On the Joint Pricing of Stocks and Bonds: Theory and Evidence *

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

First version: October 25, 2001
Last revised: January 24, 2002

*I would like to thank Zhiwu Chen, Will Goetzmann, Leonid Kogan, Stephen Ross, Ken Singleton, Werner Stanzl, Jacob Thomas, and Jiang Wang for very helpful discussions and suggestions. Also I would like to thank seminar participants at the Yale School of Management and New York University for valuable comments.
†Email: harry.mamaysky@yale.edu, phone (203) 436-0649, fax (203) 436-0630.

On the Joint Pricing of Stocks and Bonds: Theory and Evidence

Abstract

Assuming only the absence of arbitrage, I derive a dynamic model capable of jointly pricing a cross section of bonds and stocks. By providing an empirically flexible yet economically self-consistent modelling framework, this paper represents a middle ground between utility based and purely empirical approaches to asset pricing. Key implications of the model are that bond factors must be mean-reverting, that stocks must have a dependence on these bond factors, and that stocks may have additional random-walk and mean-reverting components. The model produces closed form solutions for bond prices, stock prices, and risk premia. Other quantities which may be computed (in closed form) within this framework include implied dividend processes for dividend paying securities, R^2's of forecasting regressions of returns on their conditional expectations, and a sensible measure of duration for equities.

In estimating the model using U.S. bond and stock data from the last fifty years, I find that a five factor model, with two joint bond-stock factors and three common stock factors, can adequately account for the historical behavior of the term-structure of government debt and for the behavior of a wide cross section of equity portfolios. I then study the behavior of bond and stock markets in the U.S. through the lens of this model, with an emphasis on analyzing the above mentioned theoretical constructs.

JEL Classification: G12, G13.

1 Introduction

Traditionally, empirical investigations into the behavior of stocks and bonds have proceeded along two different paths. Bonds, being a relatively well behaved asset class, have traditionally been the beneficiaries of an extremely elegant and empirically successful modelling approach. Starting with the pioneering work of Merton (1974), Vasicek (1977) and Cox, Ingersoll, and Ross (1985), researchers have had access to bond pricing models that are both economically plausible, and empirically quite successful. On the other hand, stocks have not been so fortunate. While much of the relevant theoretical work on asset pricing is extremely appealing on aesthetic grounds, such work has for the most part been unable to successfully overcome many empirical hurdles. For example, consumption asset pricing has had a very difficult time just reconciling aggregate consumption data with aggregate stock market fluctuations, not to mention stocks in the cross-section. Furthermore, as is well known, the empirical performance of the CAPM has been less than stellar. Indeed models that identify specific pricing factors based on primitive economic arguments have generally not been empirically successful. For this reason, most of the prominent (or successful) empirical work on stock pricing has proceeded by first finding pricing factors in a purely data driven way, and then resorting to an asset pricing model (traditionally, the arbitrage pricing theory of Ross (1976), or Merton's (1973) intertemporal CAPM) for an ex-post theoretical justification. The shortcomings of this approach are fairly obvious.

What is needed, then, is an economically sound model that is, at the same time, empirically flexible. In particular, we would like a model that guides us in our selection of pricing factors, and yet provides a good fit for the data. This paper delivers such a model. Interestingly, very little of the theoretical development in this paper is genuinely new, but simply combines ideas that have been floating around in the profession for quite some time. On the other hand, the end result is an arbitrage-free pricing framework that does a very good job of explaining the historical variation of U.S. government bonds and of a wide cross-section of equity returns. Such a model is useful for several reasons: First, it provides us with a lens through which we can look at the behavior of markets. By throwing some economics into the empirical mix, we may hope to impose enough structure so as to avoid common statistical problems, such as data mining for factors, or incorrect inferences due to small samples. Second, to the extent that an empirically successful model imposes restrictions on the data, we are able to learn more about how and why markets deviate from our conjectured idealized settings than we can by using a model that provides a grossly inadequate empirical platform.

It should be emphasized that this paper represents a middle ground between utility based and purely empirical pricing approaches. On the one hand the paper does not go as far as actually identifying a pricing factor (such as the market or consumption growth). It does however impose enough structure, via its assumption of no-arbitrage, to allow for factors to be extracted within the framework of the model itself, rather than in a purely data driven way. By not imposing the stringent restrictions that the presence of optimizing, risk-averse agents typically implies, the model is freed of much baggage, and is thus empirically quite flexible, though at the expense of ignoring a large number of important and interesting economic questions. The model's contribution is to show that a certain functional form for factor dynamics and for stock and bond prices is consistent with no-arbitrage. Having such restrictions allows us to estimate joint bond-stock pricing models confident of not having "missed anything," therefore enabling us to produce an economically robust decomposition of stock and bond prices into their constituent components. These components, then, can be interpreted as pricing factors in the true sense of the word.

Drawing on the technology of affine term-structure models, this paper derives a very general no-arbitrage dynamic model for bond and stock prices. As will become clear, the model can accommodate both stock prices and total return processes on equities – that is capital appreciation plus dividends reinvested back into the stock itself – though from an empirical point of view working with the total return processes proves more convenient. The paper shows that any factors that affect bond prices must be mean-reverting, and must also be present in stock prices. However, stocks may have additional random walk or mean-reverting components that are not present in bond prices. It is this extra degree of freedom for stock prices that allows the model to match up nicely with the data. Interestingly, this result can be seen as good news and bad news by both sides of the debate on stock return predictability. On the one hand, the fact that mean-reverting factors must, by no-arbitrage, be present in stock prices indicates that stock returns may indeed be predictable. On the other hand, the exact amount of this predictability may very well be quite small.

The little bit of economics that this model provides manifests itself in our ability to compute (often in closed form) several quantities of interest within the framework of the model. Most importantly bond and stock prices are obtained in closed form as functions of the state variables in the economy. This is a major benefit as it allows the model to price a cross section of bonds and stocks over time. Once we specify (an extremely flexible) pricing kernel, we are able to compute risk premia. Furthermore, by imposing a fairly benign restriction on the behavior of stock prices in the model, we are able to infer an implied dividend process from the total return process of a dividend paying security, allowing for a test of the model to be conducted. Given the factor dynamics in the model, we are able to compute the predictive power of conditional expected returns (raw or excess) for the corresponding realized returns. In particular the R^2 of a regression of actual returns on their conditional expectations can be computed as a function of the model parameters. Arguably this measure of predictability is much less susceptible to statistical problems than R^2's of actual regressions of returns on in-sample predictive variables. Also having a joint model of bond prices and stock prices allows us to derive a sensible measure of duration for stocks. For example, one fairly intuitive measure of duration for a stock is the maturity of that zero-coupon bond whose sensitivity to the joint bond-stock factors is most similar to the stock's sensitivity to those same factors. Clearly, without recourse to a joint pricing model for stocks and bonds, such analyses must necessarily be more ad-hoc. As a final benefit of the model, its continuous time setting allows us to price equity derivatives while explicitly taking into account the dependence of the underlying stocks on interest rate fluctuations (though this direction is not pursued in the paper).

Empirical estimation of the model reveals that five factors do a reasonably good job of accounting for the joint variation of stock and government bond prices in the U.S. over the last fifty years.

Two of the factors are mean-reverting joint factors between bonds and stocks, while three of the factors are stock specific. The model is then estimated using a large number of bond prices, and using the returns on three equity portfolios: the market, a small stock portfolio, and a high book to market portfolio.

I find that the implied dividend process for the market index is roughly consistent with the realized dividends for the index starting from the early 70's until today. In particular, in this time period, the best fit implied dividend series from the model does a fairly good (though not perfect) job of tracking the actual dividend yield on the CRSP value weighted index. However, over the entire sample period from 1952–2000, the implied dividend series from the model does a very poor job of tracking the actual dividend series. I find that the market risk premium has been slowly varying over the last fifty years, and has stayed in an interval between 5% and 10%. When plotted against the NBER business cycle peaks and troughs, the equity risk premium appears to be procyclical (high at peaks, and low at troughs). This is in contrast to the countercyclical behavior of the term spread.

I show that, in a particular specification of the model, no arbitrage requires that the R^2's of forecasting regressions of raw or excess returns on their conditional expectations must be a humped-shape function of time, equal to zero at very short time horizons, and again equal to zero at very long time horizons. From the estimation of the model, I find that the maximum theoretical R^2 (the model produces an R^2 measure as a function of the model parameters and of the time horizon of the forecasting regression) of a regression of realized returns on expected returns for the market index occurs at a horizon of 10 years, and has a value of roughly 13%. Surprisingly, I find much less predictability in a regression of excess returns (i.e. returns of the market minus the short-rate) on their expectations: The maximum value here is achieved at about 12 years, but the maximum R^2 is only 0.5%. Hence the model and its estimation suggest that stock predictability is due to a common component that stocks share with bonds, and that this component disappears when we look at excess returns. The news about predictability is roughly the same for the value portfolio used in the model estimation, but the predictability of expected returns for the actual returns of the small stock portfolio is far lower, peaking at only 7% at a time horizon of 10 years.

Drawing on the model's closed form solution for stock prices, and using the five factors extracted from the model estimation, I proceed to analyze how those factors enter into the prices of a large set of different equity portfolios (in particular, I use fifty portfolios representing decile sorts along five stock characteristics). I find that the joint bond-stock factors account for between 5% and 10% of the return variation of a large set of different equity portfolios. The remaining three systematic stock factors proxy for three separate sources of uncertainty in the equity markets. These are associated with the market itself, with the small stock effect, and with the value effect. The reason for this naming convention is that loadings of stocks sorted along (several) value and size dimensions are roughly monotonic with respect to the value and size factors. Since the factors used in this study come directly from a pricing model, these results seem to support the claim that size and value proxy for risk loadings.

Also, using several duration measures from the model, I find that durations for equity portfolios range from 0.5 to 3 years. The general pattern in these duration measures is that small, or growth, or distressed firms have low durations, while established, dividend paying firms have high durations. For example, the pattern in durations as we move from the low book to market deciles to the high book to market deciles has a humped-shape. However, as we move from low to high dividend to price deciles, the duration measure increases monotonically. Furthermore, I document a strong correlation between equity durations and the R^2's of regressions of those equity returns on the joint bond-stock factors.

This paper fits into a recently burgeoning field of bond-stock pricing models based on the affine term structure technology. The main differences between this paper and the related literature are (1) this paper's focus on the total return process for stocks and its derivation of an implied dividend series which can be used to test the model, (2) its ability to jointly price a cross-section of different equities (stocks and portfolios) as well as the term structure of interest rates, and (3) the fact that its empirical implementation relies jointly on a cross-section of bonds (which is usual), as well as on a cross-section of stocks (which is new).

There are several important papers which are closely related to the work presented here. Bekaert and Grenadier (2000) deals with pricing stocks and bonds in a discrete time framework. In this paper the economic primitive is the dividend growth rate, and stock pricing is done via a solution for the dividend to price ratio on the aggregate stock market as a function of the model's state variables. The dividend to price ratio in the model involves an infinite summation of exponentials of affine functions of the state variables. By making the dividend to price ratio, rather than the growth rate of dividends, the economic primitive, I find the much simpler result that log stock prices are affine functions of the state variables. Furthermore, I show that stocks may contain non-bond components, which allows for the pricing of a cross section of equities, something that the Bekaert and Grenadier model does not do. Furthermore, by formulating the model in continuous time I avoid some of the messy issues which arise in a discrete time framework (such as square-root processes which can become negative), and am able to price derivative securities thanks to the model's dynamic completeness. Ang and Liu (2001) employ ideas similar to those in Bekaert and Grenadier (2000) to derive an arbitrage-free affine model for stock pricing based on accounting information (see also Bakshi and Chen (2001)). Another related paper is by Bakshi and Chen (1997b) (which is an empirical implementation of Bakshi and Chen (1997a)) who have a very similar setup to the one in this paper, but who derive their pricing kernel from a time additive, constant relative risk aversion representative agent, and therefore have much stronger coefficient restrictions. For example, in the Bakshi and Chen (1997a,b) model, different stocks cannot have different non-zero loadings on the same stock specific risk factor. Finally, Brennan, Wang, and Xia (2001) develop an affine bond-stock model, and then use it to relate the Fama-French factors to the derived pricing factors from their framework. This paper differs from theirs by having a more general dividend process and pricing kernel, and by considering a cross-section of stocks in the theoretical development and in the estimation of the model.

The remainder of the paper proceeds as follows: Section 2 develops the ideas of this paper in a very simple example economy. Section 3 derives the general form of the model. Section 4 examines the theoretical constructs which the model allows us to compute. Section 5 specializes the model and discusses issues of estimation. Section 6 discusses the results of the model estimation. Section 7 concludes. All proofs are contained in the Appendix.

2 An Example

A great deal of insight can be gained into the main idea underlying this paper by first considering a very simple example. Section 3 develops the general form of the model, and may be read before or after this section. Let us assume that we have M stocks. Each stock m pays an instantaneous dividend given by

$$D_m(t) = \delta_m e^{a_m t - c_m Z(t)},$$

where δ_m, a_m, and c_m are constants, and where $Z(t)$ is a state variable whose dynamics are given by

$$dZ(t) = \mu dt + \sigma dW(t).$$

This dividend process has several appealing features. First it exhibits a time trend, via a_m and μ (presumably μ is negative). Second, the dividend has a random walk component, which is in keeping with empirical evidence. And finally, different stocks may have different sensitivities to the common dividend factor via their values of δ_m and c_m (as will soon be seen, a_m is not a choice variable in an arbitrage-free economy). It may help to think of these dividend paying entities as being well diversified equity portfolios, rather than stocks (for a more detailed discussion of the plausibility of pricing stocks in this setting, see Section 4.1).

Hence each stock entitles its owners to a cumulative dividend process given by

$$\int_0^t D_m(s) ds. \tag{1}$$

For the time being, the interest rate is assumed to be a constant equal to r. Of course, this renders the term structure part of our example quite boring, but the main idea underlying stock pricing will still come across. Also it should be emphasized that the general model implemented in later sections of the paper allows for a very rich dynamical behavior of interest rates and of bond prices.

We will conjecture and then show that stock prices representing claims on the cumulative dividend process in (1) are given by

$$\tilde{P}_m(t) = e^{a_m t - c_m Z(t)}, \tag{2}$$

as long as the constant a_m is appropriately chosen, given any fixed values for δ_m and c_m. With this conjectured form for the stock price, we can rewrite the instantaneous dividend as follows

$$D_m(t) = \delta_m \tilde{P}_m(t),$$

revealing δ_m to be the stock's (fixed, for now) dividend rate. Instead of dragging out the asset pricing machinery which will be used in Section 3, let us try to price stocks in this simple economy by using the classic replication-type argument.

Say we have a portfolio which at time zero holds $x(0)$ shares of stock 1 and of one share of stock 2. The value of this portfolio is given by

$$\Pi(0) = x(0)\tilde{P}_1(0) + \tilde{P}_2(0).$$

The dynamics of this portfolio are therefore given by

$$d\Pi(t) = x(t)\Big(d\tilde{P}_1(t) + \delta_1\tilde{P}_1(t)dt\Big) + \Big(d\tilde{P}_2(t) + \delta_2\tilde{P}_2(t)dt\Big) + r\Big(\Pi(t) - x(t)\tilde{P}_1(t) - \tilde{P}_2(t)\Big)dt.$$

It is easy to check that, given the dynamics of $Z(t)$ and the conjectured stock price in (2), if we chose to hold

$$x(t) = -\frac{\tilde{P}_2(t)c_2}{\tilde{P}_1(t)c_1}$$

shares of stock 1, the portfolio would be instantaneously riskless. By no arbitrage, this instantaneously riskless portfolio must earn the risk free rate, implying that

$$d\Pi(t) = r\Pi(t)dt.$$

From this observation, simple algebra reveals that this condition is equivalent to the following equation

$$\frac{a_1 + \frac{1}{2}c_1^2\sigma^2 + \delta_1 - r}{c_1} = \frac{a_2 + \frac{1}{2}c_2^2\sigma^2 + \delta_2 - r}{c_2}.$$

Denoting the value of either side of this equation as $\tilde{\mu}$, we find that no arbitrage imposes the following condition on the a_m coefficients in the model

$$a_m = r - \delta_m - \frac{1}{2}c_m^2\sigma^2 + c_m\tilde{\mu}. \quad (3)$$

This equation is exactly equivalent to (10), once the latter has been specialized to the simple economy in this example.

We therefore see that given the dividend process in (1), stock prices in this economy are of the form in (2) as long as the a_m coefficient is appropriately chosen. It is interesting to note how easy it is to derive stock prices once a convenient dividend process has been specified. To the extent that the dividend process in (1) provides an adequate description of actual dividend processes in the economy, we may hope that the model would do a reasonable job of fitting actual stock prices.

The one free parameter in (3) is $\tilde{\mu}$, which will later be seen to be the drift of Z under the risk neutral measure. Of course, in the context of our replicating argument this interpretation is not applicable. Identification of this parameter is an empirical issue which will be taken up later in the paper. It should be noted that in the ensuing analysis, no restrictions will be placed on $\tilde{\mu}$, though these can be derived by imposing further economic structure (such as utility maximization) on the model's pricing kernel.

From an empirical point of view, it proves convenient to work not with the stock process itself (which requires knowledge of the dividend process), but rather with the total return process of a given stock (which does not). The total return process of a stock is the wealth that results from reinvesting all of the dividends paid out by the stock back into the stock itself. If we denote the total return process by $P(t)$, and our holdings of the stock by $x(t)$, we find that the dynamics of the total return process are given by

$$dP(t) = x(t)\Big(d\tilde{P}(t) + \delta\tilde{P}(t)dt\Big) + r\Big(P(t) - x(t)\tilde{P}(t)\Big)dt.$$

We capture the idea that dividends are reinvested back into the stock itself, by requiring that

$$x(t) = \frac{P(t)}{\tilde{P}(t)},$$

which means that the total return process may be written as follows

$$\frac{dP(T)}{P(t)} = \frac{d\tilde{P}(t)}{\tilde{P}(t)} + \delta dt.$$

Using the dynamics of the stock price, and the form of the a_m coefficient from (3), an application of Ito's lemma allows us to conclude that

$$P_m(t) = e^{\left(r - \frac{1}{2}c_m^2\sigma^2 + c_m\tilde{\mu}\right)t - c_m Z(t)}.$$

This again provides a preview for the more general result to be found in Lemma 2.

The remainder of the paper is devoted to affixing various bells and whistles (like a stochastic interest rate) to the model outlined in this section, and to the estimation of the thusly enhanced model using U.S. data over the last fifty years.

3 A Joint Model for Stocks and Bonds

It is a well known result that under many circumstances, the lack of arbitrage opportunities is equivalent to the existence of a risk-neutral measure. The significance of this fact is that any security's current price is simply the expectation of the discounted future cash flows of that security under the risk-neutral measure. This section shows how this idea may be used to derive a tractable, arbitrage-free pricing model for bonds and stocks. The section presents the assumed factor dynamics in the economy, the bond and stock pricing equations, and the price of risk process which allows us to move from the physical to the risk-neutral measure.

3.1 The General Model

We assume the existence of two types of state variables: $Y = \{Y_1, \ldots, Y_N\}$ and $Z = \{Z_1, \ldots, Z_M\}$. The Y variables are mean reverting under the risk-neutral (and the physical) measure Q, and the Z variables are not. The reason for making this distinction will become clear shortly. The dynamics of the two sets of state variables are as follows:

$$dY(t) = \tilde{K}_Y(\tilde{\Theta} - Y(t))dt + \Sigma_Y\sqrt{S(t)}d\tilde{W}(t), \quad (4)$$

$$dZ(t) = \tilde{\mu}dt - \tilde{K}_Z Y(t)dt + \Sigma_Z\sqrt{S(t)}d\tilde{W}(t). \quad (5)$$

$\tilde{W}(t)$ is a standard $N+M$ dimensional Brownian motion under Q, Σ_Y is an $N \times (N+M)$ matrix, Σ_Z is an $M \times (N+M)$ matrix, and $[\Sigma'_Y \Sigma'_Z]'$ is of full rank. We assume that the $N \times N$ matrix $\tilde{K}_Y$ is of full rank, though the $M \times N$ matrix $\tilde{K}_Z$ may be of arbitrary rank, or even zero. $S(t)$ is an $(N+M) \times (N+M)$ diagonal matrix, with elements given by[1]

$$[S(t)]_{ii} = \alpha_i + \beta'_i Y(t). \tag{6}$$

Note that only the stationary variables $Y(t)$ are allowed to have an effect on the innovations variances of both sets of state variables.

