A no-arbitrage term structure model without latent factors

Gregory R. Duffee[*]

Haas School of Business
University of California – Berkeley

This Draft: April 26, 2004

ABSTRACT

I present a framework for modeling part of the dynamics of the term structure. The framework can be used to link the term structure to observed variables such as inflation and output. Its partial nature allows us to dispense with yield-based factors (e.g., latent factors) while retaining restrictions associated with no-arbitrage. I apply the model to the joint dynamics of inflation and the term structure. As other research has noted, both short-term and long-term bond yields adjust gradually to a change in inflation. I find that the dynamics of the price of interest rate risk needed to fit this pattern from 1983 through 2003 are implausible. An alternative interpretation is that investors were systematically surprised by the slow adjustment of short-term yields to inflation.

[*]Voice 510-642-1435, fax 510-643-1420, email duffee@haas.berkeley.edu. Address correspondence to 545 Student Services Building #1900, Berkeley, CA 94720. I thank George Pennacchi, Richard Stanton, and seminar participants at the University of Illinois for helpful comments. The most recent version of this paper is at http://faculty.haas.berkeley.edu/duffee/.

1 Introduction

Beginning with Vasicek (1977) and Cox, Ingersoll, and Ross (1985), researchers have built increasingly sophisticated no-arbitrage models of the term structure. These models specify the evolution of state variables under both the physical and equivalent martingale measures, and thus represent complete descriptions of the dynamic behavior of yields at all maturities. Much of this work focuses on latent variable settings, where the evolution of yields is described in terms of yields themselves. This rather introspective view was broadened by the important work of Piazzesi (2003) and Ang and Piazzesi (2003), who included macroeconomic variables in the workhorse latent variable framework of Duffie and Kan (1996). This extension allows us to investigate questions at the boundaries of macroeconomics and finance. For example, what does today's output gap tell us about the compensation investors demand to face interest rate risk? What does today's inflation rate tell us about the components of expected future real returns to nominal long-term bonds? These and related questions are now the focus of intensive research using models that describe the entire term structure with a combination of macroeconomic and latent factors.[1]

Yet many of these questions can be addressed without attempting to model the complete dynamics of the term structure. The logic follows the general asset-pricing approach introduced in Hansen and Singleton (1982), who noted that restrictions implied by no-arbitrage can be exploited without using (or knowing) the complete joint dynamics of asset prices and the pricing kernel. Recall that a zero-coupon bond's price is the expected value of the pricing kernel at its maturity. If we condition this expectation on the information in a specified set of macroeconomic variables and combine it with the dynamics of the same macro variables, we can exploit no-arbitrage restrictions without specifying the rest of the term structure. For example, the contemporaneous relation between a long-term bond's price and the output gap is determined by the information in the output gap for both the expected time-path of short-term interest rates over the life of the bond and expected time path of compensation investors require to face interest rate risk. In other words, it is determined by expectations conditioned on the output gap; other information is not necessarily relevant.

In this paper I describe how to build and estimate partial term structure models that use Duffie-Kan dynamics, impose no-arbitrage, and do not include yield-based factors (whether they are latent factors or yields themselves). There are latent factors in the background, but they play no role in either the model's testable restrictions or in estimation. These partial models allow us to ask what the macroeconomic factors tell us about the future evolution

[1] Recent work includes Dewachter, Lyrio, and Maes (2002), Dewachter and Lyrio (2002), Hördahl, Tristiani, and Vestin (2002), Ang and Bekaert (2003), Ang, Piazzesi, and Wei (2003), and Rudebusch and Wu (2003).

of the term structure. In particular, the covariation between bond yields and the observable factors can be decomposed into information about expected future short-term rates and information about expected future risk premia. The partial nature of the models means that certain questions cannot be addressed. For example, we cannot use them to ask what today's term structure tells us about the future evolution of the observable factors.

The tradeoffs between using a partial and a complete term structure model are standard. A complete model pins down more features of the data. If the additional restrictions imposed by a complete model are correct, then exploiting these features increases estimation efficiency and ultimately tells us more about term structure dynamics. But misspecification in these additional restrictions can distort our view of the aspects of the model that we care about. The danger of such misspecification is particularly high with complete term structure models that include both latent and observable factors. Unrestricted versions of such models have too many parameters to estimate in practice, forcing researchers to impose a priori parameter restrictions that seem sensible. In this paper I document that apparently sensible restrictions can be strongly at odds with actual term structure dynamics.

I use this framework to investigate an underappreciated issue in the relation between inflation and the nominal term structure. It is well-known that in U.S. data, an increase in inflation corresponds to an increase in short-term rates. The adjustment of short end of the yield curve to inflation occurs over a few quarters. If investors anticipate these dynamics and bond risk premia do not change when inflation changes, then long-term yields will immediately adjust to this expected gradual increase in short-term rates. But as previous research has observed, long-term yields also adjust gradually to inflation. This slow adjustment leads to complicated short-run dynamics of term premia (long-term yields less expected average short-term rates).

Armed with a flexible specification of the price of interest rate risk, we can explain all of these dynamics using the partial term structure framework. In this sense, the model is a success. However, when I focus on the most recent U.S. experience (post 1982), I find that highly unusual patterns in risk-compensation dynamics are required to fit the data. A positive shock to inflation today corresponds to an immediate, sharp decline in risk compensation that lasts for about two quarters, followed by a few quarters of relatively high risk compensation. Although this pattern is a consequence of no-arbitrage, it is hard to give it an equilibrium interpretation. It is easier to tell a story in which investors simply did not anticipate the gradual adjustment of short-term rates.

In the next section I describe the modeling framework. Section 3 describes the estimation methodology. Section 4 reviews the relevant facts about the relation between inflation and nominal bond yields. Section 5 presents results from estimation of a no-arbitrage model. It also presents some independent evidence on investors' forecasts of short-term rates. Section 6 concludes.

2 A partial term structure model

Underlying the dynamics of bond yields is some structural model that explains these dynamics in terms of the state of the macroeconomy, central bank policy, and investors' willingness to bear interest-rate risk. Although the model here uses observable macro variables, it is not a structural model like that in Buraschi and Jiltsov (2004). It is closer in spirit to a reduced-form model linking bond yields to macro variables.

2.1 The formal structure

Time is indexed by discrete periods $t, t+1, \ldots$. The length of a period is h years. The period t price and yield of a zero-coupon bond that pays a dollar at period $t+k$ are denoted $P_{k,t}$ and $y_{k,t}$. Yields are expressed as continuously-compounded annual rates. The short-term interest rate, which is equivalent to the yield on a one-period bond, is denoted r_t. An n-vector of macro variables is denoted m_t^*. The term "macro" is arbitrary here. In principle m_t^* can include any observed variable that we are interested in relating to bond yields.

Project r_t onto the history of m_t^*. To express this history in compact notation, stack lags zero through $p-1$ of m_t^* in the vector m_t. The projection is

$$r_t = \delta_0 + \delta_1' m_t + w_t. \tag{1}$$

Because this is a projection, $\text{Cov}(m_t, w_t) = 0$ by construction. Structure is imposed on the relation between r_t and the macro variables with the following two assumptions. First, there is a linear relation between r_t and m_t. This assumption allows us to replace the orthogonality relation between m_t and w_t with a stronger relation: the expectation of w_t conditioned on m_t is zero. Second, the lag length p is sufficiently long that there is no information in the history of m_t^* about the conditional mean of r_t that is not contained in m_t. Formally, the expectation of w_t conditioned on $m_{t-i}, i \geq 0$, is zero. We can reinterpret these assumptions in terms of expectations of future residuals:

$$E(w_{t+k}|m_t) = 0, \; k \geq 0. \tag{2}$$

The residual w_t captures variation in short-term rates that is unrelated to m_t. Although w_t is unforecastable with the history of m_t, other information available to investors likely will allow them to forecast some part of w_t. For example, the central bank may have recently decided to tighten monetary policy in response to events unrelated to m_t. If so, investors today expect this residual component to be greater than zero for some time. Investors may also have information about m_{t+k} that is not contained in m_t. In particular, w_t may contain information about m_{t+k}. For example, the exogeneous tightening of monetary policy may affect future realizations of the macro variables. The only implication this model has for the information in w_t is that any predictability has to go in one direction: w_t can predict future values of m_t, but m_t cannot predict future values of w_t.

