensive than the one adopted in this paper, has the obvious drawbacks (1) of not using stock return data in the estimation of the Y-type factors, which drive prices of risk in the model, and (2) of leading to incorrect standard errors in the second step of the estimation (i.e. the one for stocks). For these reasons, the model is estimated in one step, which results in an extremely large number of parameters (see Table 1) that need to be estimated.

Because we are using a Chen and Scott (1993) type approach, the five factors in the economy

$$X(t) \equiv \{Y_1(t), Y_2(t), Z_1(t), Z_2(t), Z_3(t)\}$$

are all directly observable from the prices of the two and five year zero, and from the total returns of the CRSP value weighted index, the size portfolio, and the value portfolio. Because all factor innovations are assumed to be Gaussian, the likelihood function for factor transitions is known in closed form. Let us refer to the density function for $X(s)$ conditional on $X(t)$ for $t < s$ as $f(X(s)|X(t))$. This density function is given in Section B of the Appendix.

Also let us write $h(\cdot)$ for the inverse of the price function for the above five securities (i.e. for a price vector $P(t)$ we have that $X(t) = h(P(t))$. $h(\cdot)$ is obtained by inverting the system of linear equations in (37) and (38), which then tells us what factors at time t generate the bond prices and total returns process values which we observe at that time. Then standard results (see, for example, Billingsley (1995)) allow us to conclude that the conditional density function for the $P(t)$ vector is given by

$$f_P\Big(P(s)|P(t)\Big) = f\Big(h(P(s))|h(P(t))\Big) \left|\det \frac{\partial h(P(s))}{\partial P'}\right|, \tag{39}$$

where the $|\cdot|$ term is the absolute value of the determinant of the $(N+M) \times (N+M)$ Jacobian matrix $\partial h/\partial P'$. Also let us write $f_\epsilon(\cdot)$ for the density of the pricing errors. Then the log likelihood for the sample is given by

$$\mathcal{L}(\Theta) = \sum_{t=2}^{T} \log f_P(P(t)|P(t-1)) + \sum_{t=2}^{T} \log f_\epsilon(\epsilon(t)), \tag{40}$$

where Θ is the parameters vector. Parameter estimates of the model are obtained by maximizing $\mathcal{L}$ with respect to the parameters of the model. The asymptotic covariance matrix of the parameter estimates $\hat{\Theta}$ is given by

$$\left(\sum_{t=2}^{T} \frac{\partial\phi(\hat{\Theta},t)}{\partial\hat{\Theta}} \times \frac{\partial\phi(\hat{\Theta},t)}{\partial\hat{\Theta}'}\right)^{-1}, \tag{41}$$

where $\phi(\hat{\Theta}, t) \equiv \log f_P(P(t)|P(t-1)) + \log f_\epsilon(\epsilon(t))$. Standard errors for functions of the estimated parameters are obtained using the δ-method.[11]

In order to guarantee proper estimation, we need to impose several sets of restrictions on the parameters of the model (these are in addition to the restrictions which have already been imposed in the prior discussion). First of all, as a normalization, we assume that r_Y is given by a vector of ones. Next, we assume that $\tilde{\Theta} = 0$, and estimate a value for r_0. From (7), we see that both $\tilde{\Theta}$ and r_0 determine the long-yield, and are therefore important for pricing bonds of long maturities. When the model is estimated with $r_0 = 0$ and with $\tilde{\Theta}$ free, the estimates of the separate components of the $\tilde{\Theta}$ vector are very imprecise, and the explanatory power of the model (measured by the log likelihood function) does not change (to the ninth decimal place). For this reason, $\tilde{\Theta}$ is set to zero, and r_0 is left free in the estimation. Further, we assume that $\tilde{\mu} = 0$. We do this because from the total returns process formula in (23) and from the expression for a_i in (11), we see that the $\tilde{\mu}$ vector and the drifts of the Z-type factors under the physical measure (i.e. μ) are not separately observable. Finally, we have assumed that $\tilde{K}_Y$ is a diagonal $N \times N$ matrix, and we will also assume that K_Y is a diaginal $N \times N$ matrix.

This last restriction (i.e. that K_Y is a diagonal matrix) warrants some discussion. While this restriction is not necessary for purposes of identification, it does significantly simplify many of the other calculations in the model.[12] For this reason, the K_Y restriction is very useful. On the other hand, Dai and Singleton (2001) show that in order to match certain empirical properties of bond yields (having to do with the expectations hypothesis, and with bond risk premia), it is important to allow for the K_Y matrix to have nonzero off-diagonal elements. I discuss the problems which the K_Y restriction introduces (such as potentially counter-factual implications for the behavior of the Y-factor prices of risk and of bond risk premia) in greater detail in Section 4.2.1. However, the focus of the present paper is on the empirical behavior of the prices of risk for the Z-type factors, and of the predictability implied by the model for stock returns. I argue in Section 4.2.1 that the fact that the model is able to match one important aspect of the expectations hypothesis (in a sense made precise later) and the fact that the $\tilde{k}_Z$ and k_Z matrixes are unrestricted in the present setup together allow for the model to properly account for these stock-related quantities (i.e. the Z-factor prices of risk and stock return predictability), despite the K_Y restriction. The K_Y restriction can be relaxed, but at the cost of a loss of tractability, and without much benefit for the analysis at hand, and therefore this restriction will be maintained for the remainder of the

[11]If $\sqrt{N}(\hat{\Theta} - \Theta)$ is asymptotically Normal with mean zero and variance matrix V, then $\sqrt{N}(g(\hat{\Theta}) - g(\Theta))$ converges in distribution to a mean zero Normal random variable with variance $\partial g / \partial \hat{\Theta}' \times V \times \partial g / \partial \hat{\Theta}$, where $g(\cdot)$ can be a vector valued function.

[12]Without this restriction, the likelihood function for the data is not known in closed form. Furthermore, I am already estimating such a large number of parameters that the addition of any more would make the estimation problem that much more difficult, a problem which would be compounded in the case of nonzero off-diagonal elements of the K_Y matrix by the fact that the likelihood function would need to be approximated. Perhaps more importantly, the R^2's of the forecasting regressions of Section 2.5 may not have closed form solutions.

paper.

4 Results

4.1 Estimation of the Model

Table 1 shows the results of model estimation. The value of the maximized likelihood function $\mathcal{L}(\hat{\Theta})$ in (40) is shown. Also, standard errors, computed using the outer product method in (41), are given in parentheses.[13] In all, 35 parameters are estimated using 335 monthly observations of U.S. government bond and U.S. stock data. Many of the parameter estimates have fairly small standard errors. The notable exception to this are the estimates of Θ, k_Z, and μ. Because these parameters jointly determine the behavior of the drifts of the z's under the physical measure (recall from equation (30) that the drift of the z's is given by $\mu - k_Z Y(t)$ and note that under $\mathcal{P}$ the long-run means of the Y's are given by the vector Θ), they are, unfortunately, poorly estimated. This is made clear by the large standard errors associated with Θ, k_Z, and μ in Table 1. Using a longer time series would produce tighter estimates of these parameters. However, for the reasons given in Section 3.1, a suitable longer time series may not be available. Hence, without imposing other restrictions on the parameters of the model, it is not clear how the precision of these parameter estimates may be improved.

Table 2 shows the correlation of first differences of the Y and z factors extracted from the model. We see that, with the exception of the $Y_1 - Y_2$ correlation, all of these are close to zero. This latter correlation is negative, which, as Dai and Singleton (2000) point out, seems to be a feature of U.S. term structure data. The problem, of course, is that having a diagonal K_Y matrix and independent factor innovations through the Σ_Y matrix, seems to be contradicted by the data. As has already been pointed out in Section 3.3, the K_Y assumption greatly simplifies much of the analysis, as does the assumption of independent innovations through the Σ_Y matrix. Furthermore, as will become evident later, these restrictions are unlikely to affect the focus of the analysis in the paper, namely the behavior of prices of risk for stock specific factors and of stock return predictability. Nonetheless, a specification of the present model which captures the negative $Y_1 - Y_2$ correlation present in the data is an important step for future research.

Figure 1 shows the time series behavior of the factors extracted from the model (i.e. the Y's and the z's), as well as the behavior of the short rate (given by $r_0 + Y_1(t) + Y_2(t)$).

4.1.1 Behavior of the Short Rate

Table 3 shows the relative importance of the model factors for the time series behavior of the short rate extracted from the model. The short rate in the model is given by $r(t) = r_0 + Y_1(t) + Y_2(t)$. To assess the relative contribution of each factor to the behavior of the

[13]Computing the asymptotic covariance matrix using $\left(-\frac{\partial^2 \mathcal{L}(\hat{\Theta})}{\partial \hat{\Theta}\, \partial \hat{\Theta}'}\right)^{-1}$ produces similar answers.

short rate, Table 3 regresses monthly first differences in the short rate on those of each of the factors separately, or $\Delta r(t) = \text{constant} + \Delta Y_i(t) + \text{noise}$ for $i = 1, 2$. Because, as seen from Table 2, the Y innovations are correlated, the R^2's from these regressions will not add to one. It seems that both factors are important in the evolution of the short rate, with first differences in factor one seemingly more important. Recall from (28), that the Y factors drive the behavior of prices of risk for all factors (bond and stock ones) in the model. Hence the fact that both factors account for a substantial amount of variation in the short term interest rate suggests that information from bond markets is important for the determination of prices of risk for stocks and bonds.

From Figure 1, we see that the model short rate $r(t)$ has the expected procyclical behavior which reflects Fed easing during business cycle troughs. The short rate is particularly high during the early eighties, reflecting the high interest rate environment which was in effect at the time.

4.1.2 Behavior of the CRSP Value Weighted Index

Next, Table 3 shows the relative importance of the Y factors and the z factors for the behavior of the CRSP value weighted index. From (38), we see that monthly first differences of log monthly returns of the CRSP value weighted index, $\log VW(t)$, are given by $\Delta \log VW(t) = \text{constant} - B_1 \Delta Y_1(t) - B_2 \Delta Y_2(t) - \Delta z_1(t)$. To determine the relative importance of the Y factors for the evolution of the log stock returns, the table shows the results of the regression $\Delta \log VW(t) = \text{constant} + b_1 \Delta Y_1(t) + b_2 \Delta Y_2(t) + \text{noise}$. As can be seen, the Y factors jointly account for 5% (or so) of the time series variation in continuously compounded monthly returns for the CRSP value weighted index. Because, as seen in Table 2, the Y and z factor innovations are almost uncorrelated, the remaining 95% of explanatory power is found in the regression $\Delta \log VW(t) = \text{constant} + c\Delta z_1(t) + \text{noise}$.

These results should be interpreted with some caution. In particular, these results do not suggest that the behavior of the Y factors is unimportant for the behavior of stock returns. Indeed, as seen from equations (28), (26), and (27), the Y factors exclusively drive variation in expected excess returns for stocks and bonds. Hence, the dependence of stock returns on Y factors innovations reflects, at least in part, the dependence of stock returns on expected stock returns. This dependence is analyzed at some length in Section 4.4 of the paper. An additional channel of dependence of stock returns on the Y factors comes from the fact that stock dividends (and dividend yields) depend on the Y factors (as can be seen from equation (10)). Indeed, Section 4.2.2 shows that much (though not all) of time series variation in dividend yields of the CRSP value weighted index can be explained by variation in the model extracted Y factors. Hence the results in Table 3 should be interpreted to mean that over a one month horizon, Y factor innovations explain only a small part of market returns. However, because risk premia (entirely) and dividends (partially) both depend on the behavior of the Y factors, the importance of these for stock returns must manifest itself over longer time horizons.

4.2 Diagnostics

Before proceeding with the analysis of prices of risk, risk premia, and return predictability, we need to establish that the model, in its present implementation, does a reasonable job of accounting for the empirical properties of the relevant data. We will look at this question in three ways: First, we will see the degree to which the present model is able to account for expected returns on bonds; Second, we will see the degree to which the present model accounts for the time series behavior of the actual dividends of the CRSP value weighted index; And finally, we will see the degree to which the extracted stock factors (i.e. the z's) actually proxy for differences in stock risk exposures in the cross-section (thus justifying the naming convention in Section 3.2 of calling the z's the market, size, and value factors respectively).

4.2.1 The Expectations Hypothesis

Dai and Singleton (2001) propose using a version of the Campbell and Shiller (1991) regression of changes in yields on the term spread as a means of assessing the degree to which term structure models are able to account to the behavior of risk premia. Their suggested diagnostic can be used as well in the present context. In particular, let us define $y(\tau, t)$ as the time t yield on a zero-coupon bond which matures at time $t + \tau$, i.e.

$$y_\tau(t) \equiv -\frac{\log P_{t+\tau}(t)}{\tau}. \tag{42}$$

Dai and Singleton (2001) propose running the following monthly regression (a modification of Campbell and Shiller (1991)):

$$y_{\tau-\Delta}(t+\Delta) - y_\tau(t) = \text{constant} + \phi_\tau \frac{y_\tau(t) - r(t)}{\tau/\Delta - 1} + \text{noise term.} \tag{43}$$

where τ is the time frame in question, and Δ is equal to a time span of one month, or $1/12$, and $r(t)$ is the short rate. The ϕ_τ coefficient from this regression is given by

$$\phi_\tau = \Big(\tau/\Delta - 1\Big) \frac{\mathrm{Cov}\Big(y_{\tau-\Delta}(t+\Delta) - y_\tau(t), y_\tau(t) - r(t)\Big)}{\mathrm{Var}\Big(y_\tau(t) - r(t)\Big)}. \tag{44}$$

Given the parameters of the model (under the physical measure $\mathcal{P}$), this coefficient can be computed in closed form. This is done in Section C in the Appendix. Dai and Singleton (2001) refer to ϕ_τ above as the LPY coefficient (presumably LPY is an acronym for "linear projection of yields" or something of the sort). We will use their naming convention.

