Pricing the Term Structure with Linear Regressions

Tobias Adrian and Emanuel Moench*
Federal Reserve Bank of New York†

First Version: June 4, 2008
This Version: February 13, 2009

Abstract

We suggest a method to estimate affine term structure models using linear regressions. Our approach combines cross-sectional pricing and bond return forecasting by specifying market prices of risk as affine functions of the state variables. Our method is fast, easy to implement and allows estimation of models with a large number of pricing factors. Our bond return regressions generate small in-sample pricing errors. A five factor specification outperforms commonly used three-factor specifications in out-of-sample forecasts of yields. Expected bond returns are mainly driven by factors which explain little of the contemporaneous cross-sectional variation of yields.

Keywords: Term Structure of Interest Rates, Fama-MacBeth Regressions
JEL classification: G10, G12

*The authors would like to thank Darrell Duffie, Arturo Estrella, Eric Ghysels, Paul Glasserman, Scott Joslin, Don Kim, Monika Piazzesi, Joshua Rosenberg, Jonathan Wright, Hao Wu, and seminar participants at the Federal Reserve Bank of New York and the SITE workshop at Stanford University for helpful comments and discussions.

The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

†Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045.
Contact for Adrian: `tobias.adrian@ny.frb.org`, `http://nyfedeconomists.org/adrian/`.
Contact for Moench: `emanuel.moench@ny.frb.org`, `http://nyfedeconomists.org/moench/`.

1 Introduction

Many term structure models start with three assumptions: 1) the pricing kernel is exponentially affine in the shocks driving the economy, 2) prices of risk are affine in the state variables, and 3) innovations to state variables and log yield observation errors are conditionally Gaussian and independent (see Dai and Singleton (2000, 2002, 2003), Goldstein and Dufresne (2001), Duffee (2002), Kim and Wright (2005) for examples). These assumptions give rise to log yields that are affine in the state variables, and whose coefficients on the state variables are subject to no-arbitrage constraints across maturities (see Duffie and Kan (1996), Piazzesi (2003), and Singleton (2007) for overviews). Empirically, the affine term structure literature commonly uses maximum likelihood methods to estimate coefficients and pricing factors, thus exploiting the distributional assumptions as well as the no-arbitrage constraints.

In this paper, we propose an alternative, regression based approach to estimating the time series and cross section of interest rates. We follow the literature in assuming that prices of risk are affine functions of the state variables. We depart from the literature by not making explicit assumptions about the functional form of the pricing kernel; we instead assume a particular functional form of the return generating process. Starting with observable pricing factors and no-arbitrage equations, we propose a three-stage ordinary least squares estimator. In the first stage, we decompose pricing factors into predictable components and factor innovations by regressing factors on their lagged levels. In the second stage, we estimate exposures of bond *returns* with respect to lagged levels of pricing factors and pricing factor innovations (the latter obtained in the first stage). In the third stage, we decompose the exposures of returns on the lagged pricing factors (from the second stage) into prices of risk and yield factor loadings from a cross-sectional regression. We compute standard errors that adjust for the generated regressor uncertainty from the previous stages via bootstrap.

While we can estimate all model parameters including the physical and risk neutral dynamics from our three step estimator, our model does not give rise to closed form solutions for log bond prices. Instead, we take an approximation of log returns to returns, and assume that the higher order terms of the approximation are part of the yield pricing error. Given this approximation, we show that log bond prices are affine functions of the state variables with pricing parameters that are subject to linear recursive cross-equation restrictions. In addition, *changes* in the log bond pricing error are linear functions of the return approximation error and the return pricing error. Return pricing errors are quantitatively small. Moreover, return pricing errors and yield pricing errors have no forecasting power for returns or yields. Yet, per construction, the yield pricing error is highly autocorrelated, which is not a violation of no arbitrage in our setting.

We treat pricing factors as observable. In our preferred specification, we use the first five principal components of the yield curve as pricing factors. We find that a smaller number of yield factors is not able to adequately capture the dynamics of both returns and yields. In our preferred five factor specification, the Cochrane and Piazzesi (2005) return forecasting factor does not add significant forecasting power for excess bond returns.

Our starting point of using yield principal components as factors is not essential. Indeed, we show that we can alternatively use forward rates or the parameters of the Nelson-Siegel-Svensson curve from which our zero-coupon term structure is constructed as pricing factors. The choice of pricing factors does not seem essential as long as the factors span both the time-series *and* cross-sectional variation of the term structure of interest rates.

One major point of departure from existing models is that we do not impose the cross-sectional yield no-arbitrage constraints when we estimate our model. Under the null that the model is correctly specified, imposing the constraint should not change any of the parameter estimates. Instead, we compute model implied yields from the estimated return parameters, and compare these implied yields to observed yields. We find that the resulting yield pricing errors are small across the yield curve.

Our estimation procedure is computationally fast, as it only relies on linear regressions. It is thus easy to use the model for out-of-sample forecasts. We show that our preferred five factor specification gives rise to smaller out-of sample pricing errors than A) the random walk model, B) a three factor specification, C) the Diebold and Li (2006) model. Our model can thus be readily used in real time policy analysis. It can also be estimated when the number of pricing factors equals the number of yields, as is the case in empirical applications of the Heath, Jarrow, and Morton (1992) methodology.

Using the three step regression approach comes at the cost of reduced efficiency compared to estimation via maximum likelihood. However, it is straightforward to estimate our model via two-step or iterative GMM to achieve efficiency, even if pricing errors are autocorrelated, heteroskedastic, and non-normal. This GMM estimator also has the advantage to be closely related to the empirical equity asset pricing literature (see Campbell, Lo, McKinley (1997) and Cochrane (2005) for surveys).

Instead of starting with a particular specification of the pricing kernel, we start with a model for expected bond returns that are affine in state variables. We derive this functional form for returns from two alternative sets of assumptions about the pricing kernel. Our pricing equations can be derived from an affine pricing kernel which does not require particular distributional assumptions about shocks. Alternatively, the pricing equations can be derived by assuming an exponentially affine pricing kernel, conditionally normal shocks to the state variables, and conditionally normal bond returns. These assumptions are not mutually exclusive, but they do have implications for the distribution of yield pricing errors. Finally, we also develop the regression approach to term structure estimation in the case of the usual affine set-up with conditionally Gaussian shocks and an exponentially affine pricing kernel.

Our paper is organized as follows. In Section 2, we discuss our model and the three step estimator. In Section 3, we present our main empirical findings, showing that the no-arbitrage constraint is rejected for the three factor model, but not the five factor model. In Section 4, we present specification tests and extensions. Section 5 concludes.

2 The Model

2.1 State variables and expected returns

We assume that the dynamics of a $K \times 1$ vector of state variables X_t evolve according to the following vector autoregressive process:

$$X_{t+1} = \mu + \Phi X_t + v_{t+1} \tag{1}$$

This specification of the dynamic evolution of the state variables can be interpreted as a discrete time analog to the state variable dynamics of Merton's (1973) ICAPM or Cox, Ingersoll, and Ross' (1985) general equilibrium setup. We do not necessarily assume that shocks v_{t+1} are conditionally Gaussian, identical, or independent. We do assume that the expectation of v_{t+1} conditional on the history of X_t is zero, and that the variance of v_{t+1} conditional on the history of X_t exists:

$$E_t\left(v_{t+1}|\left\{X_s\right\}_{s=0}^t\right) = E_t\left(v_{t+1}|X_t\right) = 0 \tag{2}$$

$$Var_t\left(v_{t+1}|\left\{X_s\right\}_{s=0}^t\right) = Var_t\left(v_{t+1}|X_t\right) = \Sigma_t \tag{3}$$

where $\{X_s\}_{s=0}^t$ denotes the history of X_t. Note that innovations to state variables can be expressed as $v_{t+1} = S_t \epsilon_{t+1}$ where $Var_t\left(\epsilon_{t+1} | \{X_s\}_{s=0}^t\right) = I$ and $\Sigma_t = S_t S_t'$. We denote $R_{t+1}^{(n-1)}$ the holding period return of a bond maturing in n periods, and R_t^F the risk free rate. All conditional expectations are taken relative to the history of $\{X_s\}_{s=0}^t$.

We start with the assumption that expected excess returns $E_t\left[R_{t+1}^{(n-1)}\right]$ depend linearly on prices of risk, and that prices of risk are affine functions of the state variables $(\lambda_0 + \lambda_1 X_t)$:

$$E_t\left[R_{t+1}^{(n-1)}\right] - R_t^F = Cov_t\left[R_{t+1}^{(n-1)}, X_{t+1}\right]\Sigma_t^{-1}(\lambda_0 + \lambda_1 X_t) = \beta_t^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t) \tag{4}$$

where $\beta_t^{(n-1)\prime} = Cov_t\left[R_{t+1}^{(n-1)}, X_{t+1}\right]\Sigma_t^{-1}$. In Sections A.1 and A.2 of the Appendix, we derive this return equation from alternative assumptions about the functional form of the pricing kernel and the distribution of shocks. Equation 4 is a common prediction of asset pricing models: expected returns depend on the product of exposures to risk factors, multiplied by the prices of risk of the factors.

We can then decompose excess returns into an expected and an unexpected component:

$$R_{t+1}^{(n-1)} - R_t^F = \underbrace{\beta_t^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t)}_{\substack{\text{Expected} \\ \text{Return}}} + \underbrace{R_{t+1}^{(n-1)} - E_t\left[R_{t+1}^{(n-1)}\right]}_{\substack{\text{Unexpected} \\ \text{Return}}} \tag{5}$$

The unexpected return $R_{t+1}^{(n-1)} - E_t\left[R_{t+1}^{(n-1)}\right]$ can be further decomposed into a component that is correlated with the innovations of the states, $v_{t+1} = X_{t+1} - E_t[X_{t+1}]$, and a return pricing error $e_{t+1}^{(n-1)}$ that is orthogonal to the state innovations:

$$R_{t+1}^{(n-1)} - E_t\left[R_{t+1}^{(n-1)}\right] = \gamma_t^{(n-1)\prime}(X_{t+1} - E_t[X_{t+1}]) + e_{t+1}^{(n-1)} \tag{6}$$

for some $\gamma_t^{(n-1)\prime}$. Note that, by construction, $E\left[e_{t+1}^{(n-1)} | \{X_s\}_{s=0}^{t}, v_{t+1}\right] = 0$. By replacing (6) into (4), it follows that:

$$\begin{aligned}
\beta_t^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t) &= Cov_t\left(\gamma_t^{(n-1)\prime} v_{t+1} + e_{t+1}^{(n-1)}, X_{t+1}'\right)\Sigma_t^{-1}(\lambda_0 + \lambda_1 X_t) \\
&= Cov_t\left(\gamma_t^{(n-1)\prime} v_{t+1}, v_{t+1}'\right)\Sigma_t^{-1}(\lambda_0 + \lambda_1 X_t) \\
&= \gamma_t^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t)
\end{aligned}$$

so that $\gamma_t^{(n-1)\prime} = \beta_t^{(n-1)\prime}$. It then follows that excess returns are an affine function of lagged state variables X_t, state variable innovations v_{t+1}, and return pricing errors $e_{t+1}^{(n-1)}$:

$$R_{t+1}^{(n-1)} - R_t^F = \underbrace{\beta_t^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t)}_{\substack{\text{Expected} \\ \text{Return}}} + \underbrace{\beta_t^{(n-1)\prime} v_{t+1}}_{\substack{\text{Priced Return} \\ \text{Innovation}}} + \underbrace{e_{t+1}^{(n-1)}}_{\substack{\text{Return} \\ \text{Error}}} \tag{7}$$

The excess return thus depends on the expected return, a component that is correlated with the innovations of the states, $v_{t+1} = X_{t+1} - E_t[X_{t+1}]$, and a return pricing error $e_{t+1}^{(n-1)}$ that is orthogonal to the state innovations. Therefore, the innovations to the state variables are cross-sectional pricing factors, and the levels of the states are forecasting variables. Per construction, the return pricing error $e_{t+1}^{(n-1)}$ is orthogonal to current and lagged state variables.