To complete the term structure model we need descriptions of the dynamics of m_t and w_t that are consistent with this one-way predictability. We also need to specify the dynamics of the pricing kernel. The discussion has to proceed carefully because at various points the model uses two different information sets: the complete set of information available to investors and the subset based on m_t. I use $E_t(\cdot)$ to represent expectations conditioned on the complete information set and $E(\cdot|m_t)$ to represent expectations conditioned on the more restrictive information set.

The model does not require a complete description of the dynamics of m_t. The relevant dynamics of m_t are those conditioned on m_t. I sometimes refer to these dynamics as "univariate" dynamics (but remember that m_t is a vector). I assume that the one-step-ahead expectation of m_t conditioned on m_t is given by a VAR(1):

$$m_{t+1} = c + F m_t + \epsilon_{t+1}, \quad E(\epsilon_{t+1}|m_t) = 0. \tag{3}$$

The same macro dynamics are used in Ang and Piazzesi (2003), although they assume these dynamics are based on investors' complete information sets. The linearity of the VAR is restrictive but the first-order dynamics are not because of the lags included in m_t.[2] Owing to the first-order companion form, the parameters c and F can be expressed as

$$c = \begin{pmatrix} c_1 \\ 0 \end{pmatrix}, \; F = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}, \; F_2 = \begin{pmatrix} I & 0 \end{pmatrix},$$

where c_1 is a n-vector and F_1 is an $n \times pn$ matrix. The identity matrix in F_2 has $n(p-1)$ rows and the components represented with zeros are conformable. Note that we can write the k-ahead conditional forecast as

$$E\left(m_{t+k}|m_t\right) = (I - F)^{-1}(I - F^k)c + F^k m_t \tag{4}$$

[2]This is slightly overstated; the first-order structure rules out non-Markovian dynamics.

where the identity matrices have the same dimensions as F.

Investors work with a more complete model that links the dynamics of m_t to other variables in the economy. In this more complete model investors will distinguish among different kinds of shocks to m_t. For the purposes of illustration, assume that m_t consists of current and lagged measures of inflation. An increase in inflation owing to an oil price shock may die out at a different rate than an increase in inflation owing to a policy shock. Investors will distinguish between these different kinds of shocks using information not in m_t. Equation (3) should be viewed as the result of integrating out all non-m_t information from this more complete model.

In particular, in the model used by investors the "shock" ϵ_{t+1} need not be entirely unexpected. The subscript on this innovation indicates that it enters the restricted information set at $t+1$. However, the expectation of ϵ_{t+1} conditioned on investors' complete information set at time t is not necessarily zero. Investors may have information about the evolution of macro variables that is not contained in lagged macro variables. The one assumption the model requires about investors' complete information set is that the remaining uncertainty in ϵ_{t+1} is normally distributed:

$$\epsilon_{t+1} - E_t(\epsilon_{t+1}) \sim N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & 0 \end{pmatrix}, \tag{5}$$

where Σ_{11} is an $n \times n$ positive definite matrix. (The zeros in Σ are a consequence of the companion form.) The econometrician does not know Σ and does not estimate it.

Although the non-macro component of the short-term rate w_t drops out of the testable part of the model, some structure on its dynamics is necessary in order to justify dropping it. I assume that w_t can be written as an linear function of some latent state vector x_t of arbitrary length:

$$w_t = \delta_2' x_t. \tag{6}$$

This state vector is observed by investors. Its data-generating process is affine conditional on investors' complete information set:

$$x_{t+1} = F_x x_t + \sigma(x_t)\eta_{t+1}, \tag{7}$$

$$\sigma(x_t)\sigma(x_t)' = v_0 + v_1 x_t, \tag{8}$$

$$\eta_{t+1} \sim N(0, I), \;\; \text{Cov}_t(\eta_{t+1}, \epsilon_{t+1}) = 0. \tag{9}$$

This structure implies that the expectation of x_t conditional on m_t is zero, as required by (2). The vector v_0 and matrices F_x and v_1 are free parameters but play no role in the model's testable implications.

Nothing in these dynamics rules out a correlation between x_t and investors' time-t expectations of $\epsilon_{t+k}, k > 0$. A concrete example helps to interpret these dynamics. Let the observable factor m_t^* consist of a single variable: the CPI inflation rate from $t-1$ to t. Imagine that at the end of period t OPEC announces that it will cut sales of crude oil. This will raise producer prices immediately but will not affect the period-t CPI. The announcement raises investors' expectation of inflation in $t+1$, thus the nominal short-term rate will rise to compensate investors. Hence r_t will be higher than investors expected as of the end of quarter $t-1$. Because this innovation to r_t is unrelated to the observed history of the CPI inflation rate, it shows up as a shock to latent factors η_t. This shock is correlated with next period's innovation in the CPI inflation rate ϵ_{t+1}. If oil price shocks are completely incorporated into the CPI within a period, the shock η_t will not persist. In other words, by the end of $t+1$ the OPEC action will be embedded in m_{t+1}, not in the residual component w_{t+1}.

No-arbitrage implies the existence of a pricing kernel z_{t+1} such that $P_{k,t} = E_t(z_{t+1}P_{k-1,t+1})$ for all maturities k. The key assumption about the pricing kernel is that the compensation investors require to face latent-factor risk depends only on the latent factors. The dynamics of the pricing kernel conditional on investors' complete information set are

$$z_{t+1} = \exp\left[-hr_t - \lambda_t'(\epsilon_{t+1} - E_t(\epsilon_{t+1})) - d(x_t)'\eta_{t+1} - \frac{1}{2}\text{Var}_t(\lambda_t'\epsilon_{t+1} + d(x_t)'\eta_{t+1})\right], \quad (10)$$

$$\sigma(x_t)d(x_t) = s_0 + s_1 x_t, \quad (11)$$

$$\lambda_t = \lambda_0 + \lambda_1 m_t + \lambda_2 x_t, \quad (12)$$

where the free parameters in λ_1 are in an $n \times np$ matrix λ:

$$\lambda_1 = \begin{pmatrix} \Sigma_{11}^{-1}\lambda \\ 0 \end{pmatrix}.$$

The dynamics in (12) are a special case of the essentially affine description of risk dynamics introduced in Duffee (2002). The zeros in λ_1 are not a restriction because all but the first n elements of ϵ_{t+1} are zero. The vectors s_0 and λ_0 and the matrices s_1 and λ_2 are free parameters but they play no role in the model's testable implications. Note that from the perspective of the pricing kernel, all shocks $\epsilon_{t+1} - E_t(\epsilon_{t+1})$ are alike.

These assumptions allow us to use standard techniques to solve for bond prices. Guess that log bond prices are affine in m_t and x_t:

$$\log P_{k,t} = A_k + B'_{1,k} m_t + B'_{2,k} x_t. \tag{13}$$

Because log bond prices and the pricing kernel are both normally distributed conditioned on investors' complete information sets, we can rewrite the no-arbitrage condition $P_{k,t} = E_t(z_{t+1} P_{k-1,t+1})$ as

$$\log P_{k,t} = -hr_t + E_t(\log P_{k-1,t+1}) + \frac{1}{2}\text{Var}_t(\log P_{k-1,t+1}) + \text{Cov}_t(\log P_{k-1,t+1}, z_{t+1}). \tag{14}$$

Plugging (13) into (14) and using (5), (7), (8), (9), and (10) produces

$$\begin{aligned} A_k + B'_{1,k} m_t + B'_{2,k} x_t &= -hr_t + A_{k-1} + B'_{1,k-1} E_t(m_{t+1}) + B'_{2,k-1} E_t(x_{t+1}) \\ &\quad + \frac{1}{2}\left(B'_{1,k-1} \Sigma B_{1,k-1} + B'_{2,k-1} \sigma(x_t)\sigma(x_t)' B_{2,k-1}\right) \\ &\quad - B'_{1,k-1} \Sigma \lambda_t - B'_{2,k-1} \sigma(x_t) d(x_t). \end{aligned} \tag{15}$$

Take the expectation of (15) conditioned on m_t and evaluate it using (3), (8), and (11):

$$\begin{aligned} A_k + B'_{1,k} m_t &= -h(\delta_0 + \delta'_1 m_t) + A_{k-1} + B'_{1,k-1}(c + F m_t) \\ &\quad + \frac{1}{2}\left(B'_{1,k-1} \Sigma B_{1,k-1} + B'_{2,k-1} v_0 B_{2,k-1}\right) \\ &\quad - B'_{1,k-1} \Sigma (\lambda_0 + \lambda_1 m_t) - B'_{2,k-1} s_0. \end{aligned} \tag{16}$$

Matching coefficients on m_t produces a recursive equation that defines $B_{1,k}$.

$$B'_{1,k} = -h\delta'_1 + B'_{1,k-1} F - B_{1,k-1} \Sigma \lambda_1$$

The solution is

$$B'_{1,k} = -h\delta'_1 (I - \tilde{F})^{-1}(I - \tilde{F}^k) \tag{17}$$

where $\tilde{F}$ is the equivalent-martingale counterpart to F:

$$\tilde{F} = \begin{pmatrix} F_1 - \lambda \\ F_2 \end{pmatrix}. \tag{18}$$

2.2 Comments on the model

This partial term structure model specifies two empirical relations: The univariate dynamics of m_t and the contemporaneous relation between m_t and the entire term structure. There are a total of $n(1+p)+2n^2p$ free parameters. The univariate dynamics of m_t require $n(1+np)$ parameters (c and F_1), the relation between m_t and the short rate requires np parameters in δ_1, and the price of risk λ requires n^2p parameters.