The proposed test goes as follows: Estimation of the model yields factor realizations (the Y's and Z's), from which the bond yields y_τ's and the short rate $r(t)$ (all functions of the Y-factors) can also be computed. The regression in (43) can then be run with the yields and the short rate extracted during the model estimation to produce a slope coefficient ϕ_τ

for each bond maturity τ of interest. Following the convention of Dai and Singleton (2001), we refer to these as the "fitted" coefficient values. A second method for computing the ϕ_τ's is available. Using estimates of the model parameters from Table 1, as well as the factor dynamics in (29), we can compute all of the moments in expression (44) directly, thus obtaining the value for ϕ_τ which the imposed factor dynamics and the estimated parameter values imply (see Section C of the Appendix). Dai and Singleton refer to these slope estimates as the "population" values, and we will follow their convention. Finally, the regression in (43) may be run without using a term structure model at all, using only zero-coupon bond prices directly inferred from actual bond price data. Let us refer to these as the "sample" slope estimates.[14]

The idea behind the Dai and Singleton (2001) diagnostic is to assume that the sample estimates are the correct ones, and then to see how close the fitted and the population estimates come to matching the sample ones. If the fitted estimates are close to the sample ones, then the term structure model must be doing a good job of accounting for the behavior of the observed bond prices (i.e. the term structure model produces yields which are close to the actual zero yields found from price data). However, closeness of the fitted slope values to the sample ones does not guarantee that model implied prices of risk (in equation (28), or, equivalently, model implied risk premia (in equation (26) for bonds), provide good proxies for actual prices of risk, or risk premia. The reason for this is that the prices of risk implied by the model depend heavily of the conjectured factor dynamics (in 29) and on the parameter estimates (as can be seen from equations (31) and (32)). Indeed, if the population estimates of the ϕ_τ's (in equation (54)) do not provide a good match for the sample ϕ_τ's, then the model is unlikely to have a correct specification for factor dynamics, and hence is unlikely to provide reliable estimates for either prices of risk or for expected returns.

Table 4 and Figure 3 provide the results of the Dai and Singleton (2001) diagnostic test for the present model. It turns out that the shape of the fitted ϕ_τ's produced by this model is quite close to the shape of the fitted and sample ϕ_τ curves (i.e. the slope coefficients as a function of τ) reported in the Dai and Singleton paper.[15] The closeness of the fitted curve from this paper to the sample one reported in Dai and Singleton (2001) indicates that the present model is doing a good job of capturing the empirical behavior of bond price data. Corroborating evidence in this regard is seen in Figure 2, which plots a time series of pricing errors of the model for the CRSP fixed term index bonds (keep in mind that of these, the bonds closest in maturity to ten and twenty years were used as pricing errors in the estimation, and the remaining bonds were not used in the estimation at all). As can be seen from the figure, the model does a good job of accounting for the time series behavior of

[14]This approach is not exactly model free either (though much more so than the term structure approach), because finding zero coupon bond prices from prices of traded bonds typically involves assuming some sort of functional form for how zero prices depend on time.

[15]The slight difference between the fitted values in the two papers are likely due to the slightly different sample periods in this paper and in the Dai and Singleton (2001) paper, and to the fact that bond data for the two papers are obtained from different sources. When I use data only upto the end of 1995 (which is the end of the Dai and Singleton (2001) sample), the match between the two fitted series is indeed quite close.

all seven fixed term index bonds.[16] For these reasons, in the ensuing analysis, let us simply assume that the fitted ϕ_τ's provide a reasonable baseline against which the population ϕ_τ's should be measured.

Figure 3 also shows a plot of the population slope coefficients from the regression in (43). As can be seen from the figure, the population slope curve lies outside of the 95% confidence interval around the fitted slope curve. This suggests that the term structure portion of the present model is not likely to be able to properly account for time series variation in expected returns on government bonds. The problem lies in the fact that the prices of risk for the two Y-type factors do not have enough flexibility in terms of their loadings on the Y-type factors. Recall from equation (28) that all prices of risk are driven by the Y-type factors. The prices of risk responsible for driving bond expected returns are those associated with the Y-type factors. Because the K_Y and $\tilde{K}_Y$ matrixes are forced to be diagonal (as is the stacked matrix of the Σ_Y and Σ_Z), we see from (32), that the price of risk for factor Y_n must be driven only by Y_n (rather than by some linear combination of all the Y's). Dai and Singleton (2001) point out that it is exactly this type of restriction (i.e. that it be diagonal) on the K_Y matrix which prevents Gaussian models of the sort used in this paper from properly capturing the time series variation in expected returns on bonds.

It was pointed out in Section 3.3 that the above restrictions on the Y-factor dynamics greatly simplify the analysis performed in the paper. However, this restriction has now been shown to make the model produces expected return information *for bonds* that is inaccurate. Since the focus of the paper is on expected returns for stocks, it is hoped that this inability to properly account for time variation in bond expected returns does not adversely affect the model's ability to properly account for time variation in stock expected returns. Fortunately, it is quite likely that is indeed the case!

To see that this is so, let us note that prices of risk for all factors (Y and Z-type) in the model are driven by the Y-type factors. The Y-type factors have already been shown to do a good job of accounting for the time series behavior of bond prices (though not of expected returns). By virtue of this good fit for prices, the Y-type factors are seen to adequately span the information which the bond market may contain for expected returns of any security. The problem with the expected return analysis for bonds is that the prices of risk for the Y-type factors cannot load on all the Y-factors at once. This prevents the model from properly accounting for variation in expected returns on bonds, whose expected returns are driven exclusively by the prices of risk of the Y-type factors (see equation 26).

However, expected excess returns on stocks are also driven by the prices of risk of the Z-type factors, as can be seen from equation (27). Indeed it will later be shown, in Section 4.3, that the majority of variation in stock expected returns is attributable to prices of risk

[16] A better fit (especially at the long end) could be obtained by moving to a model with three (or four) Y-type factors. This does not seem necessary for the analysis at hand, and would involve yet more parameters to estimate in what is already a difficult estimation problem. Since the focus of the analysis in this paper is not on bond, but on the stock, prices of risk, the extension of the estimation to more Y-type factors is left as a problem for future research.

of the Z-type, and not the Y-type, factors. This suggests that if the model gets the prices of risk for the Z-type factors correct, it should also do a good job of capturing the time series variation in expected excess returns of stocks. Since the $\tilde{k}_Z$ and k_Z matrixes are entirely unconstrained in the model estimation, we see from equation (32) that prices of risk for the Z-type factors can load on all of the Y-type factors, as opposed to the Y-type prices of risk, which could not do so. This freedom affords the model much more flexibility in properly accounting for the time series behavior of the prices of risk for the Z-type factors. Because these prices of risk, rather than the Y ones, account for the majority of variation in stock expected returns, it is quite likely that the present model is able to properly account for expected returns on stocks.

4.2.2 Implied Dividends

Since the model has been estimated using the total returns processes of the equity portfolios, and no dividend information (outside of that contained in the total returns processes), another diagnostic for the model is to see for how much of the variation in actual dividend yields the factors extracted from the model can account. Since dividend yields are supposed to be affine functions of the Y-type state variables, a projection of actual dividend yields onto the Y factors should produce in a "good" fit. If not, then we may conclude that the model has been misspecified.

From the analysis in Section 2.3, we see that the one constraint which needs to be imposed on the projection of actual dividend yields onto the model extracted Y-type factors is that the best-fit dividend series must satisfy the transversality condition in (16). We proceed to compute the best-fit dividend series $\hat{\delta}(t) = \hat{\delta}_0 + \hat{\delta}_Y' Y(t)$ as follows:

$$\{\hat{\delta}_0, \hat{\delta}_Y\} \equiv \arg \inf_{\delta_0, \delta_Y} \sum_{t=1}^{T} (\delta_0 + \delta_Y' Y(t) - \delta(t))^2, \tag{45}$$

subject to the constraint that the transversality condition in (24) holds for $\hat{\delta}_0$ and $\hat{\delta}_Z$. It turns out that the constraint in (24) is not binding for the time period used in this paper, and hence the minimization in (45) produces an OLS regression of the actual dividend yield on the extracted model factors. Of course, this would no longer be true were the transversality constraint to bind. The instantaneous dividend yield $\delta(t)$ is estimated to be the annual dividend of the CRSP value weighted index in the time interval $[t - 0.5, t + 0.5]$.[17]

The proximity of $\hat{\delta}(t)$ to the actual dividend series $\delta(t)$ is a measure of the information content of the Y factors extracted from the model estimation. Since the Y-type factors drive interest rates and dividend yields in the model, an inability of the extracted Y factors to properly account for the actual dividend series implies that the model has been misspecified. Such a misspecification is problematic because the model extracted factors may then not

[17]If we believe that the actual dividend yield fluctuates only slowly, then the realized annual dividend yield may be a reasonably good approximation for the instantaneous dividend yield.

span the time series behaviors of the prices of risk of the model factors, which would result in incorrect estimates of expected returns. To assess the degree of this misspecification, we run the following regression

$$\delta(t) = \alpha + \beta\hat{\delta}(t) + \epsilon(t). \tag{46}$$

Since the transversality constraint is not binding in the solution of (45), then by construction the α and β coefficients from this regression are 0 and 1 respectively. In other words, a non-binding transversality constraint in the projection in (45) implies that the best fit dividend series $\hat{\delta}(t)$ is unbiased. In this case, the R^2 of the above regression will tell us how much of the actual dividend variation is being captured by the model extracted Y factors. If the constraint in (24) does bind in solving (45), the best fit dividend series will be biased, and then the α and β coefficients in the above regression will also provide information about the degree of model misspecification.

Figure 4 shows the actual dividend series and the best fit series $\hat{\delta}$ over the entire sample period. Table 5 shows the parameter estimates from the above equations (45) and (46). As can be seen, the best fit dividend series from the model fits the realized dividend reasonably well, though clearly imperfectly. The correlation between the two series is 0.6563 (or the square root of the R^2 reported in Table 5). The transversality constraint is non-binding for the time period in question, implying that the best-fit series in unbiased. The imperfect fit of the two series, however, does suggest that a dividend yield factor is missing from the present analysis. Understanding the role that this missing factor plays in the economy is an important area for future work.[18]

4.2.3 Cross Sectional Performance of the Z-Type Factors

Following the discussion in Section 3.2, we refer to the Z-type factors as the market, size, and value factors respectively. Let us recall at this point that the model has been estimated using only three stock portfolios: the CRSP value weighted index, the 3^{rd} decile size portfolio, and the 8^{th} decile value portfolio. If the factors extracted from model estimation explain a good deal of the heterogeneity in the cross-sectional behavior of stock returns, then knowing the prices of risk associated with these factors actually provides information about the behavior of expected stock returns in the cross-section. To see that this is indeed the case, note from the equation in (27) that expected excess returns in the economy are determined jointly by factor prices of risk, as well as by security loadings (in this case, the σ_i's) on these factors. Hence nonzero loadings (or σ_i's) imply a dependence of expected excess returns on the prices of risk in queastion. If, on the other hand, the present factors do not account for a substantial portion of differences in cross-sectional return behaviors (i.e. the σ_i's with respect to the model factors are close to zero), then knowledge about the prices of risk associated with the model's Z-type factors does not say much about the behavior of expected returns for a

[18]In particular, the extent to which this missing dividend yield factor drives prices of risk in the economy determines the model's ability to properly account for time series variation in expected excess returns and prices of risk.

broad selection of stocks (but only for those stocks or portfolios similar to the ones used in the model estimation).

Another way of looking at this question is to justify the use of the "size" and "value" naming convention for Z_2 and Z_3. For example, is it the case that smaller stocks have a higher loading on the size factor than do larger stocks? Or, is it true that stocks with more valueness have a higher loading on the value factor than do stocks with lower values of valueness?[19] If it is the case that different sized stocks or stocks with different amounts of valueness have very different factor loadings, then the prices of risk associated with the Z-type factors may explain a good portion of cross-sectional variation in equity expected returns.

To assess the degree to which the model factors explain heterogeneity in cross-sectional return behavior, we do the following: Let us group stocks in portfolios by the percentile ranking of a stock within a given stock characteristic. For example, we may look at ten portfolios, where the n^{th} portfolio contains stocks which are (at some rebalancing frequency) in the n^{th} decile based on their market capitalizations. These portfolios represent a size sort. We may similarly look at portfolios sorted based on book to market, dividend to price, earnings to price, and cashflow to price ratios. These latter four sorts (i.e. except size) correspond to sorts along various proxies for value.

To see whether stocks with different characteristics do indeed display different loadings on the extracted model factors, we regress continuously compounded monthly portfolio returns on monthly first differences in factor levels, or

$$\log \frac{s(t+\tau)}{s(t)} = \alpha + \beta_1 \Delta Y_1(t+\tau) + \cdots + \beta_5 \Delta Z_3(t+\tau) + \epsilon(t+\tau). \tag{47}$$

Indeed this regression corresponds exactly to the pricing equation for stock total returns processes given in (23).[20] Each regression produces a set of factor loadings of a given portfolio on the model factors. Figure 5 shows plots of the portfolio loadings on the Z-type factors (i.e. β_3, β_4, and β_5 from (47)) for the size sort and for the for value sorts.

For the size sort, we see a monotonic increase in exposure to the size factor as we move from the smallest to the largest size decile (keep in mind that, by construction, the 3^{rd} size decile portfolio has a loading on the size factor equal to one). For the size sort, loadings on the market and value factors do not change much as we move from smaller to larger size deciles. For the four value sorts, a similar loadings pattern emerges. As we move from lower value to higher value deciles, we see a (nearly) monotonic increase in exposure to the value factor

[19] Valueness refers here to stocks which have a relatively high book to market, dividend to price, earnings to price, or cashflow to price ratio. See Section 3.1.

[20] The ϵ term in the above regression can be though to correspond to first differences of a portfolio specific Z-type factor. This characterization of the residual term suggests one minor econometric issue with the regression in (47): the ϵ's may be correlated with factor innovations, thus resulting in biased coefficient estimates. That this correlation might be nonzero follows from the discussion in Section B. However, if the results in Table 2 are representative, the monthly correlations between the Z-type and Y-type factor innovations are likely to be quite small.

(keep in mind that, by construction, the 8^{th} book to market decile has a loading on the value factor equal to one). As we move from the lowest to the highest value portfolios (for each of the four value sorts), the loadings on the market and size factors change, but (typically) by much less than do the loadings on the value factor. We also note that the R^2's for the time-series regressions in (47) typically run between 80%–100% (the one exception being the highest dividend to price decile portfolio, which has an R^2 of around 65%). Hence, in addition to generating a great deal of cross-sectional variation in loadings, the model factors seems to account for a large portion of the time-series variation in returns for (almost) all portfolios used in the analysis.

The above results therefore suggest that the two model Y-type factors and the three model Z-type factors do indeed account for a good deal of time series variation in returns, as well as for a good deal of cross-sectional variation in stock returns. These results are quite encouraging. They imply that knowledge of the prices of risk associated with the model's Z-type factors is valuable for the determination of expected excess returns for a substantial number of different stocks. Hence the prices of risk extracted in the current estimation may actually provide interesting insights about important macroeconomic determinants of expected returns in the economy.