2.2 Estimation

Based on equation (7) above, we propose the following three stage estimator for the parameters of the term structure model. For simplicity, we assume that $\beta_t^{(n)} = \beta^{(n)}\ \forall t$. Time-varying coefficients can be obtained by estimating rolling-window regressions, or by estimating $\beta_t^{(n)}$ using an appropriate filter.

1. Estimate μ and Φ via OLS. This allows the decomposition of X_{t+1} into a predictable component $\hat{\mu} + \hat{\Phi} X_t$, and an innovation v_{t+1}.

2. Regress excess returns on lagged pricing factors and contemporaneous pricing factor innovations,

$$R_{t+1}^{(n-1)} - R_t^F = a^{(n-1)} + \beta^{(n-1)\prime} \hat{v}_{t+1} + c^{(n-1)\prime} X_t + e_{t+1}^{(n-1)} \quad (8)$$

   and recover the coefficients $\hat{a}^{(n-1)}, \hat{\beta}^{(n-1)}$, and $\hat{c}^{(n-1)}$. Stacking them across maturities gives $\hat{\mathbf{a}} = (\hat{a}^{(1)}, ..., \hat{a}^{(N)})$, $\hat{\beta} = \left(\hat{\beta}^{(1)\prime}, ..., \hat{\beta}^{(N)\prime}\right)'$, and $\hat{\mathbf{c}} = \left(\hat{c}^{(1)\prime}, ..., \hat{c}^{(N)\prime}\right)'$, where $\hat{\mathbf{a}}$ is a vector of length N, and where $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$ are matrices of dimension $N \times K$, respectively.

3. Estimate the prices of risk parameters λ_0 and λ_1 via a cross sectional regression. We know from (7) that $a^{(n)} = \beta^{(n)\prime}\lambda_0$ and $c^{(n)\prime} = \beta^{(n)\prime}\lambda_1$ $\forall n$. We therefore have $\hat{\mathbf{a}} = \hat{\beta}\lambda_0$ and $\hat{\mathbf{c}} = \hat{\beta}\lambda_1$. Hence, we can obtain estimates for λ_0 and λ_1 via

$$\hat{\lambda}_0 = \left(\hat{\beta}'\hat{\beta}\right)^{-1} \hat{\beta}'\hat{\mathbf{a}} \quad (9)$$

$$\hat{\lambda}_1 = \left(\hat{\beta}'\hat{\beta}\right)^{-1} \hat{\beta}'\hat{\mathbf{c}} \quad (10)$$

   Note that when $N = K$, these estimators amount to setting $\hat{\lambda}_0 = \hat{\beta}^{-1}\hat{\mathbf{a}}$ and $\hat{\lambda}_1 = \hat{\beta}^{-1}\hat{\mathbf{c}}$.

2.3 Relation to Fama-MacBeth regressions

The last step of our three stage OLS estimation amounts to cross-sectional regressions of $\hat{\mathbf{a}}$ on $\hat{\beta}$ and $\hat{\mathbf{c}}$ on $\hat{\beta}$. Our approach thus bears a close relationship to the cross-sectional regressions of Fama and MacBeth (1973). In contrast to Fama-MacBeth regressions, we assume a particular affine form for prices of risk. Alternatively, we could follow Fama-MacBeth (1973) by regressing returns R_{t+1} on $\hat{\beta}'$ in the cross-section for each t to recover $\hat{\lambda}_{t+1}^{FM}$. We could then run a forecasting regression of those Fama-MacBeth prices of risk $\hat{\lambda}_{t+1}^{FM}$ on a constant and X_t to obtain $\hat{\lambda}_0^{FM}$ and $\hat{\lambda}_1^{FM}$. If return pricing errors $e_{t+1}^{(n)}$ are cross-sectionally uncorrelated, and there are no errors in variables, $\hat{\lambda}_0 = \hat{\lambda}_0^{FM}$, $\hat{\lambda}_1 = \hat{\lambda}_1^{FM}$, and $\hat{\lambda}_{t+1}^{FM} = \left(\hat{\lambda}_0 + \hat{\lambda}_1 X_t + \hat{v}_{t+1}\right)$. Our approach can thus be interpreted as a dynamic version of Fama-MacBeth regressions that decomposes prices of risk into a predictable component $(\lambda_0 + \lambda_1 X_t)$ and an innovation v_{t+1}. The dynamics of prices of risk are in turn linked to the dynamics of the state variables.

2.4 Standard errors

Our three-stage estimation approach uses generated regressors whose estimation uncertainty has to be taken into account in computing standard errors. We use bootstrapping methods to adjust standard errors for the two layers of generated regressor uncertainty by using the following procedure. We save the residuals $\{\hat{v}_t\}$ from the VAR(1) regression of the pricing factors on their lagged levels, and the residuals $\{\hat{e}_t\}$ from the regression of excess returns on the lagged pricing factors and their innovations. Using the stationary block bootstrap of Politis and Romano (1994a,b) which is robust to serial correlation, we then generate 1000 artificial samples of the state vector and the cross-section of returns. In each bootstrap iteration, we simulate the state equation (1) and generate a set of artificial excess returns via (8). For each of the 1000 generated samples of excess returns and factors, we then estimate the market price of risk parameters using our three-stage regression approach and report the standard deviations from the generated sample of estimates as standard errors. Standard errors for the three stage regression approach could alternatively be computed analytically (see, for example, Adrian and Rosenberg, 2008).

The three stage estimation approach also introduces a potential for errors in variables. For example, $\hat{v}_{t+1}$ is estimated in the first stage, and can potentially lead to a distortion in the estimation of $\hat{\mathbf{a}}$, $\hat{\beta}$, and $\hat{\mathbf{c}}$ in the second stage. The errors in variables problem can be addressed by estimating the three regressions jointly in one stage. We present estimates of a one stage GMM based estimator in section 4.2.

2.5 Affine yields

We fit our model to returns of nondefaultable zero coupon bonds. In this section, we can derive additional cross-sectional constraints for the term structure of log bond prices. We obtain an expression for log bond prices via a first-order approximation of holding period excess returns by log excess returns:

$$\begin{aligned} R_{t+1}^{(n-1)} - R_t^F &= \left(\frac{P_{t+1}^{(n-1)} - P_t^{(n)}}{P_t^{(n)}} + 1 \right) - \left(\frac{1 - P_t^{(1)}}{P_t^{(1)}} + 1 \right) \\ &= \ln P_{t+1}^{(n-1)} - \ln P_t^{(n)} + \ln P_t^{(1)} + \omega_{t+1}^{(n-1)} \end{aligned} \tag{11}$$

where $\omega_{t+1}^{(n-1)}$ are the higher order terms of the first-order approximation, $P_t^{(1)}$ denotes the price of a zero coupon bond that matures in one period, and $P_{t+1}^{(n-1)}$ denotes the time t price of a zero coupon bond with maturity $t+n$.

We assume that log bond prices are affine in the states X and an error term $u_t^{(n)}$:

$$\ln P_t^{(n)} = A_n + B_n' X_t + u_t^{(n)} \tag{12}$$

Instead of making assumptions about the error term $u_t^{(n)}$, we can derive its properties as a function of the return pricing error $e_t^{(n)}$ and the return approximation error $\omega_t^{(n)}$. By replacing (12) into (11), we find:

$$R_{t+1}^{(n-1)} - R_t^F = A_{n-1} + B'_{n-1}X_{t+1} + u_{t+1}^{(n-1)} - A_n - B'_n X_t - u_t^{(n)} + A_1 + B'_1 X_t + u_t^{(1)} + \omega_{t+1}^{(n-1)} \quad (13)$$

This expression for returns has to equal the return expression (7) that we derived from no arbitrage, so that we find:

$$\begin{aligned} \beta^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t + v_{t+1}) + e_{t+1}^{(n-1)} &= A_{n-1} + B'_{n-1}(\mu + \Phi X_t + v_{t+1}) + u_{t+1}^{(n-1)} \\ &\quad - A_n - B'_n X_t - u_t^{(n)} + A_1 + B'_1 X_t + u_t^{(1)} + \omega_{t+1}^{(n-1)} \end{aligned} \quad (14)$$

This equation has to hold state by state. Let $A_1 = -\delta_0$ and $B_1 = -\delta_1$ and matching terms, we obtain a system of recursive linear restrictions for the bond pricing parameters A and B:

$$A_n = A_{n-1} + B'_{n-1}(\mu - \lambda_0) - \delta_0 + \bar{\omega}^{(n-1)} \quad (15)$$

$$B'_n = B'_{n-1}(\Phi - \lambda_1) - \delta'_1 \quad (16)$$

$$A_0 = 0, \; B'_0 = 0$$

$$\beta^{(n)\prime} = B'_n \quad (17)$$

where $\bar{\omega}^{(n-1)}$ denotes the average approximation error, and $\tilde{\omega}_{t+1}^{(n-1)}$ the demeaned approximation error:

$$\omega_{t+1}^{(n-1)} = \bar{\omega}^{(n-1)} + \tilde{\omega}_{t+1}^{(n-1)} \quad (18)$$

Then we obtain the following expression for the log bond pricing errors:

$$\underbrace{u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)}}_{\substack{\text{Log Bond} \\ \text{Pricing Error}}} = \underbrace{e_{t+1}^{(n-1)}}_{\substack{\text{Return Pricing} \\ \text{Error}}} + \underbrace{\tilde{\omega}_{t+1}^{(n-1)}}_{\substack{\text{Approximation} \\ \text{Error}}} \tag{19}$$

Two remarks are in order regarding the above derivation. First note that we could perform a second-order approximation of excess returns around zero. This would provide us with cross-equation restrictions similar to traditional affine models which involve second-order terms of log bond prices. Accordingly, the second-order approximation would require estimation of additional parameters related to conditional second moments of the state variables. Our aim is to develop a bond pricing model from minimal assumptions. We therefore restrict ourselves to reporting results based on the linear cross-equation restrictions derived above. We show in Section 3 below that these provide a very good fit to the yield curve. We derive the yield pricing equations with convexity adjustment in the context of our model in section 4.3. In equation (15), we effectively introduce a convexity adjustment by adding the average return approximation error $\bar{\omega}^{(n-1)}$ to the log bond recursion. If bond returns are assumed to be conditionally normal, this average approximation error would correspond to a Jensen's inequality term.

Second, notice that expression (19) above implies that long bond pricing errors are cross-sectionally and serially correlated. Indeed, even in the absence of an approximation error, we would have that $u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)} = e_{t+1}^{(n-1)}$. By way of definition of log excess returns, the same is true in standard affine models (see Section A of the Appendix). However, estimation of these models using likelihood methods assumes independence of the yield pricing errors, an assumption that appears to be violated if return pricing errors have no autocorrelation.

3 Empirical results

In this section, we provide estimation results from our regression approach for a selection of approximate affine models with varying factor specifications. As baseline examples, we choose a $K = 3$ and a $K = 5$ factor specification where the pricing factors are computed as the first K principal components from a cross-section of yields for maturities $n = 1, \dots, 120$ months. We will show in Section 4.1 below that the good empirical performance of our approach does not hinge on this choice of pricing factors.

We start by providing in-sample estimation results for the different model specifications. We then turn to out-of-sample yield forecasts and document that the five factor specification of our model predicts yields well out-of-sample relative to benchmark models.

3.1 Data

We estimate our model based on the zero-coupon yield data constructed by Gurkaynak, Sack, and Wright (2006).[1] The construction of these data is based on the Nelson-Siegel-Svensson curve, the parameters of which are published along with the estimated zero-coupon curve. We use these parameters to back out the cross-section of yields for maturities $n = 1, \dots, 120$ months from which we then compute the principal components that we employ as pricing factors in our baseline specifications.

We estimate the models for the sample period 1987:01-2008:03 and use as inputs excess holding period returns for bonds of maturities $n = 12, 18, 24, \dots, 120$ months. This provides us with a cross-section of $N = 19$ maturities for which he have a total of $T = 255$ observations. Recall that the excess holding period return for a bond of maturity n is defined as $R^{(n)}_{t+1} - R^F_t = \frac{P^{(n)}_{t+1}}{P^{(n+1)}_t} - \frac{1}{P^{(1)}_t}$ where $P^{(n)}_t$ denotes the price in period t of a bond

[1] We thank the authors for making these data available for download on the website `http://www.federalreserve.gov/Pubs/feds/2006/200628/feds200628.xls`

which matures in $t+n$.