The instantaneous interest rate in this economy is an affine function of the state variables as follows

$$r(t) \equiv r_0 + r'_Y Y(t) + r'_Z Z(t), \tag{7}$$

for a constant r_0, an N dimensional vector r_Y, and an M dimensional vector r_Z. We will assume that the dividend rate on any security can be expressed as an affine function of the Y state variables (and may potentially be zero). Letting $\delta(t)$ denote the time t dividend rate, we require that

$$\delta(t) = \delta_0 + \delta'_Y Y(t) + \delta'_Z Z(t). \tag{8}$$

Note that under certain parameterizations of the present model, both dividend rates and interest rates may be negative.[2]

As is shown in the Appendix, absence of arbitrage implies that any security, whether it pays dividends or not, has a price process given by

$$P(t) = e^{A(t) - B(t)'Y(t) - C(t)'Z(t)}, \tag{9}$$

for some functions of time $A(t)$ (a scalar), $B(t)$ (an $N \times 1$ vector), and $C(t)$ (an $M \times 1$ vector). The following theorem gives the conditions which no arbitrage imposes on the price function.

Theorem 1 *Given the model specifications, the price process in (9) is consistent with no arbitrage as long as the following ordinary differential equations are satisfied, subject to the appropriate boundary conditions.*

$$0 = \delta_0 - r_0 + A_t(t) - \tilde{\Theta}'\tilde{K}'_Y B(t) - \tilde{\mu}'C(t) + \frac{1}{2}\sum_{i=1}^{N+M}\Big([\Sigma'_Y B(t)]_i + [\Sigma'_Z C(t)]_i\Big)^2 \alpha_i, \tag{10}$$

$$0 = \delta_Y - r_Y - B_t(t) + \tilde{K}'_Y B(t) + \tilde{K}'_Z C(t) + \frac{1}{2}\sum_{i=1}^{N+M}\Big([\Sigma'_Y B(t)]_i + [\Sigma'_Z C(t)]_i\Big)^2 \beta_i, \tag{11}$$

$$0 = \delta_Z - r_Z - C_t(t). \tag{12}$$

The t subscripts denote partial derivatives with respect to time.

[1]The notation $[\cdot]_i$ refers to the i^{th} element in a vector, and $[\cdot]_{ij}$ refers to the element in the i^{th} row and j^{th} column of a matrix.

[2]Note that this is a problem common to all affine pricing models. See Dai and Singleton (2001) for a discussion of this drawback of affine models.

Bonds.

Let us denote the time t price of a default-free zero coupon bond which matures at time T as $P(t,T)$. Such zero-coupon bonds do not pay dividends, and hence have $\delta(t) = 0$. At maturity we require that $P(T,T) = 1$, which implies the following boundary conditions for the bond's coefficients $A(t), B(t), C(t)$:

$$A(T) = B(T) = C(T) = 0. \tag{13}$$

From this and from condition (12) we conclude that $C(t) = (T-t)r_Z$. This implies that the long yield (i.e. the limit of $\log(P(t,T))/T$) must contain a dependence on the $Z(t)$ variables. However, Dybvig, Ingersoll, and Ross (1996) show that no arbitrage puts certain restrictions on the behavior of the zero coupon long rate. In particular the zero coupon long rate can never fall. The present model violates this restriction unless $r_Z = 0$. Therefore, the following restriction is imposed.

Restriction 1 $r_Z = 0$.

Hence we see that for bonds, $C(t) = 0$ for all t. With this restriction, the following lemma gives bond prices in the economy.

Lemma 1 *Assuming that Restriction 1 holds, zero coupon bond prices are given by (9), where the coefficients satisfy the following set of ODE's*

$$0 = -r_0 + A_t(t) - \tilde{\Theta}'\tilde{K}_Y'B(t) + \frac{1}{2}\sum_{i=1}^{N+M}[\Sigma_Y'B(t)]_i^2\alpha_i, \tag{14}$$

$$0 = -r_Y - B_t(t) + \tilde{K}_Y'B(t) + \frac{1}{2}\sum_{i=1}^{N+M}[\Sigma_Y'B(t)]_i^2\beta_i. \tag{15}$$

and $C(t) = 0$. The boundary conditions are $A(T) = B(T) = 0$ for zeros maturing at time T.

Therefore, despite the presence of the Z variables, zero coupon bond prices are identical to the standard affine pricing equations (see Constantinides (1992) and Duffie and Kan (1996)). However, as will be shown shortly, stock prices may depend on the Z's. Furthermore, there may be some Y factors which affect only stock prices (since the r_Y vector is free to have an arbitrary number of zeros). However, as will be shown momentarily, stock prices in general must depend on the bond factors. Hence we will refer to the set of Y factors which are used in pricing bonds as the "common bond-stock factors."

Coupon bonds, of course, can be priced as a combination of zeros. Hence the extant literature on the pricing of bonds in an affine setting applies to the present model (see, for example, Dai and Singleton (2000)). Also, the present analysis may be extended to the case of individual stocks and defaultable debt using the techniques of Duffie and Singleton (1999).

Stocks.

Consider a stock whose price is given by $\tilde{P}(t)$, and which entitles its shareholders to a cumulative dividend process $\int \delta\tilde{P}ds$. Absence of arbitrage requires that under the risk neutral-measure, the cumulative discounted gain process from a stock follows a martingale. Hence we must have that

$$\tilde{P}(t) = \mathrm{E}_t^Q \left[\int_t^T e^{-\int_t^s r(h)dh} \delta(s) \tilde{P}(s) ds + e^{-\int_t^T r(h)dh} \tilde{P}(T) \right] \tag{16}$$

for all T (see Harrison and Pliska (1981) and Dybvig and Huang (1989)). We also impose the transversality condition that

$$\lim_{T \to \infty} \mathrm{E}_t^Q \left[e^{-\int_t^T r(h)dh} \tilde{P}(T) \right] = 0. \tag{17}$$

Here E_t^Q denotes the conditional expectation operator under the risk-neutral measure.[3] Let us consider another security $P(t)$ which is generated from reinvesting the dividend flow from the stock back into the stock itself. It is then straightforward to show that under the appropriate integrability conditions, the discounted value of this total return security must be a martingale under the risk-neutral measure, or

$$P(t) = \mathrm{E}_t^Q \left[e^{-\int_t^T r(h)dh} P(T) \right], \tag{19}$$

for all T (see Lemma 1 in Dybvig and Huang (1989)). Of course, the transversality condition for the total return process is that $P(t)$ equals the expectation of the discounted future total return process over every (including the infinite) time horizon.

Going forward, it will be convenient to work with the total return security $P(t)$ rather than the dividend paying stock $\tilde{P}(t)$. For convenience, let us refer (incorrectly) to the m^{th} total return process $P^m(t)$ as the m^{th} stock (or stock portfolio). Unless otherwise noted, every reference to a stock going forward refers to the total return process on that stock. We will conjecture that the value of this total return security satisfies

$$P^m(t) = e^{A^m(t) - B^m(t)' Y(t) - C^m(t)' Z(t)}. \tag{20}$$

In order for the discounted total return security to be a martingale, $A^m(t), B^m(t), C^m(t)$ must satisfy the pricing equations in (10, 11, 12), subject to the appropriate boundary conditions.[4] We will assume that the dividend rate for stocks must be stationary, and hence that $\delta_Z = 0$. From this and the requirement that $r_Z = 0$, we find that C^m must be independent of time, and therefore a constant. We now impose another restriction on the parameters of the model.

[3] This transversality condition can be computed as follows. Define $\xi(T) \equiv e^{-\int_t^T r(h)dh} \tilde{P}(T)$. Applying Ito's lemma and the pricing condition in (51) from the Appendix, simple algebra reveals that

$$d\xi(T) = -\delta(T)\xi(T)dT + \Sigma_\xi(T) d\tilde{W}(T), \tag{18}$$

for some matrix $\Sigma_\xi(T)$, potentially a function of the state variables. It is possible to check in some cases of the model that this implies that $\mathrm{E}_t^Q[\xi(T)]$ goes to zero as T gets large.

[4] The extension from pricing a total return process on a stock to pricing the total return on portfolios is straightforward. If for all n, $P^m(t)$ satisfies the PDE implied by (19), then the price of a portfolio of stocks (i.e. $\sum_{m=1}^M x_m P^m(t)$) also satisfies the same PDE. This follows because the dividend rate for total return securities is zero, and all the other terms in the PDE are linear in P. Hence we can think of (20) as being the price for the total return on a portfolio of stocks or for an individual stock, knowing that this is consistent with all portfolio and stock total return processes satisfying (19). Also a portfolio of dividend paying stocks satisfies (16), and the associated partial differential equation, as long as each stock does, and as long as the dividend rate on the portfolio is appropriately chosen: For a portfolio dividend rate δ_Π, we require that $\delta_\Pi \sum_m x_m P^m = \sum_m x_m \delta^m P^m$.

Restriction 2 $B_t^m(t) = 0$.

This insures that the loading B^m on the mean-reverting state variables $Y(t)$ is also constant. The value of a total return process in this economy is given by the following lemma.

Lemma 2 *Assuming that Restrictions 1 and 2 hold, and that total return processes for stocks are indeed of the form in (20), the coefficients from that equation must satisfy the following restrictions:*

$$a^m = r_0 + \tilde{\Theta}'\tilde{K}_Y' B^m + \tilde{\mu}' C^m - \frac{1}{2}\sum_{i=1}^{N+M}\left([\Sigma_Y' B^m]_i + [\Sigma_Z' C^m]_i\right)^2 \alpha_i, \qquad (21)$$

$$0 = -r_Y + \tilde{K}_Y' B^m + \tilde{K}_Z' C^m + \frac{1}{2}\sum_{i=1}^{N+M}\left([\Sigma_Y' B^m]_i + [\Sigma_Z' C^m]_i\right)^2 \beta_i. \qquad (22)$$

where $a^m \equiv A_t^m(t)$ *and is a constant, and where* C^m *can be an arbitrary* $M \times 1$ *vector of constants.*

Noticing that we can then write

$$A^m(t) = A^m(0) + a^m \times t,$$

log total return processes in this economy are given by

$$\log P^m(t) = A^m(0) + a^m \times t - B^{m\prime} Y(t) - C^{m\prime} Z(t). \qquad (23)$$

It should be noted that this process for the stock total return nests virtually all empirical specifications for stock prices which have been employed in the literature.

We next need to check that total return processes in this economy do indeed satisfy the form conjectured in (20). This essentially amounts to checking a particular type of boundary condition. While the boundary condition for a zero-coupon bond (i.e. that its price should be equal to one at some future time) is very straightforward, the analogous boundary condition for the total return process on a stock is a little trickier. The total return process given in (20) must be equal at all times to the value of a portfolio which invests initially in a dividend paying stock, and then reinvests all dividends back into that stock.

Going forward, let us define two sets of Z-type variables: one set for total return processes (the Z vector), and another set for stock price processes (the Z^s vector). Under the risk-neutral measure, the dynamics of Z are given in (5), and the dynamics of Z^s are given by

$$dZ^s(t) = \tilde{\mu}dt - \tilde{K}_Z^s Y(t)dt + \Sigma_Z\sqrt{S(t)}d\tilde{W}(t). \qquad (24)$$

Note that the only difference between dZ and dZ^s is the $\tilde{K}_Z^s$ matrix. Our goal is to find a vector of state variables $Z^s(t)$ such that if total return processes have the form in (20), then the actual stock processes will have the same functional form but with the $Z(t)$ vector replaced by the $Z^s(t)$ vector. The following theorem will show that equality of the total returns process on a stock with the returns on a portfolio which reinvests all proceeds back into the stock itself places a certain restriction on the $\tilde{K}_Z$ and $\tilde{K}_Z^s$ matrixes.

Theorem 2 *Given the short rate process in (7) and any infinitely lived security whose instantaneous dividend yield is given by (8), the total return process does indeed satisfy the form conjectured in (20) as long as the following conditions hold for each stock* $m = 1, \ldots, M$*:*

$$C^{m\prime}(\tilde{K}_Z - \tilde{K}_Z^s) = \delta_Y^{m\prime}. \tag{25}$$

Let us define the following matrixes

$$\begin{aligned} C &\equiv [C^1\, C^2 \cdots C^M], \\ \delta_Y &\equiv [\delta_Y^1\, \delta_Y^2 \cdots \delta_Y^M]. \end{aligned}$$

Assuming that C *if of full-rank, we also have that*

$$\tilde{K}_Z^s = \tilde{K}_Z - (C')^{-1}\delta_Y'. \tag{26}$$

Total return processes are given by (20). This theorem tells us that given a total returns process, the associated stock price process has the same form as in (20), with the $Z(t)$ vector replaced with the $Z^s(t)$ vector whose dynamics are given in (24), with the $\tilde{K}_Z^s$ matrix from (26). This theorem, therefore, allows us to move from total returns processes to stock price processes in a way which maintains consistency of the model.

According to the above theorem, any choice of δ_0 and δ_Y which is consistent with the transversality condition in (17) is admissible. In other words, a $\tilde{K}_Z^s$ matrix can always be chosen which supports any admissible choice of δ_0 and δ_Y. Hence the model allows us some flexibility in fitting the actual dividend series. The one requirement, of course, is that the Y factors extracted from the model must be sufficiently informative about the actual dividend yields of the stocks used in the model estimation. Indeed, this observation provides one method of testing the model: Estimate the model using only total returns processes, and no dividend information, and then check how much information the extracted Y factors contain about actual dividend yields. This testing approach is pursued in the empirical part of the paper. It may also be useful to look at the dividend rate over a longer time horizon. We can define the expected dividend yield over a time horizon τ as follows

$$\begin{aligned} \bar{\delta}(t, t+\tau) &\equiv \frac{1}{\tau}\mathrm{E}_t\left[\int_t^{t+\tau} \delta(s)ds\right] \\ &= \delta_0 + \frac{\delta_Y'}{\tau}\mathrm{E}_t\left[\int_t^{t+\tau} Y(s)ds\right]. \end{aligned} \tag{27}$$

The second equality follows from the fact that in the model $\delta(t) = \delta_0 + \delta_Y' Y(t)$. This expected dividend rate can often be computed in closed form, and may be a more useful empirical entity that the instantaneous dividend rate because the latter is not observable.

3.2 Price of Risk and Expected Excess Returns

In order to implement the model empirically, we need to specify the factor dynamics under the physical measure. Let us assume that $\Lambda(Y, t)$ is the price of risk process, meaning that $d\tilde{W}(t) =$

$dW(t) + \Lambda(Y,t)dt$ where $W(t)$ is a standard Brownian motion under the physical measure. While there are numerous ways to specify the price of risk process $\Lambda(\cdot)$, I choose one which insures that the general form of the processes for Y and Z are the same under the physical and the risk-neutral measures. The price of risk process is given by

$$\Lambda(Y,t) \equiv S(Y,t)^{-\frac{1}{2}} \left(\lambda_0 + \lambda_Y Y(t)\right), \tag{28}$$

where λ_0 is an $(N+M) \times 1$ vector, and where λ_Y is an $(N+M) \times N$ matrix.[5] It is then straightforward to verify that under the physical measure, the dynamics of Y, Z, and Z^s are given by

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

$$dZ(t) = \mu dt - K_Z Y(t)dt + \Sigma_Z \sqrt{S(t)}dW(t), \tag{30}$$

$$dZ^s(t) = \mu dt - K_Z^s Y(t)dt + \Sigma_Z \sqrt{S(t)}dW(t), \tag{31}$$

where $K_Y = \tilde{K}_Y - \Sigma_Y \lambda_Y$, $\Theta = K_Y^{-1}(\tilde{K}_Y \tilde{\Theta} + \Sigma_Y \lambda_0)$, $\mu = \tilde{\mu} + \Sigma_Z \lambda_0$, $K_Z = \tilde{K}_Z - \Sigma_Z \lambda_Y$, and $K_Z^s = \tilde{K}_Z^s - \Sigma_Z \lambda_Y$. From these and the results of Theorem 2 we get that $K_Z^s = K_Z - (C')^{-1}\delta_Y'$.

A note is in order about this price of risk specification: Care must be taken to insure that arbitrage opportunities are precluded in the model.[6] For example, for the case of a single state variable which follows a Cox, Ingersoll, Ross type square root process (i.e. $dY(t) = K(\theta - Y(t))dt + \sigma\sqrt{Y(t)}dW(t)$), we must have that $\lambda_0 = 0$. Otherwise, as was shown by Cox, Ingersoll, and Ross (1985), the model will admit arbitrage opportunities. For a more detailed discussion of the restrictions which no arbitrage imposes on a price of risk process of the type shown in (28), see Duffee (2001) and Dai and Singleton (2001). In the ensuing empirical implementation of the model, the $S(t)$ matrix will simply be the identity matrix (implying Gaussian innovations), in which case no restrictions arising from these considerations need to be placed on λ_0 and λ_Y in (28).

Subject to the above caveat, given the values of K_Y, Θ, K_Z, and μ (under the physical measure), as well as the corresponding variables under the risk-neutral measure, the vector λ_0 and the matrix λ_Y may be recovered as follows:

$$\lambda_0 = \begin{bmatrix} \Sigma_Y \\ \Sigma_Z \end{bmatrix}^{-1} \begin{pmatrix} K_Y\Theta - \tilde{K}_Y\tilde{\Theta} \\ \mu - \tilde{\mu} \end{pmatrix}, \tag{32}$$

$$\lambda_Y = \begin{bmatrix} \Sigma_Y \\ \Sigma_Z \end{bmatrix}^{-1} \begin{pmatrix} \tilde{K}_Y - K_Y \\ \tilde{K}_Z - K_Z \end{pmatrix}. \tag{33}$$

Note also that any parameter constraints can be accommodated by the appropriate choice of the prices of risk. For example, the constraint that $\tilde{K}_Z = K_Z$ simply implies a particular value for λ_Y chosen according to (33).

[5] If we denote the pricing kernel by $m(t)$, then from standard results we can write innovations to the pricing kernel as

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

Keeping in mind, that Λ and W are both $(N+M) \times 1$ dimensional vectors, kernel innovations will, in general, be correlated with innovations in the Y and in the Z factors.

[6] I thank Ken Singleton for bringing this point to my attention.

Once the dynamics of the state variables are known under both measures, it is possible to derive the expected excess return process for any security in this economy. Therefore a security's expected excess return process may be extracted from an estimation of the model. The following Theorem states the relevant results, and is proved in the Appendix.

Theorem 3 *Given the price of risk process $\Lambda(\cdot)$ in (28), the instantaneous expected excess return on any security is given by*

$$\begin{aligned} \frac{\mathrm{E}[dP(t)]}{P(t)} + \delta(t) - r(t) &= -B' \left([K_Y \Theta - \tilde{K}_Y \tilde{\Theta}] + [\tilde{K}_Y - K_Y] Y(t) \right) \\ &\quad - C' \left([\mu - \tilde{\mu}] + [\tilde{K}_Z - K_Z] Y(t) \right). \end{aligned} \tag{34}$$

Note first, that bonds, by virtue of having $C = 0$, have excess returns which depend only on the first term in (34). The excess return for stocks, for which C is a constant, therefore contains a non-bond component. Furthermore, given the above $\Lambda(\cdot)$ specification, only the stationary variables Y may induce variability in expected excess returns. Hence, the excess return process is itself stationary.

4 Analysis of the Model

Our goal thus far has been to develop an empirically flexible, economically coherent framework for thinking about the joint behavior of stocks and bonds. Now that the model has been fully specified, we next discuss some of the issues which arise in the model, and also consider the types of questions which the model allows us to address.

4.1 On the Pricing of Stocks

By their nature, no-arbitrage pricing arguments rely on the existence of redundant securities. The general results say that given the existence of any one security which loads on a particular factor, any other security which loads on that same factor must have restrictions placed on the behavior of its price. While it seems easy to imagine a multiplicity of zero coupon bonds, all with different maturities, which are driven by the same set of factors, it seems harder to imagine many stocks, all of which load on the same set of Z factors, and whose prices must therefore obey Lemma 2. In particular, we have the strong prior that each stock loads on its own idiosyncratic factor, and that no other stock loads on this same factor. However, our prior with respect to the idiosyncratic behavior of portfolios must be somewhat weaker.

Let us consider an economy with $\tilde{M}$ stocks, and with $M + \tilde{M}$ factors of the Z-type. However, let us assume that of these only $M \ll \tilde{M}$ factors are systematic, in that they affect all stocks, whereas the remaining $\tilde{M}$ factors are stock specific. We can then employ an APT type (Ross (1976)) argument to say that in the limit, as $\tilde{M} \to \infty$, only the M systematic factors should be priced, in that $\mu_m = \tilde{\mu}_m$ for any factor outside of the M systematic ones. This implies that all well diversified portfolios will contain only the M systematic stock factors, as well as the N joint bond-stock factors. Since there will be an infinite number of well diversified portfolios, all of which are driven by the M stock specific factors (in addition to the N joint factors), the stock pricing arguments used in this paper will apply directly, but only to well diversified portfolios.