A complete model linking m_t and bond yields must put more structure on the residual w_t. Its dynamics are specified and are linked to the dynamics of m_t. In the Duffie-Kan framework, this additional structure is embedded in a large number of additional parameters. For concreteness, consider a model where four lags of a single macro factor are linked to the term structure. (An explicit example is examined in Section 5.) With $n = 1$ and $p = 4$, the partial term structure model has 13 free parameters. If we were to augment this model with two latent factors that follow a joint Gaussian process, the essentially affine version of the complete model would require an additional 23 parameters.[3] Estimation of the complete model is impractical unless many of these parameters are fixed. For example, Ang and Piazzesi (2003) impose independence between the macro and latent factors under both the physical and equivalent martingale measures. They also rule out the possibility that lags of the macro factors can affect the price of risk of the macro factors. Given these (and other) parameter restrictions, they are able to estimate the remaining parameters. If our goal is to express the complete dynamics of the term structure, including its links to macro variables, we have to live with substantial a priori restrictions on the parameters such as those imposed in Ang and Piazzesi. But if the questions we want to address do not require a complete model, we can use this partial framework to avoid taking a stand on the dynamics of the latent factors.

In this model, yields covary with m_t for two reasons. First, m_t is correlated with the expected path of the short-term interest rate. Second, the compensation investors require to face uncertainty in m_t^* varies with m_t. A critical assumption in the model is that the compensation investors demand to face latent-factor risk does not depend on m_t. To understand the role of this restriction, consider replacing (11) with the more general form

$$\sigma(m_t, x_t)d(m_t, x_t) = s_{20} + s_{21}m_t + s_{22}x_t.$$

This general form allows both the conditional variance of the latent factors and the price of η_t risk to depend on m_t. With this form, the testable restrictions implied by the partial term structure model disappear. The conditional covariance between the log bond price and

[3]A purely latent factor, essentially affine Gaussian model with two factors has twelve parameters. To link these factors to the macro factors, additional parameters are required. The conditional expectation of both latent factors can depend on each of the four elements of m_t, adding eight parameters. Each latent factor can affect the conditional expectation of m_t^*, adding two parameters. Finally, a constant term in the price of m_t^* risk adds one parameter. This complete model imposes the restriction, also imposed in the partial model, that macro variables do not affect the price of risk of the latent factors.

the log pricing kernel is then

$$\text{Cov}_t(\log P_{k-1,t+1}, \log z_{t+1}) = -B'_{1,k-1}\Sigma(\lambda_0 + \lambda_1 m_t + \lambda_2 x_t) - B'_{2,k-1}(s_{20} + s_{21}m_t + s_{22}x_t)$$

and its expectation conditioned on m_t is

$$E\left[\text{Cov}_t(\log P_{k-1,t+1}, \log z_{t+1})|m_t\right] = -B'_{1,k-1}\Sigma(\lambda_0 + \lambda_1 m_t) - B'_{2,k-1}(s_{20} + s_{21}m_t).$$

Replace the last line of (16) with this expected covariance and match coefficients in m_t. The result is a a recursive equation for $B_{1,k}$ that depends on $B_{2,k}$. Therefore we need a solution to the entire term structure model in order to calculate $B_{1,k}$.

Put differently, the restrictions of no-arbitrage are driven by a comparison of risk and expected excess returns across assets. The factor loadings $B_{1,k}$ give us a relative measure of macro risk across bonds, but we have no relative measure of latent factor risk across bonds without a parameterized description of the dynamics of the latent factors. Hence if m_t affects the price of latent factor risk, no-arbitrage does not allow us to say much about the relation between m_t and the term structure.[4]

If this critical restriction built into (11) is inconsistent with the data, the model will misstate what happens to the shape of the term structure when m_t changes. Since the assumption is driven by practical considerations instead of economic theory, it is important to estimate the model with a technique that evaluates the validity of the assumption with overidentifying restrictions. Estimation issues are addressed in the next section.

Other assumptions in the model are less important. In particular, the affine form for x_t is not essential. The only reason the affine form is used is to guarantee joint log-normality of bond prices and the pricing kernel. Log-normality implies (14), which is the equation that allows us to use (15) to match up the terms in m_t. A more general representation replaces the linear form of the residual (6) with some nonlinear function $f_1(x_t)$:

$$w_t = f_1(x_t), \quad E(f_1(x_t)|m_t) = 0.$$

Log-normality requires that conditional on investors' complete information set, innovations in $f_1(x_t)$ are normal:

$$f_1(x_{t+1}) - E_t(f_1(x_{t+1})) \sim N(0, V_1(x_t)).$$

[4]The same point applies to the variance-covariance matrix of latent factor shocks. If this variance depends on m_t, a Jensen's inequality term in (15) will not disappear after conditioning on m_t.

Similarly, we can replace the affine form of (13) with

$$\log P_{k,t} = A_k + B'_{1,k} m_t + f_k(x_t)$$

where again log-normality requires

$$f_k(x_{t+1}) - E_t(f_k(x_{t+1})) \sim N(0, V_k(x_t)).$$

The affine dynamics of x_t results in functional forms of $f_k(x_t)$ that are consistent with the requirement of log-normality. But since the relation between x_t and the term structure is not of direct interest, we might just as well simply assume log-normality to motivate the model's restrictions involving m_t.

2.3 Applications

The model links m_t to the the term structure through the VAR(1) description of m_t, the vector δ_1 linking the short rate to m_t, and the matrix λ that describes how the compensation investors demand to face m_t^* risk varies with m_t. The remainder of the term structure is not modeled. With this partial model of the term structure we can ask the following kinds of questions.

1. How does the expected time-path of r_t depend on m_t?

   The response of the short rate to a macroeconomic shock ϵ_t is $\delta_1' \epsilon_t$. The expected change in the short rate from t to $t+k$, conditioned on m_t, is

   $$E(r_{t+k} - r_t | m_t) = \delta_1' \left[(I - F^k)c - (I - F^k)m_t \right]. \tag{19}$$

   The final term in square brackets uses (4). One application of these expressions is to investigate Taylor-rule descriptions of the short rate. Note that this k-ahead forecast is not a minimum-variance forecast. There is additional information in the term structure (at a minimum, the current level of the short rate) that is ignored in forming this conditional expectation. Therefore this model should not be used as a forecasting tool, but rather as a way to link the macroeconomic variables with the term structure.

2. How do risk premia on bonds vary with m_t?

   The partial nature of the model does not allow us to say anything about mean excess returns to bonds. However, it does allow us to determine how variations in m_t correspond to variations in expected excess returns. The expected excess log return to a

k-maturity bond held from t to $t+1$, conditioned on m_t, is

$$E(\log P_{k-1,t+1} - \log P_{k,t} - hr_t | m_t) = \kappa_{k-1} + B'_{1,k-1} \begin{pmatrix} \lambda \\ 0 \end{pmatrix} m_t.$$

The constant term κ_{k-1} is unrestricted.