4.3 Prices of Risk and Expected Excess Returns

Given the discussion in Section 4.2.1, it is likely that the model, as implemented in this paper, is unable to properly account for the time series behavior of expected returns for bonds in the economy (because of the restrictions placed on the K_Y and $\tilde{K}_Y$ matrixes).[21] For this reason, the analysis in this section will focus of the behavior of the prices of risk of the Z-type factors and of expected excess returns for the three equity portfolios used in the model estimation (the value weighted CRSP index, the 3^{rd} decile size, and the 8^{th} book-to-market portfolios.

4.3.1 Prices of Risk

Recall that prices of risk in the model are given by (from equation (28))

$$\Lambda(t) = \lambda_0 + \lambda_Y Y(t),$$

where in the present model implementation $\lambda_0 \in \mathbb{R}^5$ and $\lambda_Y \in \mathbb{R}^{5\times 2}$. Also λ_0 and λ_Y are given by equations (31) and (32) respectively. In the current model, prices of risk turn out to be negative. This is simply an normalization, and follows from the fact that log total returns on stocks were assumed to load negatively on the pricing factors, as in equation (21). Since excess returns, from (27), are given by $\mu_i(t) + \delta_i(t) - r(t) = \Lambda(t)'\sigma_i(t)$, and since

[21]For example, from Figure 7 we see that expected excess returns on government bonds (1, 10, and 30 year) are all procyclical. This runs counter to prior findings in the literature, such as in Fama and French (1989) and Dai and Singleton (2001), who estimate a richer version of the term-structure portion of the model in the present paper.

$\sigma_i(t)' = -B_i'\Sigma_Y - C_i'\Sigma_Z$ is negative and constant, positive excess returns imply negative prices of risk. Indeed, all estimated σ's and all estimated Λ's are negative in the present model. To make the subsequent discussions more intuitive, we renormalize these so that all prices of risk and all factor loadings (σ's) are positive. Also note that the above expected excess return relationship implies that a unit increase in a security's loading on a risk factor m raises the expected excess return of that security by $\Lambda_m(t)$.

Figure 6 plots the prices of risk associated with the model factors. Also, a model implied term spread is plotted alongside the prices of risk to provide a benchmark for comparison (the term spread, as has been well documented in past work, is countercyclical). From Figure 6, we see that the price of risk associated with Z_1 (the market factor) exhibits a countercyclical behavior, tending to be low around NBER business cycle peaks, and high around NBER business cycle troughs. Because, as will be shown shortly, the Z_1 price of risk accounts for the majority of excess returns on the market, this countercyclical behavior is consistent with previous findings in the literature (for example, see Fama and French (1989) for the empirical findings, and Campbell and Cochrane (1999) for a potential theoretical justification). Similarly to that of Z_1, the price of risk for Z_2 (the size factor) exhibits a countercyclical behavior. Both of these are roughly consistent with the idea that the risk aversion of the representative investor becomes higher during difficult economic times, and therefore a higher risk premium is demanded for holding risky assets.

Interestingly, from Figure 6 we see that the price of risk associated with the value factor, Z_3, is (weakly) pro-, rather than counter-, cyclical.[22] This result is either puzzling or intuitive, depending on our interpretation of the value factor. If "valueness" proxies for distressed companies, with low stock prices, and therefore high blank over price ratios (such as book to market, dividend to price, and earnings to price), then why would the price of risk of such companies be low during recessions, when the chances of financial distress for these companies are highest? On the other hand, if "valueness" proxies for healthy firms with good assets on the books and strong earnings and high dividends, the procyclicality of the value price of risk makes more sense, as these companies are the desirable ones to hold during recessions. However, if "valueness" proxies for such healthy firms, then why would its price of risk be positive at all?[23]

Table 6 provides further results about the time series behavior of the Z-type prices of risk. First, we see that the Z_3 (value) price of risk has an almost zero correlation with the Z_1 (market) price of risk, and a strongly negative correlation with the Z_2 (size) price of risk. Table 6 also shows the steady-state value of the prices of risk implied by the model. Since prices of risk are given by $\Lambda(t) = \lambda_0 + \lambda_Y Y(t)$, and since the long-run mean of the Y's (under the physical measure) is given by Θ, the steady-state values of the prices of risk are given by

[22] We note that the Y_1 price of risk is also procyclical, yet, as has already been pointed out, it is unclear whether or not this Y_1 results is reliable.

[23] Recall that excess returns for value companies are determined by their loadings on all the factors, multiplied by the respective prices of risk. Hence the price of risk of the value factor provides excess returns above and beyond a firm's loading on the market factor.

$\Lambda_\infty = \lambda_0 + \lambda_Y \Theta$. These steady-state values are 0.3957, 0.0805, and 0.5291 for Z_1, Z_2, and Z_3 respectively. For a security which loads only on Z_1, and on none of the other factors, the Sharpe ratio would therefore be 0.3957, which is roughly consistent with past estimates of market Sharpe ratios. Table 6 also shows asymptotic standard errors for Λ_∞. As we have already mentioned, those coefficients associated with the drifts of the Z-type factors (i.e. Θ, k_Z and μ) are poorly estimated, which results in rather high standard errors for these price of risk numbers. Indeed, only the Z_3 price of risk appears significantly different from zero. However, it is difficult to believe that the market price of risk is zero, despite the high standard error produced by the model. Finding tighter standard errors for long-run price of risk estimates is an important and interesting area for future research.

Also, of the Z-type prices of risk, the market price of risk has the highest volatility, equal to approximately 0.4. Indeed this price of risk has exhibited a substantial amount of time series variability, assuming a value above 1 in the early 1990's. The size price of risk has also exhibited a substantial amount of variation. Interestingly, the steady state level of the Z_2 price of risk is quite low (lower even that the Y_1 and Y_2 prices of risk). This suggests that, even though the size factor explains substantial cross sectional variation in returns (see Figure 5, for example), the size factor price of risk is economically quite low on average. Of course, as can be seen from Figure 6, the Z_2 price of risk has been above 0.5 on occasion (and been below -0.5 on other occasions). The value price of risk is substantially more stable than either the market or the size one. It time series standard deviation from Table 6 and its extremal points from Figure 6 are both moderate compared to the analogous values associated with the Z_1 and Z_2 prices of risk. Furthermore, the high average value of the Z_3 price of risk points towards the robustness of the price of risk for the value factor: It is both economically high and exhibits extremely slow time series variation.

It should be pointed out that the negative prices of risk for Z_1 and Z_2, which occured in the early 1980's, are justifiably difficult to believe. The early 1980's were a time period characterized by extremely high nominal interest rates (see the short rate plot in Figure 1 or the term spread plot in Figure 6). Perhaps this extremely high interest rate environment represents a regime which is fundamentally different from the other time periods covered in this analysis, and as such leads to price of risk estimates (in the early year of the 1980's) which are inaccurate.

4.3.2 Expected Excess Returns

We have argued in Section 4.2.3 that because the model's Z-type factors do such a good job of accounting for cross-sectional return behavior, their associated prices of risk may provide us with important information for the behavior of asset risk premia. Furthermore, we have argued in Section 4.2.1, that the Y-type price of risk are of second order importance for explaining the beahavior of expected returns on stocks. We examine these points in more detail in this section.

First, we recall that instantaneous expeced excess returns are given in the model by

$\Lambda(t)'\sigma_i(t)$. Also, in the present Gaussian model formulation, all volatilities are constant. The top portion of Figure D shows the values of the σ_i vector for the securities used in the model estimation. For example, we see that the CRSP value weighted market index has σ_1 and σ_2 which of approximately 0.02, whereas its σ_3 is approximately 0.15. [24] We note that, by construction (i.e. from equation (38), the market's σ_4 and σ_5 are both zero, and are not displayed in the graph. The 3^{rd} decile size portfolio also has low values for σ_1 and σ_2, and has $\sigma_3 \approx 0.18$ and $\sigma_4 \approx 0.09$. For the 8^{th} decile value portfolio, we see that $\sigma_1 \approx \sigma_2 \approx 0.025$, that σ_4 is close to zero, that $\sigma_3 \approx 0.125$, and that $\sigma_5 \approx 0.07$. Hence, as expected, the small stock portfolio has a high loading on the size factor, and the high book-to-market portfolio has a high loading on the value factor.

A high value of $\sigma_{i,m}$ (i.e. for stock i and factor m) does not tell us how much of a security's expected return is due to its loading on any specific risk factor. For this we also need to know each factor's price of risk. The long run excess return on a stock i is defined as

$$\Lambda'_\infty \sigma_i.$$

The relative contribution of factor m to this total is given by

$$\frac{[\Lambda'_\infty \sigma_i]_m}{\Lambda'_\infty \sigma_i}.$$

Because all price of risk–volatility pairs are positive, the sum across $m = 1, \ldots, 5$ of these ratios is equal to 1. The labeling convention of factors corresponds to that in Footnote 4.3.2. The lower portion of Figure D shows the relative contribution of each pricing factor to the long-run excess return for each of the security's used in model estimation. For the market, we see that the majority of long-run excess returns comes from the market's loading on the Z_1 factor (labeled with the number 3 in the figure). For the 3^{rd} decile size portfolio, we see that its market loading also accounts for the majority of long-run excess returns. This is despite the fact that σ_4 (i.e. the size factor loading) for this portfolio is very high, and follows from the fact that the price of risk associated with Z_2 is very low. Finally, we see that approximately 55% of the long-run excess return on the 8^{th} decile value portflio comes its loading on the market factor, with about 40% coming from its loading on the value factor.

We note that for the three stock portfolios used in model estimation, the amount of long-run excess return which comes from loadings on the Y-type factors is quite low (less that 5% or so for all three securities). Also, the actual loadings on the Y-type factors (i.e. the σ_1 and σ_2 for stocks) are quite low relative to $\sigma_3, \sigma_4, \sigma_5$. These observations support the earlier claim that it is the behavior of the Z-type prices of risk which determined expected excess returns on stocks, and not the behavior of the Y-type prices of risk. Therefore, because the model is able to properly account for the Z-type prices of risk, though not for the Y-type prices of risk, it should be able to properly account for the behavior of expected excess stock returns.

[24] The notation of Figure D is as follows: σ_1 is the loading on Y_1, σ_2 is the loading on Y_2, σ_3 is the loading on Z_1, σ_4 is the loading on Z_2, and σ_5 is the loading on Z_3. The security speficic i subscript is suppressed.

Finally, it is important to note that the Y-type factors, though not directly responsible for the majority of the behavior of short term security returns, are crucial for the behavior of expected returns in the model because they determine the behaviors of the factor prices of risk!

Figure 7 and Table 7 show the time series behavior of the instantaneous expected excess returns for the securities used in model estimation. We see that the average market risk premium is approximately 6%, that the average 3^{rd} size decile risk premium is 8%, and that the average 8^{th} book-to-market decile risk premium is 9%. Also, all stock risk premia exhibit fairly pronounced countercyclical behavior, being relatively low at NBER business cycle peaks, and relatively high at NBER business cycle troughs. Table 7 shows the correlations, means, and standard deviations of monthly expected excess returns for a selection of zero coupon bonds, for the stocks used in model estimation, and for a model implied term spread and the model implied short rate.

The graphs in Figure 7 run only until the end of the data sample (i.e. December 2000). However, one of the important advantages of the present model is that it produces real time estimates of expected excess returns for stocks and bonds, when provided with current market prices.

4.4 Return Predictability

Because risk premia in the model are driven by the stationary Y-type factors, the time series behavior of risk premia is fundamentally tied to the degree of security returns predictability which is implied by the model. Indeed, in the present context, the question of return predictability is interesting for two reasons: First, it is obviously interesting to understand the degree to which security returns are predictable. Second, comparing model implied predictability to the return predictability observed using factor realizations from model estimation provides a robustness check for the model.

Our examination into return predictability centers around the following decomposition of the total excess returns processes for stock i:

$$R_{xr,i}(t,\tau) = \mathbb{E}_t[R_{xr,i}(t,\tau)] + \epsilon_{xr,i}(t+\tau),$$

where $R_{xr,i}(t,\tau)$ is the total excess return from time t to $t+\tau$ for stock i, $\mathbb{E}_t[R_{xr,i}(t,\tau)]$ is the time t conditional expectation of this return which is derived from the model, and $\epsilon_{xr,i}(t+\tau)$ is a residual return component which is uncorrelated with the conditional expectation. Details of this decomposition are in Section 2.5 of the text and in Section D of the Appendix. Using the notation of Section 2.5, the excess total return $R_{xr,i}(t,\tau)$ is simply $R_i(t,\tau) - R_r(t,\tau)$, or the difference of the total return on stock i, $R_i(t,\tau)$, with the return of a rolled over investment at the short rate $R_r(t,\tau)$. The model implied R^2 for the above forecasting regression is a function of the model's parameters and of the forecasting horizon τ, and is given by equation (36) (and calculated in Section D in the Appendix). The above decomposition can also be performed directly for total returns, i.e. $R_i(t,\tau)$, rather than for excess total returns. Results for both decompositions are presented.

Using this model implied $R^2(\tau)$ value allows us to perform a diagnostic of the model. This is possible because we can compute the R^2 of the above regression in an alternate way. When the model is estimated and factor realizations are extracted, we are able to compute the return conditional expectation $\mathbb{E}_t[R_i(t,\tau)]$ directly for any τ (as is done in Figure 7 for τ very small). Since returns are also known, once the model has been estimated, we can simply regress actual returns on the model implied conditional expectations over any time horizon τ. This yields another estimate for return predictability, which we will call the "fitted" R^2 and which we will label as $\hat{R}^2(\tau)$ to differentiate it from the model implied "population" $R^2(\tau)$ (this follows the naming convention of Section 4.2.1).[25] A comparison of $R^2(\tau)$ and $\hat{R}^2(\tau)$ therefore provides a test of the model's ability to capture the predictability relationships which actually exist in the data.