Indeed, the existence of such perfectly diversified portfolios may not even be necessary. All that is needed for the arguments of this paper to apply is that equity portfolios are cointegrated. In this case, no arbitrage restricts portfolio total returns processes to have the form given in Lemma 2. Bossaerts (1988) provides evidence that stock portfolios are indeed cointegrated, although whether or not this is the case is still an open question.

In order to apply the arguments of this paper to individual stocks, we would need to believe that individual stocks are also cointegrated. Though the efficacy of pairs trading strategies (see, for example, Gatev, Goetzmann, and Rouwenhorst (2001)) suggests that at least some stocks must be cointegrated, it is hard to imagine that all stocks are cointegrated. Even so, it is important to realize that this model's pricing equations for stocks are the only equations which prohibit arbitrage in the event that a redundant security were actually introduced into the economy. Hence the model's pricing system is sufficient, though not necessary, for no arbitrage to obtain amongst individual stocks. Related to this point, it may also be argued that, while redundant stocks do not literally exist, nearly redundant ones do, and that absence of "near" arbitrages (such as investments with very high Sharpe ratios) also implies the same pricing formulas as the ones presented in this paper. It would be interesting to make this latter point in a more rigorous fashion.

4.2 Forecasting Returns

In the previous section, we derived the instantaneous risk premium on securities in the model. It is also relatively straightforward to compute expected (conditional and unconditional) 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

$$R(t,\tau) \equiv \log \frac{P(t+\tau)}{P(t)} = a \times \tau - B'(Y(t+\tau) - Y(t)) - C'(Z(t+\tau) - Z(t)). \qquad (35)$$

We can therefore write

$$\mathrm{E}_t[R(t,\tau)] = a \times \tau - B'(\mathrm{E}_t[Y(t+\tau)] - Y(t)) - C'(\mathrm{E}_t[Z(t+\tau)] - Z(t)). \qquad (36)$$

In many cases of the model, these conditional expectations are rather straightforward to compute. We will do so in a later section of this paper. Obviously, bond returns are known over any desired time period and are linear in the state variables. Hence it is possible to compute the expected excess returns of stocks over bonds over any desired time horizon.

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

$$R_r(t,\tau) \equiv \int_t^{t+\tau} r(s)ds. \qquad (37)$$

Under some specifications of the model dynamics, we will later show that it is possible to compute $\mathrm{E}_t[R_r(t,\tau)]$ in closed form.

Given our discussion so far, it is possible to write the log return of an investment in any security as follows

$$R(t,\tau) = \mathrm{E}_t[R(t,\tau)] + \epsilon(t+\tau).$$

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

$$R^2(\tau) = \frac{\mathrm{Var}(\mathrm{E}_t[R(t,\tau)])}{\mathrm{Var}(\mathrm{E}_t[R(t,\tau)]) + \mathrm{Var}(\epsilon(t+\tau))}, \tag{38}$$

since the error term is orthogonal to the expectation. Also note that we are computing unconditional variances. This calculation is possible because from (36), and the factor dynamics, we see that $\mathrm{E}_t[R(t,\tau)]$ 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.

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 mis-specification). 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.

4.3 Equity Duration

In a model with a constant yield curve, zero coupon bond prices are given by $e^{-r(\tau)\tau}$. Duration in such models is the percent change in the price of a bond due to a shift in the appropriate interest rate. For a zero coupon bond with τ years left to maturity, the duration is given by

$$-\frac{dP(t,t+\tau)}{r(\tau)}\frac{1}{P(t,t+\tau)} = \tau.$$

In this framework, the duration of a portfolio of bonds is determined by assuming that the yield curve undergoes an infinitesimal parallel shift. Of course, for the parallel shift in the yield curve, the local percent change in the value of a given portfolio will be exactly equal to the percent change in the value of a zero whose maturity is equal to that portfolio's duration.

In a dynamic model with several factors Y, no single dimensional measure of duration exists.[7] Instead duration is a vector of sensitivities to the model factors. In the present framework, this vector is given by

$$-\frac{dP(t,t+\tau)}{Y(t)}\frac{1}{P(t,t+\tau)} = B(t,t+\tau),$$

[7] For a detailed discussion of the concept of duration in a multi-factor term structure model see Jeffrey (2001). A recent model of equity duration, similar in spirit to the present model, can be found in Lewin and Satchell (2001).

for a zero coupon bond of maturity τ. For stocks, this duration measure is given by

$$-\frac{dP^m(t)}{Y(t)}\frac{1}{P^m(t)} = B^m.$$

If the price of a portfolio containing x_i zeros of a given maturity τ_i is Π, we can write the multi-dimensional duration for that portfolio as

$$-\frac{d\Pi}{dY(t)}\frac{1}{\Pi(t)} = \frac{\sum_i x_i P(t,t+\tau_i)B(t,t+\tau_i)}{\sum_i x_i P(t,t+\tau_i)}.$$

Unfortunately, such multi-dimensional duration measures are very difficult to interpret. Therefore, we would like to develop a notion of duration as a single number, whose interpretation is roughly analogous to the duration concept in a static model.

With this goal in mind, one reasonable notion of duration for a portfolio in a dynamic setting is the maturity of the zero coupon bond whose sensitivity to the interest rate factors is closest to the portfolio in question. In the present framework, for a stock m this requires solving the following minimization

$$\tau_m^{(1)} = \arg\min_{\tau}(B(t,t+\tau) - B^m)'(B(t,t+\tau) - B^m).$$

Another measure of duration can be obtained as follows: For a model with N factors of type Y, fix a set of $N+1$ zeros of fixed maturities $\tau_1, \cdots, \tau_{N+1}$. Then solve the following for $x_1, \ldots, x_{N+1}$:

$$\frac{\sum_i x_i P(t,t+\tau_i)B(t,t+\tau_i)}{\sum_i x_i P(t,t+\tau_i)} = B^m.$$

Given $N+1$ zeros, this equation can be solved exactly. The duration of the stock is then given by its analog in a static framework, or

$$\tau_m^{(2)} = \frac{\sum_i x_i P(t,t+\tau_i)\tau_i}{\sum_i x_i P(t,t+\tau_i)}.$$

It turns out that these two approaches (i.e. $\tau^{(1)}$ and $\tau^{(2)}$) give numerically very similar answers, and results for both will be reported.

Keep in mind that the zeros referred to in the above analysis are priced using the pricing equations from the model (see Lemma 1). Once the model has been estimated, it is therefore possible to perform the above analysis regardless of whether a given zero actually trades or not. Indeed, it is this fact which renders the approach proposed in this paper so useful.

4.4 Extracting Stock Specific Factors

While in this paper we will estimate the joint model specified above, it is possible to use the analysis so far to develop a very simple method for separating out the equity specific component of total returns on stocks from the common component which stocks share with bonds. To see how this may be done, let us first, for convenience, us reproduce here the dynamics of the Y and Z factors under the physical measure.

$$dY(t) = K_Y(\Theta - Y(t))dt + \Sigma_Y\sqrt{S(t)}dW(t), \tag{39}$$

$$dZ(t) = \mu dt - K_Z Y(t)dt + \Sigma_Z\sqrt{S(t)}dW(t), \tag{40}$$

From our analysis so far, we know that zero coupon bond prices are given by

$$\log P(t, t+\tau) = A(\tau) - B(\tau)'Y(t),$$

and that the total return processes for stocks are given by

$$\log P^m(t) = A^m(0) + a^m \times t - B^{m\prime}Y(t) - C^{m\prime}Z(t).$$

It would seem, therefore, that one may isolate the stock specific factors Z by simply regressing the first difference of log stock prices on the first difference of the appropriate number (equal to the number of common factors) of log zero prices. However, we note that the first difference of the Z factors is given by

$$Z(t+\tau) - Z(t) = \mu\tau + \Sigma_Z \int_t^{t+\tau} \sqrt{S(h)}dW(h) - K_Z \int_t^{t+\tau} Y(h)dh.$$

It is reasonable to assume that the Σ_Z matrix renders Z innovations to be uncorrelated with innovations to Y. However, to the extent that the dividend rates on stocks depend on the common stock bond factors (which is likely), the Y integral in the first difference of the Z factors will be correlated with $Y(t+\tau) - Y(t)$. Therefore, the error term of a regression of log stock prices on log bond prices will be correlated with the regressors, leading to biased estimates.

Of course, the magnitude of this bias depends on the magnitude of the correlation between first differences in the Y's and first differences in the Z's. As will be seen later, this correlation for the U.S. market seems to be quite low, suggesting that the bias from this regression based approach will be quite small. However, it should be emphasized that this is an empirical result, and though it may have been the expected result, it was not ex-ante obvious that this would be the case.

5 Model Implementation

In this section we will discuss how the model of the previous section is implemented and estimated. First, we will discuss the restrictions which are made on the factors dynamics, then we will derive expected returns and R^2's in closed form, and finally we will discuss an estimation strategy for the model.

5.1 The Factor Dynamics

This section describes a parameterization of the general model of Section 3.1 in which the factors innovations have constant volatilities. This leads to a particularly tractable version of the model in which both the pricing equations and the factor transition probabilities are known in closed form.[8]

We will assume that the K matrixes in (4,5) are diagonal under both the risk-neutral and the empirical measures. The matrix $S(t)$ is assumed to be the identity matrix. This forces innovation

[8]The bulk of empirical evidence suggests that the assumption of constant innovation volatilities is contradicted for both fixed-income and equity markets. The imposition of this constraint is potentially justified only on the grounds that it so dramatically simplifies the analysis of the model and of the factor transition densities. This being a first attempt at this type of analysis, this tractability seems especially useful. Of course, an important area for future work is to extend the analysis in this paper to incorporate stochastic volatilities for both the stock and bond factors.

volatilities to remain constant. Also, the $(M+N) \times (M+N)$ matrix $[\Sigma_Y' \Sigma_Z']'$ is assumed to be diagonal (which implies that all factor innovations are uncorrelated). We will assume that $r_Y = 1$ (which is a normalization).

The price of a stock or bond in this model is given by (9). Given this specification, the coefficients for the price of a zero-coupon bond maturing at time T are the standard Vasicek ones, and are given by

$$
\begin{aligned}
B_i(t) &= \frac{r_{Yi}}{\tilde{K}_{ii}}\left(1 - e^{\tilde{K}_{ii}(t-T)}\right), \\
A(t) &= -(r_0 + \tilde{\Theta}' r_Y)(T-t) + \tilde{\Theta}' B(t) + \frac{1}{2}\sum_{i=1}^{N} \sigma_{Yi}^2 \int_t^T B_i^2(s) ds.
\end{aligned}
$$

for $i = 1, \ldots, N$. For a stock m, the coefficients are as follows

$$B^m = (\tilde{K}_Y')^{-1}(r_Y - \tilde{K}_Z' C^m), \tag{41}$$

$$a^m = r_0 + \tilde{\Theta}' \tilde{K}_Y' B^m + \tilde{\mu}' C^m - \frac{1}{2}\sum_{i=1}^{N} \sigma_{Yi}^2 B_i^{m2} - \frac{1}{2}\sum_{j=1}^{M} \sigma_{Zj}^2 C_j^{m2}, \tag{42}$$

$$C^m = \text{Arbitrary } M \times 1 \text{ Vector.} \tag{43}$$

The log bond and stock prices are therefore given by

$$
\begin{aligned}
\log P^m(t) &= A^m(0) + a^m \times t - B^{m\prime} Y(t) - C^{m\prime} Z(t), \\
\log P(t,T) &= A(t,T) - B(t,T)' Y(t).
\end{aligned}
$$

Note that all of the stock and bond coefficients are known in closed form. Also, the stock contains a mixture of persistent and mean-reverting components, whereas the bond contains only mean-reverting components. Interestingly, this specification represents a generalization of the one used in Fama and French (1988). Instantaneous expected excess returns on a given security are given by (34). Estimation of the B and C matrixes for total return processes in this economy implies a particular K_Z matrix. We will assume that $K_Z = \tilde{K}_Z$.

The following Theorem establishes that the transversality condition in (17) is satisfied as long as a particular parameter restriction holds.

Theorem 4 *Given the assumptions on factor dynamics made in this section, the transversality condition in (17) holds as long as the following parameter restriction is satisfied*

$$\delta_0 + \delta_Y' \tilde{\Theta} - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{\delta_{Yi}\sigma_{Yi}}{\tilde{K}_{ii}}\right)^2 > 0. \tag{44}$$

The proof is in the Appendix.

A note is in order about the number of factors which are assumed by the model. Since we will employ only $N+M$ factors, the model will be able to exactly price at most $N+M$ securities at any given time. The way this problem is resolved in the bond market is by assuming that all the other bond prices are observed with error (the error, of course, corresponding to the disagreement between the model and the observed price). This approach was introduced by Chen and Scott (1993).

As we have already discussed, this interpretation is invalid for the stock market. Instead, let us think of the model as follows. We will estimate the model using only M equity portfolios which are thought to be sufficiently diverse. This exercise will yield M systematic factors. In order for the model to account for the exact prices of any number of other stocks or stock portfolios we will introduce an idiosyncratic return factor for each stock. Of course to the extent that the idiosyncratic factors thus derived have commonalities, the model either has too few systematic factors or is misspecified.

5.2 Predictability

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 theorems state the relevant results.

Theorem 5 *With the assumptions about factors dynamics which were made in Section 5.1 we can decompose returns on a rolled-over investment at the short-rate as follows:*

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

where

$$\mathrm{E}_t[R_r(t,\tau)] = r_0 \times \tau + \sum_{i=1}^{N} r_{Yi}\left[\theta_i \times \tau + \frac{Y_i(t)-\theta_i}{K_{ii}}\left(1-e^{-K_{ii}\tau}\right)\right].$$

Furthermore, we have that

$$\mathrm{Var}\Big(\mathrm{E}_t[R_r(t,\tau)]\Big) = \sum_{i=1}^{N} r_{Yi}^2 \frac{\sigma_{Yi}^2}{2K_{ii}^3}\left(1-e^{-K_{ii}\tau}\right)^2,$$

$$\mathrm{Var}\Big(\epsilon_r(t+\tau)\Big) = \sum_{i=1}^{N} r_{Yi}^2 \frac{\sigma_{Yi}^2}{K_{ii}^2}\left[\tau - \frac{2}{K_{ii}}\left(1-e^{-K_{ii}\tau}\right) + \frac{1}{2K_{ii}}\left(1-e^{-2K_{ii}\tau}\right)\right].$$

The R^2 for a forecasting regression over a time τ horizon is given by (38).

The next theorem states the relevant results for both stock returns, and for excess stock returns (i.e. returns above a rolled over investment at the short rate).

Theorem 6 *With the assumptions about factors dynamics which were made in Section 5.1 we can decompose total returns on stocks as follows:*

$$R(t,t+\tau) = \mathrm{E}_t[R(t,\tau)] + \epsilon(t+\tau).$$

Let us define $\phi_0 \equiv a^m$ and $\phi_Y \equiv K_Z' C^m$. Then we have that

$$\mathrm{E}_t[R(t,\tau)] =$$

$$\phi_0\tau - \sum_{i=1}^{N} B_i^m \theta_i (1-e^{-K_{ii}\tau}) - \sum_{j=1}^{M} c_j \mu_j \tau + \sum_{i=1}^{N} \phi_{Yi}\left[\theta_i\tau - \frac{\theta_i}{K_{ii}}\left(1-e^{-K_{ii}\tau}\right)\right]$$
$$-\sum_{i=1}^{N}\left[\left(B_i^m + \frac{\phi_{Yi}}{K_{ii}}\right)\left(e^{-K_{ii}\tau}-1\right)Y_i(t).\right],$$

where we have used $[\cdot]_i$ to indicate the i^{th} elements of a vector.

$$\begin{aligned}\mathrm{Var}\Big(\mathrm{E}_t[R(t,\tau)]\Big) &= \sum_{i=1}^{N}\left(B_i^m + \frac{\phi_{Yi}}{K_{ii}}\right)^2\left(e^{-K_{ii}\tau}-1\right)^2\frac{\sigma_{Yi}^2}{2K_{ii}},\\ \mathrm{Var}\Big(\epsilon(t+\tau)\Big) &= \sum_{i=1}^{N}\frac{\phi_{Yi}^2\sigma_{Yi}^2}{K_{ii}^2}\Big[\tau - \frac{2\xi_i}{K_{ii}}\Big(1-e^{-K_{ii}\tau}\Big) + \frac{\xi_i^2}{2K_{ii}}\Big(1-e^{-2K_{ii}\tau}\Big)\Big] + \sum_{j=1}^{M}\sigma_{Zj}^2 C_j^{m2}\tau,\end{aligned}$$

where
$$\xi_i \equiv 1 + \frac{B_i^m K_{ii}}{\phi_{Yi}}.$$

The R^2 for a forecasting regression over a time τ horizon is given by (38).

The excess returns of stocks above a rolled over investment in the short-rate can be decomposed as follows:
$$R_{xr}(t,\tau) \equiv R(t,\tau) - R_r(t,\tau) = \mathrm{E}_t[R_{xr}(t,\tau)] + \epsilon_{xr}(t+\tau),$$

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

$$\begin{aligned}\phi_0 &\equiv a^m - r_0,\\ \phi_Y &\equiv K_Z' C^m - r_Y.\end{aligned}$$

With these theorems 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 3 *Given the assumptions of the previous two theorems, the following is true of forecasting regressions for stocks:*
$$\lim_{\tau\to 0} R^2(\tau) = 0 \quad \textit{and} \quad \lim_{\tau\to\infty} R^2(\tau) = 0.$$

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 on stocks, 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. The results of the Lemma bear a comment. The limiting results as τ becomes large are not surprising because obviously as the forecasting horizon gets large it becomes more difficult to say anything about the future (recall that the short rate infinitely far in the future is Normally distributed, even though the yield on an infinitely lived zero coupon bond is constant). Also, the fact that the R^2 for rolled-over investments at the short-rate approaches 1 for short time horizons is obvious since the current short-rate is known, and since the short-rate exhibits some persistence. The one potentially surprising result is that the R^2 for the stock forecasting regressions go to 0 as τ gets small. On the one hand, this is consistent with abundant empirical evidence which suggests that over a short-time horizon all forecasting variables (e.g. dividend yields, price to earnings ratios, etc.) do very poorly. On the other hand, why do we find the same results within the model? The answer lies in the fact that the variances of $\mathrm{E}_t[R(t,\tau)]$ and $\mathrm{E}_t[R_{xr}(t,\tau)]$ go to zero much faster than do the variances of the residual terms $\epsilon(t+\tau)$ and $\epsilon_{xr}(t+\tau)$ as τ becomes small. Hence both variances go to 0, but so does their ratio.

Of course, the relative importance of the mean-reverting and random walk components in stock prices determines the behavior of the R^2 of the forecasting regression in between horizons of 0 and infinity. For this reason, an analysis of this intermediate forecasting behavior must wait until we have estimated the model.

5.3 Model Estimation

Since we have assumed in Section 5.1 that innovations variances are constant, 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 $Y(t+T)$ is given by

$$Y_i(t+\tau) \sim N\left(e^{-K_{ii}\tau}Y_i(t) + \theta_i(1-e^{-K_{ii}\tau}), \frac{\sigma_{Yi}^2}{2K_{ii}}(1-e^{-2K_{ii}\tau})\right).$$

Recall that

$$Z(t+\tau) = Z(t) + \mu\tau + \Sigma_Z(W(t+\tau) - W(t)) - K_Z \int_t^T Y(s)ds.$$

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

$$Z_j(t) + \mu_j\tau - \sum_{i=1}^{N}[K_Z]_{j,i}\left[\theta_i\tau + \frac{Y_i(t)-\theta_i}{K_{ii}}\left(1-e^{-K_{ii}\tau}\right)\right].$$

and a variance of

$$\sigma_{Zj}^2\tau + [K_Z]_j V [K_Z]_j'.$$

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

$$\frac{\sigma_{Yi}^2}{K_{ii}^2}\left[\tau - \frac{2}{K_{ii}}\left(1-e^{-K_{ii}\tau}\right) + \frac{1}{2K_{ii}}\left(1-e^{-2K_{ii}\tau}\right)\right].$$

This conditional distribution of Z follows from Lemma 4 in the Appendix. Because we observe prices, and not the factors directly, we need to use the standard change of variable formula to transform the factor likelihood function into a likelihood function for the observed prices. Let us say that $h(\cdot)$ is the inverse price function. Then the vector of factors may be written as $Y = h(P)$, where P is the vector of prices. If $f(\cdot)$ is the density function for the factors, then the price density function is given by

$$f_P(P) = f(h(P)) \times \left| \det \frac{\partial h(P)}{\partial P'} \right|,$$

where the $|\cdot|$ term is the absolute value of the determinant of the Jacobian matrix (see Billingsley (1995), for example).