3. How does the term structure react to an innovation in a macroeconomic variable?

   Consider the expectation of the k-maturity yield conditional on the contemporaneous m_t. (In other words, we observe m_t but we have no observations of $x_{t-s}, s \geq 0$.) The expectation is

$$y_{k,t} = E(y_{k,t}|m_t) + \nu_{k,t}, \tag{20}$$

$$E(y_{k,t}|m_t) = a_k + (1/k)\delta_1'(I - \tilde{F})^{-1}(I - \tilde{F}^k)m_t, \tag{21}$$

$$a_k = -(1/k)A_k, \quad \nu_{k,t} = -(1/k)B'_{2,k}x_t. \tag{22}$$

   The constant term a_k is unrestricted. The effect on $y_{k,t}$ of the macroeconomic shock ϵ_t is $(1/k)\delta_1'(I - \tilde{F})^{-1}(I - \tilde{F}^k)\epsilon_t$.

4. What does m_t tell us about the future evolution of the term structure?

   The k-period-ahead forecast of the change in the yield on a bond with constant maturity q is

$$\begin{aligned} &E(y_{q,t+k} - y_{q,t}|m_t) = \\ &\frac{1}{q}\delta_1'(I - \tilde{F})^{-1}(I - \tilde{F}^q)\left[(I - F)^{-1}(I - F^k)c - (I - F^k)m_t\right]. \end{aligned} \tag{23}$$

5. Is the empirical failure of the expectations hypothesis associated with m_t?

   Campbell and Shiller (1991) estimated regressions of the form

$$y_{l-s,t+s} - y_{l,t} = b_0 + b_1 \frac{s}{l-s}(y_{l,t} - y_{s,t}) + e_{t+s,l,s} \tag{24}$$

   for maturities $l > s$. Under the weak form of the expectations hypothesis the coefficient b_1 should equal one, but in the data it is often negative. A common interpretation of this result is that bond risk premia and the slope of the term structure are positively correlated. Is this also true of the variability in the term structure that is associated with variations in m_t? Consider estimating (24) using m_t as instruments. The conditional expectation of yield spread on the right of (24) can be expressed as

$$E(y_{l,t} - y_{s,t}|m_t) = \rho_{l,s} + \left(-\frac{1}{l}B_{1,l} + \frac{1}{s}B_{1,s}\right)' m_t$$

where $\rho_{l,s}$ is an unrestricted constant. The conditional expectation of the left side of (24) can be expressed as

$$E(y_{l-s,t+s} - y_{l,t}|m_t) = \phi_{l,s} + \frac{s}{l-s}E(y_{l,t} - y_{s,t}|m_t) - \frac{1}{l-s}B'_{1,l-s}\left(F^s - \tilde{F}^s\right) m_t \quad (25)$$

where $\phi_{l,s}$ is an unrestricted constant. If $\lambda = 0$, then $F = \tilde{F}$ and the final term in (25) is identically zero. In this case, the population estimate of b_1 from IV estimation of (24) is one. More generally, the population regression coefficient is

$$\begin{aligned}\hat{b}_1 &= 1 - \frac{1}{s}\left[\left(-\frac{1}{l}B_{1,l} + \frac{1}{s}B_{1,s}\right)' \mathrm{Var}(m) \left(-\frac{1}{l}B_{1,l} + \frac{1}{s}B_{1,s}\right)\right]^{-1} \times \\ & \left(-\frac{1}{l}B_{1,l} + \frac{1}{s}B_{1,s}\right)' \mathrm{Var}(m) \left(F^s - \tilde{F}^s\right)' B_{1,l-s}\end{aligned}$$

where $\mathrm{Var}(m)$ is the unconditional variance-covariance matrix of m_t. Given this variance and the parameters of the term structure model, the regression coefficient can be computed.

3 Econometric methodology

I consider a setting in which the econometrician observes, at each date $1, 2, \ldots, T$, the macroeconomic vector m_t^* and yields on J bonds with constant maturities $k_j, j = 1, \ldots, J$. The natural technique to estimate the parameters of the model is the Generalized Method of Moments (GMM) of Hansen (1982). One obvious set of moments are the OLS moment conditions associated with the macroeconomic vector autoregression. The $n(1 + np)$ moments associated with observation t are

$$[m_t^* - c_1 - F_1 m_{t-1}] \otimes (1 \;\; m'_{t-1})'. \quad (26)$$

The choice of moment conditions to identify the term structure parameters is not as clear. One approach is to relate the level of m_t to the level of bond yields using (20) and (21). More precisely, we can introduce additional free parameters a_1 through a_J to pick up constant terms in the J yields and use the following $np + 1$ moments associated with bond

j at time t:

$$\left[y_{k_j,t} - a_j - \frac{-1}{k_j} B'_{1,k_j} m_t\right] \otimes (1 \;\; m'_t)'. \tag{27}$$

Unfortunately, typical choices of macroeconomic variables produce residuals $\nu_{k,t}$ in (20) that are highly serially correlated. For example, the residuals from Taylor-rule regressions of the short-term interest rate are close to a random walk.[5] Although this behavior does not invalidate the moment conditions, it creates substantial problems for statistical inference.

Both yield spreads and first-differences of yields are less persistent than are yields. We can use the following np moments for the first difference of the k_j-maturity yield:

$$\left[(y_{k_j,t} - y_{k_j,t-1}) - \frac{-1}{k_j} B'_{1,k_j}(m_t - m_{t-1})\right] \otimes (m_t - m_{t-1}). \tag{28}$$

Another advantage of (28) relative to (27) is that constant terms do not need to be added to the set of free parameters. If there is a nonzero correlation between x_t and $m_{t+s}, s > 0$, this moment condition may not have expectation zero because of the possibility that m_t is correlated with the residual of the bond yield observed at $t - 1$. In this case the right side of the Kronecker product can be replaced with functions of the history of m_t that do not include m^*_t. An obvious candidate is simply m_{t-1}.

An $np + 1$ moment vector for the yield spread between the k_j-maturity bond and the k_k-maturity bond is

$$\left[(y_{k_j,t} - y_{k_k,t}) - a_{j,k} - \left(\frac{-1}{k_j} B'_{1,k_j} - \frac{-1}{k_k} B'_{1,k_k}\right) m_t\right] \otimes (1 \;\; m'_t)' \tag{29}$$

This moment condition requires the additional free parameter $a_{j,k}$.

In practice I use a combination of (26) and (28). The model's parameters are exactly identified if there is the number of observed bond yields is one greater than the number of macroeconomic factors. Additional bond yields allow overidentifying tests of the model.

We can also test for restrictions on the dynamics of the price of risk. There are two special cases of interest. The first is when the compensation required by investors to face macro risk depends only on the current value of the macroeconomic variables. In this case we can write λ as

$$\lambda = (\lambda_{11} \mid 0)$$

where λ_{11} is an n by n matrix. Another special case is when compensation depends only on

[5] See, e.g., Rudebusch (2002).

the one-step-ahead forecast of the macroeconomic variables. In this case we can write λ as

$$\lambda = (\lambda_{11} \mid 0)\, F$$

Both of these cases impose $(p-1)n$ additional restrictions on the parameter vector. (Neither of these special cases is meaningful if the number of lags p is one.)

4 Inflation and the term structure

Short-term interest rates react slowly to inflation. (By reaction, I refer to a statistical relation. Causation is a big issue that is not central to the discussion here.) Most of the relevant evidence is found in the literature on the Taylor rule. The description of the central bank's reaction function in Taylor (1993) is that the current Fed funds rate is based in part on inflation over the past year. Put differently, current short-term rates depend on lags zero through three of quarterly inflation. There is an ongoing debate over whether the Fed reacts more slowly to inflation (and output) than is implied by this rule. Clarida, Galí, and Gertler (2000) recommend adjusting the Taylor rule to account for the the desire of the Fed to smooth rates. Rudebusch (2002) argues that the evidence is consistent with the Taylor rule combined with persistent policy shocks, while English, Nelson, and Sack (2003) claim that a better description of the data is slow reaction to inflation and output combined with persistent policy shocks. A recent review of the evidence is in Sack and Wieland (2000).

The literature on macroeconomic vector autoregressions contains similar evidence, although the primary focus of that literature is on identifying the effects of monetary policy shocks. For example, Leeper, Sims, and Zha (1996) contains some results on the responsiveness of short-term rates to inflation shocks. A related literature beginning with Fama (1975) reverses the question of predictability by considering the forecast power of interest rates for future inflation. This direction of predictability cannot be addressed by the partial term structure model developed in this paper, thus I do not consider it further.