Table 8 shows the results of the population $R^2(\tau)$ computed using the model parameter estimates from Table 1. Asymptotic standard errors, computed using the δ-method, are reported in parentheses. Results are shown for total returns and for excess returns of the CRSP value weighted index, of the 3^{rd} size decile, and of the 8^{th} value decile portfolios. As can be seen (and as has been shown in Lemma D.1 in the Appendix), all of the population $R^2(\cdot)$'s are hump-shaped functions of time. Systematically, the excess return R^2's are higher than their raw return counterparts. Furthermore, the highest degree of predictability seems to be present for excess returns on the CRSP value weighted index: At a time horizon of 5 years, this maximal R^2 value for the excess return on the market is 13.5%. Unfortunately, all R^2's are very poorly estimated according to the standard errors in Table 8. Given the sensitivity of R^2 of actual regressions to the sample used, this imprecision is not surprising. An interesting area for future work is the determination of reasonable model restrictions which would allow for more precise estimate of the population $R^2(\tau)$'s.

Figure 9 provides a graphical illustration of the results of Table 8, with the population $R^2(\tau)$'s and the model standard error bands plotted for the three equity portfolios used in model estimation, for excess and raw returns, for time horizons, τ's, from zero to ten years. Also shown in Figure 9 are the fitted $\hat{R}^2(\tau)'s$. As can be seen the fitted and population $R^2(\tau)$'s for the excess return regressions are reasonably close to one another. Indeed, for the excess returns regressions, all three fitted series lie inside the 95% confidence interval for the population R^2's. Furthemore, all three fitted excess return $\hat{R}^2$ series exhibit a shape roughly similar to that obtained for the population series. These results are a comforting testament to the model's ability to capture the predictability relationships which are present in the

[25]The R^2's of the actual regressions of excess returns on model conditional expectations are highly sensitive to the sample period used for the regression. To determine representative values of these fitted $\hat{R}^2(\tau)$ values we use the following strategy: For any time horizon $\tau = 1, \ldots, 10$, estimate a separate regression for each 15 year (overlapping) subsample which exists in our sample. This yields a collection of $\hat{R}^2_a(\tau)$'s, for each 15 year subsample a which exists in the full sample. The actual fitted $\hat{R}^2(\tau)$ which is used in the analysis in the paper is the median over the subsamples, the a's, of the $\hat{R}^2_a(\tau)$ values.

data for excess returns.

Unfortunately, the match between the fitted and population R^2's for raw returns is not nearly as good as the match for excess returns. Indeed, for all three securities, we see from Figure 9 that the fitted $\hat{R}^2$'s fall outside of the population 95% confidence interval at medium term forecasting horizons. This inability of the model to match the fitted $\hat{R}^2$ and population R^2 series for raw returns provides an interesting challenge for future research.

5 Conclusion

This paper has estimated a joint bond–stock pricing model. The paper's premise has been that information from both bond and stock markets should be incorporated into the estimation of prices of risk and of expected returns for both bonds and stocks. The paper has argued that while its implications for bond expected returns are not reliable, its implications for the expected returns of stocks are robust. The paper confirms the existence of at least two important pricing factors for stocks: the market factor and a value factor. A size factor seems to be an important determinant of cross-sectional return behavior, but does not carry an economically meaningful price of risk. Finally, the paper has argued that its measures for excess return predictability are trustworthy, whereas its measures for raw return predictability do not appear to be so.

The present analysis suggests many interesting areas for future work:

- It would be useful for the present model to be estimated using the empirical specification for the term-structure portion of the model from Dai and Singleton (2001) which would allow the present model to better capture the expected return behavior of bonds. In this setting, the computation of model implied R^2's is an important goal. Furthermore, perhaps the model's inability to account for the predictability of raw stock returns can be resolved by this enhancement.

- The present model should be modified to account for stochatic volatility present in both stock and bond returns. Commonality of stochastic volatility, and its associated price of risk, across the two asset classes can then be analyzed.

- By virtue of being a joint bond–stock model, the present model allows for useful measures of duration for stocks. For example, one such measure is the maturity of the zero-coupon bond whose loadings on the Y-type factors are closest (in a sum of squares sense) to those of the stock. The usefulness of such a duration measure for interest rate risk management of joint stock–bond portfolios can then be assessed.

- An ever present and important issue is the determination of an economic justification for the presence of multiple prices of risk in equity markets? Also, what other priced risk factors have been missed in the present analysis? That some have been left out is hinted at by the low R^2 of the time series regressions of high dividend to price portfolio returns on the model's five factors (see Section 4.2.3), and also by the inability of the model's Y-type factors to properly account for the historical behavior of market dividend yields (see Section 4.2.2).

- Since the present model is set in a continuous time setting, it allows for the pricing of options in a setting with stochastic interest rates, as well as stochastic volatility. Derivation of option prices in the context of the present model would allow for tests of the implied joint restrictions between bond, stock, and options markets. Such an analysis may point out pricing incosistencies among the three markets. Also note that because the present model does not incorporate default risk, its option pricing implications are for either index options, or for options on high grade credits.

6 Appendix

A Preliminary Results

The following lemma proves useful.

Lemma A.1 *Given a factor Y with dynamics*

$$dY(t) = K(\theta - Y(t))dt + \sigma dW(t),$$

where W is a vector of independent standard Brownian motions. We then have that

$$\int_t^T Y(s)ds \tag{48}$$

is Normally distributed with a mean of

$$\theta(T-t) + \frac{Y(t)-\theta}{K}\left[1 - e^{-K(T-t)}\right],$$

and with a mean zero component given by

$$\frac{\sigma}{K}\int_t^T \left[1 - e^{-K(T-u)}\right] dW(u), \tag{49}$$

whose variance is

$$\frac{\sigma\sigma'}{K^2}\left[(T-t) - \frac{2}{K}\Big(1 - e^{-K(T-t)}\Big) + \frac{1}{2K}\Big(1 - e^{-2K(T-t)}\Big)\right].$$

This result is standard, and therefore the proof is omitted.

B The Factor Likelihood Function

The state variables in this model are conditionally Gaussian. Hence their transition probabilities are known in closed form. It is straightforward to check that conditional on time t information, the distribution of $Y_n(t+\tau)$ is given by

$$Y_n(t+\tau) \sim N\left(e^{-[K_Y]_{nn}\tau}Y_n(t) + \theta_n(1 - e^{-[K_Y]_{nn}\tau}), \frac{\sigma^2_{Yn}}{2[K_Y]_{nn}}(1 - e^{-2[K_Y]_{nn}\tau})\right). \tag{50}$$

Recall that the Y-type factors in the model have independent innovations by assumption. From equation (30) we have that

$$z(t+\tau) = z(t) + \mu\tau + \Sigma_Z\Big(W(t+\tau) - W(t)\Big) - k_Z\int_t^T Y(s)ds. \tag{51}$$

The distribution of the m^{th} element of $z(t+\tau)$ conditioned on time t information is therefore Normal with a mean of

$$z_m(t) + \mu_m \tau - \sum_{n=1}^{N} [k_Z]_{m,n} \left[\theta_n \tau + \frac{Y_n(t) - \theta_n}{[K_Y]_{nn}} \left(1 - e^{-[K_Y]_{nn}\tau}\right) \right].$$

and a variance of

$$\sigma_{Zm}^2 \tau + [k_Z]_m \, V \, [k_Z]'_m.$$

where $[\cdot]_m$ indicates the m^{th} row of a matrix, and where V is a diagonal $N \times N$ matrix with diagonal elements given by

$$\frac{\sigma_{Yn}^2}{[K_Y]_{nn}^2} \left[\tau - \frac{2}{[K_Y]_{nn}} \left(1 - e^{-[K_Y]_{nn}\tau}\right) + \frac{1}{2[K_Y]_{nn}} \left(1 - e^{-2[K_Y]_{nn}\tau}\right) \right].$$

This follows directly from Lemma A.1. We observe that conditional on time t information, the covariance between the n^{th} Y factor, and the m^{th} z factor is given by

$$-\frac{[k_Z]_{mn} \, \sigma_{Yn}^2}{[K_Y]_{nn}^2} \left(\frac{1}{2} - e^{-[K_Y]_{nn}\tau} + \frac{1}{2} e^{-2[K_Y]_{nn}\tau} \right). \tag{52}$$

This follows from the dynamics of z given in (51), from the evolution of the Y's, and from Lemma A.1. In particular, we notice that the mean-zero part of $Y_n(t+\tau)$ (under the physical measure $\mathcal{P}$) conditional on $Y_n(t)$ can be written as

$$\sigma_{Yn} e^{-[K_Y]_{nn}\,(t+\tau)} \int_t^{t+\tau} e^{[K_Y]_{nn}\,u} \, dW_n(u).$$

From this and from the dynamics of $\int Y(t)dt$ in (49) we see that

$$\begin{aligned} &\mathrm{Cov}\Big(Y_n(t+\tau) - Y_n(t), z_m(t+\tau) - z_m(t)\Big) = \\ &\quad -\frac{[k_Z]_{mn} \, \sigma_{Yn}^2}{[K_Y]_{nn}} \int_t^{t+\tau} e^{-[K_Y]_{nn}\,(t+\tau-u)} \left(1 - e^{-[K_Y]_{nn}(t+\tau-u)}\right) du. \end{aligned}$$

The result in (52) follows from the evaluation of this integral.

Finally, z innovations from time t to $t+\tau$, conditional on time t information, are also correlated by virtue of the $\int Y(t)dt$ term in (51), with a covariance between $z_m(t+\tau) - z_m(t)$ and $z_{m'}(t+\tau) - z_{m'}(t)$ given by

$$\sum_{n=1}^{N} [k_Z]_{m,n} [k_Z]_{m',n} \mathrm{Var}_t \left(\int_t^{t+\tau} Y_n(s)ds \right) = [k_Z]_m \, V \, [k_Z]'_{m'}. \tag{53}$$

for the V matrix defined above. With this, the conditional distribution of the Y and z vector is fully specified.

C Derivation of the LPY Coefficient

We first note that from equation (4) we have that the yield on a zero-coupon bond with maturity τ is given by

$$y_\tau(t) = -\frac{A_{t+\tau}(t)}{\tau} + \frac{B_{t+\tau}(t)'}{\tau} Y(t).$$

Also note that in the present framework (from equation (50)), $Y_n(t+\tau)$ is given by

$$Y_n(t+\tau) = e^{-[K_Y]_{nn}\tau} Y_n(t) + \theta_n(1 - e^{-[K_Y]_{nn}\tau}) + N\left(0, \frac{\sigma_{Yn}^2}{2[K_Y]_{nn}}(1 - e^{-2[K_Y]_{nn}\tau})\right),$$

where the error term is independent of $Y_n(t)$. The short rate is given by

$$r(t) = r_0 + r_Y' Y(t).$$

Therefore, we can rewrite $y_{\tau-\Delta}(t+\Delta) - y_\tau(t)$ as

$$y_{\tau-\Delta}(t+\Delta) - y_\tau(t) =$$

$$\text{constant} + \left(\begin{bmatrix} e^{-[K_Y]_{11}\Delta} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{-[K_Y]_{NN}\Delta} \end{bmatrix} \frac{B_{t+\tau}(t+\Delta)}{\tau - \Delta} - \frac{B_{t+\tau}(t)}{\tau} \right)' Y(t) + \text{noise}$$

where the noise term is independent of $Y(t)$. Also, we have that

$$y_\tau(t) - r(t) = \text{constant} + \left(\frac{B_{t+\tau}}{\tau} - r_Y\right)' Y(t).$$

Since the constants and the noise term in the above do not matter for the covariance and variance calculations, we see that ϕ_τ from (44) is given by

$$\phi_\tau = \left(\tau/\Delta - 1\right) \frac{\sum_{n=1}^N \left(\frac{[B_{t+\tau}(t+\Delta)]_n}{\tau-\Delta} e^{-[K_Y]_{nn}\Delta} - \frac{[B_{t+\tau}(t)]_n}{\tau}\right)\left(\frac{[B_{t+\tau}]_n}{\tau} - [r_Y]_n\right)}{\sum_{n=1}^N \left(\frac{[B_{t+\tau}]_n}{\tau} - [r_Y]_n\right)^2}. \tag{54}$$

D Derivation of Return Predictability Relationships

Given the assumptions we have made about factor dynamics, it is possible to compute expected returns and the model R^2's in closed form. The following propositions state the relevant results.

Proposition D.1 *With the assumptions about factors dynamics which were made in Section 2 we can decompose returns on a rolled-over investment at the short-rate (from equation (35)) as follows:*

$$R_r(t,\tau) = \mathbb{E}_t[R_r(t,\tau)] + \epsilon_r(t+\tau),$$

where

$$\mathbb{E}_t[R_r(t,\tau)] = r_0 \times \tau + \sum_{n=1}^{N} r_{Yn} \left[\theta_n \times \tau + \frac{Y_n(t) - \theta_n}{[K_Y]_{nn}}\left(1 - e^{-[K_Y]_{nn}\tau}\right)\right].$$

Furthermore, we have that

$$\mathrm{Var}\Big(\mathbb{E}_t[R_r(t,\tau)]\Big) = \sum_{n=1}^{N} \frac{r_{Yn}^2\,\sigma_{Yn}^2}{2[K_Y]_{nn}^3}\left(1 - e^{-[K_Y]_{nn}\tau}\right)^2,$$

and that

$$\mathrm{Var}\Big(\epsilon_r(t+\tau)\Big) =$$

$$\sum_{n=1}^{N} \frac{r_{Yn}^2\,\sigma_{Yn}^2}{[K_Y]_{nn}^2}\left[\tau - \frac{2}{[K_Y]_{nn}}\left(1 - e^{-[K_Y]_{nn}\tau}\right) + \frac{1}{2[K_Y]_{nn}}\left(1 - e^{-2[K_Y]_{nn}\tau}\right)\right].$$

The R^2 for a forecasting regression over a time τ horizon is given by equation (36).

Proof. Since $r(t) = r_0 + r_Y' Y(t)$, we have that

$$\int_t^{t+\tau} r(s)ds = \tau r_0 + \sum_{n=1}^{N} r_{Yn} \int_t^{t+\tau} Y_n(s)ds.$$

Given the behavior of the Y integral in (48), and the fact that the asymptotic variance of $Y_n(t)$ is $\sigma_{Yn}^2/(2[K_Y]_{nn})$, the results of Proposition D.1 follow immediately.

Q.E.D.

The next proposition states the relevant results for stock total returns processes, and for excess returns of total returns processes over a rolled over investment at the short rate.