3.2 In sample estimation

We estimate the parameters $(\mu, \Phi, \lambda_0, \lambda_1)$ using our three step estimation approach. We further obtain estimates of the short rate parameters δ_0 and δ_1 by regressing the log one-month yield on the pricing factors. We then feed all parameters into the recursions (15) and (16) to obtain the pricing parameters under no-arbitrage and use the latter to compute the model-implied yields.

3.2.1 Three Factor Specification

Table 1 below reports the time series properties of the yield pricing errors implied by the three factor specification of our model. As these results show, the three factor model explains yields of intermediate maturities very well, implying maximum average pricing errors of the order of about 2 basis points across the entire maturity spectrum. Moreover, yield pricing errors are skewed for most maturities. Figure 1 shows a plot of the average yield curve as observed and fitted by the three factor model. As can be seen from this plot, the three factor specification provides a very good fit of the average yield curve. This finding is underscored by Figure 2, which shows the unconditional standard deviations of observed and model-implied yields. While these almost coincide for intermediate maturities, there is some notable difference at the short and long end of the curve. Figures 3 and 4 show the time series of observed and fitted yields and excess holding period returns for a set of different maturities. These plots provide additional evidence that the three factor specification of our approximate affine model gives a close fit to the term structure of interest rates.

Recall that we use the first three principal components of the yield data as the pricing factors in the three factor specification of our approximate affine model. According to their loadings on individual bond yields, these are commonly labeled "level", "slope", and "curvature". Our arbitrage-free term structure model allows us to derive the risk premia associated with shocks to these components. Figure 5 provides a plot of the yield loadings $-\frac{1}{n}B_n$ implied by our model. Shocks to the first principal component affect yield of all maturities by about the same amount, so it can clearly be interpreted as a level factor. Moreover, the yield loadings of the second and third factor show that these can readily be viewed as representing slope and curvature. Hence, the three principal components keep their interpretation when used as state variables in our no-arbitrage model.

While the loadings of yields of different maturities on the level, slope, and curvature factors have often been documented in the literature, relatively little attention has previously been given to their impact on expected excess returns. In our arbitrage-free model, this is straightforward to analyze. Figure 6 shows the loadings $B_n'\lambda_1$ of expected returns on the three factors. The slope factor represents the most important driving force behind variations in risk premia, its impact on expected excess returns linearly decreasing with maturity. The curvature factor features a hump-shaped impact on expected returns and exerts the strongest influence on expected returns at maturities between three and four years. Finally, note that the yield curve level affects expected excess returns at the long end of the curve more strongly than at the short end.

3.2.2 Five Factor Specification

We have seen in the previous section that a three factor specification of our pricing model, although implying small pricing errors, still leaves some of the variation of interest rates at the very short end of the yield curve unexplained. We now show that a five factor specification of our model fits the yield curve about equally well in-sample. Since traditional term structure models are estimated imposing non-linear cross-equation restrictions, estimation of these models with more than three factors becomes computationally very demanding. In contrast, our estimation approach is based on simple linear regressions, and it therefore comes at no cost to add pricing factors to the model.

Table 3 reports the time series properties of the yield pricing errors implied by the five factor specification of our model. The average yield pricing errors are very small, ranging between 1.3 and 2.7 basis points. Moreover, the standard deviation of the pricing errors is very small, not exceeding 1.8 basis points for any of the maturities. However, consistent with our decomposition of yield pricing errors, we still find some evidence for serial correlation. Figures 7 to 10 visualize these results. As can be seen from these plots, the five factor specification provides a very good fit to the yield curve across the entire maturity spectrum. Taking into account that the pricing factors are the principal components of yields, this may not seem surprising at first sight. It is important to keep in mind, though, that the model has been fitted to returns and that all parameters have been obtained via linear regressions without imposing the cross-equation restrictions. These have only been used ex-post to back out the model-implied yield curve.

While the first three principal components of yields have a common interpretation as level, slope, and curvature, higher order principal components of yields have traditionally not been given much attention in the term structure literature as they only explain minimal shares of the cross-sectional variation of yields. In Figure 11, we plot the yield loadings implied by our five factor approximate affine model. According to this plot and consistent with the in-sample pricing errors, the fourth and fifth principal component carry a very small amount of explanatory power at the very short end of the curve, but explain little to no variation in longer maturities. Turning now to the effects of these two factors on expected excess returns shown in Figure 12, this picture changes dramatically.

Indeed, shocks to the fourth and fifth principal components are the main driving forces behind movements in risk premia. While the expected return loadings for level, slope, and curvature almost do not change, the fourth principal component has a strong positive hump-shaped effect on expected one-month excess returns. In contrast, the fifth principal component exhibits strongly negative coefficients on expected excess returns. In spirit, these results are in line with the findings in Cochrane and Piazzesi (2005) who document that a factor which has a negligible contemporaneous effect on the yield curve predicts future excess returns.

3.3 Out-of sample forecasts

We have seen above that both a three factor and a five factor specification of our approximate affine term structure model fit the yield curve very well in-sample. While the additional factors in the do not noticeably enhance in-sample fit, they seem to be the key driving force behind model-implied expected excess returns. In this subsection, we thus assess whether the added pricing factors carry additional predictive power.

To this end, we perform the following recursive out-of-sample forecast exercise. We use the subperiod 1987:01-2002:12 as our training sample and recursively re-estimate the model for all months from 2003:01 until 2008:03 using data from 1987:01 to the month when the forecast is made. That is, we extract the first K principal components from the yields up to date t, use these as pricing factors, estimate the model parameters with our three-stage regression approach, and then predict yields according to

$$y^{(n)}_{t+h|t} = \hat{a}_n + \hat{b}_n \hat{X}_{t+h|t}$$

where $\hat{a}_n$ and $\hat{b}_n$ have been back out from (15) and (16) based on estimates of the model parameters obtained using data up to date t, and where $\hat{X}_{t+h|t}$ is the h-step ahead conditional forecast of the state variables implied by their VAR(1) representation.

Table 5 reports the root mean squared forecast errors for the yield predictions implied by the three and five factor specifications of our model at forecast horizons 1-month, 6-months, and 12-months ahead. We further provide results obtained from the five factor specification augmented by an estimate of the Cochrane-Piazzesi return forecasting factor. All RMSEs are stated relative to those implied by a simple random walk model for yield levels. As a means of comparison, we also compute yield forecasts based on the dynamic extension of the Nelson-Siegel (1987) model that has been suggested by Diebold and Li (2006). Nelson and Siegel's model is not arbitrage-free, but simply decomposes variations in the cross-section of yields at every point in time into three factors with pre-specified loadings. Diebold and Li (2006) give a dynamic interpretation to these time-varying fitted parameters, estimate their dynamics with a simple VAR, and show that this generates good out-of-sample yield forecasts.

Let's first compare the out-of-sample predictive ability at the one-month ahead horizon. The Nelson-Siegel model is outperformed by the random walk as indicated by relative RMSEs that are larger than one. The same is true for the three factor specification of our approximate affine model for all maturities but the 12-month yield. In contrast, the five factor specification outperforms all other models including the random walk for maturities up to five years. The improvement relative to the RW forecast appears to be relatively small, though.

However, the improvement is considerably larger for forecasts six and twelve months ahead. Indeed, we find that our five factor specification achieves a reduction in RMSEs with respect to the random walk of the order of 20% for maturities up to five years. This is comparable in magnitude to the reduction in yield forecast errors that can be achieved using principal components extracted from a large cross-section of macroeconomic variables as pricing factors as in Moench (2008). Our model also clearly outperforms the Nelson-Siegel model. We interpret these results as evidence that factors which may have little explanatory power in the cross-section, might well be useful for predicting future yields and hence represent important state variables. This is again in line with the results in Cochrane and Piazzesi (2005) who find that a linear combination of forward rates which only explains a small share of the cross-sectional variation of yields is a strong predictor for one-year excess holding period returns. As a rough check of whether the fourth and fifth principal components span the information in the CP-factor, we also compare the results of our five factor specification to one that is augmented by an estimate of the CP-factor from our data sample.[2] These results show that the CP-factor only marginally reduces forecasts errors, and in some cases even slightly deteriorates yield predictions at the long end of the curve. We therefore conclude that the CP-factor does not carry important predictive information beyond the first five principal components of yields in our sample.

4 Specification Analysis and Extensions

4.1 Alternative factors

The empirical results of our model documented in the previous sections have been based on principal components of yields as pricing factors. This choice has merely been one of convenience. Indeed, our approach produces very similar in-sample and out-of-results using a set of forward spreads, individual yields or the time-series of Nelson-Siegel param-

[2] We construct the CP factor by regressing log one-year excess holding returns onto the one-year yield, as well as the four and seven year forward rates. This gives a similar tent-shaped pattern of regression coefficient as in Cochrane and Piazzesi. We use the linear dependence to construct the return-forecasting factor.

eters as pricing factors, respectively.[3]

4.2 Generalized Method of Moments

The three step regression approach introduces potential errors in variables, and does not necessarily provide efficient parameter estimates. Instead of running three stages of OLS regressions, we can estimate the model via two stage or iterative GMM to obtain efficient estimates, following Hansen (1982). Consider:

$$g(\theta) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} (X_{t+1} - \mu - \Phi X_t) \\ vec\left((X_{t+1} - \mu - \Phi X_t) X_t'\right) \\ \left(R_{t+1} - R_t^F - (b'(\lambda_0 - \mu + (\lambda_1 - \Phi) X_t + X_{t+1}))\right) \\ vec\left(\left(R_{t+1} - R_t^F - (b'(\lambda_0 - \mu + (\lambda_1 - \Phi) X_t + X_{t+1}))\right) X_t'\right) \\ vec\left(\left(R_{t+1} - R_t^F - (b'(\lambda_0 - \mu + (\lambda_1 - \Phi) X_t + X_{t+1}))\right) X_{t+1}'\right) \end{bmatrix} \tag{20}$$

where $\theta = \{\mu, \Phi, b', \lambda_0, \lambda_1\}$. Then minimizing

$$\theta = \arg\min g'(\theta)\, W g(\theta)$$

where W is the identity matrix in the first stage, and the inverse of the Newey-West (1987) estimate of the variance-covariance matrix of $g(\theta)$. The second stage provides an efficient estimator of θ. Note that the first stage parameter estimates are numerically identical to our three stage OLS estimates, while the second stage estimates are generally different.

[3]These results are available from the authors upon request.