Research on the Taylor rule, as well as macro vector autoregressions, typically concentrates on the dynamics of short-term rates rather than the dynamics of long-term rates. An exception in the Taylor rule literature is Rudebusch (2002), who uses information in forward rates combined with the assumption that risk premia do not depend on the level of short-term rates. Exceptions in the vector autoregression literature are Bernanke, Gertler, and Watson (1997), Evans and Marshall (1998), and Evans and Marshall (2002). They examine how both short-term and long-term yields respond to various shocks. The vector autoregression structure allows them to construct expected short-term interest rates, which in turn allows them to discuss the response of term premia (long-term yields less expected average short-term rates) to these shocks.[6] Ang and Piazzesi (2003) also reports VAR evidence on the reaction of bond yields to inflation. A broad conclusion from this work is that both short-term and long-term yields adjust slowly to inflation shocks.

Some regression evidence relating short-term rates, long-term rates, and inflation will illustrate the nature of the slow adjustment and the puzzle that it raises. The data I use are quarterly from 1953Q1 through 2003Q4. Inflation in quarter t is measured by the change in the log of the personal consumption expenditure (PCE) chained price index from $t-1$ to t. I define quarter-t bond yields as yields as of the end of the first month in quarter $t+1$. This is not a standard choice in the literature on the Taylor rule, and thus deserves some motivation. The puzzle highlighted in the results that follow is that investors do not seem to react immediately to the news in inflation about future short-term interest rates. We therefore need to be sure that investors really have the information before we look for bond-price reactions to the information. The PCE index is announced in the first month of the next quarter, hence bond prices at the end of this month should reflect fully the information in the index. The short-term rate is the three-month yield from the Center for Research in Security Prices (CRSP) riskfree rate file. To illustrate the behavior of long-term yields I use the five-year zero-coupon yield from the CRSP Fama-Bliss file. All data are continuously compounded and expressed as annual rates.

I estimate a univariate AR(4) for inflation. I link inflation to bond yields by regressing yields on lags zero through three of inflation. Owing to the high serial correlation of the residuals, I estimate these yield regressions in two ways. First, I estimate the regression in levels assuming that the residual follows an AR(1) with parameter ρ. The parameters are estimated with Hildreth-Lu. Second, I estimate the regression with OLS using first differences (without a constant term), adjusting the variance-covariance matrix of the parameter estimates for generalized heteroskedasticity and four lags of moving-average residuals as in Newey and West (1987a).

An alternative approach is to use a vector autoregression of inflation and bond yields, but the results here are easier to compare with those of the term structure model presented in the next section.

The regressions are estimated over two periods: 1953 through 1982 and 1983 through 2003.[7] The former period includes a variety of monetary policy regimes, while the latter

[6] Diebold, Rudebusch, and Aruoba (2003) also use vector autoregressions to jointly model the yield curve and macroeconomic variables, but they do not consider the contemporaneous effect of inflation shocks on the term structure.

[7] To account for inflation lags, all regressions using the earlier sample are estimated over 1954Q1 through 1982Q4 (116 observations) and regressions using the later sample are estimated over 1984Q1 through 2003Q3

(which begins after the end of the Fed's monetarist experiment) is a more homogeneous sample. The motivation for splitting the sample is addressed in more detail later. The estimated inflation dynamics are reported in Table 1. Regressions of yields on contemporaneous and lagged inflation are reported in Table 2.

Results for the regressions in level form are displayed in Panel A. The first two rows concern the response of short-term rates to inflation. In both samples there is a slow response to inflation. In the earlier sample the largest response is to the second lag of inflation, while in the later sample the largest response is to the first lag. A joint test that the coefficients on all lags equal zero is rejected at only the ten percent level, but the lack of a strong rejection should not be taken too seriously. If the coefficients on all four lags are restricted to equal each other (i.e., short-term rates are regressed on the average of the past four quarters of inflation), the hypothesis that the coefficient equals zero is strongly rejected in both sample periods. The point estimates for the first-differenced versions of the regression (the first two rows of Panel B) are very close to those of the level versions. With first differences the hypothesis that all four coefficients equal zero is strongly rejected.

The combination of the AR(4) description of inflation and the regression of the three-month yield on inflation allow us to construct an impulse response of the three-month rate to a change in inflation. The responses for each sample period are displayed with solid lines in Figure 1. In both samples, the three-month yield increases immediately by about 10 basis points in reaction to a one percentage point increase in quarterly inflation. The yield continues to increase over the next few quarters.

A simple comparison of the solid lines in Panels A and B suggests that interest rates were more responsive to inflation in the earlier period than in the later period. This conflicts with the view (discussed in more detail in Section 5.2) that the Fed has been more aggressive in reacting to inflation since the early 1980s. The reason for the apparent discrepancy lies in the focus here on quarterly inflation instead of annual inflation commonly used in the Taylor rule literature. As Table 1 documents, inflation was much more persistent in the early sample than in the later sample. The point estimates from this table imply that in the early sample, a one percentage point shock to quarter t's inflation corresponded to a 72 basis point shock to the annual inflation rate measured beginning in quarter t. The corresponding shock to annual inflation in the later sample is only 51 basis points. Thus when measured in terms of shocks to annual inflation, short-term rates were less responsive to inflation in the early period than in the later period.

Armed with the impulse response of short-term rates to quarterly inflation shocks, we can calculate the expected change in the average three-month yield over the next five years.

(79 observations; recall that the last interest rate observation is missing owing to its definition).

Absent a change in term premia, this should correspond to the change in the yield on a five-year bond. The implied "constant term premia" responses are plotted in Figure 1 with dotted lines. The shape of this response differs from the shape of the response of the three-month yield because forward-looking investors anticipate the delayed response of the short-term yield.

The actual response of the five-year yield, based on the regression results in the third and fourth rows of Panel B, is displayed with dashed lines in Fig. 1. It differs substantially from the constant term premia response. In the early sample the actual response is much weaker than the constant term premia response. In fact, it is statistically indistinguishable from zero in either level form or first-difference form. In the later sample the actual response exceeds the constant term premia response. The consistent pattern across the two samples is that the long-term bond yield reacts gradually to the change in inflation. In this sense, the time pattern of the long-term bond yield is similar to that of the short-term yield.

What accounts for the apparent delayed reaction of the long-term bond yield? One possibility is that it is just an artifact of sampling error in the point estimates. To test this hypothesis we need to estimate these equations jointly. We also want to exploit information from other points on the yield curve. Another possibility is that the estimated dynamics are accurate, investors know these dynamics, and the curious behavior of the long-term yield is produced by complicated dynamics of the price of interest rate risk. Both of these possibilities are formally investigated in the next section using the partial term structure model of Section 2. I defer discussion of other possible explanations until these two have been examined.

5 Estimates from a partial term structure model

In this section I use the framework presented in Section 2 to examine the relation between inflation and the term structure. The vector of observed macroeconomic variables is

$$m_t = \begin{pmatrix} \pi_t & \pi_{t-1} & \pi_{t-2} & \pi_{t-3} \end{pmatrix}$$

where π_t is the PCE inflation rate. Quarter-t yields on four zero-coupon bonds are observed. (As in the previous section, this is actually the yield as of the end of the first month of quarter $t+1$.) Their constant maturities are three months, one year, three years, and five years. Yield data are from CRSP.

There are 13 parameters to estimate. Five of these correspond to the AR(4) description of inflation from (3). They are the scalar c and the four-element vector F_1. There are also four elements of the vector δ_1 in (1). (The scalar δ_0 in this equation could also be estimated but plays no role in the model.) Finally, there are four elements in the vector λ that represent the difference between the physical and equivalent martingale AR(4) dynamics of inflation in (18).

To put this model in context, it is helpful to compare it to two alternative models in the literature. Ang and Piazzesi (2003) model the term structure using two macro variables—measures of inflation and real activity—and three latent variables. By including both inflation and real activity, their approach is closer in spirit to Taylor rule regressions. I decided to focus exclusively on inflation because including another macro factor adds a large number of free parameters to the model. The partial term structure model with four lags of two macro variables has 42 free parameters. Since there are only 79 observations of quarterly data in the 1983—2003 period after dropping necessary lags and leads, some parameter restrictions are necessary to draw any reliable statistical inferences from the data. Many restrictions may seem sensible. For example, Ang and Piazzesi allow the price of interest rate risk to depend on current inflation and output, but not lagged inflation and output. This seemingly innocuous assumption would set twelve of the parameters to zero. But as we will see, risk premia appear to fluctuate substantially with lags of inflation; the assumption may be sensible but it is contradicted by the data. A nice feature of the term structure framework presented in this paper is that we do not require that the variables in m_t capture all, or even much, of the dynamics of the term structure. Therefore excluding information about real activity does not invalidate the modeled relation between inflation and the term structure.