In order to take advantage of cross-sectional information available from the prices of securities which have not been used in the inversion, I follow the approach in Chen and Scott (1993) and assume that the prices of other fixed-income securities are observed with error. I assume that these observational errors are Normally distributed with zero means, zero correlations, and variances which need to be estimated.

Letting $\epsilon(t)$ denote these observational errors, and letting $f_\epsilon(\cdot)$ denote the Normal density function, we can write the log-likelihood function for the sample as follows:

$$L = \sum_{t=2}^{T} \log f_P(P(t)) + \sum_{t=2}^{T} \log f_\epsilon(\epsilon(t)).$$

Since this function is known in closed form, I estimate the model using (conditional) maximum likelihood.

The data needed for estimation of the model are a time series of N bond prices, the total returns on a set of M equity portfolios, and a time series of bond prices which will be used to compute the pricing errors $\epsilon(t)$'s (say we have N_ϵ such bonds). The estimation begins with a guess for the values of the model's parameters. This is equivalent to starting with a guess for the $h(\cdot)$ function. Using this initial parameters vector, we use the associated $h(\cdot)$ function to extract a time series of $N+M$ factors from the $N+M$ securities prices. Using the extracted factors, we then compute the ϵ's for the N_ϵ bonds used for this purpose. This gives us a time series of factors and observation errors. Using these we are then able compute the value of the likelihood function associated with our initial guess for the model parameters. With the goal of maximizing the likelihood function, we then update our initial guess of the model's parameters using standard numerical search methods, and proceed until convergence is achieved.

6 Empirical Results

With the model now specified and with the estimation strategy established, it is time to look at the results.

6.1 Number of Factors and Data Description

I use five factors in this study. Two of these are joint bond-stock prices. A sizable literature has been devoted to the determination of exactly how many factors are needed to describe the behavior of the term structure of interest rates.[9] Suffice it to say that two factors is not an unreasonable number, though certainly others may disagree. Three of the five factors are the systematic stock specific factors. A sizable literature has also been devoted to answering the question of how many stock factors are enough, and again three seems to be a reasonable number.[10] I should point out that the majority of the empirical analysis which follows will not be adversely affected by the presence of an additional factor or two. For our present purposes, we will satisfy ourselves with the claim that a five factor model is sufficiently parsimonious so as to avoid overfitting the data, and will leave the determination of the exact number of factors to future research. Note as well that for the present implementation, the three stock specific factors are Z type factors (of course, the joint factors must all be Y type factors).

Given our five factors, we need to have a time series of five securities prices in order to invert the model and obtain factor realizations. The securities which are used in the estimation are a two and a five year zero (obtained from the Fama-Bliss discount bonds file available from CRSP), and the returns on the value-weighted CRSP index, and on two other equity portfolios. The first is a portfolio consisting of the third smallest decile of U.S. firms by market capitalization, and the second is a portfolio consisting of the eighth decile of U.S. firms by their book to market ratios (i.e. a value portfolio).[11] Obviously, these additional stock portfolios are chosen for a reason: We know that much of cross sectional variation in stock returns is associated with differences in firm size and in book to market value. The extreme portfolios in each sort are not chosen in order to guard against data mining.[12]

Since the C^m vector for each stock in the model is completely arbitrary, we need to impose some structure here. Let us associate C^1 with the CRSP value-weighted index, C^2 with the small stock portfolio, and C^3 with the value portfolio. We will then assume that these vectors have the following form:

$$\begin{bmatrix} C^{1\prime} \\ C^{2\prime} \\ C^{3\prime} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ c_1 & 1 & 0 \\ c_2 & c_3 & 1 \end{bmatrix}. \tag{45}$$

This is a natural choice for the following reason: It is the most restrictive choice of the C vectors which allows for all the stock factor innovations to be uncorrelated. The first stock factor is the portion of returns of the CRSP value-weighted portfolio which is uncorrelated with the two joint bond-stock factors. The second stock factor is the part of returns on small stocks which is uncorrelated with the joint factors and with the first stock factor. Finally, the third stock factor is the part of the value portfolio which is uncorrelated with all the remaining factors. This

[9] Litterman and Scheinkman (1991) is the classic paper in this area. They find that three factors suffice, though the third factor is not particularly important. Also see Dai and Singleton (2000), and the references contained therein.

[10] The classic papers here are Fama and French (1992,1993), who formulate the small stock and the value factors. Jegadeesh and Titman (1993) find the so-called momentum affect, which has been used as a factor in subsequent studies. Together with the market portfolio, these three factors seem to account for a significant portion of cross-sectional return variation in U.S. stocks over the last 40 years or so.

[11] Returns on both of these stock portfolios are obtained from Ken French's website (at mba.tuck.dartmouth.edu), as are the returns on all equity portfolios used in the empirical portion of this paper.

[12] Nonetheless, this is the part of the empirical analysis in the paper which suffers most from the possibility of data mining – we have chosen to work with stock portfolios which we know ex-ante to be unusual. An interesting area for future work is to use equity portfolios sorted along non-size and non-value dimensions.

decomposition allows us to refer to the first stock factor as the "equity" factor, to the second stock factor as the "small stock" factor, and to the third stock factor as the "value" factor. This is the naming convention we will follow in the remainder of the paper.

Finally, we need to specify the bonds which will be used to generate the observational errors ϵ's. Of the five bonds used for this purpose, two are taken from the CRSP fixed term indices file, and in every month represent the U.S. government coupon bonds whose maturities are closest to 10 and 20 years respectively. The remaining three bonds are the one, three, and four year zeros from the Fama-Bliss discount files.

All the data are monthly, and run from June 1952 until December 2000. June 1952 is the earliest date for which the Fama-Bliss zero coupon prices are available in CRSP.

6.2 Parameter Estimates

Table 1 reports the parameter estimates and the associated standard errors. We see that the average value of the short rate over the forty eight and a half year period ending in December 2000 is roughly 6.1% (i.e. $r_0 + \theta_1 + \theta_2$). Since in the model, the short-rate is given by $r(t) = r_0 + Y_1(t) + Y_2(t)$, the annual standard deviation of the short rate (assuming the Y innovations are independent, which they are not) is about 4%. Given the K estimates, the annual AR(1) coefficients for the two joint factors (i.e. e^{-K}) are 0.3810 and 0.8938 respectively. Hence the first joint factor is strongly mean-reverting, while the second one is less so.

The loading of the small stock portfolio on the equity factor (i.e. the first stock specific factor) is 1.18 and the loadings of the value portfolio on the the equity factor and the small-stock factor are 0.88 and 0.14 respectively. The standard deviation of annual innovations of the three stock factors are respectively 14.4%, 9.0%, and 7.1%.

Figure 1 shows the time series of the five model factors, and the short rate. The fluctuations in the first joint factor Y_1 around the late 70's and early 80's reflects the change in Fed monetary policy which took place around that time. Also, the three stock specific factors have a negative drift because the loading of log stock prices on these were negative by construction. Another interesting observation about the equity specific factors is that the small stock factor Z_2 reached its peak value shortly before 1985, and has since gone back to a value closer to its level in the mid-70's. This reflects the poor performance of small stocks in the last decade. The equity factor Z_1 has shown steady gains since 1975, though it remained relatively flat from the 50's until then. The value factor Z_3 has shown steady gains in the entire sample period.

Table 2 reports the empirical correlations of the monthly factor innovations. We see that the two joint stock-bond factors are actually negatively correlated (the correlation is -0.35), which is a known fact about bond factors in the U.S. (see, for example, Dai and Singleton (2000)). Hence our assumption of uncorrelated factor innovations is not correct. The stock factors show close to zero correlations with respect to themselves, as well as the common bond-stock factors. These low correlations imply that the portion of the variance of Z innovations which comes from drift of the Z (i.e. from its dependence on the Y's) is quite small relative to the innovations coming from the Brownian motion component.

The fact that the estimated volatilities of the two joint factors are 2.4% and 1.2% respectively suggests that both are important determinants of returns in the model. However, because of the negative correlation of the extracted factors, we must be careful in how we interpret these standard deviations. To verify that both of the joint factors are indeed important for describing the behavior of bond prices, we can regress short rate innovations on innovations in the two factors. Table 3 shows the results of this regression. When the second joint factor is excluded, the R^2 of the regression falls to 69%, and when the first factor is excluded the R^2 falls to 5%. When both are included, the R^2 is of course equal to 1. Hence both factors seem to be important in the determination of interest rate behavior.

The bottom row of table 2 shows Dickey-Fuller tests for the five factors extracted from the model. As can be seen, we can reject that the first joint factor Y_1 has a unit root at the 99% level. However, the unit root null cannot be rejected for any of the other four factors (including the joint factor Y_2). While this inability to reject a unit root is not surprising for the three stock specific factors, it is surprising for the second joint factor Y_2, which is mean-reverting by construction. Probably this inability to reject is due to the low power of the test.

According to the model, continuously compounded total returns on the three stock portfolios used in estimating the model have the following form

$$\log P(t+T) - \log P(t) = \alpha T + \beta_1(Y_1(t+T) - Y_1(t)) + \cdots + \beta_5(Z_3(t+T) - Z_3(t)). \quad (46)$$

Factor innovations are stationary, and hence this specification can be estimated by a simple regression. Table 3 reports the results of several such regressions for the monthly total returns on the CRSP value-weighted index. The first row in the table shows the betas from the regression which includes all five factors. The coefficient estimates and the fact that the regression fit is perfect are so by construction (since this index was one of the equity portfolios used in the model estimation). The second set of rows shows the regression in (46) which only includes the joint factors and the equity stock factor. By construction, these three factors explain 100% of the return in the value-weighted index. The next set of results are from a regression of the index return on only the joint factors. We see that these account for roughly 4% of the variation in the index returns. Since the correlation between these and the equity factor are close to zero, the next regression of the index returns on first differences in the equity factor accounts for the remaining 96% of the variation.

From Table 1, we see that the drifts of the equity factors (under both the physical and the risk-neutral measures) are time varying functions of the joint factors. Note that the K_Z matrix is implied from estimates of the B and C matrixes for stocks, via equation (41). We see that the equity factor has the most volatile drift (its loadings on the two joint factors are 0.33 and 0.97 respectively).

6.3 The Cross-Section of Bond Returns

Much work has been done in assessing the fit of no-arbitrage models for the term structure of U.S. government debt (for a recent example, see Dai and Singleton (2000)). Hence, we will not devote much time to re-examining this question in the context of the present paper. The reason for this choice is that estimating a joint bond-stock model is likely to reveal new information about the behavior of stock prices, but is unlikely to reveal much new information about the behavior of bond prices. After all, as we have shown, the stock specific factors do not affect bond prices in the theoretical development, whereas the bond factors must affect stock prices.

Figure 3 shows the pricing errors of the model for the bonds available in the CRSP fixed term index file. This file contains in every month the seven U.S. government coupon bonds with maturities which are closest to 1, 2, 5, 7, 10, 20, and 30 years respectively. While the bonds change from month to month, we see from the graphs in Figure 3 that over the time period used in this study, the average maturities of bonds in each category are close to the target maturities (the graphs report the average maturity of bonds in each series). Keeping in mind that the model was estimated using much information about the short end of the term structure, and using relatively less information about the long end, we see that the model does a reasonably good job of pricing the shorter end of the term structure. The average pricing errors for the seven bond categories are -0.03, 0.00, -0.02, 0.08, 0.78, -0.27, and -0.96 percent respectively. A more descriptive measure of the model's fit is the standard deviation of the pricing errors for each bond maturity. Respectively, these standard deviations are 0.29, 0.25, 0.88, 1.25, 2.14, 6.32, and 7.10 percent. Hence we see that the model does a good job with the short end of the term structure, but provides a considerably worse fit for maturities beyond 10 years.

From the figure we see that for the long end of the term structure, the pricing errors are highest for dates prior to the early 70's, and much lower after that. As it happens, the first term structure models (such as Merton (1974) and Vasicek (1977)) probably came out as working papers at around this time. One is left to wonder whether or not this is just a coincidence?

In any case, the main conclusion to be drawn from this discussion is that the present model does a reasonably good job of accounting for the variation in the term structure of U.S. government debt, and is certainly good enough for purposes of the questions addressed in this study. Of course, the model is missing important characteristics present in that data, such as stochastic volatility and a third bond factor. Hence an interesting area for future research would be to repeat the analysis of this paper, but with a more realistic specification for the behavior of the bond factors.

6.4 The Cross-Section of Stock Returns

Using data from three equity portfolios and from the term structure of U.S. government debt, we have estimated a no-arbitrage asset pricing model, and have thereby extracted a set of five factors. We should have some confidence in referring to these extracted quantities as "factors" because they are the underlying sources of uncertainty in an economically self-consistent statistical model of the bond and stock markets. It should be emphasized that data from only three equity portfolios (a market index, a small stock portfolio, and a value portfolio) have been used. It becomes interesting, therefore, to understand how other equity portfolios depend on these extracted factors.

According to the pricing equation (9), the factor structure of returns for a given stock or stock portfolio can be extracted by running the following regression:

$$\log P(t+\tau) - \log P(t) =$$

$$\alpha_0 + \alpha_1 \tau + \beta_1 (Y_1(t+\tau) - Y_1(t)) + \cdots + \beta_5 (Z_3(t+\tau) - Z_3(t)) + \epsilon(t+\tau). \quad (47)$$

The α_1 term is present because not all observations are evenly spaced (i.e. τ varies from month to month).[13] Note that the epsilons from the regression in (47) are interpreted as innovations to the idiosyncratic Z factor associated with this security. Furthermore, the drift of this idiosyncratic factor will likely depend on the Y factors (recall that this occurs because the appropriate row in K_Z must be manipulated to reconcile the empirical B^m and C^m estimates for a given security).[14] Indeed these estimates of the B^m and C^m coefficients for a given total return process imply a particular implied dividend process for that security (recall Theorem 2). Hence one potential test of the model is to look at the implied dividend rates from a cross-section of equity portfolios, and to compare these to the actual dividend rates of those portfolios.

The regression in (47) was run for characteristics sorted portfolios. The five stock characteristics used for sorting are: company size, book to market ratio, dividend to price ratio, earnings to price ratio, and cash flows to price ratio. For each category, a set of 10 decile portfolios is available. For example, for the size category, we have 10 portfolios, each containing companies with progressively larger market capitalizations. The data and data descriptions are available on Ken French's website (see footnote 11).[15] These portfolios are useful because they represent sorts along dimensions which are known to affect returns. Hence the return series for the 10 portfolios within each category have distinct behaviors.

Tables 5 – 9 and Figures 8 and 9 report the results of these regressions. The tables also report the mean and standard deviation of monthly returns for each of the fifty equity portfolios. The figures show the coefficient estimates from the tables plotted across the ten portfolios of each characteristics sort. For example, in the upper left graph of Figure 8, we see plots of the loadings of the ten decile size portfolios on the stock specific factors (the Z's). As the graph makes clear, the loading on the size factors (Z_2) is monotonically decreasing as we move from the smallest to the largest size portfolio. The upper left graph in figure 9 shows the R^2's from the regressions in (47) applied to the ten size portfolios. Note that the R^2 of the 3^{rd} size portfolio is equal to one by construction, as this portfolio was used in the estimation of the model.

The beta estimates from (47) are reported for decile portfolios sorted along five firm characteristic dimensions: size, book to market, dividend to price, earnings to price, and cashflows to price. A large body of work has documented the fact that portfolios sorted along these dimensions (i.e. size and value) will have very different return series. Hence an interesting question to ask is how

[13]However, because the variation in τ from month to month is very small, the estimates of α_0 and α_1 are very imprecise. Much can be gained in estimate precision by dropping the α_1 term (i.e. by assuming that all τ are the same). Neither estimation strategy has a sizable effect on the β estimates in (47). Furthermore, since τ varies from month to month, the residuals from this regression are heteroskedastic, even if we assume that innovation shocks have a constant variance.

[14]The hope is that, similarly to the Z factors extracted from the model, the innovations to this idiosyncratic Z factor will have a low correlation with first differences in the Y's. While this is likely to be the case, OLS estimation of the regression in (47) is, strictly speaking, susceptible to bias. The solution is to perform maximum likelihood estimation of the model for each of the fifty stock portfolios. Unfortunately, this strategy is extremely computationally intensive, and therefore will not be pursued.

[15]Portfolios are formed at the end of June in each calendar year t, using accounting information from the end of the previous calendar year $t-1$. See the website for more details.

this return heterogeneity is being generated. For now, we will focus on differences in loadings on the stock specific factors across these characteristic sorts. We will later look at the differences in the loadings of these portfolios on the joint bond-stock factors.

From Table 5 (and the upper left graph of Figure 8) we see that the return variability in size portfolios is due to a monotonic increase in loadings on the second equity factor (Z_2) as we move from the smallest to the largest decile portfolio. The bottom row in the table reports the standard deviation of beta estimates across the ten portfolios for each of the five model factors. For example, the standard deviation of the ten β_4 estimates is 0.4846. Furthermore, we see that this standard deviation is much higher than the standard deviations for β_3 and β_5, suggesting that the return pattern across size sorted portfolios is largely caused by different loadings on the size stock factor.

Table 6 and the upper right graph of Figure 8 show the decile portfolios sorted by their book to market values. We see that the majority of variation in the stock specific portion of these portfolios is coming from changes in loading on Z_3, appropriately named the value factor. However, it is surprising to note that this dependence is non-monotonic. The eighth decile portfolio has a larger (in magnitude) loading on the value factor than the ninth and tenth decile portfolios. Hence variation in returns on these portfolios is being caused by something other than their "valueness" characteristic. The more "value" portfolios have lower loadings on the value factor than the less "value" portfolios.

Tables 7 – 9 (and the lower three graphs of Figure 8) report results of the regressions for portfolios sorted by their dividend to price ratios, their earnings to price ratios, and their cash flow to price ratios, respectively. Arguably, all three sorts are also proxying for the value effect. The results across the three sorts are similar. Most of the return variation in the stock specific portion of those portfolio returns is captured by variation in loadings on the value factor Z_3. Furthermore, contrary to the book to market portfolios, the loadings on this third factor are monotonic as we move from the lowest to the highest decile portfolios. This suggests that each of these three sorts better captures the "value" effect than does the book to market sort. Also across all the value sorts we see a mild increase in the loading on the small stock factor as we move from the growth (decile 1) to the value (decile 10) portfolios.

From the R^2's of all fifty regressions (see Figure 9), we see that the five factor model captures a good portion of variability in the cross section of equity returns. It is surprising to note the sharp drop-off in R^2's as we move from the low dividend-to-price to the high dividend-to-price deciles. This seems to suggest that a factor (either joint or stock specific) is missing, and that this factor is important for the higher dividend-to-price portfolios.

Furthermore, variation in loadings on the size factor is associated with the size sorted portfolios, and variation in loadings on the value factor is associated with different types of value sorts. These results provide evidence in support of the view that return variation across size and value portfolios is attributable to different risk loadings. However, even though the results of this paper rely on factors which are by construction consistent with a pricing model, they may still be open to the criticism that return variation is characteristics, and not, factor based (see Daniel and Titman (1997)).

The above results do not rely on long-short factor proxy portfolios (as in Fama and French (1993)). Instead, the model in this paper suggests a decomposition of equity returns into (nearly) orthogonal bond-stock and stock specific factors. Along these lines, it may be argued that it is not necessary to go through the pains of estimating the relatively burdensome decomposition suggested in this paper, when a much simpler one may suffice. Subject to a caveat, this turns out to be the case. Consider, the following decomposition of the returns on the value weighted, the small stock, and the value stock portfolios:

$$\begin{aligned} R^{VW}(t, t+\tau) &= a_1 + f_1(t+\tau), \\ R^{Small}(t, t+\tau) &= a_2 + c_1 f_1(t+\tau) + f_2(t+\tau), \\ R^{Value}(t, t+\tau) &= a_3 + c_1 f_1(t+\tau) + c_2 f_2(t+\tau) + f_3(t+\tau), \end{aligned}$$

where the f_i's are uncorrelated, and the R's are continuously compounded returns.[16] Hence the f_2 is the residual of a regression of the small stock portfolio on the value weighted index, and f_3 is the residual from a regression of the value portfolio on the other two portfolios. This decomposition of equity returns is exactly analogous to the one implied by the C matrix shown in (45). The one difference is that the component of equity returns which is common to stocks and bonds has not been taken out in this benchmark decomposition.