4.3 Convexity adjustment

In section 2.5, we use a first order approximation to yields in order to generate the cross-sectional no-arbitrage recursions. Alternatively, one can make the following convexity adjustment:

$$\begin{aligned} R_{t+1}^{(n-1)} - R_t^F &= \left(\frac{P_{t+1}^{(n-1)} - P_t^{(n)}}{P_t^{(n)}} + 1\right) - \left(\frac{1 - P_t^{(1)}}{P_t^{(1)}} + 1\right) \\ &= \ln P_{t+1}^{(n-1)} - \ln P_t^{(n)} + \ln P_t^{(1)} - \frac{1}{2} Var_t \left(\ln P_{t+1}^{(n-1)}\right) + \tilde{\omega}_{t+1}^{(n-1)} \end{aligned} \tag{21}$$

where $\tilde{\omega}_{t+1}^{(n-1)}$ are the higher order terms relative to the second-order approximation. We assume that log bond prices are affine in the states X and an error term $\tilde{u}_t^{(n)}$:

$$\ln P_t^{(n)} = A_n + B_n' X_t + \tilde{u}_t^{(n)}. \tag{22}$$

Moreover, following the literature on affine models with CIR type state variable processes, we assume that the conditional variances and covariances Σ_t of the state variable innovations can be expressed as affine functions of the state variables themselves, i.e.

$$vec(\Sigma_t) = vec(S_t S_t') = S_0 + S_1 X_t \tag{23}$$

Then, instead of making assumptions about the error term $\tilde{u}_t^{(n)}$, we can derive its properties as a function of the return pricing error $e_t^{(n)}$ and the return approximation error $\tilde{\omega}_t^{(n)}$. We find:

$$\begin{aligned} R_{t+1}^{(n-1)} - R_t^F &= A_{n-1} + B_{n-1}' X_{t+1} + \tilde{u}_{t+1}^{(n-1)} - A_n - B_n' X_t - \tilde{u}_t^{(n)} + A_1 + B_1' X_t + \tilde{u}_t^{(1)} + \tilde{\omega}_{t+1}^{(n-1)} \\ &\quad - \frac{1}{2}\left(B_{n-1}' \otimes B_{n-1}'\right)(S_0 + S_1 X_t) - \frac{1}{2} Var_t \left(\tilde{u}_{t+1}^{(n-1)}\right) \end{aligned} \tag{24}$$

If we assume that the higher order terms of the approximation are uncorrelated with the state variables, this expression for returns has to equal the return expression that we derived from no arbitrage (7), so that we find:

$$
\begin{aligned}
& B'_{n-1}\left(\lambda_0+\lambda_1 X_t+v_{t+1}\right)+e_{t+1}^{(n-1)} \\
= \quad & A_{n-1}+B'_{n-1}X_{t+1}+\tilde{u}_{t+1}^{(n-1)}-A_n-B'_nX_t-\tilde{u}_t^{(n)} \\
& +A_1+B'_1X_t+\tilde{u}_t^{(1)}+\tilde{\omega}_{t+1}^{(n-1)}-\frac{1}{2}\left(B'_{n-1}\otimes B'_{n-1}\right)\left(S_0+S_1X_t\right)-\frac{1}{2}Var_t\left(\tilde{u}_{t+1}^{(n-1)}\right)
\end{aligned} \tag{25}
$$

This equation has to hold state by state. Letting $A_1=-\delta_0$ and $B_1=-\delta_1$ and matching terms, we obtain a system of recursive linear restrictions for the bond pricing parameters A and B :

$$A_n = A_{n-1}+B'_{n-1}\left(\mu-\lambda_0\right)-\frac{1}{2}\left(B'_{n-1}\otimes B'_{n-1}\right)S_0-\delta_0 \tag{26}$$

$$B'_n = B'_{n-1}\left(\Phi-\lambda_1\right)-\frac{1}{2}\left(B'_{n-1}\otimes B'_{n-1}\right)S_1-\delta_1 \tag{27}$$

$$e_{t+1}^{(n-1)} = \tilde{u}_{t+1}^{(n-1)}-\tilde{u}_t^{(n)}+\tilde{u}_t^{(1)}+\tilde{\omega}_{t+1}^{(n-1)}-\frac{1}{2}Var_t\left(\tilde{u}_{t+1}^{(n-1)}\right) \tag{28}$$

In comparison to equations (15) and (16), we thus have additional Jensen's inequality terms in this higher order approximation to the log yields. We show in section A.3 of the appendix that a traditional affine term structure model with an exponentially linear pricing kernel gives the same set of cross-equation restrictions as this second-order approximation of excess returns.

5 Conclusion

We outline an empirical approach to the estimation of a particular class of affine term structure models. Our approach is computationally fast, gives rise to small pricing errors, and very good out-of-sample forecasts compared to benchmark models. Our contribution is readily implantable for absolute pricing applications—such as in the macro-finance literature—and to other asset classes. Most importantly, our method is practical for real time analysis.

A Appendix: Pricing Kernel Assumptions

In sections A.1 and A.2 of this appendix, we derive equation (4) from two alternative assumptions about the pricing kernel and shocks. We then derive a regression based approach to traditional conditionally Gaussian term structure model with exponentially affine pricing kernel in section A.3.

A.1 Derivation of equation (4) from an affine pricing kernel

In this section of the appendix, we make no distributional assumptions about the state variable shocks v_{t+1}. We derive equation (7) by assuming that the pricing kernel $\frac{M_{t+1}}{M_t}$ is affine:

$$\frac{M_{t+1}}{M_t} = \frac{1-\lambda_t'\epsilon_{t+1}}{R_t^F} \quad \text{(A-1)}$$

$$S_t\lambda_t = \lambda_0 + \lambda_1 X_t \quad \text{(A-2)}$$

This price of risk formulation encompasses the price of risk specifications of the completely affine model of Dai and Singleton (2000) and the essentially affine model of Duffee (2002). The assumption of no-arbitrage implies (see Dybvig and Ross (1987)):

$$1 = E_t\left[\frac{M_{t+1}}{M_t}R_{t+1}^{(n)}\right] \quad \text{(A-3)}$$

$$R_t^F = 1/E_t\left[\frac{M_{t+1}}{M_t}\right] \quad \text{(A-4)}$$

Using the definition of covariance, the no-arbitrage condition (A-3) can be written as:

$$E_t\left[R_{t+1}^{(n)}\right] - R_t^F = -Cov_t\left[R_{t+1}^{(n)}, M_{t+1}/E_t\left[M_{t+1}\right]\right] \quad \text{(A-5)}$$

Using the specification of the pricing kernel (A-1) and replacing $\epsilon_{t+1} = S_t^{-1}\left(X_{t+1} - E_t\left[X_{t+1}\right]\right)$ and $\lambda_t = S_t^{-1}\left(\lambda_0 + \lambda_1 X_t\right)$, we find:

$$\begin{aligned} E_t\left[R_{t+1}^{(n)}\right] - R_t^F &= Cov_t\left[R_{t+1}^{(n)}, X_{t+1}\right]\Sigma_t^{-1}\left(\lambda_0 + \lambda_1 X_t\right) \\ &= \beta_t^{(n)\prime}\left(\lambda_0 + \lambda_1 X_t\right) \end{aligned} \quad \text{(A-6)}$$

where $\beta_t^{(n)\prime} = Cov_t\left[R_{t+1}^{(n)}, X_{t+1}\right]\Sigma_t^{-1}$ is the vector of conditional OLS betas of regressing returns on the vector of state variables X_{t+1}.

A.2 Derivation of equation (4) from an exponentially affine pricing kernel

In this section of the appendix, we make particular distributional assumptions about the state variable shocks v_{t+1} and the holding period returns $R_{t+1}^{(n)}$. To start, we assume that ϵ_{t+1} has a Gaussian distribution:

$$\epsilon_{t+1} \sim N\left(0, I_K\right) \quad \text{(A-7)}$$

The pricing kernel is exponentially affine, and prices of risk are affine:

$$\frac{M_{t+1}}{M_t} = \exp\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1}\right) \quad \text{(A-8)}$$

$$S_t\lambda_t = \lambda_0 + \lambda_1 X_t \quad \text{(A-9)}$$

where r_t is the continuously compounded risk free rate. Following the same logic as in appendix A.1, we find that the no-arbitrage condition $E_t\left[\frac{M_{t+1}}{M_t}R_{t+1}^{(n)}\right] = 1$ can be written as:

$$E_t\left[R_{t+1}^{(n)}\right] - R_t^F = -Cov_t\left[R_{t+1}^{(n)}, M_{t+1}/E_t\left[M_{t+1}\right]\right] \quad \text{(A-10)}$$

We assume that holding period returns $R_{t+1}^{(n)}$ are conditionally Gaussian. Because $M_{t+1}/E_t\left[M_{t+1}\right]$ is a continuously differentially function of Gaussian shocks ϵ_{t+1}, we can apply Stein's Lemma:

$$E_t\left[R_{t+1}^{(n)}\right]-R_t^F=Cov_t\left[R_{t+1}^{(n)},\lambda_t'\epsilon_{t+1}\right] \tag{A-11}$$

So we find equation (4):

$$E_t\left[R_{t+1}^{(n)}\right]-R_t^F=\beta_t^{(n)\prime}\left(\lambda_0+\lambda_1 X_t\right) \tag{A-12}$$

A.3 Regression approach to the standard exponentially-affine yield curve model with Gaussian shocks

In this section, we show that traditional affine term structure models with an exponentially linear pricing kernel and Gaussian shocks give rise to return regressions that are similar to the ones that we derive from our expected return beta representation.

As in our model, assume that there is a K-dimensional vector X of state variables that govern the evolution of yields. The joint dynamics of the state vector are described by the affine vector autoregressive process with conditional hetereoskedasticity

$$X_{t+1}=\mu+\Phi X_t+S_t\epsilon_{t+1}$$

where we assume that $Var_t[\epsilon_{t+1}|X_t]=I$ and that $vec(S_tS_t')=S_0+S_1X_t.$, i.e. conditional variances and covariances are linear functions of the states. This is the discrete-time analog of the Cox-Ingersoll-Ross (1985) specification of state-variable dynamics. Note that in the homoskedastic case $S_0=vec(\Sigma)$ and $S_1=0_{K^2\times 1}$.

Assume that the vector of shocks ϵ_{t+1} is conditionally normally distributed, and that the pricing kernel has the following functional form:

$$M_{t+1}/M_t = \exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1}) \tag{A-13}$$

where r_t denotes the continuously compounded short rate. The short rate is assumed to be given by the following affine function of the states X:

$$r_t = \delta_0 + \delta_1' X_t. \tag{A-14}$$

To obtain closed-form bond prices, one needs to define market prices of risk such that $S_t\lambda_t$ is affine in X_t. Following Duffee (2002), we let $\lambda_t = S_t^{-1}\lambda_0 + S_t^{-1}\lambda_1 X_t$ where λ_0 is a $K \times 1$ vector and λ_1 a $K \times K$ matrix. Note that to prevent market prices of risk to go to infinity when the conditional volatility S_t of the pricing factors approaches zero, additional restrictions need to be imposed on the matrix λ_1. For a discussion of these restrictions see Duffee (2002).

We follow the literature by assuming that log bond prices are an affine function of state variables:

$$\ln P_t^{(n)} = A_n + B_n' X_t + u_t^{(n)} \tag{A-15}$$

The absence of arbitrage-opportunities requires:

$$P_t^{(n)} = E_t\left[\frac{M_{t+1}}{M_t} P_{t+1}^{(n-1)}\right] \tag{A-16}$$

Then we find the following restrictions on parameters and the distribution of $u_t^{(n)}$:

$$A_n = A_{n-1} + B'_{n-1}(\mu - \lambda_0) + \frac{1}{2}(B'_{n-1} \otimes B'_{n-1})S_0 - \delta_0 \quad \text{(A-17)}$$

$$B'_n = B'_{n-1}(\Phi - \lambda_1) + \frac{1}{2}(B'_{n-1} \otimes B'_{n-1})S_1 - \delta'_1 \quad \text{(A-18)}$$

$$A_1 = -\delta_0 \text{ and } B'_1 = -\delta'_1 \quad \text{(A-19)}$$

$$\exp u_t^{(n)} = E_t\left[\exp u_{t+1}^{(n-1)}\right] \quad \text{(A-20)}$$

Note that $\exp u_t^{(n)} = E_t\left[\exp u_{t+1}^{(n-1)}\right]$ is a statement that the exponential of the bond pricing error u follows a martingale. Note that the log bond pricing error $u_t^{(n)}$ is usually not replaced in the no arbitrage recursion, as it is assumed to be an *i.i.d.* observation error. The problem with this assumption is that it leads to autocorrelated return errors.