Proposition D.2

(Part i.) With the assumptions about factors dynamics which were made in Section 2, we can decompose $R_j(t,\tau)$ from equation (33) as follows:

$$R_j(t,\tau) = \mathbb{E}_t[R_j(t,\tau)] + \epsilon_j(t+\tau),$$

where $\epsilon_j(t+\tau)$ is independent of $\mathbb{E}_t[R_j(t,\tau)]$. Let us define $\phi_0 \equiv a_j^{TR}$ and $\phi_Y \equiv k_Z' C_j$ where the total returns process of stock j is given by $s_j(t) = \exp(a_j^{TR}\, t - B_j' Y(t) - C_j' z(t))$. Then we have that

$$\mathbb{E}_t[R_j(t,\tau)] =$$

$$\phi_0 \tau - \sum_{n=1}^{N} [B_j]_n \theta_n (1 - e^{-[K_Y]_{nn}\tau}) - \sum_{m=1}^{M} [C_j]_m \mu_m \tau$$

$$+ \sum_{n=1}^{N} \phi_{Yn} \left[\theta_n \tau - \frac{\theta_n}{[K_Y]_{nn}}\left(1 - e^{-[K_Y]_{nn}\tau}\right)\right]$$

$$- \sum_{n=1}^{N} \left[\left([B_j]_n + \frac{\phi_{Yn}}{[K_Y]_{nn}}\right)\left(e^{-[K_Y]_{nn}\tau} - 1\right) Y_n(t)\right],$$

where we have used $[\cdot]_n$ to indicate the n^{th} element of a vector. Also we have the following

$$\begin{aligned}
\mathrm{Var}\Big(\mathbb{E}_t[R_j(t,\tau)]\Big) &= \sum_{n=1}^{N}\left([B_j]_n+\frac{\phi_{Yn}}{[K_Y]_{nn}}\right)^2\left(e^{-[K_Y]_{nn}\tau}-1\right)^2\frac{\sigma_{Yn}^2}{2[K_Y]_{nn}},\\
\mathrm{Var}\Big(\epsilon_j(t+\tau)\Big) &= \sum_{n=1}^{N}\frac{\phi_{Yn}^2\,\sigma_{Yn}^2}{[K_Y]_{nn}^2}\times\\
&\qquad\left[\tau-\frac{2\xi_n}{[K_Y]_{nn}}\left(1-e^{-[K_Y]_{nn}\tau}\right)+\frac{\xi_n^2}{2[K_Y]_{nn}}\left(1-e^{-2[K_Y]_{nn}\tau}\right)\right]\\
&\quad+\sum_{m=1}^{M}\sigma_{Zm}^2[C_j]_m{}^2\,\tau,
\end{aligned}$$

where

$$\xi_n\equiv 1+\frac{[B_j]_n[K_Y]_{nn}}{\phi_{Yn}}.$$

The R^2 for a forecasting regression over a time τ horizon is given by (36).

(Part ii.) The excess returns of the total returns process of stock j over a rolled over investment in the short-rate can be written as

$$R_{xr}(t,\tau)\equiv R_j(t,\tau)-R_r(t,\tau)=\mathbb{E}_t[R_{xr}(t,\tau)]+\epsilon_{xr}(t+\tau).$$

The expectation and the variances for the R_{xr} process are obtained by redefining the ϕ_0 and ϕ_Y from the above formulas as follows

$$\begin{aligned}
\phi_0 &\equiv a_j^{TR}-r_0,\\
\phi_Y &\equiv k_Z'C_j-r_Y.
\end{aligned}$$

Proof. The total returns process for a stock j is given by

$$s_j(t)=\exp\left(a_j^{TR}t-B_j{}'Y(t)-C_j{}'z(t)\right).$$

which allows us to write that

$$\log\frac{s_j(t+\tau)}{s_j(t)}=a_j^{TR}\tau-B_j{}'(Y(t+\tau)-Y(t))-C_j{}'(z(t+\tau)-z(t)). \tag{55}$$

From the dynamics of Y we can write

$$\begin{aligned}
&Y_n(t+\tau)-Y_n(t)=\\
&\quad\left(e^{-[K_Y]_{nn}\tau}-1\right)Y_n(t)+\theta_n\left(1-e^{-[K_Y]_{nn}\tau}\right)+\sigma_{Yn}e^{-[K_Y]_{nn}(t+\tau)}\int_t^{t+\tau}e^{[K_Y]_{nn}u}dW_n(u).
\end{aligned}$$

From the dynamics of z we can write

$$z_m(t+\tau) - z_m(t) = \mu_m \tau + \sigma_{Zm} \left(W_m(t+\tau) - W_m(t) \right) - [k_Z]_m \int_t^{t+\tau} Y(u)du,$$

where $[\cdot]_m$ indicates the m^{th} row of a matrix. Using the result for the Y integral in (48), we can collect terms in (55) to get the results in the first part of Proposition D.2.

Note that the excess returns of the total returns process of a stock over the short rate are given by

$$\log \frac{s_j(t+\tau)}{s_j(t)} - \int_t^{t+\tau} r(u)du.$$

This has exactly the same functional form as the total returns on a stock, but with different coefficients on τ and on the Y integral. The second part of Proposition D.2 follows from this observation.

Q.E.D.

With these propositions in hand, it is possible to establish some limiting results for the R^2 of forecasting regressions for stocks, for investing at the short rate, and for the excess returns on stocks. The following Lemma states the relevant result.

Lemma D.1 *Given the assumptions of the Propositions D.1 and D.2, the following is true of forecasting regressions for the total returns process of any stock j:*

$$\lim_{\tau \to 0} R^2(\tau) = 0 \quad \textit{and} \quad \lim_{\tau \to \infty} R^2(\tau) = 0.$$

For rolled-over investments at the short rate, the following is true:

$$\lim_{\tau \to 0} R^2(\tau) = 1 \quad \textit{and} \quad \lim_{\tau \to \infty} R^2(\tau) = 0.$$

For the excess returns of total returns of stock j over the short rate, the following is true:

$$\lim_{\tau \to 0} R^2(\tau) = 0 \quad \textit{and} \quad \lim_{\tau \to \infty} R^2(\tau) = 0.$$

The proof of this Lemma involves some straightforward applications of L'Hopital's rule. Of course, the relative importance of the mean-reverting and random walk components in stock prices determines the behavior of the R^2 of forecasting regressions for horizons between zero and infinity. An analysis of the empirical properties of this behavior is in the text.

References

Ang, A. and G. Bekaert, 2001, "Stock return predictability: Is it there?" working paper.

Ang, A. and J. Liu, 2001, "A general affine earnings model," working paper.

Arnott, R.D., H.N. Hanson, M.L. Leibowitz, and E.H. Sorensen, 1989, "A total differential approach to equity duration," *Financial Analysts Journal*, 45 (5), 30–37.

Babbs, S.H. and K.B. Nowman, 1999, "Kalman filtering of generalized Vasicek term structure models," *Journal of Financial and Quantitative Analysis*, 34 (1), 115–130.

Bakshi, G.S. and Z. Chen, 1997a, "An alternative valuation model for contingent claims," *Journal of Financial Economics*, 44, 123–165.

Bakshi, G.S. and Z. Chen, 1997b, "Asset pricing without consumption or market portfolio data," working paper.

Bakshi, G.S. and Z. Chen, 2001, "Stock valuation in dynamic economies," working paper.

Bekaert, G. and S. Grenadier, 2000, "Stock and bond pricing in an affine economy," working paper.

Billingsley, P., 1995, *Probability and Measure*, Wiley-Interscience, New York, New York.

Bossaerts, P., 1988, "Common nonstationary components of asset prices," *Journal of Economic Dynamics and Control*, 12 (2/3), 347–364..

Brennan, M.J., A.W. Wang, and Y. Xia, 2001, "Intertemporal capital asset pricing and the Fama-French three-factor model," working paper.

Campbell, J.Y., 1999, "Asset prices, consumption, and the business cycle," *Handbook of Macroeconomics, Volume 1*, ed. by J.B. Taylor and W. Woodford, Elsevier Science.

Campbell, J.Y. and J.H. Cochrane, 1999, "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 107 (2), 205-251.

Campbell, J.Y. and A.S. Kyle, 1993, "Smart money, noise trading and stock price behavior," *Review of Economic Studies*, 60, 1-34.

Campbell, J.Y. and R.J. Shiller, 1991, "Yield spreads and interest rate movements: A bird's eye view," *The Review of Economic Studies*, 58, 495–514.

Chen, R.-R. and L. Scott, 1993, "Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates," *The Journal of Fixed Income*, December 1993, 14–31.

Constantinides, G.M., 1992, "A theory of the nominal term structure of interest rates," *The Review of Financial Studies*, 5 (4), 531–552.

Cox, J.C., J. Ingersoll, and S. Ross, 1985, "A theory of the term structure of interest rates," *Econometrica*, 53, 385–408.

Dai, Q. and K.J. Singleton, 2000, "Specification analysis of affine term structure models," *Journal of Finance*, 55 (5), 1943–1978.

Dai, Q. and K.J. Singleton, 2001, "Expectation puzzles, time-varying risk premia, and affine models of the term structure," working paper.

Daniel, K. and S. Titman, 1997, "Evidence on the characteristics of cross sectional variation in stock returns," *Journal of Finance*, 52 (1), 1–33.

Duffee, G.R., 2001, "Term premia and interest rate forecasts in affine models," working paper.

Duffie, D., 2001, *Dynamic Asset Pricing Theory*, Third Edition, Princeton University Press.

Duffie, D. and R. Kan, 1996, "A yield-factor model of interest rates," *Mathematical Finance*, 6, 379–406.

Duffie, D. and K.J. Singleton, 1997, "An econometric model of the term structure of interest rate swap yields," *Journal of Finance*, 52, 1287–1321.

Duffie, D. and K.J. Singleton, 1999, "Modeling the term structures of defaultable bonds," *The Review of Financial Studies*, 12 (4), 687–720.

Dybvig, P.H. and C.-f. Huang, 1989, "Nonnegative wealth, absence of arbitrage, and feasible consumption plans," *The Review of Financial Studies*, 1 (4), 377–401.

Dybvig, P.H., J.E. Ingersoll, and S.A. Ross, 1996, "Long forward and zero-coupon rates can never fall," *Journal of Business*, 69 (1), 1–25.

Elton, E.J., M.J. Gruber, C.R. Blake, 1999, "Common factors in active and passive portfolios," *European Finance Review*, 3 (1).

Fama, E.F. and K.R. French, 1988, "Permanent and temporary components of stock prices," *Journal of Political Economy*, 96 (2), 246–273.

Fama, E.F. and K.R. French, 1989, "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, 25, 23–49.

Fama, E.F. and K.R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47 (2), 427–465.

Fama, E.F. and K.R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

Gatev, E., W.N. Goetzmann, and K.G. Rouwenhorst, 2001, "Pairs trading: Performance of a relative value arbitrage trade," working paper.

Goetzmann, W.N. and P. Jorion, 1993, "Testing the predictive power of dividend yields," *Journal of Finance*, 48 (2), 663–679.

Harrison, M. and D, Kreps, 1979, "Martingales and arbitrage in multi-period securities markets," *Journal of Economic Theory*, 20, 381–408.

Harrison, J.M. and S.R. Pliska, 1981, "Martingales and stochastic integrals in the theory of continuous trading," *Stochastic Processes and their Applications*, 11, 215–260.

Hodrick, R.J., 1992, "Dividend yields and expected stock returns: Alternative procedures for inference and measurement," *The Review of Financial Studies*, 5 (3), 357–386.

Jeffrey, A., 2001, "An alternative way of measuring duration and convexity in the context of interest rate risk management," working paper, Yale School of Management.

Jegadeesh, N. and S. Titman, 1993, "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, 48 (1), 65-90.

Kirby, C., 1997, "Measuring the predictable variation in stock and bond returns," *The Review of Financial Studies*, 10 (3), 579–630.

Leibowitz, M.L, 1986, "Total portfolio duration: A new perspective on asset allocation," *Financial Analysts Journal*, 42 (5), 18–29 and page 77.

Lewin, R.A. and S.E. Satchell, 2001, "The derivation of a new model of equity duration," working paper, University of Cambridge.

Litterman, R. and J. Scheinkman, 1991, "Common factors affecting bond returns," *The Journal of Fixed Income*, June, 54–61.

Mamaysky, H., 2001, "Interest rates and the durability of consumption goods," working paper.

Mamaysky, H., 2002a, "A model for pricing stocks and bonds," working paper.

Mamaysky, H., 2002b, "A model for pricing stocks and bonds with default risk," working paper.

Merton, R.C., 1973, "An intertemporal capital asset pricing model," *Econometrica*, 41, 867–887.

Merton, R.C., 1974, "On the pricing of corporate debt: The risk structure of interest rates," *Journal of Finance*, 29, 449–470.

Oksendal, B., 1998, *Stochastic Differential Equations*, Springer-Verlag.

Richardson, M., 1993, "Temporary components of stock prices: A skeptic's view," *Journal of Business and Economic Statistics*, 11 (3), 199–207.

Ross, S., 1976, "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13, 341–360.

Shiller, R., 1981, "Do stock prices move too much to be justified by subsequent changes in dividends," *American Economic Review*, 71, 421–436.

Vasicek, O., 1977, "An equilibrium characterization of the term structure of interest rates," *Journal of Financial Economics*, 5, 177–188.

Table 1: **Maximum Likelihood Estimation of the Model Parameters.** Standard errors are computed using the inverse of the information matrix for the likelihood function. Variables with no standard errors reported are either not estimated (e.g. set to zero), or are computed from the parameter estimates using equations from the text (e.g. the $\tilde{k}_Z$ matrix uses (22) and the long rate y_∞ uses equation (7)).