The R^2's of regressions of f_1, f_2, and f_3 on the common bond-stock factors are 4%, 2%, and 0.7% respectively. Since the orthogonal return components extracted from the above regressions have relatively little dependence on the common bond-stock factors, it turns out that if we run the following regression

$$\log P(t+\tau) - \log P(t) = \alpha_0 + \alpha_1 \tau + \beta_1 f_1(t+\tau) + \cdots + \beta_3 f_3(t+\tau) + \epsilon(t+\tau), \qquad (48)$$

we observe a pattern in betas and stock characteristics similar to the one just discussed (the results of these regressions are not reported here). Had the joint bond-stock factors represented a larger portion of the f_2 and f_3 orthogonal components, the results from the regressions in (47) and (48) may well have been quite different. Indeed, the f_i's cannot even be interpreted as factors since they do not come from any asset pricing model. Hence the results from the analysis in (48) could not be trusted without recourse to the analysis in (47).

6.5 Empirical Equity Durations

Tables 5 – 9 report the B^m estimates for each of the fifty equity portfolios. Following the analysis of Section 4.3, we are able to compute the duration for each of these portfolios. Figures 6 and 7 report the results of this analysis. The first figure shows the B^m loadings across the decile portfolios for each of the characteristics sorts. For example, upper left graph shows these coefficients for the ten size portfolios, starting with the smallest and going to the largest portfolios by market cap. The circles represent the B vector of that zero coupon bond whose loadings on the joint bond-stock

[16]The idea of extracting factors from residuals of regressions is not new. For example, see Elton, Gruber, and Blake (1999).

factors are most similar to those of the equity portfolio in question. The square-lines in Figure 7 show the maturity of this zero. The circle-lines show the duration measure computed using the bond-matching approach outlined in Section 4.3. The duration level is displayed on the left axis. The maturities of the three bonds used for this approach are 6 months, 5 years, and 10 years respectively. Finally, the solid lines in the graphs in Figure 7 show the R^2's of regressions of the continuously compounded returns of the equity portfolio in question on first differences of the joint bond-stock factors.

The duration for the CRSP value weighted index is 2.107 years, and the durations for the size sorted portfolios range from 0.5 to 2.5 years, and are nearly monotonic as we go from the small stock (low durations) to the large stocks (high durations). Hence the returns on smaller stocks tend to be less sensitive to the macroeconomic conditions that drive changes in government interest rates.

The duration for value sorted (i.e. the book-to-market, earnings to price, and cash flow to price) portfolios peaks for the middle decile portfolios, and is lowest for the extreme (low and high) portfolios. For example, for the book to market sorted portfolios the durations for the low book to market portfolios start at 2 years, peak at 2.5 years for the middle deciles, and fall to 1.5 years for highest book to market portfolios. This suggests that across the value sort, the extreme growth and extreme value stocks are least sensitive to interest rate changes.

When looking at the behavior of durations for the dividend to price sort, we find that those stocks which pay a higher dividend rate are more sensitive to interest rate fluctuations than stocks which pay dividends at a lower rate. Indeed the pattern is monotonic as we go from the low dividend portfolio (duration of under 2 years) to the highest dividend portfolio (duration of 3 years).

A striking feature of all the figures is the high correlation of the levels of the duration measures and the R^2 from the regressions of returns on the joint factors. This results is comforting in the sense that the duration measure is capturing the degree to which equity returns depend on variation in the joint bond-stock factors. The duration analysis is, however, far more informative than the R^2. After all, what does an R^2 of 7% mean? The duration figure translates this into a far more economically significant quantity: the maturity of the zero coupon bond whose response to fluctuations in the joint factors is closest to that of the equity portfolio in question (or the duration of a bond portfolio which exactly replicates the equity sensitivity to the joint factors).

Two notes are in order at this point. First, the magnitude of durations reported in this paper (i.e. around 0.5 to 3 years, and 2.107 for the value weighted index) is consistent with the classic work in this area (namely Leibowitz (1986)), though the present model may be said to have a more sound economic foundation.[17] Second, the general results of this section run counter to the intuition that returns on stocks which pay no current dividends should be more sensitive to interest rate fluctuations because those dividends will be received so far in the future that the stock should behave like a long dated zero-coupon bond. This paper's results suggest that the uncertainty about future cash flows associated with stocks which are currently not paying dividends (such as growth

[17]In particular, Leibowitz finds that, between January of 1980 and November of 1985, the duration of the S&P500 index is 2.186 (see page 23). He also reports (same page) that during the same time period the duration of a broad bond index is 4.51. Also see Arnott, Hanson, Leibowitz, and Sorensen (1989).

stocks, small stocks, or low dividend to price ratio stocks) explains much more of return variation than do term structure fluctuations. Interest rate fluctuations seem to matter more for stocks of established, dividend paying firms.

One issue with the duration measures presented here is whether governments are the appropriate set of bonds against which stock durations ought to be measured. For example, stocks of distressed firms may comove much more with the debt of those distressed firms, than they do with the government term structure. Hence, another measure of duration would be to compare stock sensitivities to common factors with the appropriate set of corporate bonds. Within the present model it is fairly straightforward to derive pricing formulas for stocks and bonds when default is possible. Then durations can be computed as in this paper, except the matching bonds can come from the appropriate subset of corporate debt. This is an interesting topic for future research.

6.6 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), a test of the model proceeds as follows. 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 result in a "good" fit. If not, then we may conclude that the model has been misspecified.

From the analysis in Theorem 2, we see that the we are free to proceed with the projection in the sense that any choice of δ_0 and δ_Y is consistent with some no-arbitrage price process. The one constraint which does need to be imposed on the dividend series comes from the transversality condition in (17). We proceed to compute the "best fit" dividend series $\hat{\delta}(t) = \hat{\delta}_0 + \hat{\delta}_Y' Y(t)$ as follows:

$$\{\hat{\delta}_0, \hat{\delta}_Y\} \equiv \arg \inf_{\delta_0, \delta_Y} \sum_{t=1}^{T} (\delta_0 + \delta_Y' Y(t) - \delta(t))^2, \tag{49}$$

subject to the constraint that the transversality condition in (44) holds for $\hat{\delta}_0$ and $\hat{\delta}_Z$. For the case when the constraint in (44) is not binding, the minimization in (49) results in an OLS projection of the actual dividend yield onto the extracted model factors. Of course, this will no longer be true when the transversality constraint in binds. The instantaneous dividend yield $\delta(t)$ is estimated to be the annual dividend of the CRSP value weighted index in the time interval $[t - 0.5, t + 0.5]$.[18]

The proximity of $\hat{\delta}(t)$ to $\delta(t)$ is a measure of the information content of the Y factors extracted from the model estimation. An inability of these extracted Y series to properly account for the actual dividend series implies that the model has been misspecified. To assess the degree of this misspecification, we run the following regression

$$\delta(t) = \alpha + \beta \hat{\delta}(t) + \epsilon(t). \tag{50}$$

If the transversality constraint was not binding in solving (49), then by construction the α and β coefficients from this regression are 0 and 1 respectively. In other words, a non-binding transversality

[18] 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.

constraint in the projection in (49) 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 (44) does bind in solving (49), 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.

The top graph in Figure 2 shows the actual dividend series and the best fit series $\hat{\delta}$ over the entire sample period. As can be seen, the best fit dividend series from the model is much smoother than the actual dividend series. This is roughly the opposite of Shiller's finding that realized dividend payouts are not volatile enough to justify the observed fluctuations in stock prices (see Shiller (1981) and Campbell and Kyle (1993), for example). Here, the actual dividend series is more volatile than the one predicted by the model. The inability of the model implied dividend series to match the actual dividend series can also be seen from the fact that the R^2 of the regression in (50) over the entire sample period is 2.85% (see Table 4) – recall that for a correctly specified model the population value of this parameter is 100%.

This inability of the model implied dividend series to match the actual dividend yield of the CRSP value weighted index may be due to a structural break which may have taken place in the early 1970's (although I do not conduct any formal tests of this conjecture). One reason to suspect that something happened in this time period is the sudden drop in the pricing errors for the long end of the term structure which took place approximately in 1973 (see Figure 3). Further evidence can be seen in the bottom two graphs of Figure 2.

The middle graph in the Figure shows the implied and actual dividend yields from 6/1952 – 12/1969. This represents a repeat of the analysis performed above, but in the early part of the sample. The bottom graph shows both dividend series from 1/1970 – 12/2000. As can plainly be seen, the match is far better in each subperiod than it is in the entire sample. The reason can be seen from the regressions in Table 4. In the early subperiod, the $\hat{\delta}_Y$ vector is negative, whereas the vector is positive in the later subperiod. Since the short rate in the model is given by $r(t) = r_0 + Y_1(t) + Y_t(2)$, dividend yields were negatively correlated with (all) interest rates until the early seventies, and which point the correlation became positive.

Hence estimation of (49) over the entire sample produces a poor-fitting implied dividend series. Indeed in both subperiods the R^2's of the regressions in (50) are both quite high (53% and 43% respectively). In the early subperiod, the transversality constraint binds in the projection in (49), implying that $\hat{\delta}(t)$ is biased.

Of course, this subperiod analysis is not consistent with the model itself, which requires that δ_0 and δ_Y stay constant. However, the fact that the extracted factors seem to work well within each subperiod provides hope that a better fitting specification of the model may be found. Indeed, the following questions remain to be answered: Why did dividend yield correlations with the short rate change around the early 1970's? What missing factor can account for this time varying correlation structure?

A related issue is why the $\tilde{K}_Z^s$ matrix implied from the projection in (49) is non-zero (recall equation (26) from Theorem 2). Why should the drift of the stock specific factors present in the actual stock prices depend on the joint bond-stock factors?

6.7 Expected Returns

Equation (34) gives the instantaneous excess returns on any security in the model. Furthermore, using (36) we can compute the expected return on stocks over any time period, and we can convert this to an expected excess return by simply subtracting off the returns to a rolled-over investment at the short rate. Figure 4 shows plots of the instantaneous excess returns, and of the one year raw returns, for the three stock portfolios used in model estimation. Recall that expected excess returns in the model are driven entirely by fluctuations in the common factors Y_1 and Y_2.

The three series correspond to the value weighted market index (bottom), the third size decile (middle), and the eighth value decile (top) portfolios. As can be seen from the bottom graph, the average instantaneous expected excess return on the value weighted index was around 7%. This number seems to be quite plausible, given the historical evidence from other sources. For example, Campbell (1999, Table 2) reports that from 1947–1996, the continuously compounded real return on the U.S. market has been 7.6%, with an average 3-month real money market rate of 0.8%. The small stock risk premium was slightly higher, and the value risk premium was the highest at just over 10%.[19]

An interesting feature of this figure is that all market risk premia reached their peak in the middle of the 80's, and have been steadily declining since then. Indeed, by the end of 2000, all risk premia were close to their historical lows reached in the mid-50's. This finding is in keeping with the conventional wisdom that the markets were over-valued in the late 90's, as well as in the year 2000. Finally, the variations in expected returns have been quite pronounced. For the market index, the lowest point was in the mid-50's at a level of just under 6%, and the peak was in the early 80's at about 10%.

The bottom graph of Figure 4 shows the implied term spread from the model, and is equal to the difference between the ten year zero yield and the short rate. As is expected, the term spread is countercyclical, in the sense that it is low during NBER business cycle peaks, and is high during NBER troughs. This pattern has previously been documented in the literature (see Fama and French (1989)). Researchers have argued that this pattern may be due to a time varying risk premium, which is high during recessions and low during peaks (for example, see Mamaysky (2001) for a model which generates this pattern). However, as we see from the other two graphs in Figure 4, the behavior of expected excess or raw stock return is exactly the opposite: These seem to be high during peaks, and low during troughs, and hence procyclical. What could account for this unexpected behavior? After all, if the effect is arising only from variation in the risk premium, then expected returns on stocks should also be countercyclical.

One mechanism which is consistent with both effects is the following: Dividend expectations and risk premia are negatively correlated. A high risk premium (during recessions) causes the short

[19]In results not reported here, annual expected excess returns are similar in level and variation to their instantaneous counterparts, but are on average 1% lower. This difference in levels between the annual and the instantaneous risk premia arises due to the volatility adjustment which needs to be applied when moving between drifts and expected log returns.

rate to fall, and the term structure is therefore upward sloping (high term spread). On the other hand, dividend expectations are low, but these dividends are being discounted at the low short term interest rates, and hence prices stay at nearly the same levels. Then, with low dividend expectations and with no commensurate decrease in prices, equity expected (raw and excess) returns are also low during times when the term spread is high. The same logic, in reverse, shows that during times where the term spread is low, the expected returns on equity should be high. Of course, there are competing effects in this story (like dividend expectations and levels of interest rates), and which effect dominates is a question in need of a theory. Indeed such a theory is an interesting topic for future research.

6.8 Predictability

At this point, the interested reader may be curious about the ability of the expected return series to predict the actual return series. There are well known problems with addressing this question from a purely empirical point of view. Among these are data mining, survivorship bias, and the use of overlapping returns in a small sample. For example, many forecasting variables and time horizons have been examined using the same data sample – hence the time horizons and forecasting variables which are commonly used are likely to be order statistics of sorts. Furthermore, long horizon regressions require a long time series of data, and since one is not available researchers have traditionally used overlapping returns.[20] To the extent that the present model provides an accurate description of the data, we are able to use the model to overcome some of these difficulties. First of all, within the framework of the model we are able to compute the actual conditional expected return, as well as the theoretical R^2 over all forecasting horizons, and are thus able to entirely avoid issues of data mining. Furthermore, statements about predictability at long time horizons can be made due to the structure of the economic environment which the model imposes. The model parameters are estimated using maximum likelihood with monthly returns. Hence small sample problems with using overlapping data are avoided.

Of course, no approach is perfect, and the present one does suffer from its own drawbacks. First of all, the model may be misspecified, which would render its R^2 calculations wrong, or at best inaccurate. Furthermore, the theoretical R^2 curve (i.e. a plot of the R^2 of forecasting regressions versus time) is itself an estimate. However, deriving standard error bounds around this curve is difficult to do analytically, and very time consuming to do via simulations. Hence the results reported in this paper show only the point estimate of the R^2 curve, but not the associated standard errors.

From Theorems 5 and 6 we have formulas for the R^2's of forecasting regressions of returns on their conditional expectations. Figure 5 shows a plot of the theoretical R^2's of these forecasting regressions as a function of the forecasting horizon. The theoretical R^2's are shown for predictive regressions for the CRSP value weighted index, the third size decile portfolio, and the eighth value decile portfolio. Note again that these are not R^2's from actual regressions, but come from

[20] Relevant papers include Hodrick (1992), Richardson (1993), Goetzmann and Jorion (1993), Kirby (1997), and more recently Ang and Bekaert (2001), among many others.

computations done within the framework of the model itself.

The short and long term behaviors of these forecasting regression have already been established in Lemma 3. From our estimate of the model's parameters,we find that the predictive power of the regression for the rolled over short-rate falls from 100% at the short end, to about 20% at a time horizon of 30 years. The predictive power for the market index regression goes from 0 to a maximum of about 13% at around 10 years, and then begins to slowly decay. At 50 years, the forecasting power of the expected return for the realized market return is about 5%. The story is roughly similar for the value portfolio. Interestingly, though, the amount of predictive ability for the small stock portfolio is considerably lower than for the other two. The R^2's for the small stock portfolio peak at a horizon of roughly ten years, but at a maximum value of only 7%. This appears to be attributable to the fact that small stock portfolios are less dependent on the joint bond-stock factors as can be seen from Figure 6.

An interesting result is that, given the parameter estimates in this paper, there is virtually no predictive power for the market excess returns for any of the three portfolios. Presumably, since the predictable part of the price process itself is coming from the price's dependence on the joint bond-stock factors, when the short rate is subtracted, this dependence is largely cancelled out, and no predictability remains. Indeed, the R^2 for a predictive regression for excess returns for all three portfolios peaks at under 0.5% at a time horizon of around 12 years. Given the virtually interchangeable use of raw returns versus excess returns evident in the literature, this result is both puzzling and troubling. On the one hand, why is there such a large difference in forecasting ability for raw versus excess returns? On the other hand, why is this distinction not readily apparent in the literature?[21] These questions clearly deserve more study: For example, Monte Carlo experiments can be conducted to see whether empirical R^2's of forecasting regressions fall within the confidence intervals of actual R^2's generated using simulated time-series from the model (in samples of the same length as those which are used in practice).

7 Conclusion

This paper has developed an empirically flexible, and economically self-consistent model of bond and stock prices. Estimation of this model shows that five factors (two joint and three stock) do a reasonable, though imperfect, job of describing the behavior of U.S. bond and stock markets over the last fifty years. This paper raises a number of very interesting questions for future research.

For example, the present empirical analysis should be extended to take into account stochastic volatilities for both the stock and the bond factors. Furthermore, it would be interesting to know if there isn't a sixth or a seventh factor which may be an important determinant of asset prices. More work should be done to understand what economic restrictions can be placed on the pricing

[21] It is possible to find examples of high predictability for both excess and raw stock returns. In a well known paper, Fama and French (1989) find R^2's of predictive regressions using overlapping observations for excess returns to be as high as 60% between 1941–1987, though this value falls to 25% over the entire sample period from 1927–1987 (see Tables 2 and 3). Goetzmann and Jorion (1993) show that overlapping predictive regressions for raw stock returns yield R^2's as high as 39% in the time period between 1927–1990. They proceed to argue that such high R^2's are implausible, and are indeed consistent with the null of no return predictability in a small sample.

kernel employed in this paper, while at the same time allowing the model to retain its empirical flexibility. Along these lines, option prices may be used to derive further testable restrictions on the model parameters. Also, it would be interesting to understand how the extracted factors from this paper price a cross section of individual stocks (rather than portfolios), and how they are related to macroeconomic variables or announcements. Instead of using a size and a value portfolio to estimate the model, the analysis of this paper can be repeated with a model estimated using equity portfolios formed along non-size and non-value dimensions. Would the resultant stock specific factors still proxy for the size and value effects? Also how can the model be modified so that its implied dividend series provides a better fit to the actual dividend yields? What economic forces account for the procyclical variation of expected (raw and excess) stock returns and for the countercyclical variation of the term spread? Finally, how can the model's notion of equity duration be used to understand and manage the interest rate risk of portfolios with fixed income and equity investments?

In general, it may be hoped that the model presented in this paper will provide a rigorous theoretical basis for future investigations into the empirical behavior of asset prices.

8 Appendix

8.1 Proof of Theorem 1

Let ξ be an $(N+M)\times 1$ vector given by $[Y'\,Z']'$. If D[$\cdot$] is the associated Ito operator for ξ then no-arbitrage implies that the price of a non-dividend paying security must satisfy

$$\mathrm{D}[P(\xi,t)] + P_t(\xi,t) + \delta(\xi)P(\xi,t) = r(\xi)P(\xi,t). \tag{51}$$

This results follows from an application of the Feynman-Kac formula to the general pricing equation

$$P(t) = \mathrm{E}_t^Q\left[\int_t^T e^{-\int_t^s r(h)dh}\delta(s)P(s)ds + e^{-\int_t^T r(h)dh}P(T)\right]. \tag{52}$$

These are standard results in the literature on asset pricing (see Harrison and Kreps (1979), Harrison and Pliska (1981), and Dybvig and Huang (1989); for a textbook treatment see Duffie (1996)).

Given the factor dynamics in (4,5), the conjectured form of a security's price in (9), the coefficient equations in (10,11,12) follow from Ito's lemma and (51) above.

Q.E.D.

8.2 Proof of Lemmas 1 and 2

These follow directly from (10,11,12) and the arguments presented in the main body of the text.

Q.E.D.