If we assume that $u_t^{(n)}$ is Gaussian we find:

$$u_t^{(n)} = E_t\left[u_{t+1}^{(n-1)}\right] + u_t^{(1)} + \frac{1}{2}Var_t\left[u_{t+1}^{(n-1)}\right] \quad \text{(A-21)}$$

So under the additional assumption that $u_t^{(n)}$ is Gaussian, $u_t^{(n)}$ is a random walk with a drift that is proportional to half its conditional variance. Thus in the presence of pricing errors, the assumption that $u_t^{(n)}$ is *i.i.d.* is clearly not consistent with no arbitrage. If we further assume that $Var_t\left[u_{t+1}^{(n-1)}\right] = \sigma^2_{(n-1)}$, and $e_{t+1}^{(n-1)} \sim N\left(0, \sigma^2_{(n-1)}\right)$, the process for $u_t^{(n-1)}$ reduces to:

$$u_{t+1}^{(n-1)} = u_t^{(n)} - u_t^{(1)} - \frac{1}{2}\sigma^2_{(n-1)} + e_{t+1}^{(n-1)} \quad \text{(A-22)}$$

We define the one-period log excess return as

$$rx_{t+1}^{(n-1)} = \ln P_{t+1}^{(n-1)} - \ln P_t^{(n)} + \ln P_t^{(1)} \quad \text{(A-23)}$$

$$= (A_{n-1} - A_n + A_1) + \left(B'_{n-1}X_{t+1} + B'_1X_t - B'_nX_t\right) + \left(u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)}\right)$$

Then we find:

$$rx_{t+1}^{(n-1)} = \underbrace{a_{n-1} + c'_{n-1}X_t}_{\text{Expected Return}} + \underbrace{b'_{n-1}v_{t+1}}_{\text{Return Innovation}} + \underbrace{e_{t+1}^{(n-1)}}_{\text{Return Pricing Error}} \tag{A-24}$$

where

$$a_{n-1} = -\frac{1}{2}(B'_{n-1} \otimes B'_{n-1})S_0 - \frac{1}{2}\sigma^2_{(n-1)} + B'_{n-1}\lambda_0 \tag{A-25}$$

$$b'_{n-1} = B'_{n-1} \tag{A-26}$$

$$c'_{n-1} = -\frac{1}{2}(B'_{n-1} \otimes B'_{n-1})S_1 + B'_{n-1}\lambda_1 \tag{A-27}$$

$$e_{t+1}^{(n-1)} = u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)} + \frac{1}{2}\sigma^2_{(n-1)} \tag{A-28}$$

We can stack the regression coefficients for different maturities n to obtain

$$rx_{t+1} = a + CX_t + Bv_{t+1} + e_{t+1} \tag{A-29}$$

where a and σ^2 are vectors of size N, and where B and C are matrices of dimension $N \times K$, respectively. According to the relationship between a_n, b_n, and c_n above, we have:

$$a = B\lambda_0 - \frac{1}{2}B^*S_0 - \frac{1}{2}\sigma^2 \tag{A-30}$$

$$C = B\lambda_1 - \frac{1}{2}B^*S_1 \tag{A-31}$$

where B^* is a $N \times K^2$ matrix with $n-$th row given by $(B'_n \otimes B'_n)$. One can thus easily get B^* from B. Reorganizing terms, the above relationships show that

$$B'a = B'B\lambda_0 - \frac{1}{2}B'B^*S_0 - \frac{1}{2}B'\sigma^2 \tag{A-32}$$

$$B'C = B'B\lambda_1 - \frac{1}{2}B'B^*S_1 \tag{A-33}$$

and hence

$$\hat{\lambda}_0 = (B'B)^{-1} B'a + \frac{1}{2} (B'B)^{-1} B'B^* S_0 - \frac{1}{2} (B'B)^{-1} B'\sigma^2 \quad \text{(A-34)}$$

$$\hat{\lambda}_1 = (B'B)^{-1} B'C + \frac{1}{2} (B'B)^{-1} B'B^* S_1 \quad \text{(A-35)}$$

This implies that given the parameters S_0 and S_1, one can obtain estimates of the market price of risk parameters λ_0 and λ_1 from the coefficients of the regressions of one-period excess holding returns on the model states and their innovations.

B Tables and Figures

Table 1: **Three Factor Model: Yield Pricing Errors**

This table summarizes the time series properties of the yield pricing errors implied by the three factor specification of our approximate affine model. The sample period is 1987:01-2008:03. "mean", "std", "skew", and "kurt" refer to the sample mean, standard deviation, skewness, and kurtosis of the yield errors; $\rho(1), \rho(6)$, and $\rho(12)$ denote the autocorrelation coefficients of order one, six, and twelve, respectively.

	n = 12	n = 24	n = 36	n = 60	n = 84	n = 120
mean	0.023	0.018	0.007	0.004	0.013	0.003
std	0.062	0.025	0.017	0.028	0.020	0.042
skew	-0.043	0.682	-0.151	-0.883	-0.276	-0.346
kurt	5.953	3.910	5.520	3.668	2.929	2.963
$\rho(1)$	0.864	0.838	0.831	0.863	0.921	0.895
$\rho(6)$	0.631	0.306	0.592	0.371	0.564	0.552
$\rho(12)$	0.316	0.003	0.296	0.112	0.327	0.410

Table 2: **Three Factor Model: Market Prices of Risk**

This table summarizes the estimates of the market price of risk parameters λ_0 and λ_1 for the three factor specification of our Approximate Affine Model. t-statistics are reported in brackets. The standard errors have been computed according to the bootstrap procedure laid out in Section 4.1.

	λ_0		λ_1	
F1	0.039	-0.021	-0.015	-0.008
t-stat	(1.592)	(-1.206)	(-1.098)	(-0.927)
F2	-0.020	-0.003	0.006	0.013
t-stat	(-0.804)	(-0.129)	(0.447)	(0.960)
F3	0.013	-0.035	-0.023	0.022
t-stat	(0.227)	(-0.674)	(-0.559)	(0.569)

Table 3: **Five Factor Model: Yield Pricing Errors**

This table summarizes the time series properties of the yield pricing errors implied by the five factor specification of our approximate affine model. The sample period is 1987:01-2008:03. "mean", "std", "skew", and "kurt" refer to the sample mean, standard deviation, skewness, and kurtosis of the yield errors; $\rho(1), \rho(6)$, and $\rho(12)$ denote the autocorrelation coefficients of order one, six, and twelve, respectively.

	n = 12	n = 24	n = 36	n = 60	n = 84	n = 120
mean	0.019	0.014	0.013	0.016	0.018	0.027
std	0.018	0.006	0.004	0.005	0.010	0.018
skew	-0.042	1.178	-0.288	-0.292	-0.074	-0.889
kurt	3.525	4.456	3.328	2.144	2.783	3.596
$\rho(1)$	0.891	0.851	0.712	0.967	0.941	0.912
$\rho(6)$	0.481	0.470	0.362	0.843	0.705	0.662
$\rho(12)$	0.265	0.199	0.174	0.669	0.548	0.432

Table 4: **Five Factor Model: Market Prices of Risk**

This table summarizes the estimates of the market price of risk parameters λ_0 and λ_1 for the five factor specification of our Approximate Affine Model. t-statistics are reported in brackets. The standard errors have been computed according to the bootstrap procedure laid out in Section 4.1.

	λ_0	λ_1				
F1	0.032	-0.024	-0.017	-0.010	0.021	-0.027
t-stat	(1.428)	(-1.289)	(-1.303)	(-1.072)	(1.939)	(-2.389)
F2	-0.020	-0.001	0.002	0.008	0.000	0.008
t-stat	(-0.752)	(-0.033)	(0.126)	(0.532)	(0.001)	(0.561)
F3	0.019	-0.069	-0.010	0.045	-0.158	0.014
t-stat	(0.344)	(-1.385)	(-0.228)	(1.532)	(-3.978)	(0.352)
F4	0.005	0.099	-0.006	-0.021	-0.068	0.069
t-stat	(0.059)	(1.229)	(-0.072)	(-0.393)	(-1.209)	(1.230)
F5	0.045	0.144	-0.039	-0.026	-0.006	-0.208
t-stat	(0.440)	(1.333)	(-0.462)	(-0.361)	(-0.065)	(-2.849)

Table 5: **RMSEs from Approximate Affine Specifications and Nelson-Siegel Model Relative to RW: 200301-200803**

This table summarizes the root mean squared forecast errors relative to those implied by a random walk for yields. The models have been estimated using data from 1987:01 until the period when the forecast is made. The forecasting period is 2003:01-2008:03. "AA(3)" and "AA(5)" refer to the three factor and five factor specifications of our Approximate Affine Model, "AA(5)+CP" denotes the five factor specification augmented by the Cochrane-Piazzesi return forecasting factor, and "NS" denotes the Diebold-Li (2006) version of the three-factor Nelson-Siegel model with VAR factor dynamics.

	n = 12	n = 24	n = 36	n = 60	n = 84	n = 120
			1 months ahead			
$AA(3)$	0.938	1.008	1.010	1.008	1.015	1.038
$AA(5)$	0.890	0.949	0.982	1.003	1.013	1.030
$AA(5)+CP$	0.896	0.926	0.960	0.997	1.015	1.039
NS	1.069	1.050	1.034	1.051	1.039	1.043
			6 months ahead			
$AA(3)$	0.976	1.028	1.031	1.028	1.052	1.091
$AA(5)$	0.844	0.874	0.904	0.969	1.031	1.096
$AA(5)+CP$	0.842	0.858	0.889	0.956	1.023	1.093
NS	0.955	0.978	0.971	0.979	1.003	1.023
			12 months ahead			
$AA(3)$	1.019	1.051	1.051	1.058	1.114	1.190
$AA(5)$	0.786	0.772	0.774	0.832	0.945	1.096
$AA(5)+CP$	0.778	0.755	0.761	0.830	0.959	1.126
NS	0.871	0.884	0.876	0.909	1.004	1.108

Figure 1: **Three Factor Model: Average Observed and Model-Implied Yields**

This figure plots average observed yields against those implied by the Three Factor Approximate Affine Model.

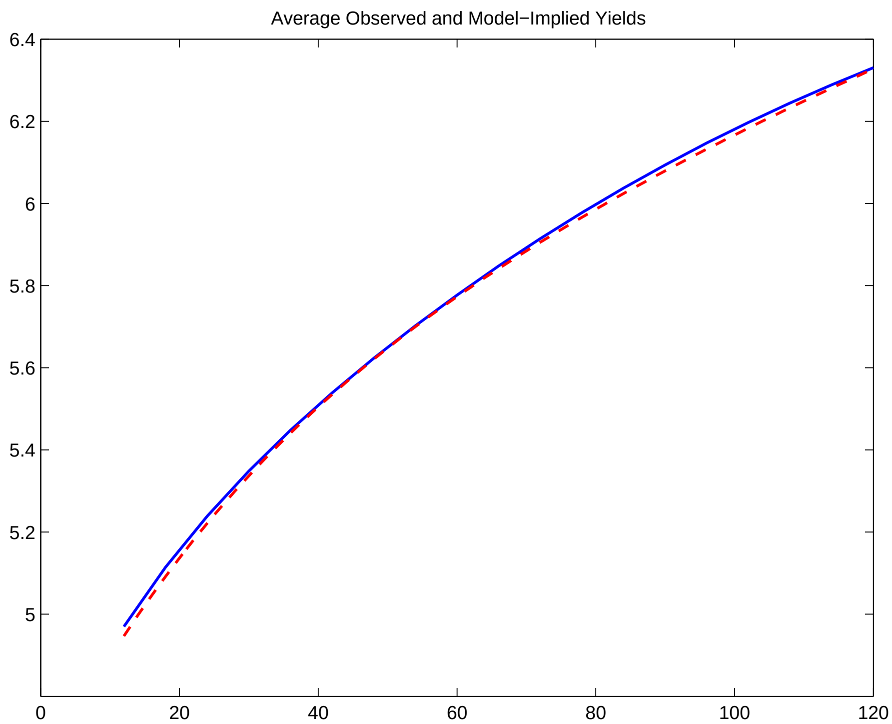


Figure 2: **Three Factor Model: Standard Deviation of Observed and Model-Implied Yields**

This figure plots the unconditional standard deviations of observed yields against those implied by the Three Factor Approximate Affine Model.