Parameter Estimates					
Σ_Y	0.02758962	0.01478133	$\mathcal{L}(\hat{\Theta})$	8988.16071799	
	(0.00057161)	(0.00049003)			
$\tilde{K}$	0.63547846	0.03453723			
	(0.01997152)	(0.00283593)			
K	0.82557277	0.13191938			
	(0.24556313)	(0.08909490)			
$\tilde{\Theta}$	0.00000000	0.00000000			
Θ	-0.00759292	-0.05888181			
	(0.00910123)	(0.02729782)			
r_0	0.13289364		y_∞	0.04036664	
	(0.00516293)				
r_Y	1.00000000	1.00000000			
$A^n(0)$	0.00000000	0.00000000	0.00000000		
$[C]_1$	1.00000000	0.00000000	0.00000000		
$[C]_2$	1.17520689	1.00000000	0.00000000		
	(0.03202933)				
$[C]_3$	0.78923754	0.12890930	1.00000000		
	(0.02605679)	(0.03406288)			
$\tilde{\mu}$	0.00000000	0.00000000	0.00000000		
μ	0.06163533	-0.03090398	-0.03101848		
	(0.08167795)	(0.07326182)	(0.04513316)		
$[B]_1$	0.80779468	2.12159960			
	(0.44049864)	(0.65249038)			
$[B]_2$	0.83297712	1.31463013			
	(0.60267562)	(0.86753637)			
$[B]_3$	1.01361217	2.18766869			
	(0.38311156)	(0.56961926)			
$[k_Z]_1$	-2.27650541	-0.81245726	$[\tilde{k}_Z]_1$	0.48666388	0.92672582
	(1.68975915)	(1.35722267)			
$[k_Z]_2$	-1.15960728	0.39467550	$[\tilde{k}_Z]_2$	-0.10126976	-0.13449826
	(1.15433574)	(1.08668145)			
$[k_Z]_3$	0.22127485	0.05098627	$[\tilde{k}_Z]_3$	-0.01516749	0.21037525
	(0.90703693)	(0.66060377)			
Σ_Z	0.15603665	0.09665511	0.07296794		
	(0.00454621)	(0.00302044)	(0.00275254)		
σ^2_ϵ	0.00034291	0.00072606	0.00000390	0.00000402	0.00000713
	(0.00004066)	(0.00008034)	(0.00000574)	(0.00000426)	(0.00000575)
μ_ϵ	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000

Table 2: **Correlations of Factor Innovations.** This table reports the correlations of monthly first differences of the values of the model factors. The means and standard deviations are also reported.

	Y_1	Y_2	Z_1	Z_2	Z_3
Y_1	1.0000				
Y_2	-0.3488	1.0000			
Z_1	0.0017	0.0061	1.0000		
Z_2	0.0040	0.0005	0.0068	1.0000	
Z_3	0.0030	-0.0016	0.0004	-0.0008	1.0000
Mean	0.0000	-0.0001	0.0000	-0.0016	-0.0022
S.D.	0.0077	0.0049	0.0453	0.0280	0.0210

Table 3: **Variation in Short Rate and in Value Weighted Index.** The left side of the table shows regressions of the first differenced short rate on first differences in the two joint bond-stock factors. The right side of the table shows regressions of the first differenced log value-weighted CRSP index returns on first differences in the two common and the three stock specific factors. Data are monthly. Standard errors are in parentheses.

Y_1	Y_2	R^2	Y_1	Y_2	Z_1	Z_2	Z_3	R^2
0.7785		0.6313	-0.8078	-2.1216	-1.0000	0.0000	-0.0000	1.0000
(0.0642)			(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
	0.4506	0.0852	-0.8692	-2.2164				0.0531
	(0.1031)		(0.3263)	(0.6010)				
1.0000	1.0000	1.0000			-1.0024			0.9532
(0.0000)	(0.0000)				(0.0130)			

Table 4: **Results from LPY Regressions.** This table shows the results from the regression in (43), run for various maturities (τ's), using the model implied zero coupon bond yields and short rate. OLS standard errors and R^2's (on a scale of zero to one) are reported. Results are shown for the full sample, and for the time period ending in December 1995. The regressions are run using monthly data.

Maturity	0.25	0.75	1.00	2.00	3.00	4.00	5.00	7.00	10.00
19730228	0.236	0.169	0.100	-0.205	-0.515	-0.819	-1.112	-1.661	-2.377
19951229									
S.E.	(0.569)	(0.748)	(0.770)	(0.822)	(0.884)	(0.964)	(1.057)	(1.262)	(1.578)
R^2	0.001	0.000	0.000	0.000	0.001	0.003	0.004	0.006	0.008
19730228	0.219	0.171	0.114	-0.141	-0.406	-0.669	-0.926	-1.414	-2.064
20001229									
S.E.	(0.500)	(0.659)	(0.679)	(0.730)	(0.790)	(0.868)	(0.957)	(1.153)	(1.454)
R^2	0.001	0.000	0.000	0.000	0.001	0.002	0.003	0.005	0.006

Table 5: **Dividend Yield Forecasting Regressions.** This table shows the projection of the actual dividend yield of the CRSP value weighted index onto the Y factors extracted from estimation of the model by subperiods. The projection was done subject to the transversality condition constraint in (24). The best fit dividend series from the model solves (45), and is given by $\hat{\delta}(t) \equiv \hat{\delta}_0 + \hat{\delta}_{Y1} Y_1(t) + \hat{\delta}_{Y2} Y_2(t)$. The table also shows the regressions of the actual dividend yield series of the CRSP value weighted index onto the best fit dividend series from the model. This regression is given by $\delta(t) = \alpha + \beta \hat{\delta}(t) + \epsilon(t)$, where $\delta(t)$ is the actual dividend yield of the CRSP value weighted index, and is estimated over the time interval $[t - 0.5, t + 0.5]$. Data are monthly. Newey-West standard errors using a twelve month lag are in parentheses.

Period	α	β	R^2	$\hat{\delta}_0$	$\hat{\delta}_{Y1}$	$\hat{\delta}_{Y2}$
	-0.0000	1.0000	0.4307	0.0585	0.0927	0.3634
	(0.0078)	(0.2028)				

Table 6: **Moments of Prices of Risk.** The table shows the empirical correlation matrix for the five price of risk processes in the estimation and the term spread (given by the difference of the model yield for a ten year zero and the model short rate). The next part of the table shows the empirical means and standard deviations for the six time series, computed at a monthly frequency. The final part of the table shows the theoretical long-run means for the five price of risk series. These are given by $\Lambda_\infty \equiv \lambda_0 + \lambda_Y \Theta$, and the standard errors (given in parentheses) for the latter are computed using the standard errors from Table 1 and the δ-method.

	Y_1	Y_2	Z_1	Z_2	Z_3	10Yr-SR
Y_1	1.0000					
Y_2	-0.2496	1.0000				
Z_1	-0.7401	-0.4665	1.0000			
Z_2	-0.8915	0.6612	0.3551	1.0000		
Z_3	0.8364	-0.7395	-0.2504	-0.9940	1.0000	
10Yr-SR	-0.9752	0.0292	0.8705	0.7692	-0.6944	1.0000
Mean	0.1725	0.1735	0.3410	0.1141	0.5161	0.0114
S.D.	0.1468	0.1696	0.4132	0.3012	0.0994	0.0174
LR Means	0.1749	0.1376	0.3957	0.0805	0.5291	
	(0.2105)	(0.0618)	(0.4760)	(0.2611)	(0.2327)	

Table 7: **Moments of Expected Excess Returns.** The table shows the correlations between model instantaneous expected excess returns for the following: a 1 year zero, a 10 year zero, a 30 year zero, the value weighted CRSP index, the 3^{rd} decile size portfolio (small stocks), and the 8^{th} decile book to market portfolio (value stocks), the ten year yield – short rate term-spread (both implied from the model), and the model implied short rate. The risk premium for bonds is from equation (26) and the risk premium for stocks is in equation (27). The table also shows the means and standard deviations for the five time series. Results are computed using monthly data.

	1 Yr	10 Yr	30 Yr	Mkt	Size	Value	10Yr-SR	SR
1 Yr	1.0000							
10 Yr	0.7435	1.0000						
30 Yr	0.6205	0.9858	1.0000					
Mkt	-0.9959	-0.6800	-0.5472	1.0000				
Size	-0.9302	-0.4462	-0.2895	0.9596	1.0000			
Value	-0.9999	-0.7322	-0.6073	0.9973	0.9363	1.0000		
10Yr-SR	-0.8457	-0.2718	-0.1062	0.8904	0.9825	0.8545	1.0000	
SR	0.9698	0.8842	0.7931	-0.9438	-0.8126	-0.9656	-0.6899	1.0000
Mean	0.0060	0.0292	0.0554	0.0625	0.0809	0.0915	0.0114	0.0715
S.D.	0.0034	0.0206	0.0457	0.0596	0.0867	0.0448	0.0174	0.0290

Table 8: **Model R^2's for Forecasting Regressions.** This table shows the R^2's (in equation (36)) of forecasting regressions of returns on their conditional expectations using the parameter estimates in Table 1. Results are shown for raw returns (RAW) and for returns above the riskfree rate (XR). Asymptotic standard errors, computed using the standard errors from Table 1 and the δ-method, are in parentheses.

Years	0.25	0.75	1.00	2.00	3.00	4.00	5.00	7.00	10.00
Mkt Raw	0.011	0.024	0.027	0.032	0.032	0.031	0.030	0.029	0.027
	(0.027)	(0.058)	(0.067)	(0.086)	(0.095)	(0.102)	(0.107)	(0.112)	(0.111)
Small Raw	0.021	0.041	0.045	0.045	0.039	0.033	0.029	0.023	0.018
	(0.034)	(0.067)	(0.074)	(0.079)	(0.075)	(0.073)	(0.071)	(0.070)	(0.067)
Value Raw	0.004	0.008	0.009	0.011	0.010	0.010	0.010	0.009	0.008
	(0.015)	(0.033)	(0.039)	(0.050)	(0.057)	(0.061)	(0.065)	(0.068)	(0.068)
Mkt XR	0.040	0.086	0.099	0.123	0.131	0.134	0.135	0.131	0.120
	(0.053)	(0.109)	(0.124)	(0.160)	(0.180)	(0.192)	(0.200)	(0.204)	(0.196)
Small XR	0.043	0.085	0.094	0.103	0.100	0.095	0.091	0.085	0.076
	(0.052)	(0.101)	(0.113)	(0.134)	(0.145)	(0.153)	(0.159)	(0.165)	(0.161)
Value XR	0.027	0.061	0.071	0.094	0.103	0.108	0.110	0.109	0.101
	(0.044)	(0.096)	(0.113)	(0.154)	(0.178)	(0.193)	(0.201)	(0.207)	(0.198)

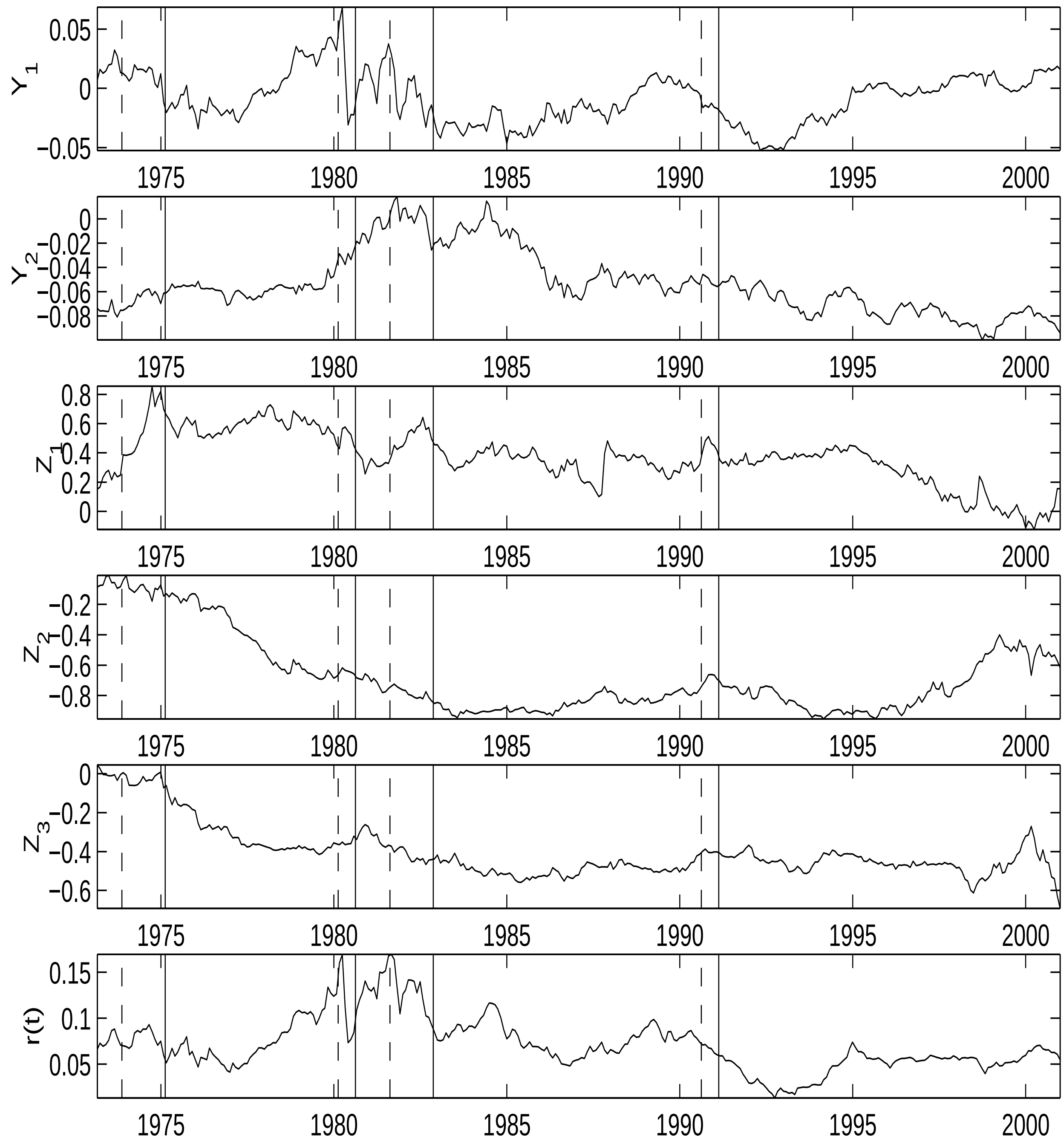

Figure 1: **Factors and the Short Rate.** This figure shows the five extracted model factors, and the short rate. The solid line shows the x-axis. The dashed vertical lines represent NBER business cycle peaks, and the solid vertical lines represent NBER business cycle troughs.