8.3 Proof of Theorem 2

A total return process may be thought of as a stock with no-dividends. We conjecture that for the m^{th} stock, the form of this process is given by

$$P^m = e^{A^m(t) - B^{m\prime}Y(t) - C^{m\prime}Z(t)}. \tag{53}$$

Going forward we will suppress the m superscript. For the discounted version of this process to be a martingale, the total return process must satisfy equation (51) with $\delta(t)=0$. Assuming that $r(t) = r_0 + r_Y'Y(t)$, the above coefficients must satisfy

$$A_t(t) \quad = \quad r_0 + \tilde{\Theta}'\tilde{K}_Y'B + \tilde{\mu}'C - \frac{1}{2}\sum_{i=1}^{N+M}\Big([\Sigma_Y'B]_i + [\Sigma_Z'C]_i\Big)^2\alpha_i, \tag{54}$$

$$0 \quad = \quad -r_Y + \tilde{K}_Y'B + \tilde{K}_Z'C + \frac{1}{2}\sum_{i=1}^{N+M}\Big([\Sigma_Y'B]_i + [\Sigma_Z'C]_i\Big)^2\beta_i. \tag{55}$$

We will now check that, subject to certain conditions, the total return process in (53) is consistent with the self-financing wealth process of a portfolio which owns a single share of the actual stock, and reinvests all dividend proceeds back into the stock itself. Letting $x(t)$ be the number of shares of the stock held at time t, the wealth process must satisfy

$$d\omega(t) = x(t)(d\hat{P}(t) + \delta(t)\hat{P}(t)dt) + r(t)(\omega(t) - x(t)\hat{P}(t))dt,$$

where $\hat{P}$ is the no-arbitrage stock price, and δ is the time t dividend rate (assumed to be of the form $\delta(t) = \delta_0 + \delta_Y' Y(t)$). This implies that

$$\frac{d\omega(t)}{\omega(t)} = \frac{x(t)\hat{P}(t)}{\omega(t)} \left(\frac{d\hat{P}(t)}{\hat{P}(t)} + \delta(t)dt \right) + r(t) \left(1 - \frac{x(t)\hat{P}(t)}{\omega(t)} \right) dt.$$

Full reinvestment implies that the fraction of wealth invested in the stock is equal to 1, and hence

$$x(t) = \frac{\omega(t)}{\hat{P}(t)}.$$

Applying Ito's formula to the wealth process, we find that

$$\omega(T) = \omega(t) \exp\left\{ \int_t^T \frac{d\hat{P}(s)}{\hat{P}(s)} + \delta(s)ds - \frac{1}{2} \frac{d\langle \hat{P}(s) \rangle}{\hat{P}^2(t)} \right\},$$

where $\langle \cdot \rangle$ indicates the quadratic variation process. For the no-arbitrage condition in (16) to be satisfied, the price process must satisfy

$$\hat{P} = e^{\hat{A}(t) - \hat{B}' Y(t) - \hat{C}' Z^s(t)},$$

where Z^s is the vector of the stock specific variable with dynamics given in (24). The coefficients must solve

$$\hat{A}_t(t) = r_0 - \delta_0 + \tilde{\Theta}' \tilde{K}_Y' \hat{B} + \tilde{\mu}' \hat{C} - \frac{1}{2} \sum_{i=1}^{N+M} \left([\Sigma_Y' \hat{B}]_i + [\Sigma_Z' \hat{C}]_i \right)^2 \alpha_i, \quad (56)$$

$$0 = \delta_Y - r_Y + \tilde{K}_Y' \hat{B} + (\tilde{K}_Z^s)' \hat{C} + \frac{1}{2} \sum_{i=1}^{N+M} \left([\Sigma_Y' \hat{B}]_i + [\Sigma_Z' \hat{C}]_i \right)^2 \beta_i. \quad (57)$$

for some $M \times 1$ vector $\hat{C}$. An application of Ito's lemma to $\hat{P}$, and some algebra, reveal that

$$\omega(T) = \omega(t) \exp\left\{ \hat{A}(T) - \hat{A}(t) - \hat{B}'(Y(T) - Y(t)) - \hat{C}'(Z^s(T) - Z^s(t)) + \int_t^T (\delta_0 + \delta_Y' Y(s))\, ds \right\}.$$

From (53), we can write the total return process as

$$P(T) = P(t) \exp\Big\{ A(T) - A(t) - B'(Y(T) - Y(t)) - C'(Z(T) - Z(t)) \Big\}.$$

We see that $P(T) = \omega(T)$, as long as

$$\hat{A}(T) - \hat{A}(t) + \delta_0(T - t) = A(T) - A(t), \quad (58)$$

$$\hat{B} = B, \quad (59)$$

$$\hat{C} = C, \quad (60)$$

$$\hat{C}'(Z^s(T) - Z^s(t)) - \delta_Y' \int_t^T Y(s)ds = C'(Z(T) - Z(t)). \quad (61)$$

The last condition, combined with the dynamics of Z^s and Z, implies that

$$\left(\hat{C}'\tilde{K}_Z^s + \delta_Y'\right)\int_t^T Y(s)ds = C'\tilde{K}_Z\int_t^T Y(s)ds.$$

From this we conclude that for all $m = 1, \ldots, M$

$$C^{m\prime}(\tilde{K}_Z - \tilde{K}_Z^s\Big) = \delta_Y^{m\prime}.$$

Furthermore, we see that the coefficient restrictions in (58,59,60,61) are consistent with the ODE's in (54,55) and in (56,57).

Q.E.D.

8.4 Proof of Theorem 3

From (51), we have that

$$\mathrm{E}[dP(t)] = (r(t) - \delta(t))P(t),$$

under the risk-neutral measure. When we move to the physical measure, since $d\tilde{W}(t) = dW(t) + \Lambda(Y,t)dt$, we get that

$$\mathrm{E}[dP(t)] = (r(t) - \delta(t))P(t) + \Sigma_P\Lambda(Y,t)dt,$$

where Σ_P is the $1 \times (N+M)$ vector of the price's loading on the $N+M$ Brownian motions in the economy. An application of Ito's lemma shows that

$$\Sigma_P = (P_Y'\Sigma_Y + P_Z'\Sigma_Z)\sqrt{S(t)}, \tag{62}$$

where P_Y and P_Z indicate vectors of derivatives. Given the form of the price process in (9) and the form of the price of risk Λ in (28), the result in (34) follows after some algebra.

Q.E.D.

8.5 Proof of Theorem 4

Let us define a new variable $\xi(t)$ as follows

$$\xi(t) \equiv e^{-\int_0^t r(s)ds}P(t).$$

An application of Ito's lemma and using the no-arbitrage condition in (51), we find that

$$d\xi(t) = -\delta(t)\xi(t)dt + e^{-\int_0^t r(s)ds}\Sigma_P d\tilde{W}(t),$$

for Σ_P given in (62). Applying Ito's lemma to $\xi(t)\,e^{\int_0^t \delta(s)ds}$ allows us to conclude that

$$\xi(T) = e^{-\int_0^T \delta(s)ds}\xi(0) + e^{-\int_0^T \delta(s)ds}\int_0^T e^{\int_0^t(\delta(s)-r(s))ds}\Sigma_P d\tilde{W}(t).$$

The second term on the right-hand side is a martingale as long as the following condition holds (see, for example, Oksendal (1998))

$$\mathrm{E}_0^Q\left[\int_0^T \left(e^{-\int_0^T \delta(s)ds + \int_0^t(\delta(s)-r(s))ds}\right)^2 \Sigma_P\Sigma_P' dt\right] < \infty.$$

By Tonelli's Theorem, this expectation is equivalent to

$$\int_0^T \mathrm{E}_0^Q \left[\left(e^{-\int_0^T \delta(s)ds + \int_0^t (\delta(s) - r(s))ds} \right)^2 \Sigma_P \Sigma_P' \right] dt,$$

which is finite because the term inside the expectation is log-normally distributed for all t, and so has finite second moments. With this, we see that

$$\mathrm{E}_0^Q [\xi(T)] = \mathrm{E}_0^Q \left[e^{-\int_0^T \delta(s)ds} \right] \xi(0).$$

Note that

$$\zeta(T) \equiv \int_0^T \delta(s)ds = \delta_0 T + \delta_Y' \int_0^T Y(s)ds.$$

By Lemma 4 below, we see that $\int_0^T Y_i(s)ds$ is Normally distributed for all i and T. Hence we have that

$$\mathrm{E}_0^Q \left[e^{-\zeta(T)} \right] = \exp \left\{ -\mathrm{E}_0^Q [\zeta(T)] + \frac{1}{2} \mathrm{Var}_0^Q (\zeta(T)) \right\}.$$

Using the distribution of $\int_0^T Y_i(s)ds$ in Lemma 4 (keep in mind we are now working under the risk-neutral measure) and taking the limit as $T \to \infty$, we see that the expectation on the left-hand side goes to zero as long as the condition in (44) is satisfied.

Q.E.D.

8.6 Proof of Theorems 5 and 6

The following lemma proves useful.

Lemma 4 *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

$$\int_t^T Y(s)ds \tag{63}$$

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

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

whose variance is

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

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

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

$$\int_t^{t+\tau} r(s)ds = \tau r_0 + \sum_{i=1}^{N} r_{Yi} \int_t^{t+\tau} Y_i(s)ds.$$

Given the behavior of the Y integral in (63), and the fact that the asymptotic variance of $Y_i(t)$ is $\sigma_i^2/(2K_{ii})$, the results of Theorem 5 follow immediately.

From the form of the total return process in (53), we can write that

$$\log \frac{P^m(t+\tau)}{P(t)} = a^m \tau - B^{m\prime}(Y(t+\tau) - Y(t)) - C^{m\prime}(Z(t+\tau) - Z(t)). \tag{64}$$

From the dynamics of Y we can write

$$Y_i(t+\tau) - Y_i(t) =$$
$$\left(e^{-K_{ii}\tau} - 1\right) Y_i(t) + \theta_i \left(1 - e^{-K_{ii}\tau}\right) + \sigma_i e^{-K_{ii}(t+\tau)} \int_t^{t+\tau} e^{K_{ii}u} dW_i(u).$$

From the dynamics of Z we can write

$$Z_j(t+\tau) - Z_j(t) = \mu_j \tau + \sigma_j \left(W_j(t+\tau) - W_j(t)\right) - [K_Z]_j \int_t^{t+\tau} Y(s)ds,$$

where $[\cdot]_j$ indicates the j^{th} row of a matrix. Using the result for the Y integral in (63), we can collect terms in (64) to get the results in the first part of Theorem 6. Note that the excess total returns on a stock are given by

$$\log \frac{P^m(t+\tau)}{P^m(t)} - \int_t^{t+\tau} r(s)ds.$$

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 Theorem 6 follows from this observation.

Q.E.D.

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

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., 1996, *Dynamic Asset Pricing Theory*, 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, 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.

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.

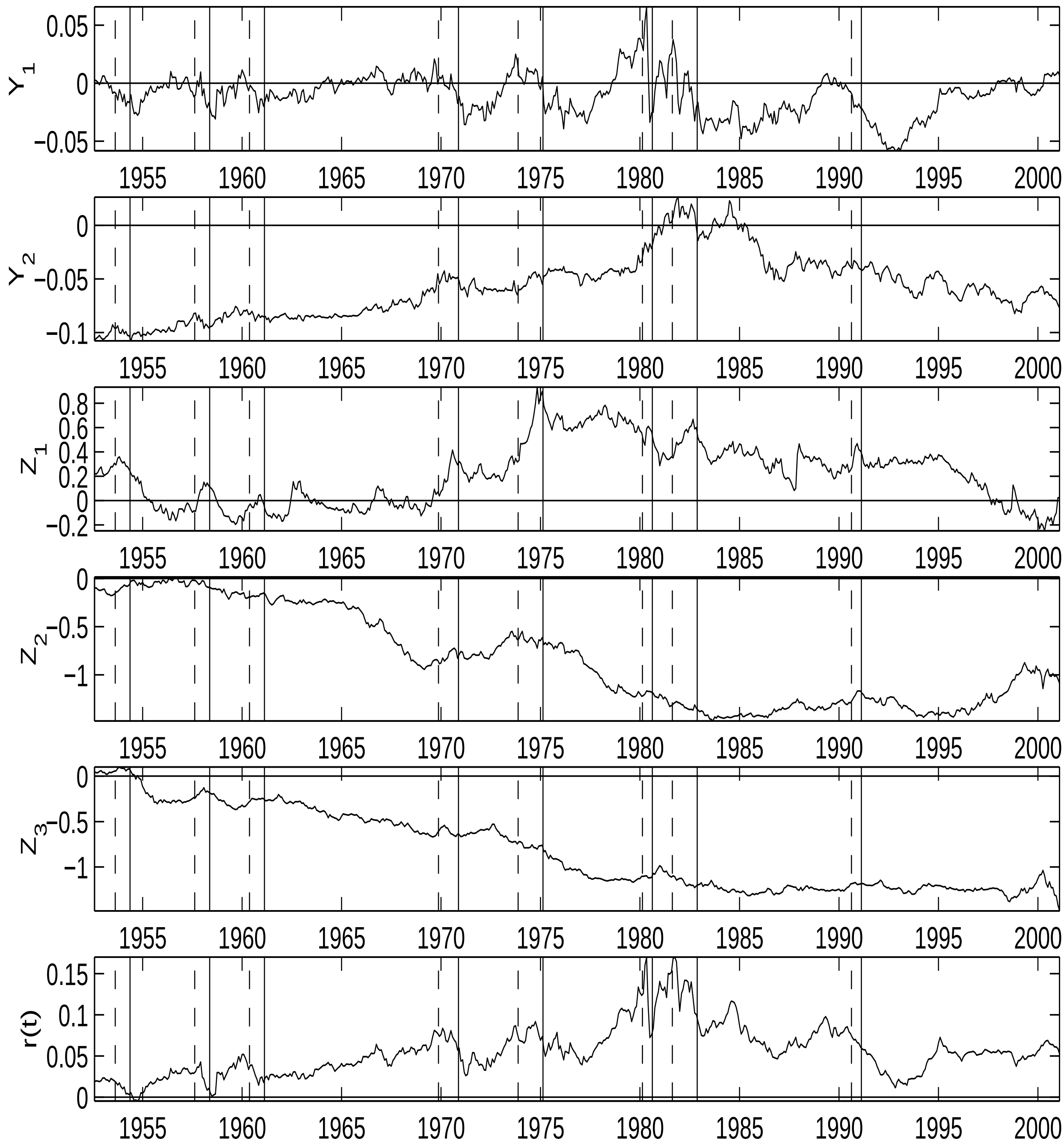

Figure 1: **Factors and the Short Rate.** This figure shows the five extracted model factors, and the short rate. The solid line shows the x-axis. The dashed vertical lines represent NBER business cycle peaks, and the solid vertical lines represent NBER business cycle troughs.

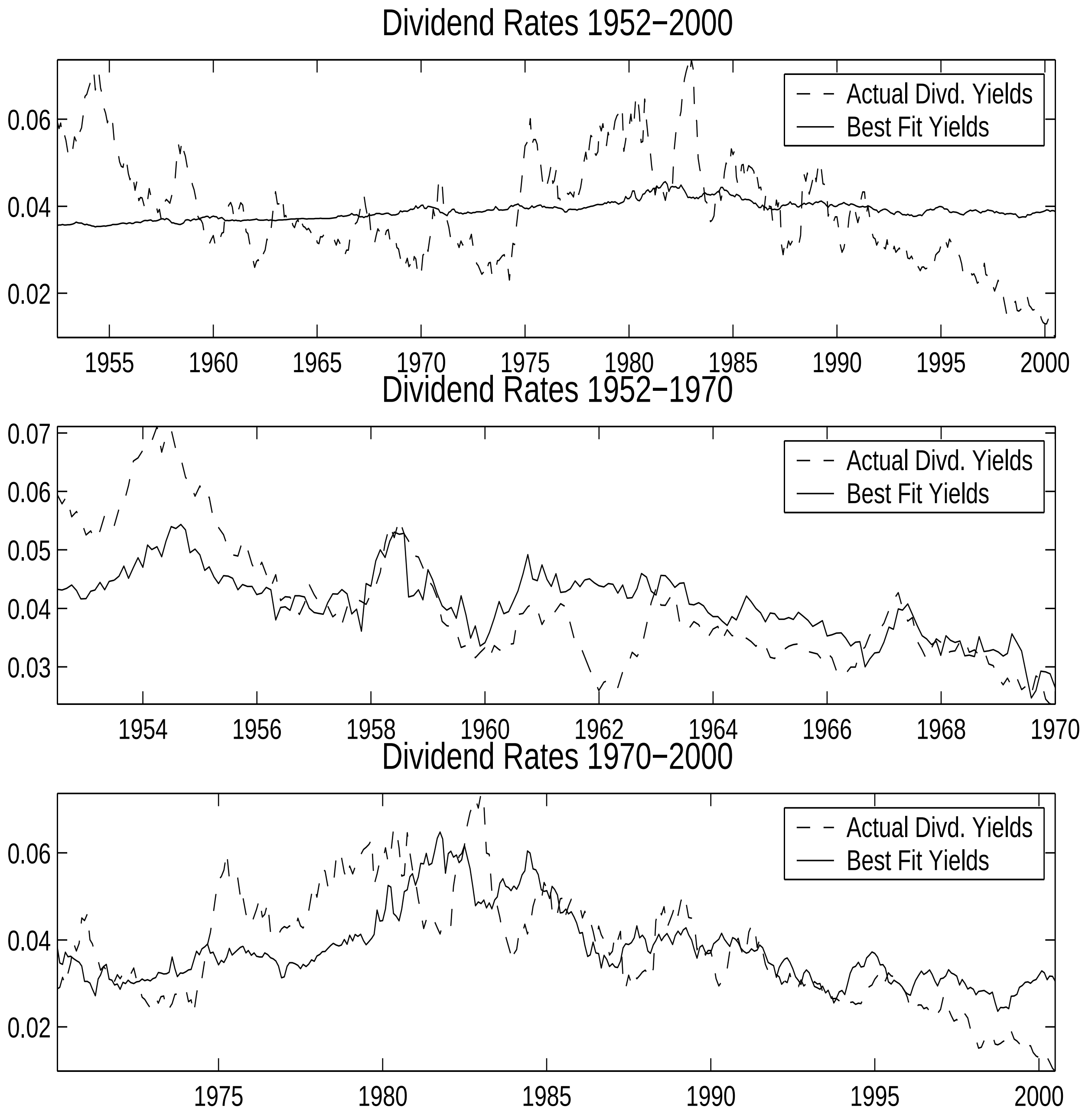

Figure 2: **The Actual and Implied Dividend Rates.** The figure shows the actual dividend rate of the CRSP value-weighted index, as well as the best fit dividend series from the model. The graphs show the full sample, as well as two subsamples. In month t the actual dividend rate is the realized dividend rate in the time interval $[t - 0.5, t + 0.5]$. This latter rate is 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 (49) and is given by $\hat{\delta}(t)$. It depends on the month t values of the Y factors. Each graph is labeled with its respective time period.

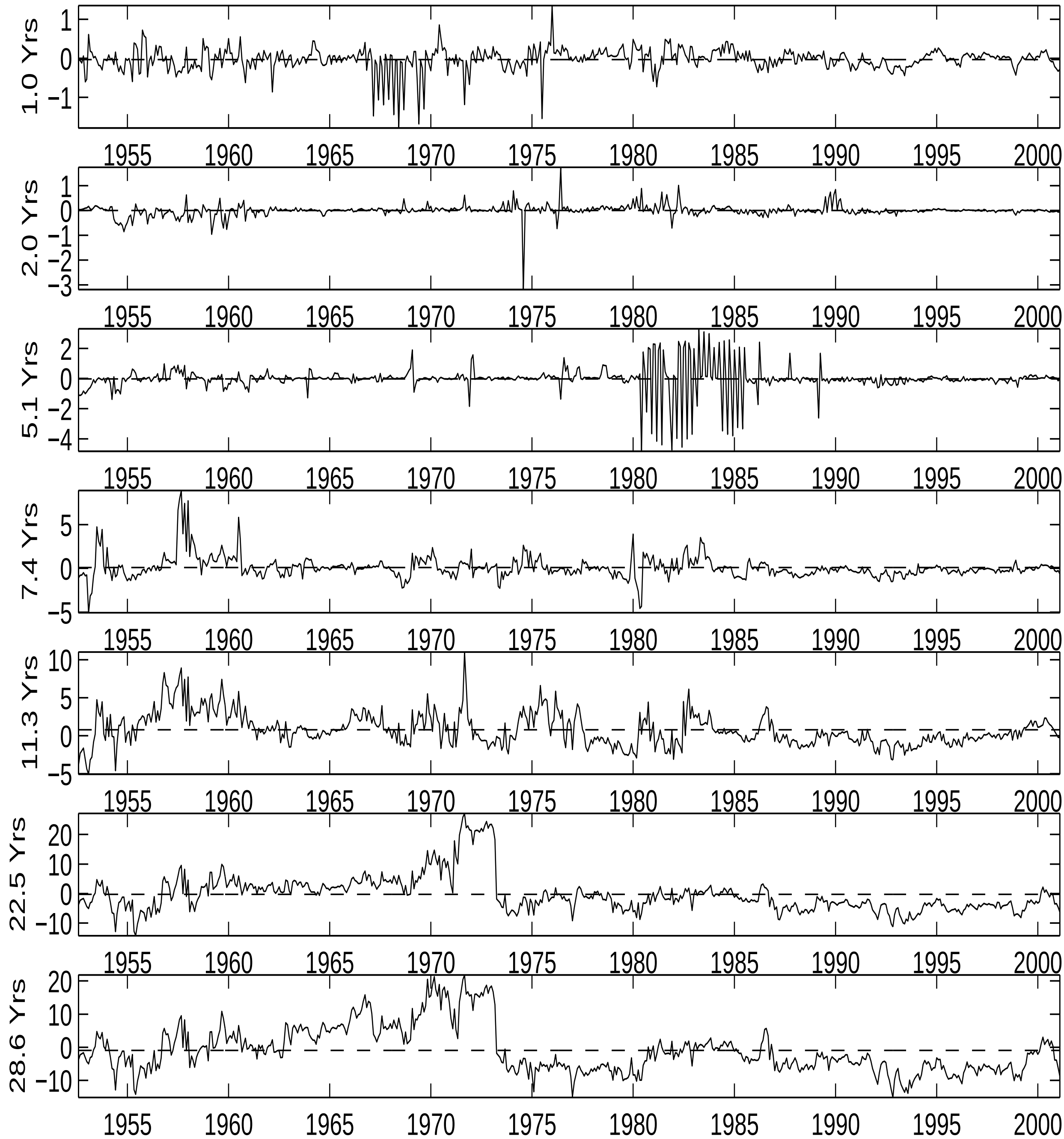

Figure 3: **Bond Pricing Errors.** The pricing errors (in percent) for bonds in the CRSP Fixed Term Indices file. The average maturity of bonds in each series is shown to the left of each graph. The dashed line shows the average level of the pricing error over entire sample period.