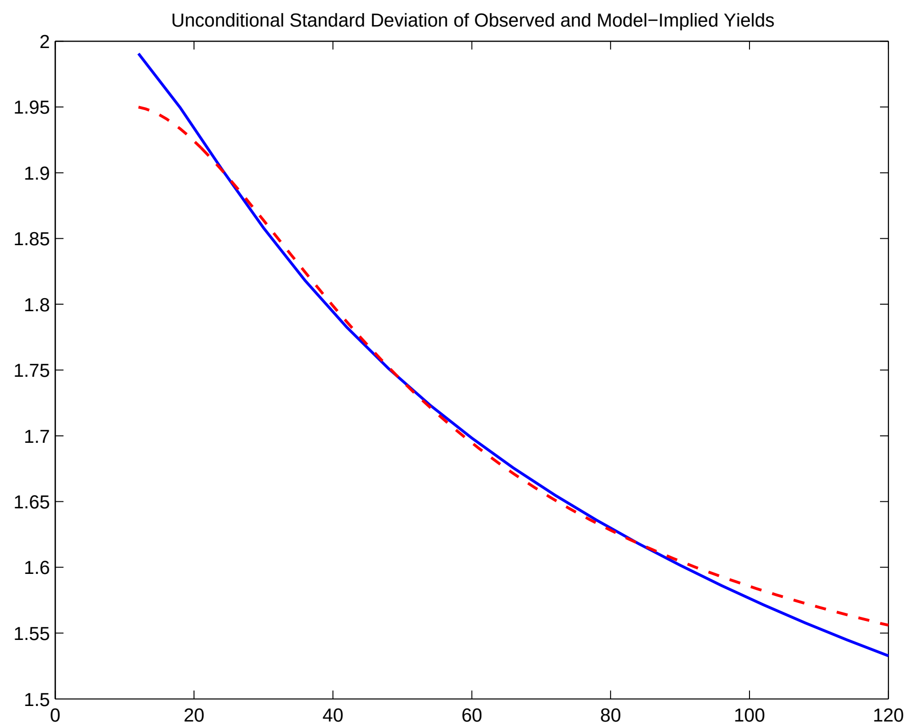


Figure 3: **Three Factor Model: Observed and Model-Implied Yields**

This figure provides plots of the observed and fitted yields for the 1-year, 2-year, 5- and 10-year maturities. The observed yields are plotted by solid lines, whereas dashed lines correspond to yields implied by the three factor approximate affine model.

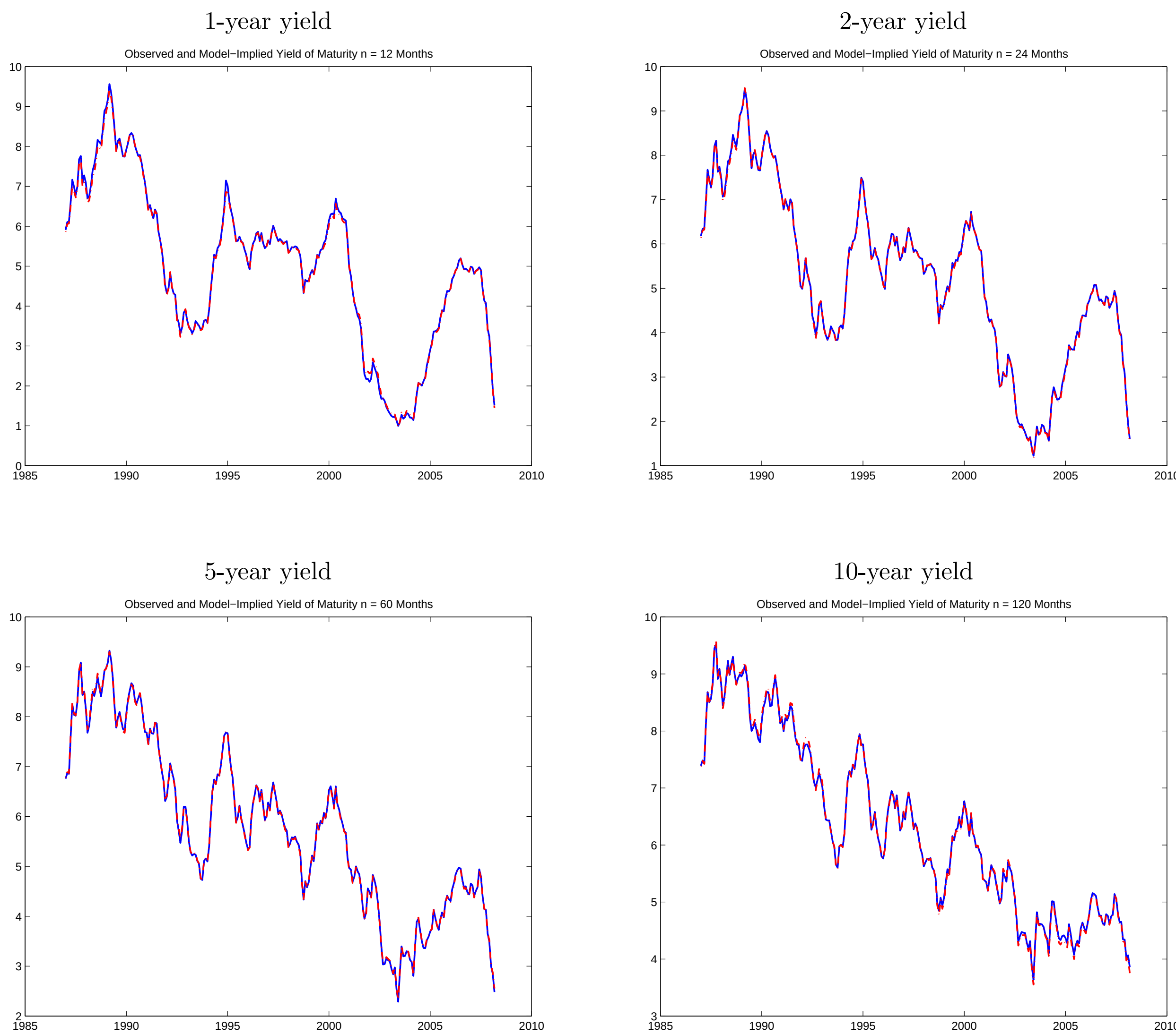
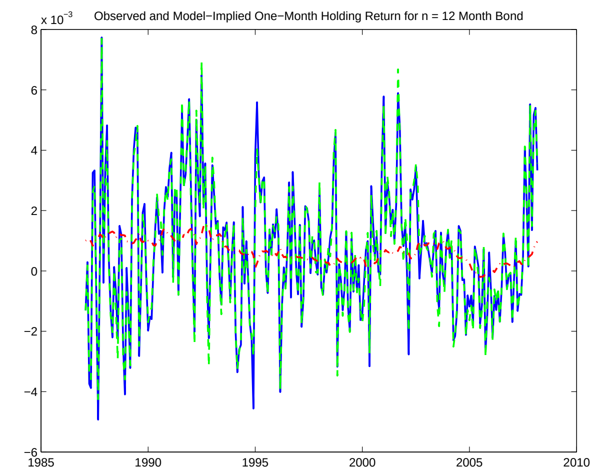


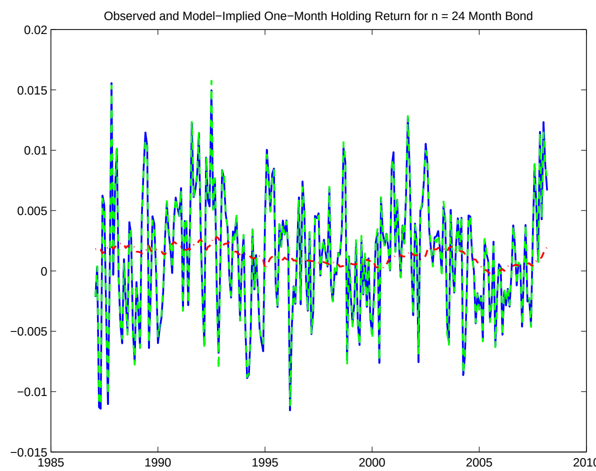
Figure 4: **Three Factor Model: Observed and Model-Implied Excess One-Month Holding Returns**

This figure provides plots of the observed and model-implied excess one-month holding returns for the 1-year, 2-year, 5- and 10-year maturities. The observed returns are plotted by solid lines, whereas dashed green lines correspond to actual model-implied returns, and dash-dotted red lines to model-implied expected one-month excess holding returns.

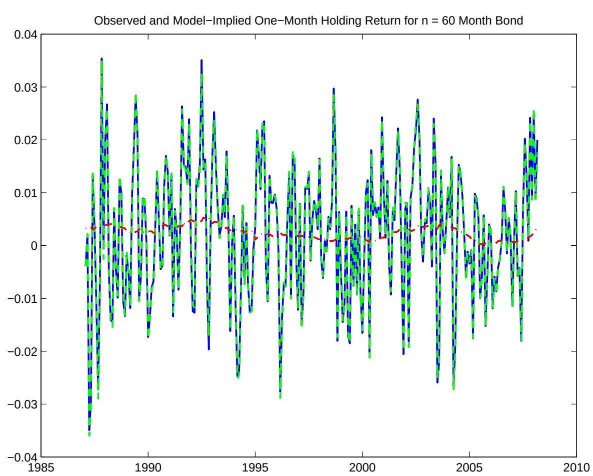
1-year bond



2-year bond



5-year bond



10-year bond

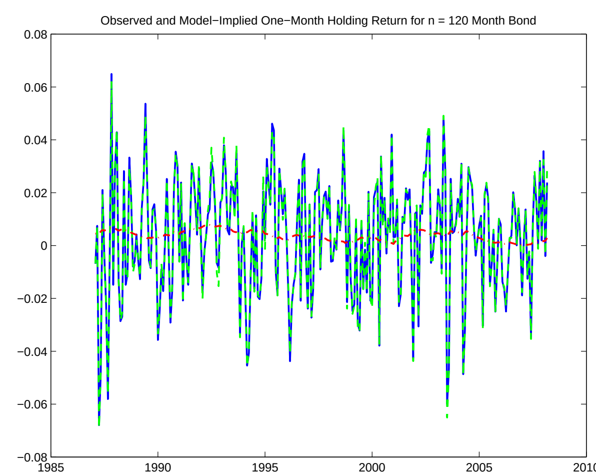


Figure 5: **Three Factor Model: Factor Loadings for Yields**

This figure provides a plot of the yield loadings $-\frac{1}{n}B_n$ implied by the Three Factor Approximate Affine Model. The coefficients can be interpreted as the response of the n-month yield to a contemporary shock to the respective factor.

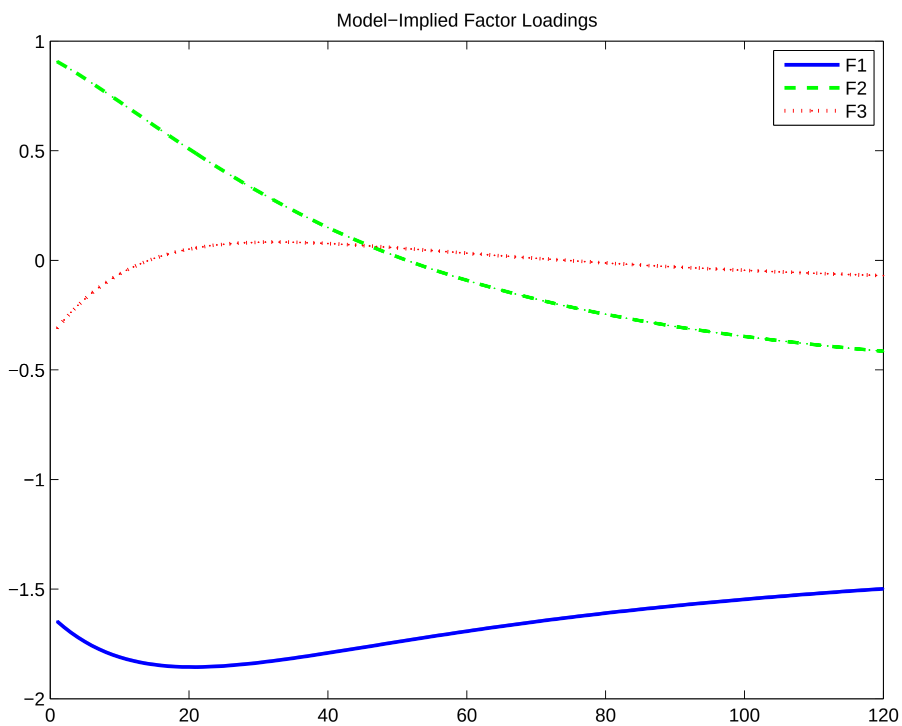


Figure 6: **Three Factor Model: Factor Loadings for Expected Returns**

This figure provides a plot of the expected return loadings $B_n'\lambda_1$ implied by the Three Factor Approximate Affine Model. The coefficients can be interpreted as the response of the expected one-month excess holing return on an n-month bond to a contemporary shock to the respective factor.