Ang and Bekaert (2003) use a no-arbitrage, regime-switching framework to describe the joint dynamics of inflation and the term structure. The model includes a single inflation factor (current inflation), two latent factors, and either one or two variables characterizing the regime. The small number of inflation-related state variables in their model limits its use in investigating the questions addressed here. Yet as they note, the case for regime-switching is compelling. The evidence presented later in this section certainly supports their view. Unfortunately, tractable bond pricing in a regime-switching framework requires a number of restrictions on the nature of the regime switching; not all of the components of the dynamics are allowed to switch regimes. The requirement of tractability leads to a variety of nonnested regime-switching models. (Compare, for example, Ang and Bekaert with Dai, Singleton, and Yang (2003)). In this paper I do not attempt to extend the partial term structure framework to a regime-switching setting. Instead, I estimate the (one-regime) model over two separate time periods. Differences between the two estimated models can help guide the further development of regime-switching term structure models. In particular, we will see that the response of nominal short-term yields to inflation shocks has varied substantially over time.

Estimation is with GMM. I use 21 moment conditions. Five are moment conditions of the AR(4) description of inflation from (26). For each of the four Treasury securities there are an additional four moment conditions from (28) that represent the ability of differenced inflation to explain differenced yields.[8] Stack the moment conditions for quarter t in the vector f_t. For T observations of this vector, the parameter estimates solve the problem

$$\min J = T g_T' W g_T, \quad g_T = \frac{1}{T} \sum_{i=1}^{T} f_t, \tag{30}$$

for some weighting matrix W.

I use two iterations of GMM. For the first iteration, the weighting matrix is the inverse of the sample covariance matrix of the moments evaluated at "OLS/risk-neutral" parameters. This means that c and F_1 are taken from the OLS estimation results in Table 1, the vector δ_1 is taken from Panel B of Table 2, and λ is set to zero. The resulting GMM parameter estimates are then used to construct an asymptotically efficient weighting matrix and the parameters are estimated again. The covariance matrices are estimated using the robust method of Newey and West (1987a) with four moving average lags to account for the negative serial correlation in first differences of yields. I demean the moment vectors prior to calculation of the covariance matrices.

The solution to (30) requires nonlinear optimization. To find the global minimum, I randomly generated 100 starting values. For each starting value, I used Simplex (IMSL routine dumpol) to get in a well-behaved neighborhood of the parameter estimates. I then used a derivative-based algorithm (IMSL routine dbconf) to improve the accuracy of the estimates.

5.1 Results for 1953 through 1982

Table 3 displays the results of estimating the model over the sample period 1953Q1 through 1982Q4. Because the first year of data is reserved for lags, there are 116 quarterly observations. Panel A reports the parameter estimates and Panel B reports overidentifying tests of the model. The parameter estimates of inflation dynamics (F_1) are close to the corresponding estimates in Table 1, although the estimates here are more precise. However, the estimated reaction of the short rate to inflation is much more muted here than in the corresponding

[8] I experimented with the instrumental variables version of this moment condition, as discussed in the text after (28). When m_t is used as the instrument vector, the model's parameter estimates do not correspond to our intuition about the values of these parameters. In particular, the elements of δ_1 are generally negative, indicating that short-term rates fall when inflation rises. I therefore report results using (28) instead of its IV version.

regression estimates in Table 2. This point is best made visually. Figure 2 is the model-based counterpart to Figure 1. The solid line in Panel A is the expected time-path of the change in the three-month yield given a one percentage point shock to inflation, based on the point estimates in Panel A of Table 3. It is much closer to zero than is the corresponding line in Figure 1.

The dotted line in the figure is the expected time-path of change in the five-year yield implied by the model, which takes into account the response of the price of interest rate risk. For ease of comparison, the figure also displays the regression-based response of the five-year yield that was discussed in the context of Figure 1. The model's point estimates imply that the five-year yield is fairly insensitive to fluctuations in inflation. This is consistent with the regression evidence in Table 2 that could not allow us to reject the hypothesis that the five-year yield did not react to inflation. The reason for this insensitivity (other than the muted effect of inflation on short-term yields) is that bond risk premia move inversely with inflation. This is seen in Panel A of Figure 3. A one percentage point increase in inflation lowers quarterly excess returns to the five-year yield by about ten basis points per quarter over the next few years.

The estimated model effectively implies that the slow adjustment of long-term yields to inflation over the 1953—1982 period is a consequence of sampling error. Based on the model's parameter estimates, the reaction of long-term yields to inflation should be initially small and die out slowly over time. The $\chi^2(8)$ test of the adequacy of the model (the first row in Panel B) does not come close to rejecting the model, indicating that the slow adjustment of yields observed in the data is not statistically strong enough to overturn this interpretation of the evidence.

The sensitivity of expected excess returns to inflation is statistically very strong. The second row of Panel B reports that the restriction $\lambda = 0$ is overwhelmingly rejected by the likelihood ratio test of Newey and West (1987b). Moreover, risk premia are more closely associated with lagged inflation than current inflation. A positive shock to inflation in quarter t raises expectations of future short-term yields and has no immediate effect on expected excess returns. Nonetheless the quarter-t response of long-term yields is small because investors understand that risk premia will soon fall and remain lower for the next few years. If we force all action in risk premia to work through current inflation (formally, only the first element of λ is nonzero), the model's performance is substantially worse. The third row of Panel B reports that this functional form of λ is overwhelmingly rejected in favor of the more general form.

On balance, this partial model provides an economically sensible, statistically robust description of the relation between inflation and the term structure the 1953—1982 period.

Unfortunately, the model's performance with more recent data is not as favorable.

5.2 Results for 1983 through 2003

There is considerable evidence the joint dynamics of inflation and the term structure changed regimes in the early 1980s. Clarida et al. (2000) argue that since the appointment of Volcker and Greenspan, the Fed has responded more aggressively to inflation. Their view, based on Taylor-rule regressions, is corroborated in a Bayesian estimation setting by Goto and Torous (2003). I take another look at this issue by estimating the partial term structure model over the period 1983 through 2003. Because the first year of data is reserved for lags and the last observation of r_t is 2003Q3, there are 79 quarterly observations.

The results of the model are summarized in Table 4 and in the second panels of Figures 2 and 3. I will focus on the visual evidence. The solid line in Panel B of Figure 2 indicates that the three-month bill yield increases by more than 25 basis points for every one percentage point increase in quarterly inflation. The peak response occurs about one year after the shock.[9] Note that in contrast to the earlier period, here the implied time-path of short-term rates is quite close to the corresponding time-path displayed in Figure 1 that is based on regression evidence. The dotted line in Figure 2, which displays the implied response of the five-year bond yield, is also similar to the response based on regression evidence. The model reproduces the slow reaction of long-term yields to the inflation shock through complicated dynamics of risk premia. The χ^2 test of the overidentifying restrictions reported in Table 4 does not come close to rejecting the model.

Yet the way that the model fits the data is economically unsatisfactory. Expected returns to long-term bonds fluctuate dramatically in response to inflation. The evidence is in Panel B of Figure 3. For concreteness, consider the implications of this figure given a quarter-t positive shock of one percentage point in the inflation rate. The shock immediately lowers the expected quarterly excess return of a five-year bond by about 80 basis points. Over the next two quarters the expected excess return rises by 120 basis points, so that by $t+2$ it is 40 basis points above pre-shock value. It bounces around above this pre-shock value for a few quarters and dies off after a couple of years. The model produces this variation in expected

[9]This response may appear to be too small, given the typical Taylor rule evidence that short-term rates have risen more than one-for-one with inflation in the Volcker-Greenspan era. There are four reasons for the discrepancy. First, I use quarterly inflation instead of annual inflation. Second, I use inflation calculated from the PCE deflator rather than the GDP deflator used in Clarida et al. (2000) and Rudebusch (2002). The former measure is more volatile than the latter. Third, I do not include the output gap. Over this sample period, inflation and the output gap (actual minus potential GDP as a fraction of potential GDP) are negatively correlated, and the Taylor rule evidence is that short-term rates are positively associated with both inflation and the output gap. Finally, I measure interest rates at the end of the first month of the subsequent quarter rather than as an average of rates during the quarter.

excess returns because it allows the price of interest rate risk to depend on both current and lagged inflation. The restriction that risk premia respond only to current inflation is strongly rejected (the third row of Panel B in the table).