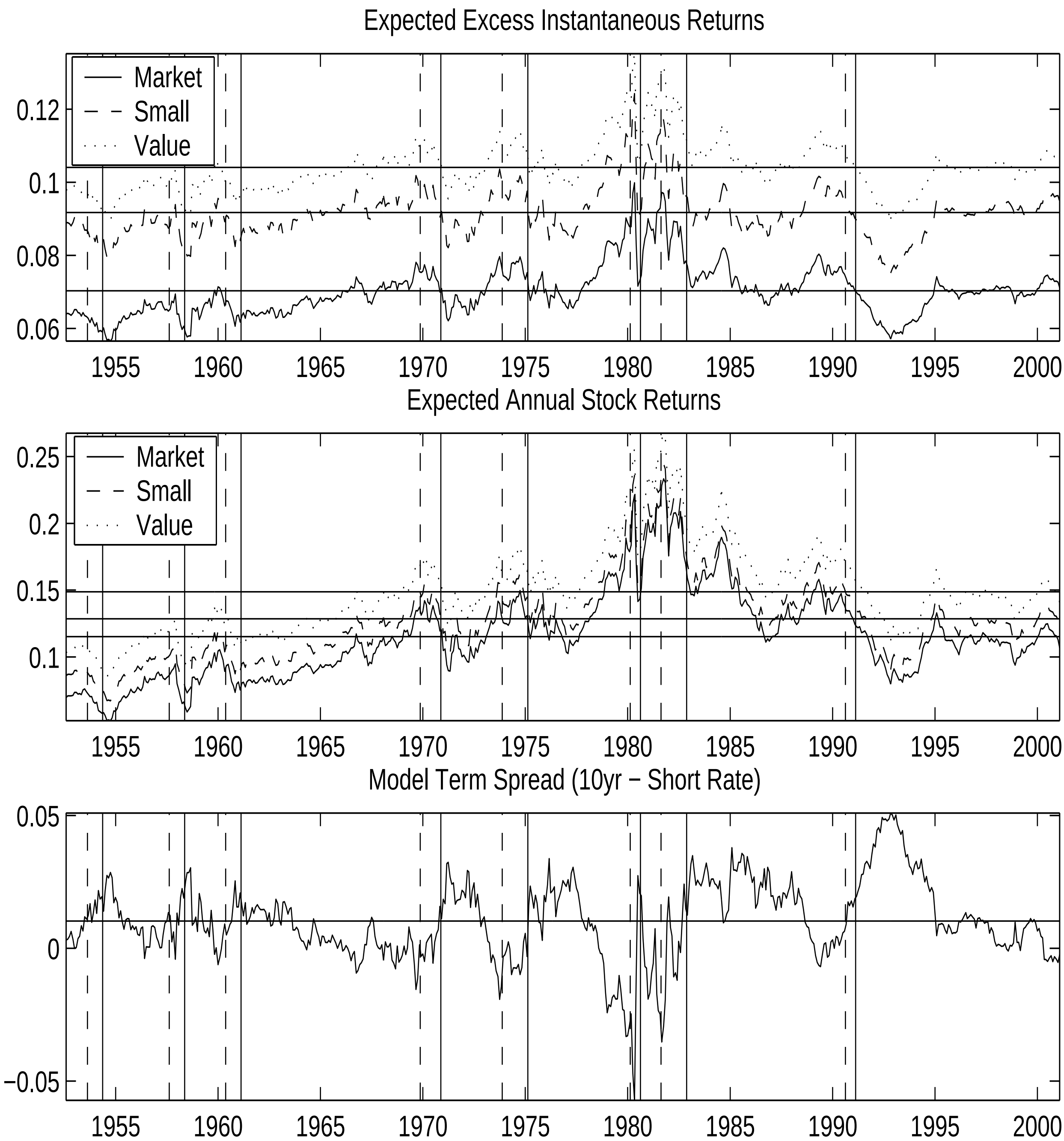

Figure 4: **Expected Returns.** The top graph shows the instantaneous expected excess returns on the CRSP value weighted index, the 3^{rd} decile size portfolio, and the 8^{th} decile value portfolio (the lowest, middle, and top series respectively). The next graph shows the annual expected raw returns for the same portfolios, and in the same order. The last graph shows the term spread implied from the model. The solid horizontal lines represent the mean level of expected excess returns over the sample period. The dashed vertical lines represent NBER business cycle peaks, and the solid vertical lines represent NBER business cycle troughs.

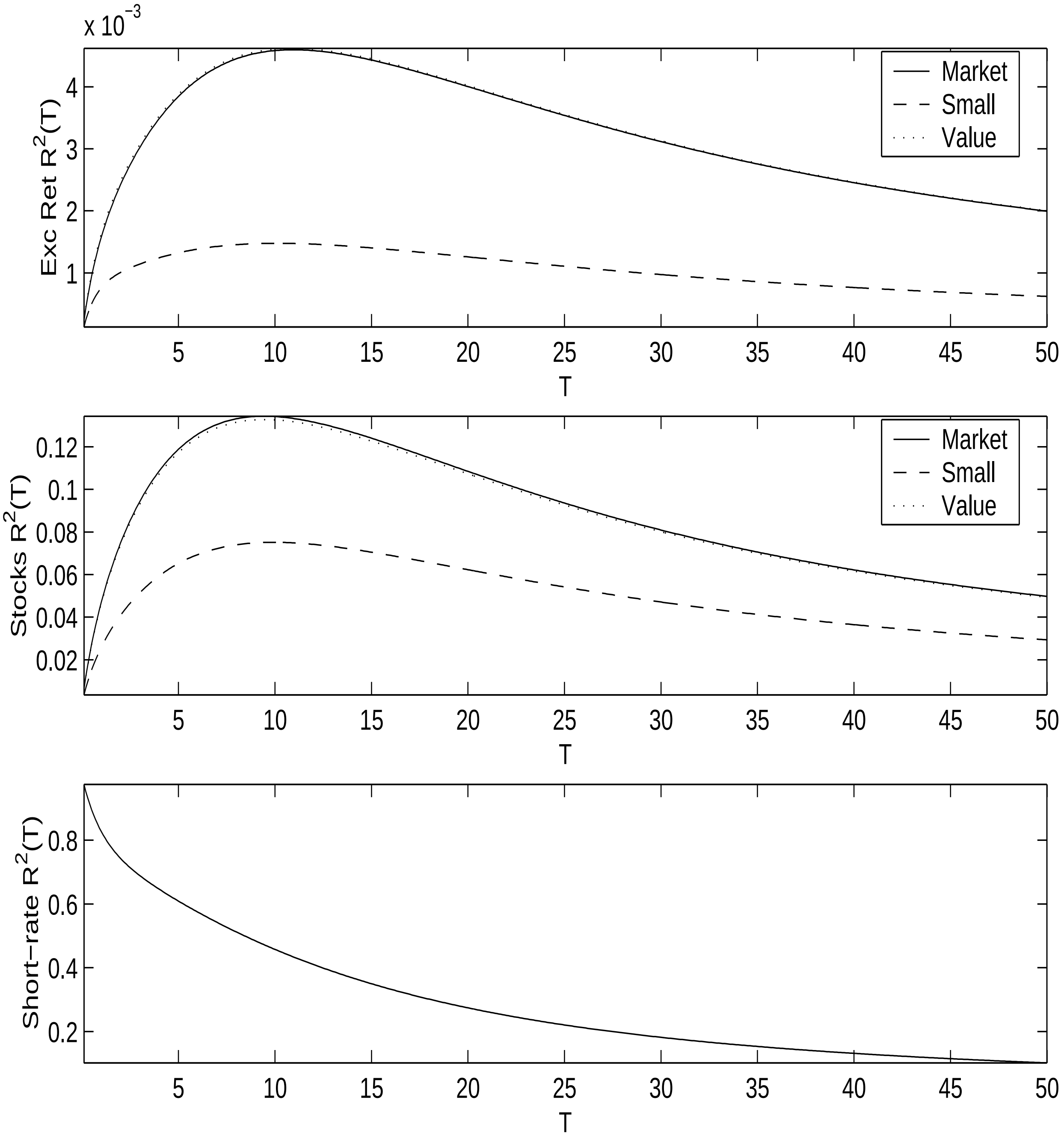

Figure 5: **Theoretical R^2's for Predictive Regressions.** These figures show the theoretical R^2's of a regression of realized returns on expected returns as a function of the forecasting horizon. The top figure shows the R^2's of the forecasting regressions for stock portfolios above the short rate, the middle figure shows the R^2's of forecasting regressions for stock portfolio returns, and the bottom figure shows the R^2's of forecasting regressions for a rolled-over investment at the short rate. The portfolios used are the CRSP value weighted index, the 3^{rd} decile size portfolio, and the 8^{th} decile value portfolio.

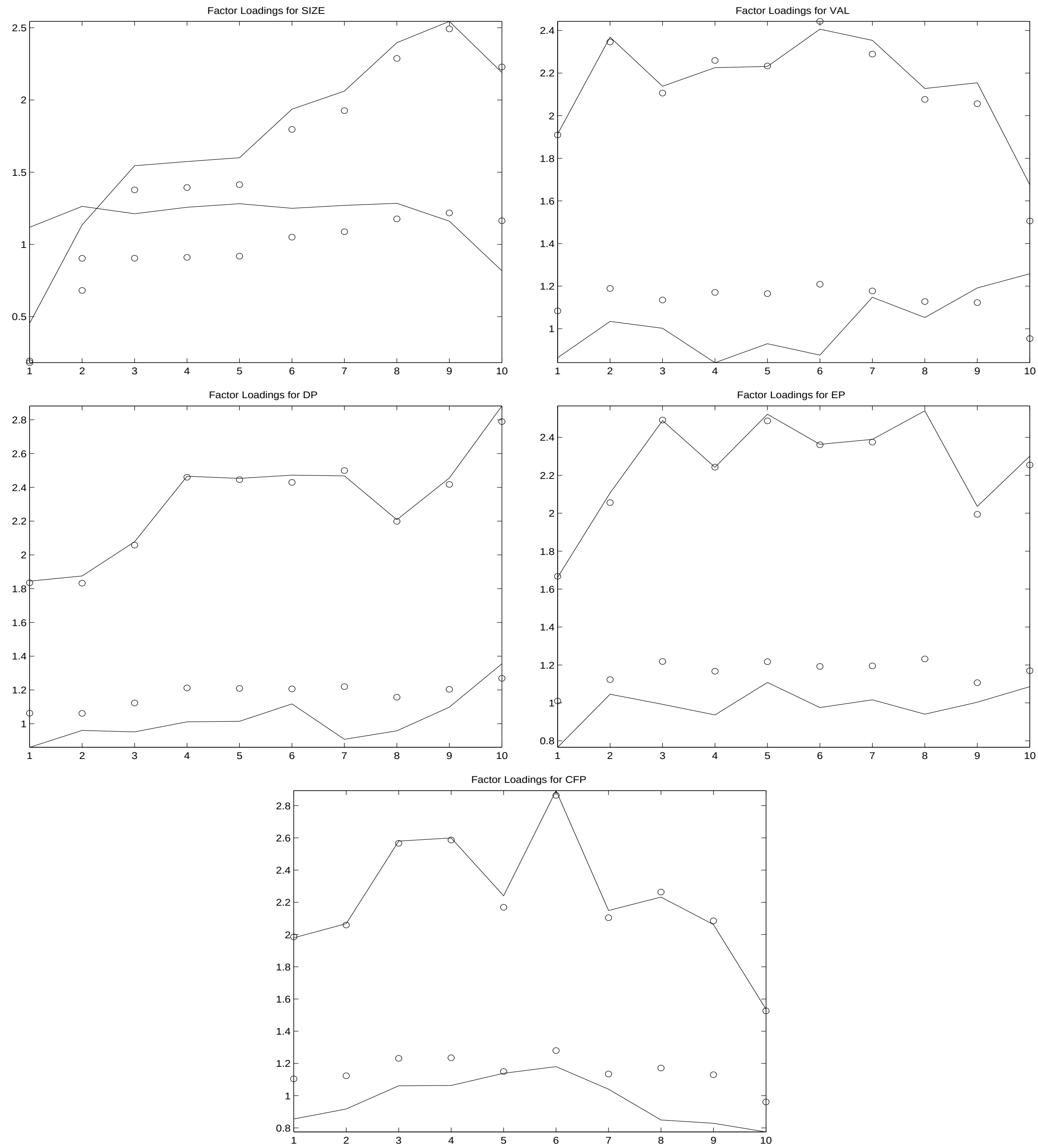

Figure 6: **Equity Loadings on Joint Factors** For the five equity characteristics sorts, these show the loadings of the portfolios on the joint bond-stock factors. The circles represent the loadings on these joint factors of the zero coupon bond with the closest loadings to the equity portfolios. The x-axis corresponds to the decile portfolios in the characteristics sorts.

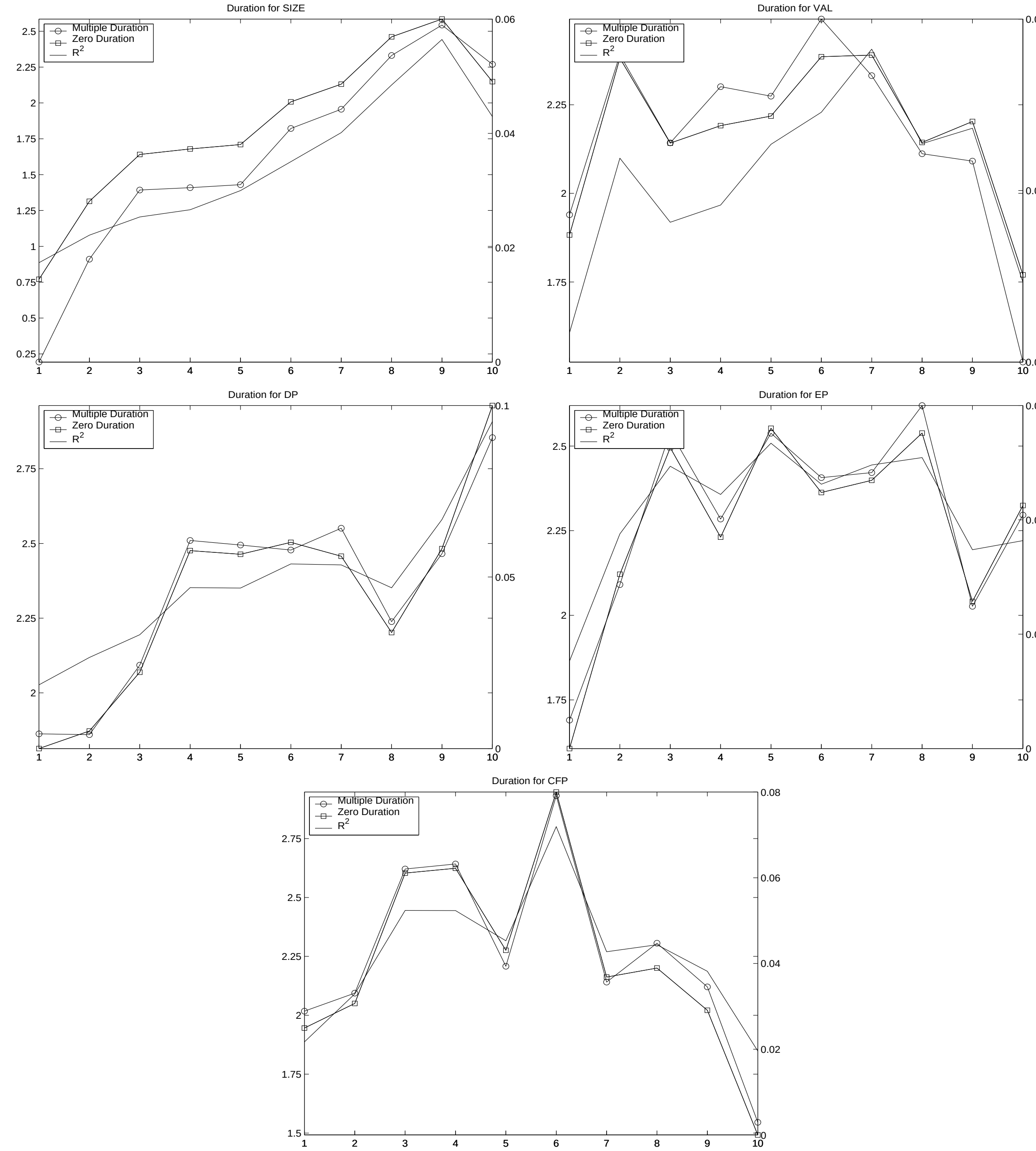

Figure 7: **Durations and R^2's of Equity Portfolios.** For the five equity characteristics sorts, these show durations of the equity portfolios (left axis) using the two duration measures discussed in the text, as well as the R^2's of the regression of continuously compounded equity returns on innovations in the joint bond-stock factors (right axis). The x-axis corresponds to the decile portfolios in the characteristics sorts.

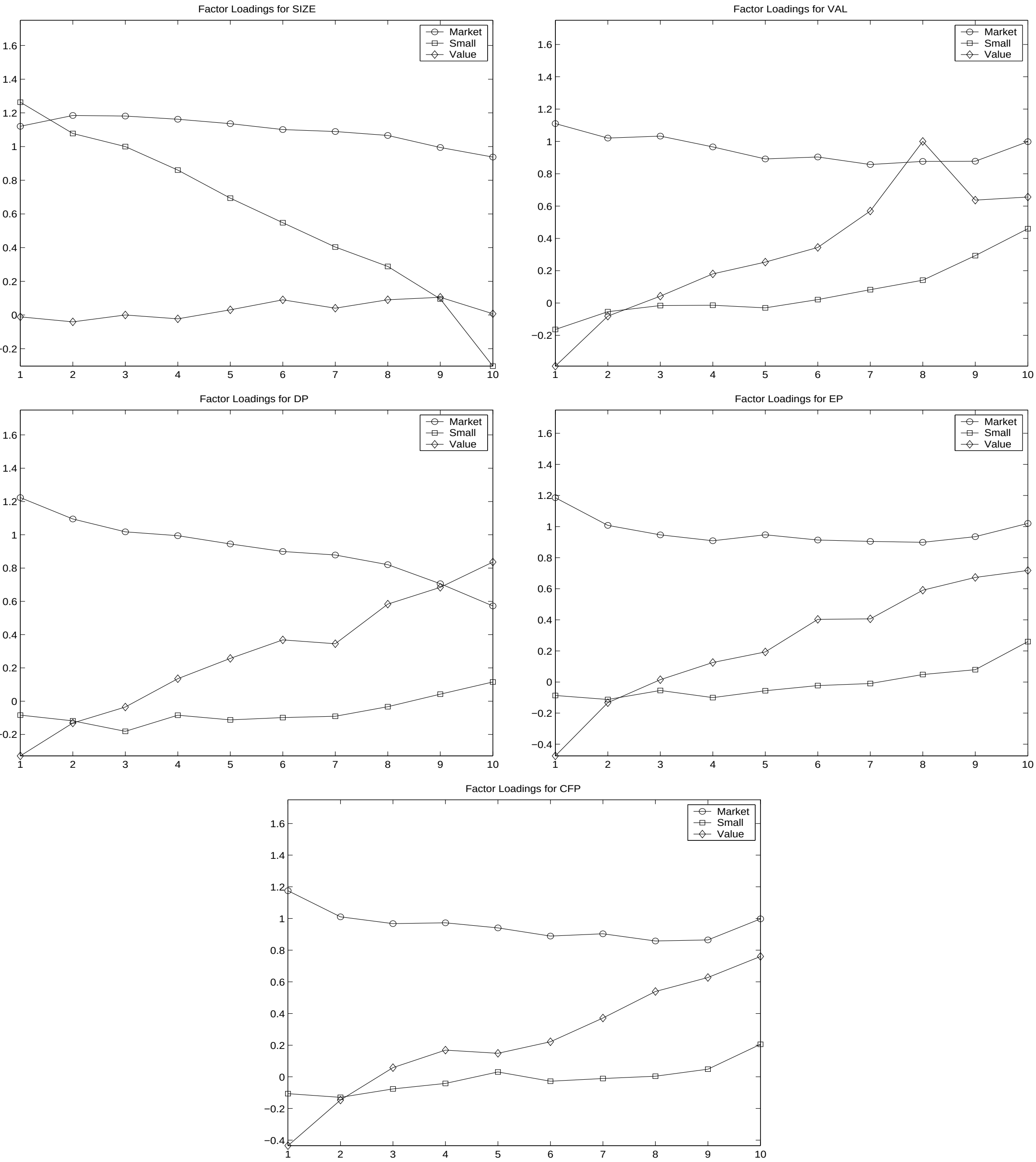

Figure 8: **Equity Loadings on Stock Factors.** For the five equity characteristics sorts, these show the loadings of the equity portfolios on the stock specific factors (the Z's). The x-axis corresponds to the decile portfolios in the characteristics sorts.