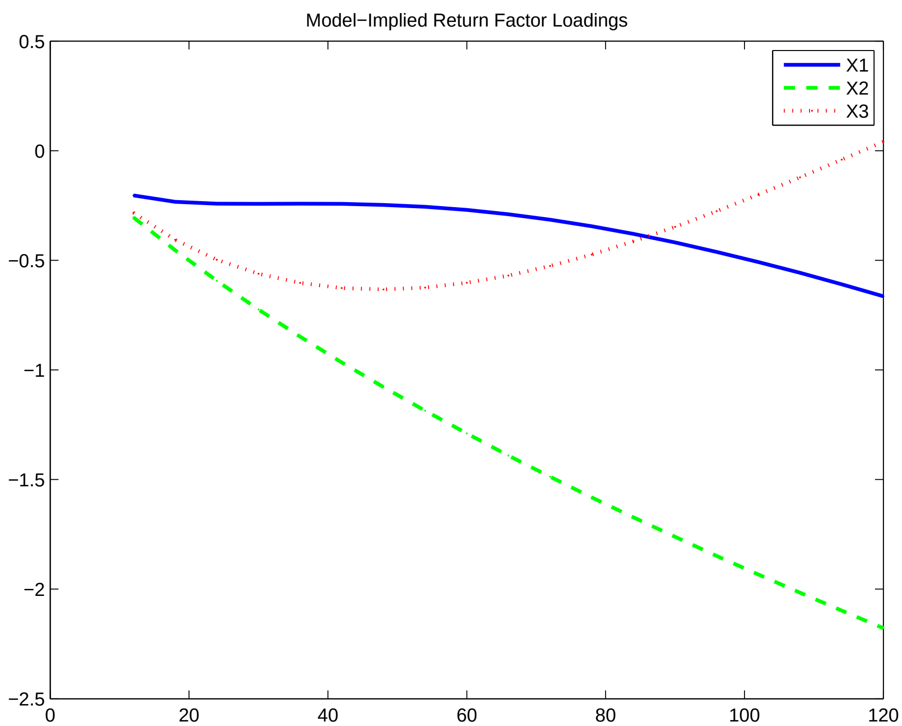


Figure 7: **Five Factor Model: Average Observed and Model-Implied Yields**

This figure plots average observed yields against those implied by the Five Factor Approximate Affine Model.

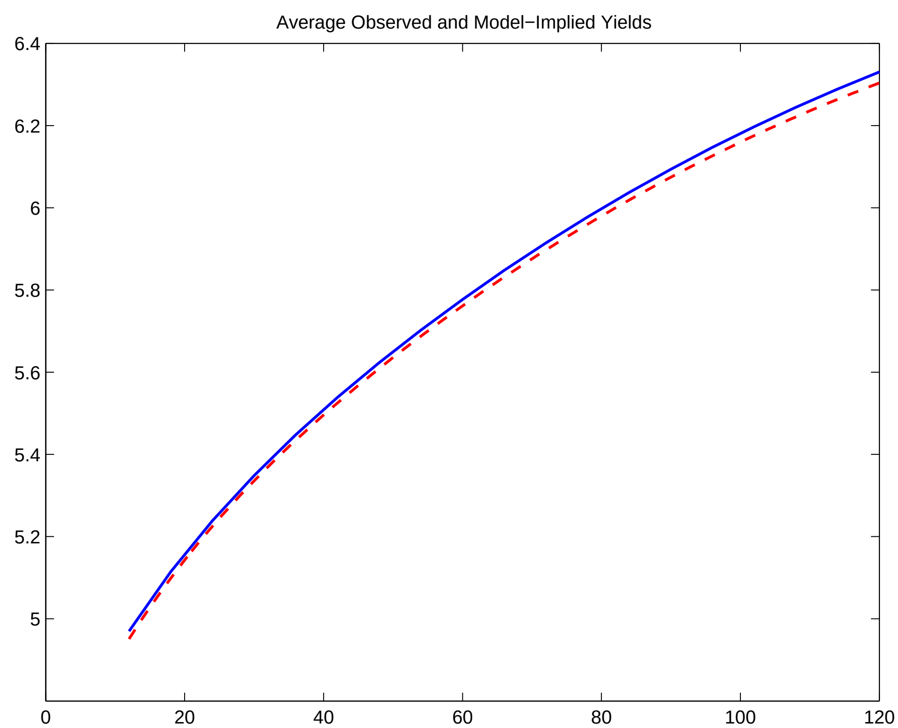


Figure 8: **Five Factor Model: Standard Deviation of Observed and Model-Implied Yields**

This figure plots the unconditional standard deviations of observed yields against those implied by the Five Factor Approximate Affine Model.

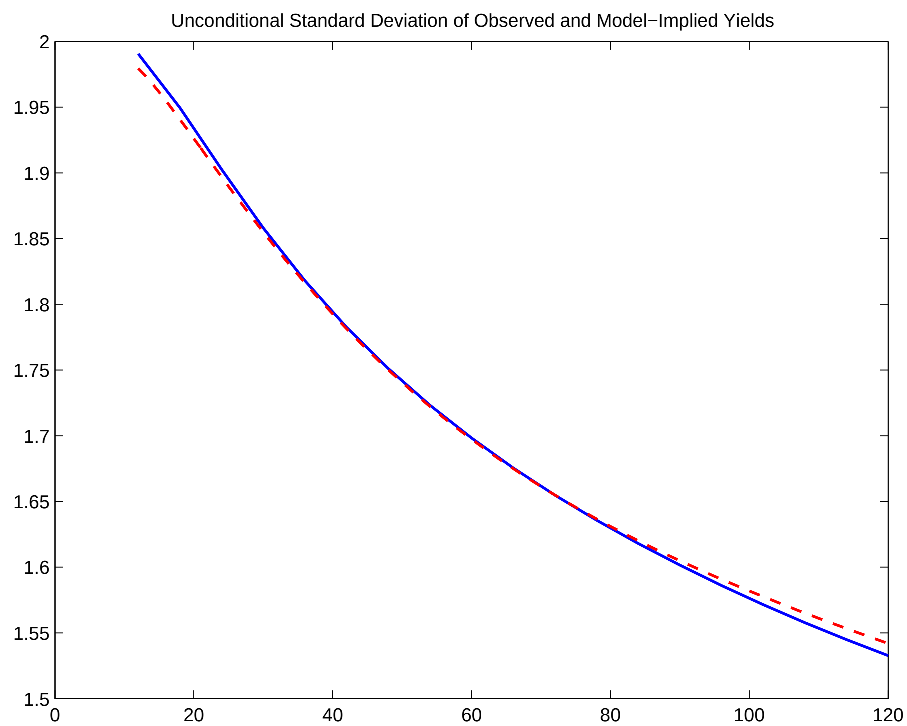


Figure 9: **Five Factor Model: Observed and Model-Implied Yields**

This figure provides plots of the observed and fitted yields for the 1-year, 2-year, 5- and 10-year maturities. The observed yields are plotted by solid lines, whereas dashed lines correspond to yields implied by the three factor approximate affine model.

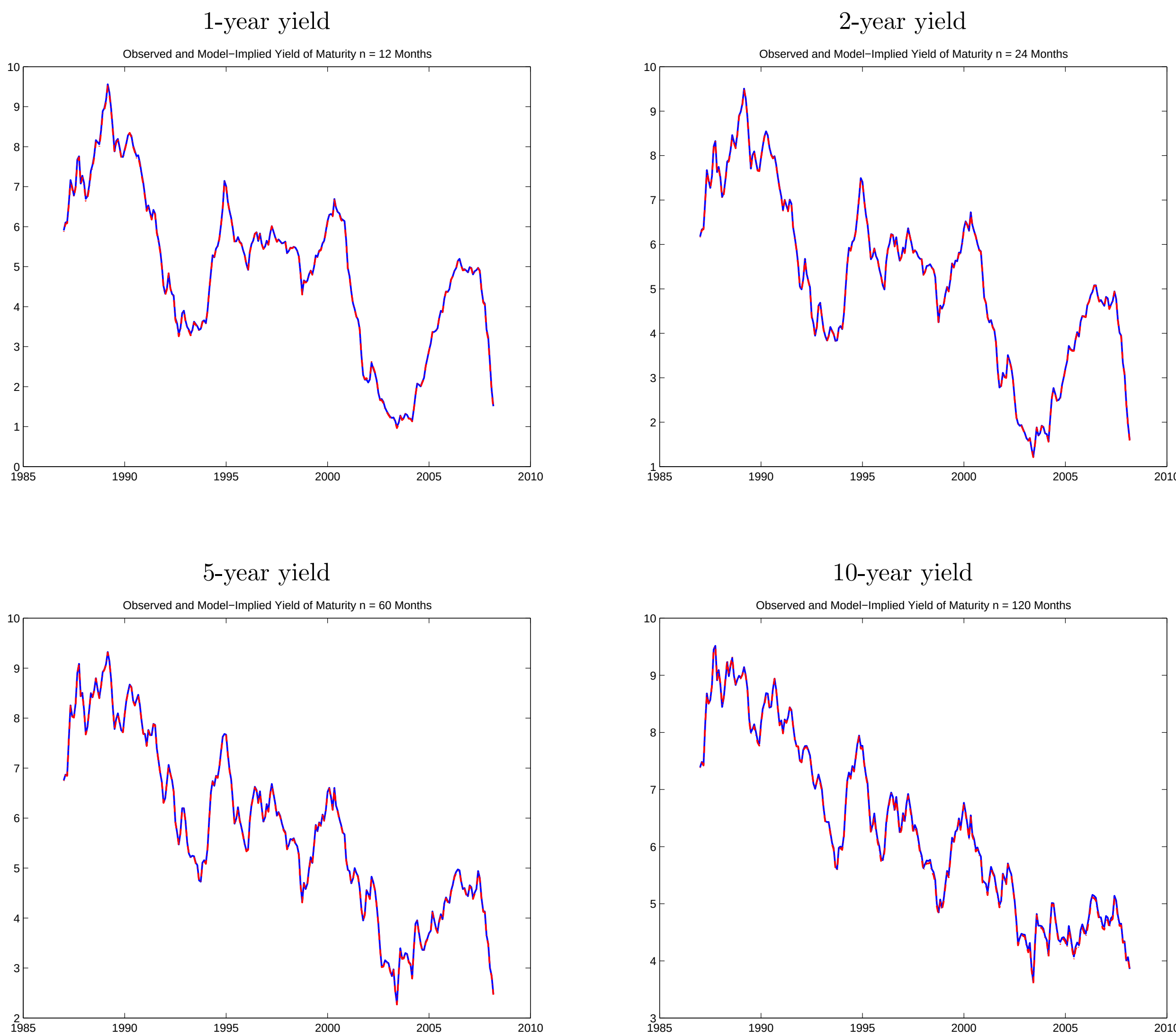


Figure 10: **Five Factor Model: Observed and Model-Implied Excess One-Month Holding Returns**

This figure provides plots of the observed and model-implied excess one-month holding returns for the 1-year, 2-year, 5- and 10-year maturities. The observed returns are plotted by solid lines, whereas dashed green lines correspond to actual model-implied returns, and dash-dotted red lines to model-implied expected one-month excess holding returns.