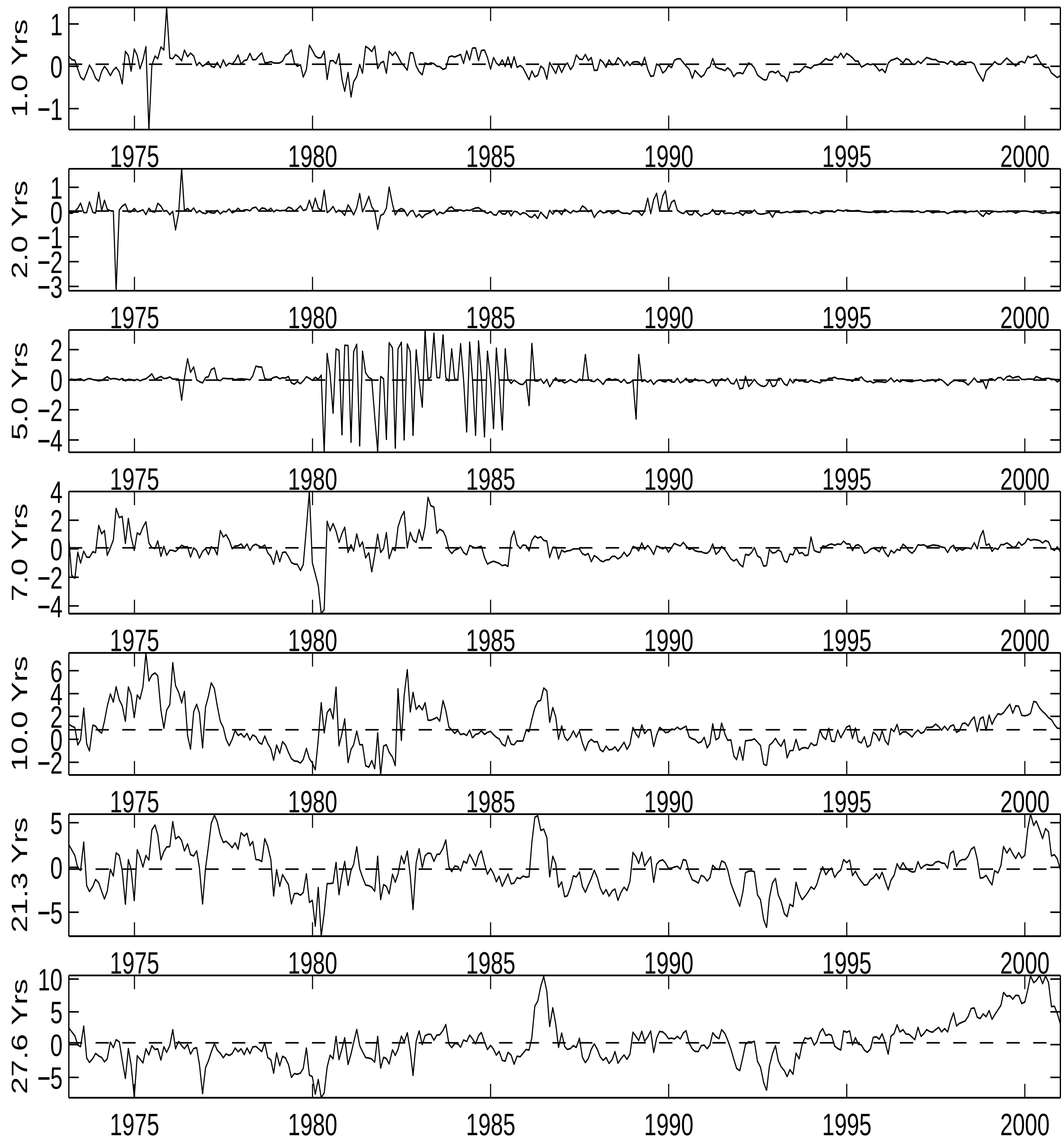

Figure 2: **Bond Pricing Errors.** The pricing errors (in percent) for bonds in the CRSP Fixed Term Indices file. The average maturity of bonds in each series is shown to the left of each graph. The dashed line shows the average level of the pricing error over entire sample period.

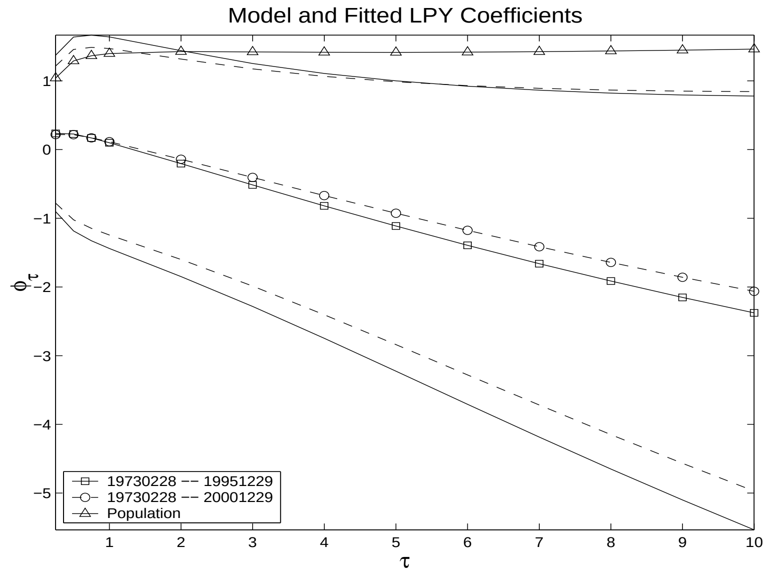

Figure 3: **Fitted and Model Values of LPY.** The figure shows the LPY coefficient from the regression in (43). The population line shows the value of the ϕ_τ coefficient implied from the relationship in (54) using estimates for the model parameters given in Table 1. The fitted lines (labeled with dates) represent the values of ϕ_τ from regressions using the pricing relationship of the model and the extracted realization of the model factors. Also shown are 95% OLS confidence intervals for the parameter estimates.

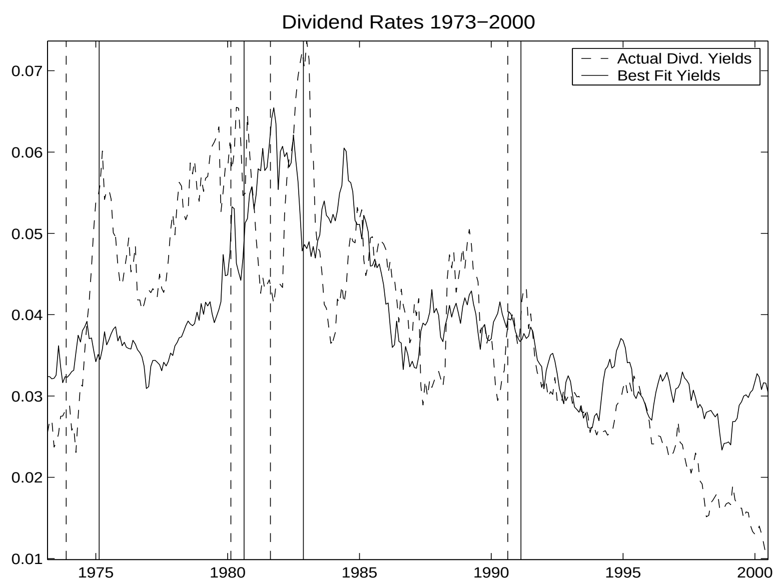

Figure 4: **The Actual and Implied Dividend Rates.** The figure shows the actual dividend rate of the CRSP value-weighted index, as well as the best fit dividend series from the model. In month t the actual dividend rate is the realized dividend rate in the time interval $[t-0.5, t+0.5]$, computed as the difference between the annual total return on the CRSP value weighted index and the annual return excluding dividends on the same index. The best fit dividend series from the model solves (45) and is given by $\hat{\delta}(t)$.

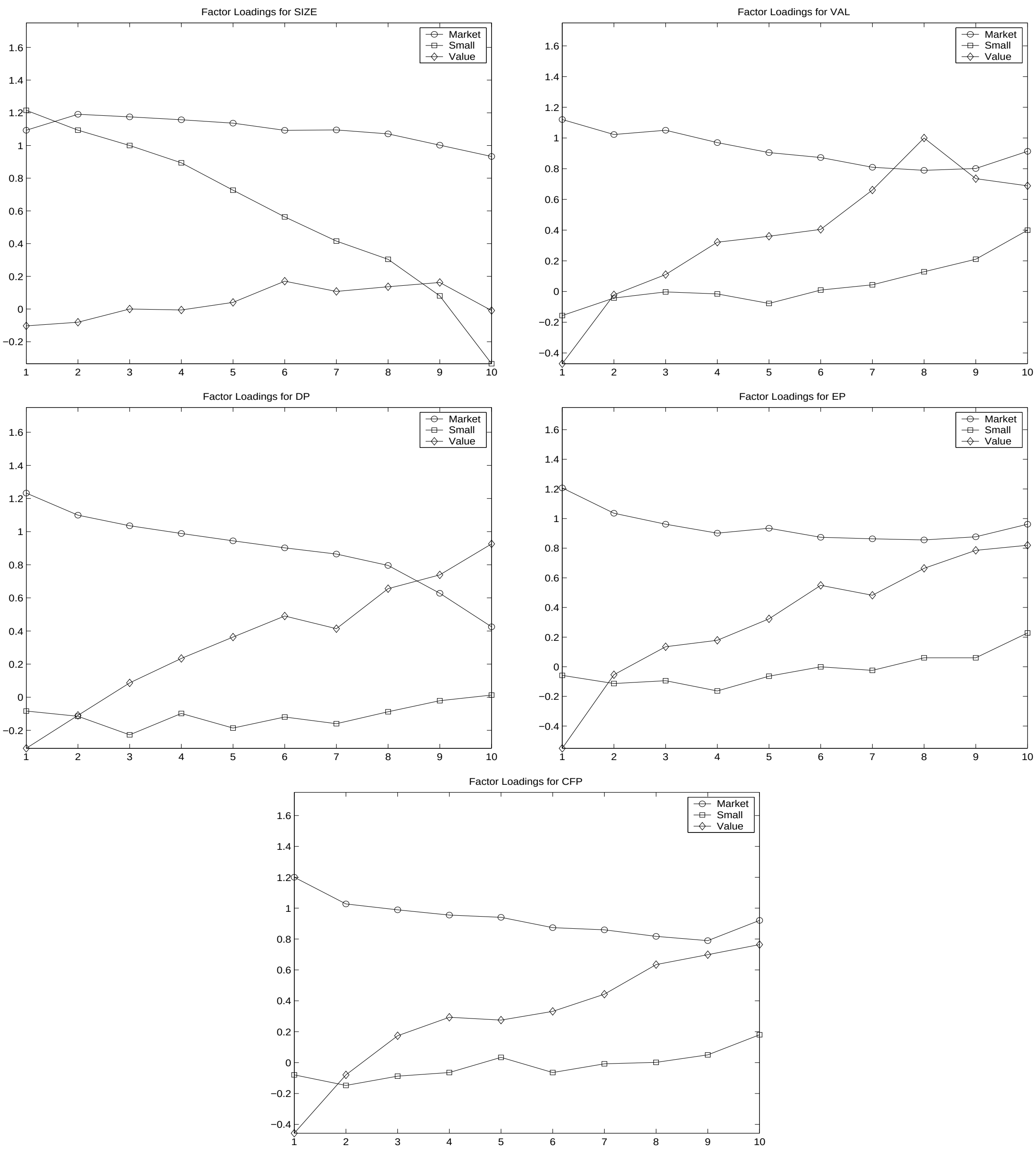

Figure 5: **Equity Loadings on Stock Factors.** For the five equity characteristics sorts, these show the loadings of the equity portfolios on the stock specific factors (the Z's). The x-axis corresponds to the decile portfolios in the five characteristics sorts (i.e. size (smallest to largest), book to market (low to high), dividend to price (low to high), earnings to price (low to high), and cashflow to price (low to high)).

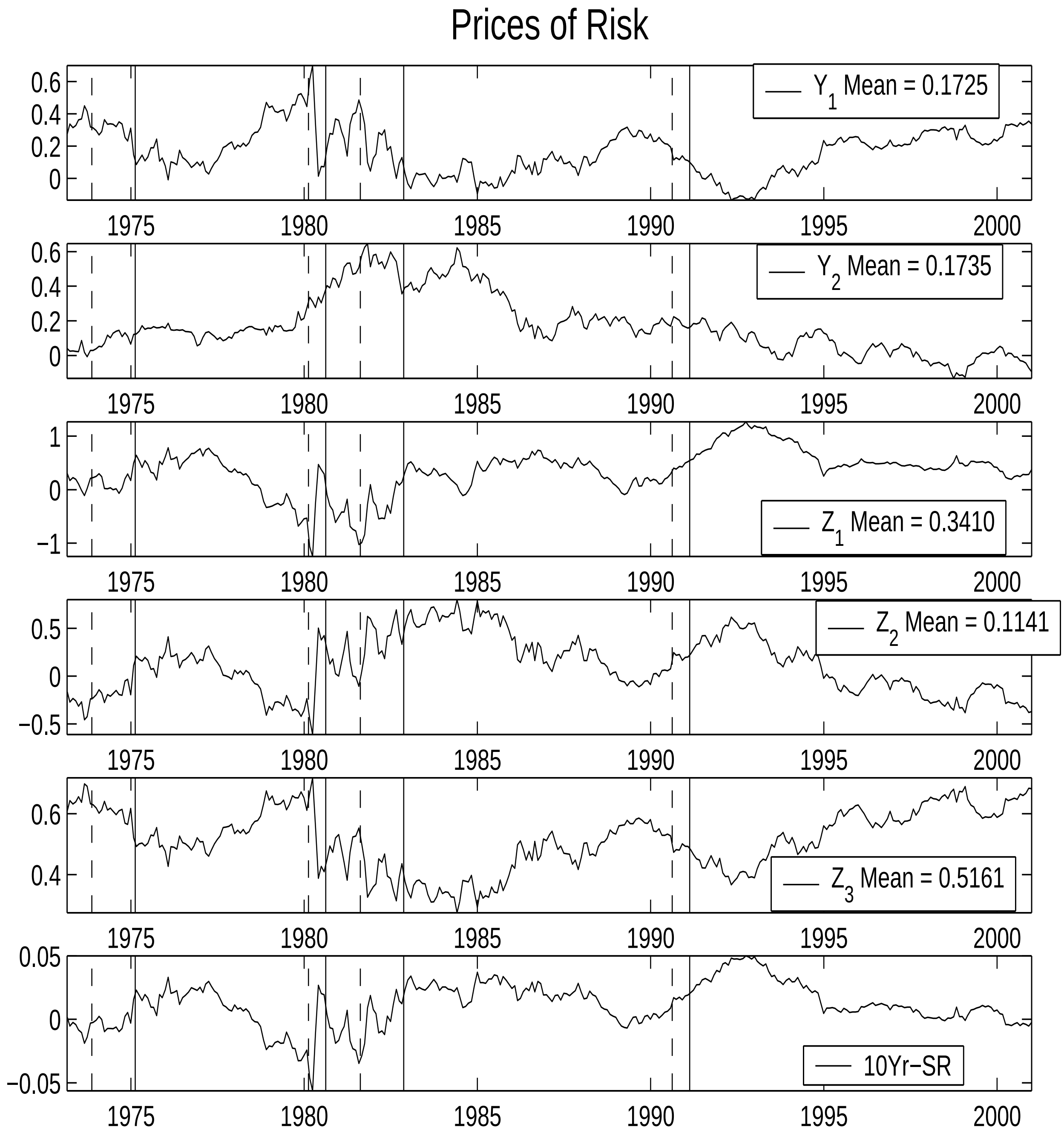

Figure 6: **Model Prices of Risk.** The graphs show the price of risk processes derived for the five model factors (two Y-type and three Z-type). Also the term spread between the model yield for a ten year zero and the model short rate is shown. The captions show the empirical means of the time series. The dashed vertical lines represent NBER business cycle peaks, and the solid vertical lines represent NBER business cycle troughs.