Expected excess bond returns behave this way because the maintained hypothesis of the model is that investors foresee a sharp increase in long-term bond yields from $t+1$ to $t+3$ that is subsequently reversed. To accept these fluctuations, investors in equilibrium must be willing to face an expected capital loss at first, then reap substantial gains. Although there is always some general equilibrium model that can support arbitrary dynamics of the short-term rate and the price of risk, it is difficult to envision a sensible economic setup that supports these dynamics.

It is easier to interpret the results as a consequence of investors underestimating the magnitude of the delayed reaction of short-term interest rates to inflation. This conclusion is supported by direct evidence on investors' expectations of interest rate dynamics. I now turn to this evidence.

5.3 Independent evidence on forecast errors

Beginning in 1981, the *Survey of Professional Forecasters* has collected forecasts of three-month Treasury bill yields. In roughly the middle of quarter t, participants are asked for their forecasts of the average bill rate for each quarter from t to $t+4$. At the time the responses are collected, participants know the average price level during quarter $t-1$. If forecasts are rational, forecast errors (r_{t+k} less forecasters' expectation of r_{t+k} formed in the middle of quarter t) will be orthogonal to the change in inflation from quarter $t-4$ to quarter $t-1$. But if investors underestimate the magnitude of the relation between inflation and future interest rates, forecast errors will be positively correlated with this change in inflation.

There is another way to think about investors' errors in forecasting interest rates. The slow adjustment of interest rates to inflation induces positive serial correlation in first-differenced interest rates. If investors do not fully recognize this slow adjustment, forecast errors will be positively correlated with the lagged change in short-term interest rates.

I test both hypotheses using the following regression.

$$r_{t+k} - E_{t-(1/2)}(r_{t+k}) = b_0 + b_1(\pi_{t-1} - \pi_{t-4}) + b_2(E_{t-(1/2)}(r_t) - r_{t-1}) + e_{t+k} \quad (31)$$

The notation $E_{t-(1/2)}(\cdot)$ refers to an expectation formed in the middle of quarter t. It would be cleaner to use $r_t - r_{t-1}$ as the second regressor, but r_t is not known at time $t-(1/2)$. Following common practice, I use the median forecast from the survey as a proxy for the true expectation. The three-month bill yield is measured by the average three-month bill yield during quarter t, to conform with the rate that forecasters are asked to predict. For comparability with the results presented in the previous subsection, I estimate this regression using quarterly data from 1983 through 2003. Because the first year of data is reserved for lags, the number of observations is $80-k$. Asymptotic t statistics are adjusted for generalized heteroskedasticity and k lags of moving-average residuals using the technique of Newey and West (1987a).

The results are displayed in Table 5. The first set of regressions focuses on the forecast power of lagged inflation. By itself, the lagged change in inflation has a statistically weak power to predict the forecast errors. Only one of the four reported t-statistics exceeds two. But from an economic perspective this forecast power is striking. Over this sample period, the actual predictive power of inflation for future interest rates greatly exceeded the predictive power that forecasters assigned to inflation. The easiest way to see this is to decompose (31) into two separate regressions: A regression of r_{t+k} on lagged inflation and a regression of $E_{t-(1/2)}(r_{t+k})$ on lagged inflation. Because the left-hand-side of (31) is the difference between these two variables, the coefficient estimates for (31) are identically equal to the difference in the coefficient estimates of the separate regressions. Consider, for example, the relation between lagged inflation and the two-quarter-ahead short-term rate. The coefficient in Table 5 is 0.098, which equals the difference between 0.127 (the coefficient with r_{t+2} on the left) and 0.029 ($E_{t-(1/2)}(r_{t+2})$ on the left). In other words, the responsiveness of short-term rates to inflation in the sample was more than four times what forecasters expected.

The rest of the results in Table 5 document that professional forecasters failed to predict the positive serial correlation of changes in short-term rates that characterized the 1983 through 2003 period. The coefficients on the lagged change in the three-month yield are strongly positive, both economically and statistically. (They also overestimated the level of yields; the constant terms in the regressions are strongly negative.) This forecasting failure captures much of the forecasting failure associated with lagged inflation.

5.4 Interpreting the evidence

The peculiar pattern of implied bond risk premia, combined with the nature of forecasters' errors in predicting future short-term rates, indicates that the one-regime, no-arbitrage model is an inappropriate description of inflation and interest rate dynamics during 1983 through 2003. Since the financial industry devotes considerable resources to forecast interest rate dynamics, we should probably throw out the one-regime part of the model rather than the no-arbitrage part. We can then attribute the forecast errors as rational errors made by investors as they learn about the new regime. But casual inference does not easily lead to an intuitive description of the regime that investors *thought* they were in.

The differences between the 1953—1982 and 1983—2003 periods suggest that regime-switching models need to incorporate time-varying sensitivity of short-term rates to inflation. (Formally, this means that different regimes have different values of δ_1.) This conclusion is consistent with Clarida et al. (2000) and Goto and Torous (2003). But the evidence here contradicts the view of Goto and Torous that after the early 1980s, investors knew that they were in a new regime with well-understood dynamics.

6 Concluding comments

This paper makes two contributions to the term structure literature. The first contribution is a methodological framework to investigate the relation between the term structure and other, non-yield variables. The framework imposes no-arbitrage without requiring a complete description of the term structure's dynamics. Therefore it can be used to describe the dynamics of expected returns to bonds conditional on the non-yield variables. The framework is simple to implement with GMM.

The second contribution is to use this framework to interpret the slow response of both short-term and long-term bond yields to inflation. Over the 1983—2003 period, shocks to quarterly inflation were followed by strong movements in short-term yields over the next year. Given the assumption that investors understood these dynamics, I find that they willingly accepted large swings in expected excess returns to long-term bonds. Although this behavior is consistent with no-arbitrage, the necessary dynamics of the price of interest rate risk are so peculiar that alternative interpretations of the evidence are likely to be more palatable. One reasonable interpretation is that investors mistakenly believed that some other regime was operating during much of this period. To investigate this hypothesis, the framework here should be extended to multiple regimes for the dynamics of the non-yield variables.

References

Ang, Andrew, and Geert Bekaert, 2003, The term structure of real rates and expected inflation, Working paper, Columbia Business School.

Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745-787.

Ang, Andrew, Monika Piazzesi, and Min Wei, 2003, What does the yield curve tell us about GDP Growth?, Working paper, Columbia Business School.

Bernanke, Ben S., Mark Gertler, and Mark Watson, Systematic monetary policy and the effects of oil price shocks, 1997, *Brookings Papers on Economic Activity* 1997, 91-142.

Buraschi, Andrea, and Alexei Jiltsov, 2004, Time-varying inflation risk premia and the expectations hypothesis: A monetary model of the Treasury yield curve, *Journal of Financial Economics*, forthcoming.

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

Clarida, Richard, Jordi Galí, and Mark Gertler, 2000, Monetary policy rules and macroeconomic stability: evidence and some theory, *Quarterly Journal of Economics* 115, 147-180.

Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Dai, Qiang, Kenneth J. Singleton, and Wei Yang, 2003, Regime shifts in a dynamic term structure model of U.S. Treasury bond yields, Working paper, Stern School of Business.

Dewachter, Hans, and Marco Lyrio, 2002, Macro factors and the term structure of interest rates, Working paper, Catholic University of Leuven.

Dewachter, Hans, Marco Lyrio, and Konstantijn Maes, 2002, Estimation of a joint model for the term structure of interest rates and the macroeconomy, Working paper, Catholic University of Leuven.

Diebold, Francis X., Glenn D. Rudebusch, and S. Boragan Aruoba, 2003, The macroeconomy and the yield curve: A nonstructural analysis, Working paper, The Wharton School.

Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405–443.

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

English, William B., William R. Nelson, and Brian P. Sack, 2003, Interpreting the significance of the lagged interest rate in estimated monetary policy rules, *Contributions to Macroeconomics* 3, Article 5.

Evans, Charles L., and David A. Marshall, 1998, Monetary policy and the term structure of nominal interest rates: evidence and theory, *Carnegie-Rochester Conference Series on Public Policy* 49, 53-111.

Evans, Charles L., and David A. Marshall, 2002, Economic determinants of the nominal Treasury yield curve, Working paper, Federal Reserve Bank of Chicago.

Fama, Eugene F., 1975, Short-term interest rates as predictors of inflation, *American Economic Review* 65, 269-282.