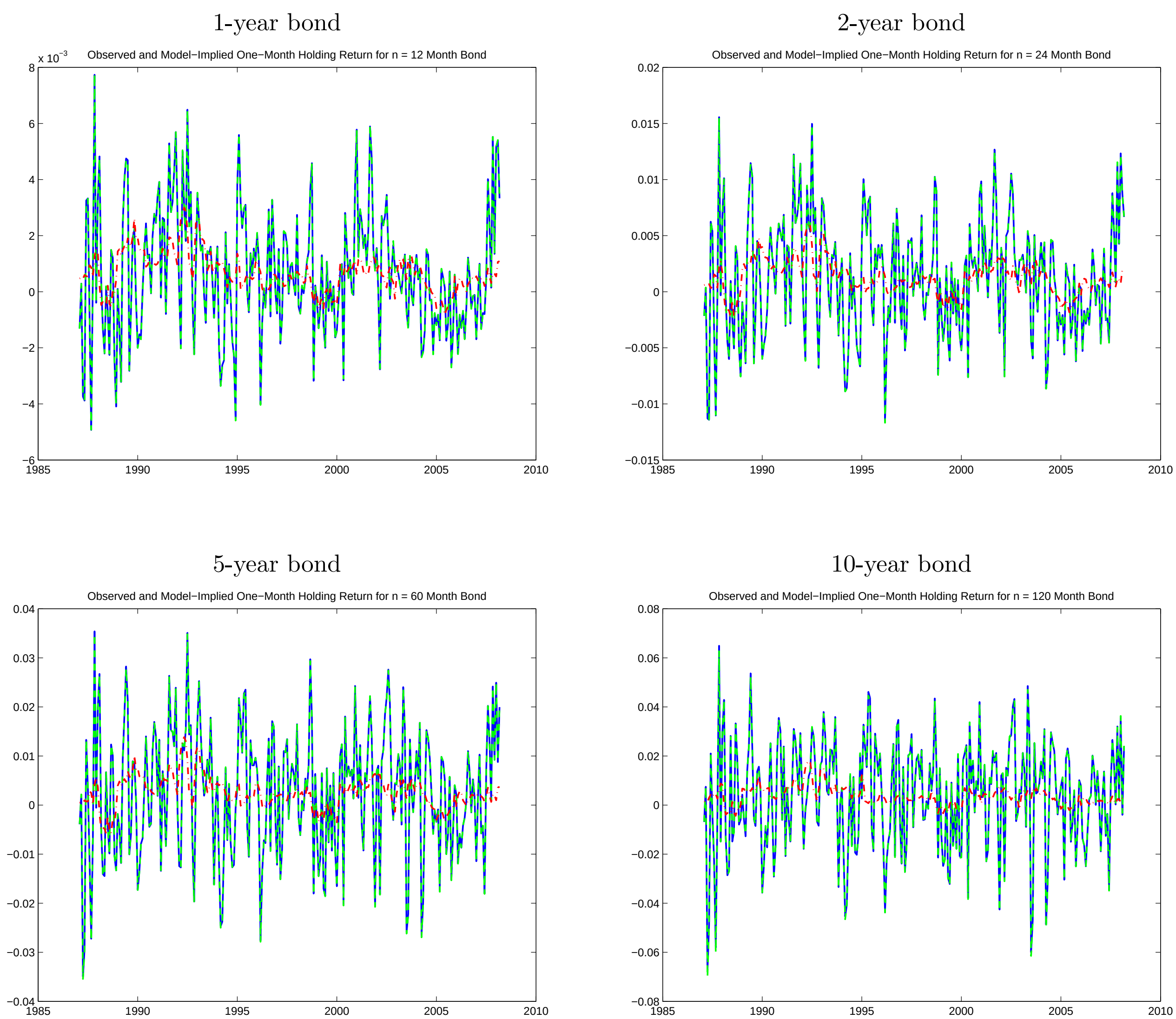


Figure 11: **Five Factor Model: Factor Loadings for Yields**

This figure provides a plot of the yield loadings $-\frac{1}{n}B_n$ implied by the Five Factor Approximate Affine Model. The coefficients can be interpreted as the response of the n-month yield to a contemporary shock to the respective factor.

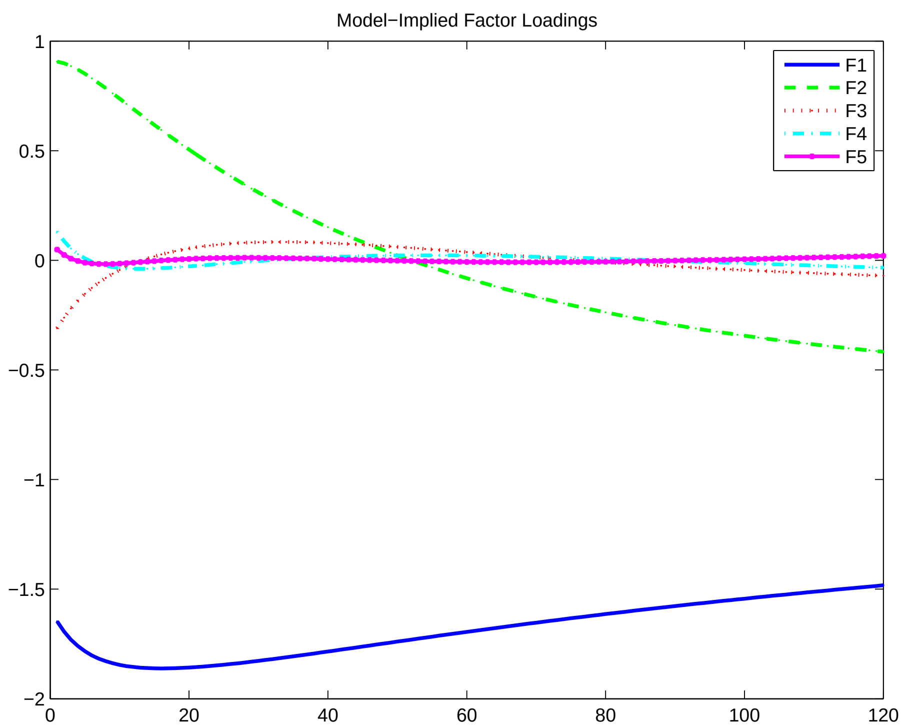
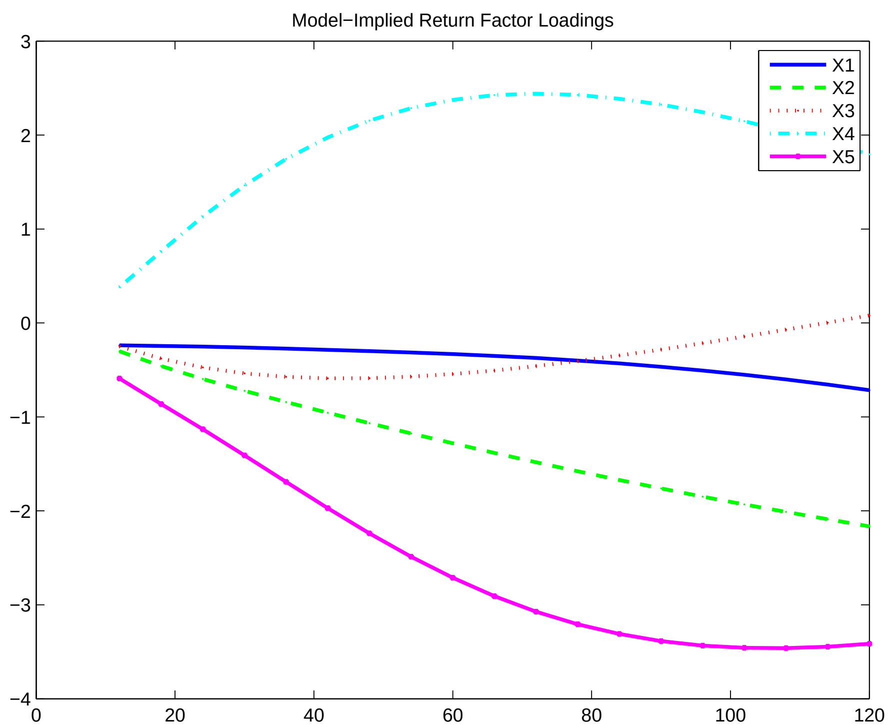


Figure 12: **Five Factor Model: Factor Loadings for Expected Returns**

This figure provides a plot of the expected return loadings $B_n'\lambda_1$ implied by the Five Factor Approximate Affine Model. The coefficients can be interpreted as the response of the expected one-month excess holing return on an n-month bond to a contemporary shock to the respective factor.



C References

Adrian, Tobias, and Joshua Rosenberg, 2008, "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *Journal of Finance*, forthcoming.

Campbell, John Y., Lo, Andrew, and A. Craig MacKinley, 1997, *The Econometrics of Financial Markets*, Princeton University Press.

Cochrane, John, 2005, *Asset Pricing*, Princeton University Press, Revised Edition.

Cochrane, John, and Monika Piazzesi, 2005, "Bond risk premia," *American Economic Review* 95, pp. 138-160.

Collin-Dufresne, Pierre, and Robert S. Goldstein, 2001, "Do Bonds Span Fixed Income Markets? Theory and Evidence for 'Unspanned' Stochastic Volatility," *Journal of Finance*, pp. 1685-1730.

Cox, John C., Ingersoll, Jonathan E., and Stephen A. Ross, 1981, "A Theory of the Term Structure of Interest Rates", *Econometrica* 53(2), pp. 385-407.

Dai, Qiang, and Kenneth Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance* 55, pp. 1943-1978.

Dai, Qiang, and Kenneth Singleton, 2002, "Expectations Puzzle, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics* 63, pp. 415-441.

Dai, Qiang, and Kenneth Singleton, 2003, "Term Structure Dynamics in Theory and Reality", *Review of Financial Studies* 16, pp. 631-678.

Diebold, Frank X., and C. Li, 2006, "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics* 130, pp. 337-364.

Duffie, Darrell, and Rui Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance* 6, pp. 379-406.

Duffee, Gregory R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance* 57, pp. 405-443.

Dybvig, Philip H., and Stephen A. Ross, 1987, “Arbitrage,” *The New Palgrave: a Dictionary of Economics*, pp. 100-106.

Evans, Martin, 2003, ”Real Risk, Inflation Risk, and the Term Structure,” *Economic Journal* 113, pp. 345-389.

Fama, Eugene F., and James D. MacBeth, 1973, ”Risk, return, and equilibrium: Empirical tests,” *Journal of Political Economy* 81, pp. 607-636.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2006, ”The U.S. Treasury Yield Curve: 1961 to the Present,” *Journal of Monetary Economics* 54, pp. 2291-2304.

Hamilton, James D., 1994, *Time Series Analysis*, Princeton University Press.

Hansen, Lars-Peter, and Ravi Jagannathan, 1993, “Assessing Specification Errors in Stochastic Discount Factor Models,” *Journal of Finance* 48, pp. 1091-1092.

Jarrow, Robert A., David Heath, and Andrew Morton, 1992, ”Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation,” *Econometrica* 60, pp. 77-105.

Kim, Don, and Jonathan H. Wright, 2005, ”An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates,” *Board of Governors of the Federal Reserve System Finance and Economics Discussion Series* 2005-33.

Merton, Robert. C., 1973, ”An intertemporal asset pricing model,” *Econometrica* 41, pp. 867-887.

Moench, Emanuel, 2008, ”Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach”, *Journal of Econometrics*, forthcoming.

Nelson, Charles R. and Andrew F. Siegel, 1987, ”Parsimonious Modeling of Yield Curves”, *Journal of Business* 60(4), 473-489.

Piazzesi, Monika, 2003, "Affine Term Structure Models," *Handbook of Financial Econometrics.*

Politis, Dimitris N., Romano, Joseph P., 1994a, "Limit theorems for weakly dependent hilbert space valued random variables with applications to the stationary bootstrap," *Statistica Sinica* 4 (2), 461–476.

Politis, Dimitris N., Romano, Joseph P., 1994b, "The stationary bootstrap," *Journal of the American Statistical Association* 89 (428), 1303–1313.

Singleton, Kenneth J., 2006, *Empirical Dynamic Asset Pricing*, Princeton University Press.

Smith, Richard J., 1992, "Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments," *Econometrica* 60, pp. 973-980.