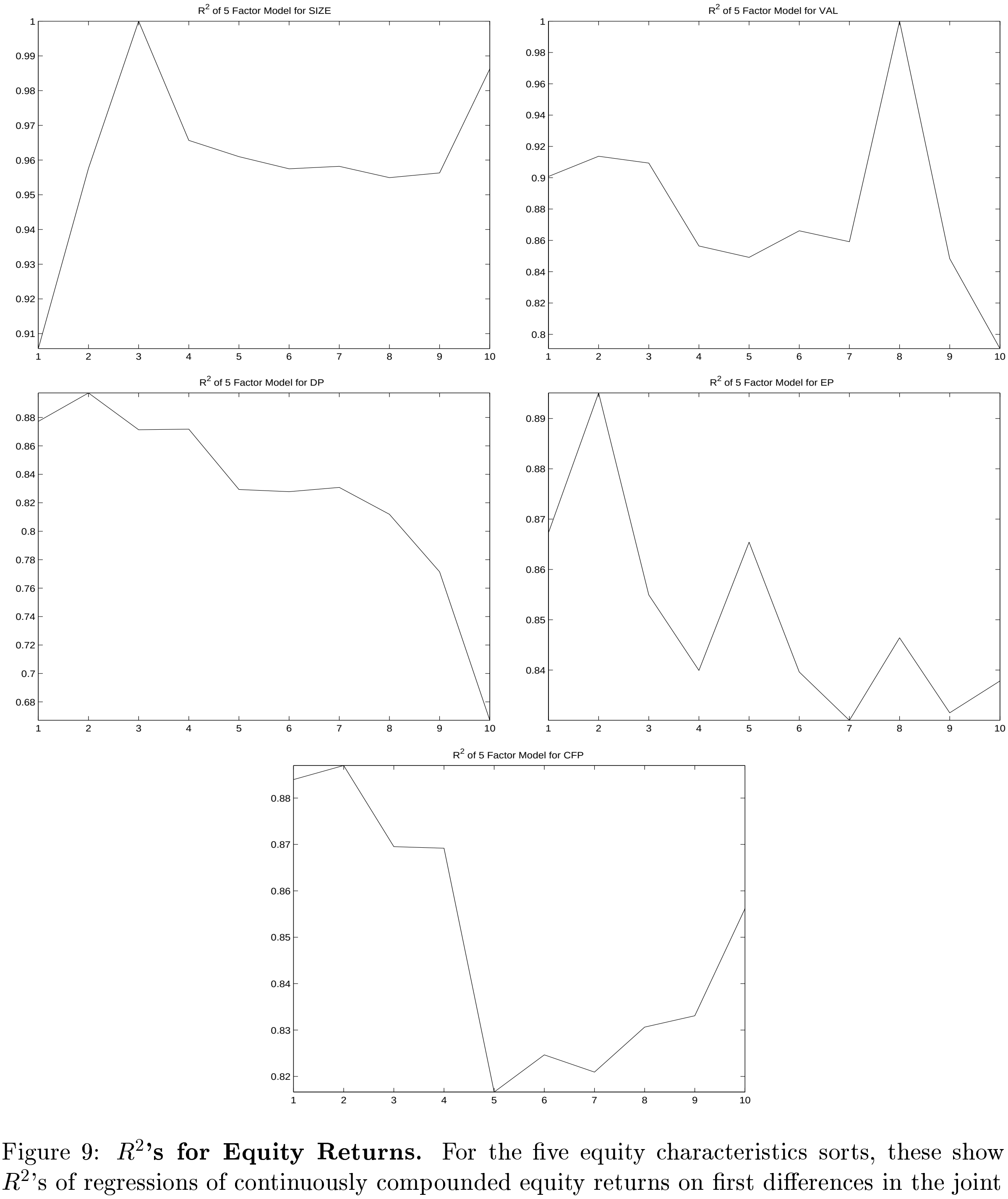

Figure 9: **R^2's for Equity Returns.** For the five equity characteristics sorts, these show the R^2's of regressions of continuously compounded equity returns on first differences in the joint and stock-specific factors. The x-axis corresponds to the decile portfolios in the characteristics sorts.

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 not estimated. The K_Z matrix is computed using the relationship in (41).

	Parameter Estimates				
Σ_Y	0.02453069	0.01204956			
	(0.00075223)	(0.00029038)			
$\tilde{K}$	0.67246967	0.01604500			
	(0.01917058)	(0.00116372)			
K	0.96492578	0.11230279			
	(0.20836571)	(0.05982335)			
$\tilde{\Theta}$	0.00000000	0.00000000			
Θ	-0.01000561	-0.05193591			
	(0.00366470)	(0.01694422)			
r_0	0.12309002				
	(0.00615245)				
r_Y	1.00000000	1.00000000			
$A^n(0)$	0.00000000	0.00000000	0.00000000		
$[C]_1$	1.00000000	0.00000000	0.00000000		
$[C]_2$	1.18132127	1.00000000	0.00000000		
	(0.02598111)				
$[C]_3$	0.87656805	0.14160487	1.00000000		
	(0.02080910)	(0.03267899)			
$\tilde{\mu}$	0.00000000	0.00000000	0.00000000		
μ	-0.06298961	-0.00869968	-0.03991872		
	(0.02075715)	(0.01304387)	(0.01029536)		
$[B]_1$	0.98838566	2.10195785	$[K_Z]_1$	0.33534062	0.96627409
	(0.27240396)	(0.45517053)			
$[B]_2$	1.21258198	1.54457729	$[K_Z]_2$	-0.21156961	-0.16626288
	(0.36638890)	(0.61901620)			
$[B]_3$	1.05244414	2.12752207	$[K_Z]_3$	0.02827365	0.14240255
	(0.27305563)	(0.45604879)			
Σ_Z	0.14402203	0.09026597	0.07116300		
	(0.00422220)	(0.00264650)	(0.00208656)		
σ_ϵ^2	0.00044151	0.00368130	0.00000377	0.00000533	0.00000919
	(0.00002878)	(0.00023122)	(0.00000149)	(0.00000296)	(0.00000294)
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 bottom row of the table reports Dickey-Fuller unit root tests, with the $*$ indicating rejection of the unit root null at the 99% level.

	Y_1	Y_2	Z_1	Z_2	Z_3
Y_1	1.0000				
Y_2	-0.3491	1.0000			
Z_1	-0.0082	0.0010	1.0000		
Z_2	-0.0037	-0.0025	0.0033	1.0000	
Z_3	0.0042	-0.0003	-0.0022	0.0023	1.0000
DF	11.8489*	2.0448	1.8010	0.9780	2.0150

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
0.8002		0.6900	-0.9884	-2.1020	-1.0000	-0.0000	0.0000	1.0000
(0.0468)			(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
	0.3902	0.0537	-0.9884	-2.1020	-1.0000			1.0000
	(0.0960)		(0.0000)	(0.0000)	(0.0000)			
1.0000	1.0000	1.0000	-0.9478	-2.0826				0.0415
(0.0000)	(0.0000)		(0.2594)	(0.5398)				
					-0.9991			0.9574
					(0.0094)			

Table 4: **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 (44). The best fit dividend series from the model solves (49), 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	$\hat{\delta}_0$	$\hat{\delta}_{Y1}$	$\hat{\delta}_{Y2}$	α	β	R^2
6/52–12/00	0.0435	0.0230	0.0743	-0.0000	1.0000	0.0285
				(0.0355)	(0.8977)	
6/52–12/69	0.0177	-0.4211	-0.2498	-0.0166	1.4004	0.5313
				(0.0097)	(0.2659)	
1/70–12/00	0.0545	0.0762	0.3764	-0.0000	1.0000	0.4282
				(0.0073)	(0.1951)	

Table 5: **Variation in Size Sorted Portfolios.** This table shows regressions of the first differenced log returns of 10 size sorted portfolios on first differences in the two common and the three stock specific factors. Data are monthly. The bottom row shows the standard deviation of the factor loadings across the characteristic sorts. Newey-West standard errors are in parentheses.

Type	Y_1	Y_2	Z_1	Z_2	Z_3	R^2	Mean	S.D.
Small	-1.1184	-0.4541	-1.1202	-1.2635	0.0108	0.9057	0.0099	0.0603
	(0.0911)	(0.2118)	(0.0175)	(0.0485)	(0.0571)			
	-1.2634	-1.1351	-1.1845	-1.0776	0.0406	0.9576	0.0103	0.0584
	(0.0655)	(0.1301)	(0.0115)	(0.0331)	(0.0380)			
	-1.2126	-1.5446	-1.1813	-1.0000	0.0000	1.0000	0.0107	0.0561
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)			
	-1.2572	-1.5736	-1.1622	-0.8605	0.0227	0.9657	0.0105	0.0548
	(0.0572)	(0.1076)	(0.0121)	(0.0266)	(0.0291)			
	-1.2818	-1.6001	-1.1359	-0.6937	-0.0310	0.9610	0.0108	0.0522
	(0.0599)	(0.1100)	(0.0126)	(0.0225)	(0.0280)			
	-1.2498	-1.9358	-1.1011	-0.5479	-0.0903	0.9575	0.0103	0.0498
	(0.0698)	(0.1257)	(0.0116)	(0.0229)	(0.0326)			
	-1.2701	-2.0611	-1.0894	-0.4035	-0.0403	0.9582	0.0104	0.0483
	(0.0555)	(0.1078)	(0.0112)	(0.0223)	(0.0323)			
	-1.2844	-2.3965	-1.0661	-0.2883	-0.0908	0.9549	0.0104	0.0471
	(0.0569)	(0.1036)	(0.0108)	(0.0225)	(0.0254)			
	-1.1612	-2.5442	-0.9947	-0.0957	-0.1049	0.9563	0.0101	0.0436
	(0.0718)	(0.1157)	(0.0090)	(0.0184)	(0.0231)			
Big	-0.8176	-2.1897	-0.9378	0.3033	-0.0077	0.9863	0.0095	0.0408
	(0.0299)	(0.0534)	(0.0049)	(0.0112)	(0.0112)			
SD(β)	0.1423	0.6259	0.0802	0.4846	0.0515			

Table 6: **Variation in Value Sorted Portfolios.** This table shows regressions of the first differenced log returns of 10 value sorted portfolios on first differences in the two common and the three stock specific factors. Data are monthly. The bottom row shows the standard deviation of the factor loadings across the characteristic sorts. Newey-West standard errors are in parentheses.

Type	Y_1	Y_2	Z_1	Z_2	Z_3	R^2	Mean	S.D.
Growth	-0.8639	-1.9136	-1.1107	0.1635	0.3901	0.9008	0.0086	0.0501
	(0.1049)	(0.2152)	(0.0211)	(0.0282)	(0.0362)			
	-1.0345	-2.3686	-1.0208	0.0538	0.0809	0.9137	0.0096	0.0454
	(0.0849)	(0.1504)	(0.0203)	(0.0329)	(0.0565)			
	-1.0022	-2.1381	-1.0328	0.0154	-0.0428	0.9094	0.0097	0.0458
	(0.0702)	(0.1400)	(0.0200)	(0.0368)	(0.0516)			
	-0.8408	-2.2255	-0.9659	0.0135	-0.1807	0.8564	0.0092	0.0444
	(0.1310)	(0.1955)	(0.0272)	(0.0476)	(0.0777)			
	-0.9299	-2.2310	-0.8915	0.0296	-0.2532	0.8492	0.0105	0.0416
	(0.1152)	(0.1966)	(0.0291)	(0.0436)	(0.0566)			
	-0.8763	-2.4057	-0.9038	-0.0215	-0.3441	0.8661	0.0107	0.0421
	(0.0877)	(0.1759)	(0.0217)	(0.0340)	(0.0421)			
	-1.1479	-2.3534	-0.8570	-0.0825	-0.5696	0.8591	0.0105	0.0417
	(0.1003)	(0.1711)	(0.0210)	(0.0376)	(0.0489)			
	-1.0524	-2.1275	-0.8766	-0.1416	-1.0000	1.0000	0.0124	0.0428
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)			
	-1.1914	-2.1544	-0.8773	-0.2935	-0.6368	0.8484	0.0124	0.0439
	(0.1030)	(0.2106)	(0.0218)	(0.0391)	(0.0513)			
Value	-1.2580	-1.6755	-0.9984	-0.4601	-0.6561	0.7909	0.0122	0.0516
	(0.2075)	(0.3088)	(0.0354)	(0.0528)	(0.0554)			
SD(β)	0.1450	0.2230	0.0850	0.1837	0.4088			

Table 7: **Variation in D/P Sorted Portfolios.** This table shows regressions of the first differenced log returns of 10 D/P sorted portfolios on first differences in the two common and the three stock specific factors. Data are monthly. The bottom row shows the standard deviation of the factor loadings across the characteristic sorts. Newey-West standard errors are in parentheses.

Type	Y_1	Y_2	Z_1	Z_2	Z_3	R^2	Mean	S.D.
Low	-0.8606	-1.8447	-1.2235	0.0835	0.3284	0.8772	0.0099	0.0552
	(0.1607)	(0.1841)	(0.0243)	(0.0498)	(0.0679)			
	-0.9603	-1.8750	-1.0950	0.1183	0.1308	0.8973	0.0089	0.0488
	(0.1036)	(0.1657)	(0.0203)	(0.0359)	(0.0624)			
	-0.9518	-2.0777	-1.0178	0.1813	0.0351	0.8713	0.0101	0.0463
	(0.1127)	(0.1864)	(0.0265)	(0.0501)	(0.0842)			
	-1.0111	-2.4657	-0.9946	0.0842	-0.1352	0.8718	0.0102	0.0455
	(0.1138)	(0.1982)	(0.0256)	(0.0414)	(0.0747)			
	-1.0144	-2.4532	-0.9450	0.1121	-0.2571	0.8293	0.0093	0.0447
	(0.1164)	(0.1914)	(0.0317)	(0.0513)	(0.0804)			
	-1.1172	-2.4721	-0.8997	0.0983	-0.3685	0.8278	0.0103	0.0433
	(0.1097)	(0.1562)	(0.0265)	(0.0579)	(0.0727)			
	-0.9079	-2.4682	-0.8787	0.0903	-0.3451	0.8308	0.0103	0.0421
	(0.0952)	(0.1774)	(0.0254)	(0.0585)	(0.0646)			
	-0.9581	-2.2083	-0.8204	0.0328	-0.5834	0.8119	0.0116	0.0411
	(0.1082)	(0.1832)	(0.0236)	(0.0586)	(0.0705)			
	-1.0983	-2.4550	-0.7058	-0.0423	-0.6844	0.7715	0.0109	0.0386
	(0.1159)	(0.2159)	(0.0241)	(0.0411)	(0.0545)			
High	-1.3549	-2.8822	-0.5727	-0.1157	-0.8364	0.6671	0.0099	0.0388
	(0.1492)	(0.2675)	(0.0383)	(0.0523)	(0.0683)			
SD(β)	0.1405	0.3186	0.1876	0.0858	0.3706			

Table 8: **Variation in E/P Sorted Portfolios.** This table shows regressions of the first differenced log returns of 10 E/P sorted portfolios on first differences in the two common and the three stock specific factors. Data are monthly. The bottom row shows the standard deviation of the factor loadings across the characteristic sorts. Newey-West standard errors are in parentheses.

Type	Y_1	Y_2	Z_1	Z_2	Z_3	R^2	Mean	S.D.
Low	-0.7660	-1.6613	-1.1856	0.0864	0.4749	0.8673	0.0083	0.0543
	(0.1335)	(0.2284)	(0.0241)	(0.0390)	(0.0444)			
	-1.0455	-2.1065	-1.0079	0.1121	0.1330	0.8951	0.0081	0.0453
	(0.0839)	(0.1754)	(0.0207)	(0.0325)	(0.0568)			
	-0.9924	-2.4881	-0.9465	0.0542	-0.0153	0.8549	0.0096	0.0437
	(0.0975)	(0.1578)	(0.0258)	(0.0511)	(0.0719)			
	-0.9365	-2.2420	-0.9088	0.1004	-0.1258	0.8399	0.0093	0.0424
	(0.1111)	(0.1854)	(0.0268)	(0.0455)	(0.0802)			
	-1.1078	-2.5219	-0.9473	0.0561	-0.1934	0.8654	0.0101	0.0438
	(0.0843)	(0.1641)	(0.0237)	(0.0511)	(0.0723)			
	-0.9754	-2.3627	-0.9134	0.0230	-0.4031	0.8396	0.0116	0.0434
	(0.0987)	(0.1729)	(0.0288)	(0.0438)	(0.0700)			
	-1.0164	-2.3893	-0.9044	0.0099	-0.4064	0.8300	0.0120	0.0434
	(0.0969)	(0.1825)	(0.0193)	(0.0434)	(0.0603)			
	-0.9405	-2.5397	-0.8989	-0.0484	-0.5908	0.8464	0.0129	0.0439
	(0.1192)	(0.1752)	(0.0230)	(0.0418)	(0.0523)			
	-1.0033	-2.0360	-0.9350	-0.0794	-0.6726	0.8315	0.0134	0.0461
	(0.1373)	(0.2215)	(0.0251)	(0.0434)	(0.0538)			
High	-1.0855	-2.3012	-1.0206	-0.2598	-0.7174	0.8378	0.0136	0.0505
	(0.1453)	(0.2129)	(0.0253)	(0.0601)	(0.0581)			
SD(β)	0.0958	0.2708	0.0875	0.1119	0.3811			

Table 9: **Variation in CF/P Sorted Portfolios.** This table shows regressions of the first differenced log returns of 10 CF/P sorted portfolios on first differences in the two common and the three stock specific factors. Data are monthly. The bottom row shows the standard deviation of the factor loadings across the characteristic sorts. Newey-West standard errors are in parentheses.

Type	Y_1	Y_2	Z_1	Z_2	Z_3	R^2	Mean	S.D.
Low	-0.8552	-1.9797	-1.1756	0.1061	0.4344	0.8840	0.0084	0.0534
	(0.1243)	(0.2311)	(0.0226)	(0.0332)	(0.0419)			
	-0.9175	-2.0676	-1.0103	0.1294	0.1465	0.8870	0.0093	0.0456
	(0.0973)	(0.1910)	(0.0218)	(0.0314)	(0.0593)			
	-1.0612	-2.5802	-0.9675	0.0766	-0.0582	0.8695	0.0091	0.0444
	(0.0952)	(0.1923)	(0.0213)	(0.0460)	(0.0689)			
	-1.0632	-2.5995	-0.9728	0.0417	-0.1685	0.8692	0.0092	0.0447
	(0.0762)	(0.1632)	(0.0237)	(0.0494)	(0.0650)			
	-1.1390	-2.2406	-0.9405	-0.0306	-0.1486	0.8166	0.0110	0.0445
	(0.1230)	(0.1922)	(0.0274)	(0.0552)	(0.0707)			
	-1.1801	-2.8937	-0.8892	0.0280	-0.2216	0.8246	0.0096	0.0428
	(0.0880)	(0.1604)	(0.0306)	(0.0642)	(0.0798)			
	-1.0401	-2.1491	-0.9034	0.0107	-0.3716	0.8209	0.0111	0.0433
	(0.1307)	(0.2177)	(0.0306)	(0.0462)	(0.0740)			
	-0.8494	-2.2326	-0.8577	-0.0039	-0.5391	0.8306	0.0114	0.0419
	(0.1227)	(0.1941)	(0.0248)	(0.0391)	(0.0551)			
	-0.8287	-2.0621	-0.8648	-0.0483	-0.6271	0.8331	0.0135	0.0427
	(0.1246)	(0.1862)	(0.0273)	(0.0388)	(0.0493)			
High	-0.7752	-1.5367	-0.9974	-0.2056	-0.7602	0.8562	0.0134	0.0486
	(0.1269)	(0.1942)	(0.0262)	(0.0505)	(0.0508)			
SD(β)	0.1426	0.3806	0.0935	0.0949	0.3617			