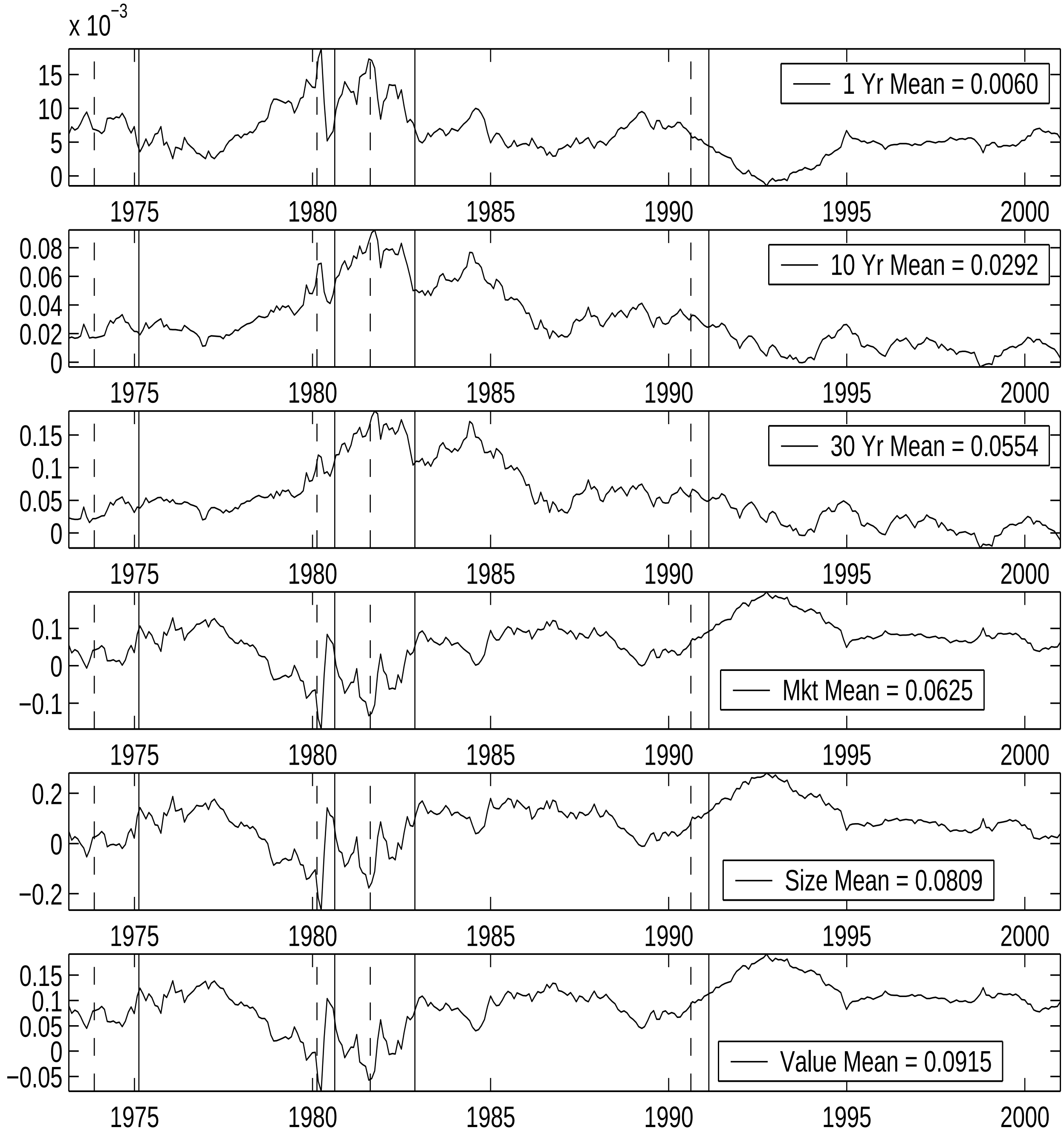

Figure 7: **Instantaneous Expected Excess Returns.** The graphs show the instantaneous expected excess returns derived from the model for the following securities: a 1 year zero, a 10 year zero, a 30 year zero, the value weighted CRSP index, the 3^{rd} decile size portfolio (small stocks), and the 8^{th} decile book to market portfolio (value stocks). The captions show the empirical means of the time series. The dashed vertical lines represent NBER business cycle peaks, and the solid vertical lines represent NBER business cycle troughs.

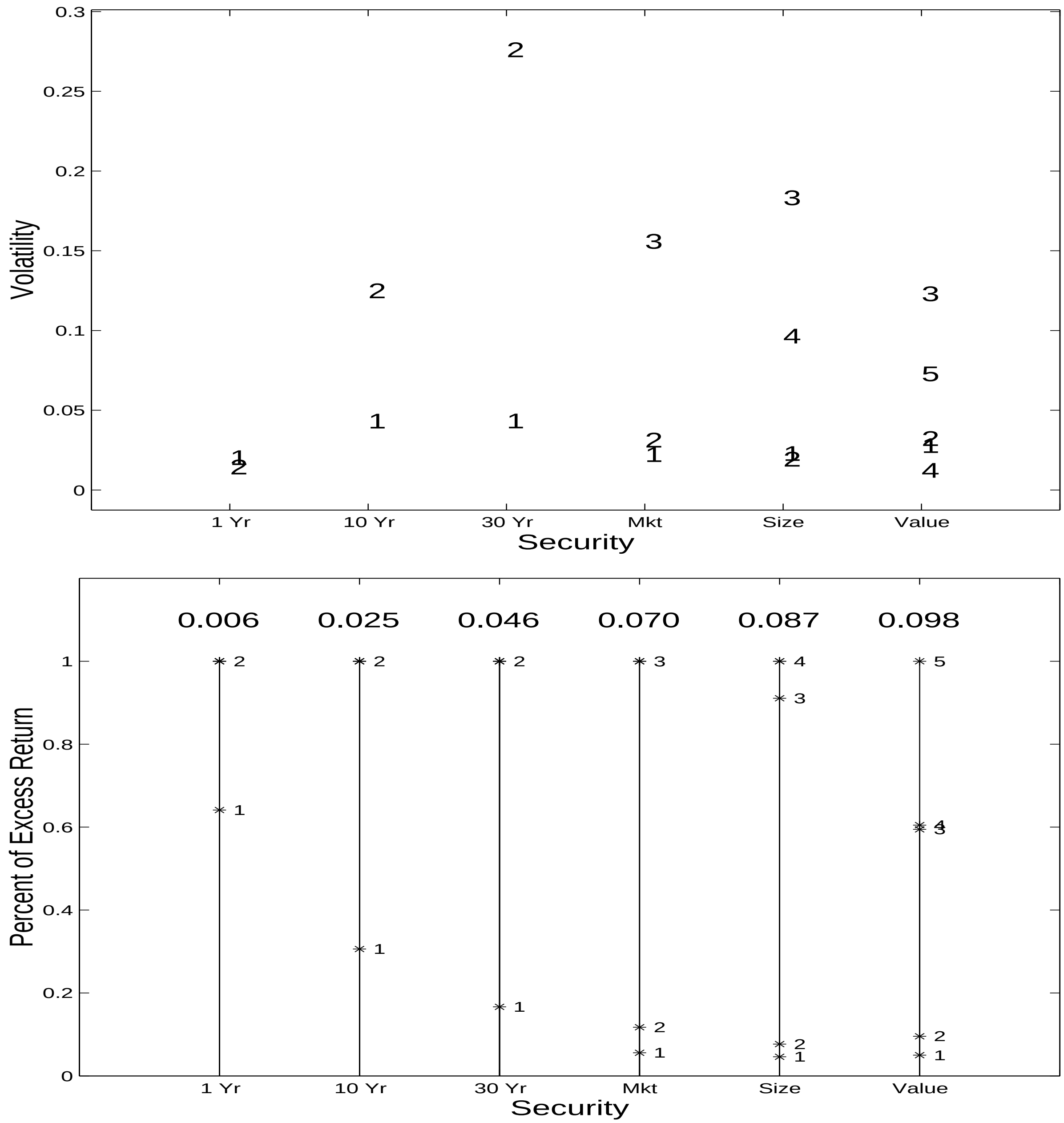

Figure 8: **Risk Loadings of Securities.** The first graph shows the loadings on the five prices of risk of the following securities: a 1 year zero, a 10 year zero, a 30 year zero, the value weighted CRSP index, the 3^{rd} decile size portfolio (small stocks), and the 8^{th} decile book to market portfolio (value stocks). These loadings are the σ's from the risk premia $\Lambda' \sigma$ of equations (26) and (27) for bonds and stocks respectively. Here $\sigma_1, \ldots, \sigma_5$ correspond to loadings on Y_1, Y_2, Z_1, Z_2, Z_3 respectively. The second graph shows the relative contribution of each of the prices of risk to the long-run excess return of each security. The long run excess return for security j is given by $\Lambda'_\infty \sigma_j$ where $\Lambda_\infty \equiv \lambda_0 + \lambda_Y \Theta$. The contribution of price of risk process i is the i^{th} element of the vector $\Lambda'_\infty \sigma_j$ divided by the sum of the elements of the vector.

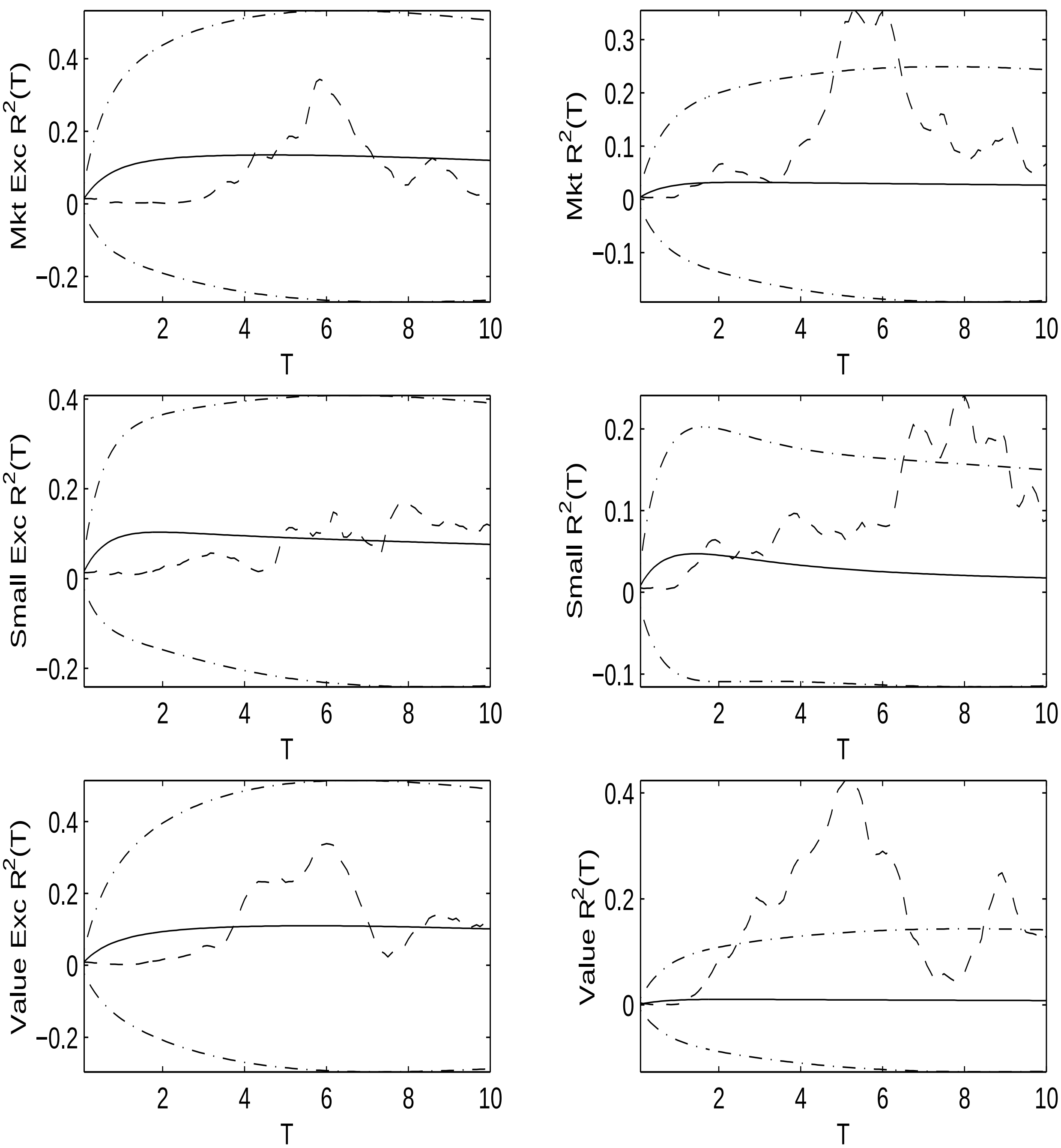

Figure 9: **Model and Fitted R^2's for Predictive Regressions.** These figures show the R^2's of a regression of realized returns on expected returns as a function of the forecasting horizon. The solid line shows the theoretical R^2's implied from the model parameters in Table 1 using equation (36) (a 95% confidence interval of the model R^2's is given by the dash-dotted line). The dashed line shows the median of actual R^2's from regressions of realized returns on the expected returns from equation (34) obtained using factor values extracted in the model estimation. These fitted regressions were conducted in rolling 15 year windows. The graphs on the left report results for excess returns, and the graphs on the right report results for raw returns.