Goto, Shingo, and Walter Torous, 2003, The conquest of U.S. inflation: Its implications for the Fisher hypothesis and the term structure of nominal interest rates, Working paper, University of South Carolina.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moment estimators, *Econometrica* 50, 1029-1053.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimators of nonlinear rational expectations models, *Econometrica* 50, 1269-1286.

Hördahl, Peter, Oreste Tristani, and David Vestin, 2002, A joint econometric model of macroeconomic and term structure dynamics, Working paper, European Central Bank.

Leeper, Eric M., Christopher A. Sims, and Tao Zha, 1996, What does monetary policy do?, *Brookings Papers on Economic Activity* 1996, 1-63.

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

Newey, Whitney K., and Kenneth D. West, 1987b, Hypothesis testing with efficient method of moments estimation, *International Economic Review* 28, 777-787.

Piazzesi, Monika, 2003, Bond yields and the Federal Reserve, *Journal of Political Economy*, forthcoming.

Rudebusch, Glenn D., 2002, Term structure evidence on interest rate smoothing and monetary policy inertia, *Journal of Monetary Economics* 49, 1161-1187.

Rudebusch, Glenn D., and Tao Wu, 2003, A macro-finance model of the term structure, monetary policy, and the economy, Working paper, Federal Reserve Bank of San Francisco.

Sack, Brian, and Volker Wieland, 2000, Interest-rate smoothing and optimal monetary policy: A review of recent empirical evidence, *Journal of Economics and Business* 52, 205-228.

Taylor, John B., 1993, Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy* 39, 195-214.

Vasicek, Oldrich A., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.

Sample period	Lag 1	2	3	4	SEE
1953–1982	0.647	0.218	0.062	−0.002	1.324
	(0.104)	(0.101)	(0.114)	(0.118)	
1983–2003	0.396	0.095	0.273	0.029	0.948
	(0.095)	(0.118)	(0.147)	(0.101)	

Table 1: An AR(4) description of quarterly inflation

Inflation is measured by the quarterly change in the PCE chain-weighted price index. Estimation of the AR(4) is with OLS. Standard errors, adjusted for generalized heteroskedasticity, are in parentheses. The column labeled "SEE" reports the standard error of the estimate.

Panel A. Level form with AR(1) residuals

Sample Period	Maturity	Lag 0	Lag 1	Lag 2	Lag 3	Joint test [p-val]
1953–1982	3 month	0.084	0.080	0.192	0.060	2.318
		(0.066)	(0.068)	(0.068)	(0.065)	[.061]
1983–2003	3 month	0.088	0.194	0.107	0.136	2.486
		(0.062)	(0.070)	(0.070)	(0.061)	[.051]
1953–1982	5 year	−0.029	0.054	0.061	0.014	1.000
		(0.041)	(0.043)	(0.043)	(0.041)	[.411]
1983–2003	5 year	0.120	0.064	0.130	0.075	1.152
		(0.072)	(0.080)	(0.081)	(0.071)	[.339]

Panel B. First differences

Sample Period	Maturity	Lag 0	Lag 1	Lag 2	Lag 3	Joint test [p-val]
1953–1982	3 month	0.074	0.058	0.169	0.035	19.487
		(0.058)	(0.098)	(0.056)	(0.052)	[.001]
1983–2003	3 month	0.105	0.182	0.130	0.143	10.778
		(0.049)	(0.058)	(0.072)	(0.061)	[.029]
1953–1982	5 year	−0.025	0.056	0.061	0.012	3.666
		(0.045)	(0.044)	(0.043)	(0.035)	[.453]
1983–2003	5 year	0.128	0.038	0.143	0.073	11.512
		(0.078)	(0.092)	(0.086)	(0.049)	[.021]

Table 2: Regressions of bond yields on contemporaneous and lagged inflation

Bond yields are regressed on lags zero through three of inflation. The data are quarterly. Bond yields are from CRSP and inflation is the log change in the PCE chain-weighted price index. The regressions in Panel A are estimated with Hildreth-Lu and those in Panel B are estimated with OLS. Standard errors are in parentheses. In Panel B they are adjusted for generalized heteroskedasticity and four lags of moving average residuals. The column labeled "Joint test" reports the value of the statistic that the coefficients on inflation are all zero. In Panel A this is an F statistic and in Panel B this is a $\chi^2(4)$ statistic.

Panel A. Parameter estimates

Parameter	Element of the vector 1	2	3	4
δ_1	0.106	−0.064	0.101	0.023
	(0.020)	(0.023)	(0.021)	(0.030)
F_1	0.678	0.221	0.023	0.017
	(0.070)	(0.079)	(0.073)	(0.088)
λ	−0.142	1.369	−0.904	1.320
	(0.149)	(0.164)	(0.155)	(0.148)

Panel B. Tests of overidentifying restrictions

Model	LR statistic	d.f.	p-val
Unrestricted	3.760	8	0.878
$\lambda = 0$	43.170	4	0.000
$\lambda_{[2:4]} = 0$	28.582	3	0.000

Table 3: Estimates of a no-arbitrage model of inflation and the term structure, 1953-1982

The table reports parameter estimates of the partial term structure model described in Section 5. In the model, The three-month yield is the sum of a constant and δ_1 times lags zero through three of quarterly inflation. Quarterly inflation follows an AR(4) with parameters F_1. Under the equivalent martingale measure the AR(4) of inflation has parameters $F_1 - \lambda$. The data are quarterly, from 1953Q1 through 1982Q4. Estimation is with GMM. Standard errors are in parentheses. The likelihood ratio test statistics for the restricted models in Panel B are tests relative to the unrestricted model.

Panel A. Parameter estimates

Parameter	Element of the vector 1	2	3	4
δ_1	0.072	0.226	0.163	0.120
	(0.024)	(0.046)	(0.051)	(0.034)
F_1	0.343	0.122	0.369	-0.051
	(0.068)	(0.063)	(0.102)	(0.086)
λ	1.542	0.055	-1.126	-0.583
	(0.165)	(0.175)	(0.129)	(0.267)

Panel B. Tests of overidentifying restrictions

Model	LR statistic	d.f.	p-val
Unrestricted	7.415	8	0.493
$\lambda = 0$	16.810	4	0.002
$\lambda_{[2:4]} = 0$	9.078	3	0.028

Table 4: Estimates of a no-arbitrage model of inflation and the term structure, 1983-2003

The table reports parameter estimates of the partial term structure model described in Section 5. In the model, The three-month yield is the sum of a constant and δ_1 times lags zero through three of quarterly inflation. Quarterly inflation follows an AR(4) with parameters F_1. Under the equivalent martingale measure the AR(4) of inflation has parameters $F_1 - \lambda$. The data are quarterly, from 1983Q1 through 2003Q4. Estimation is with GMM. Standard errors are in parentheses. The likelihood ratio test statistics for the restricted models in Panel B are tests relative to the unrestricted model.

Quarters ahead (k)	Constant	Lagged change in inflation	Lagged change in interest rates	SEE	R^2
1	−0.168	0.063		0.475	0.025
	(−2.70)	(1.48)			
2	−0.356	0.098		0.832	0.019
	(−2.68)	(1.33)			
3	−0.565	0.160		1.135	0.028
	(−2.64)	(1.65)			
4	−0.724	0.219		1.408	0.034
	(−2.44)	(2.11)			
1	−0.150	0.032	0.343	0.456	0.111
	(−2.65)	(0.77)	(2.92)		
2	−0.356	0.047	0.551	0.806	0.093
	(−2.67)	(0.65)	(2.24)		
3	−0.523	0.091	0.752	1.098	0.102
	(−2.66)	(0.96)	(2.40)		
4	−0.681	0.145	0.792	1.377	0.088
	(−2.49)	(1.35)	(2.16)		

Table 5: Regressions of errors in T-bill yield forecasts on lagged changes in inflation and interest rates, 1983-2003

The table reports results of estimating equation (31) in the text, in which the error in the k-quarter ahead forecast of the three-month Treasury bill yield is regressed on lagged nine-month change in quarterly inflation and on the lagged change in the three-month Treasury bill yield. Forecasts are median forecasts from the Survey of Professional Forecasters. The sample period is 1983 through 2003. Estimation is with OLS. The column labled “SEE” reports standard errors of the estimate. Asymptotic t statistics are in parentheses. They are adjusted for generalized heteroskedasticity and k lags of moving-average residuals.

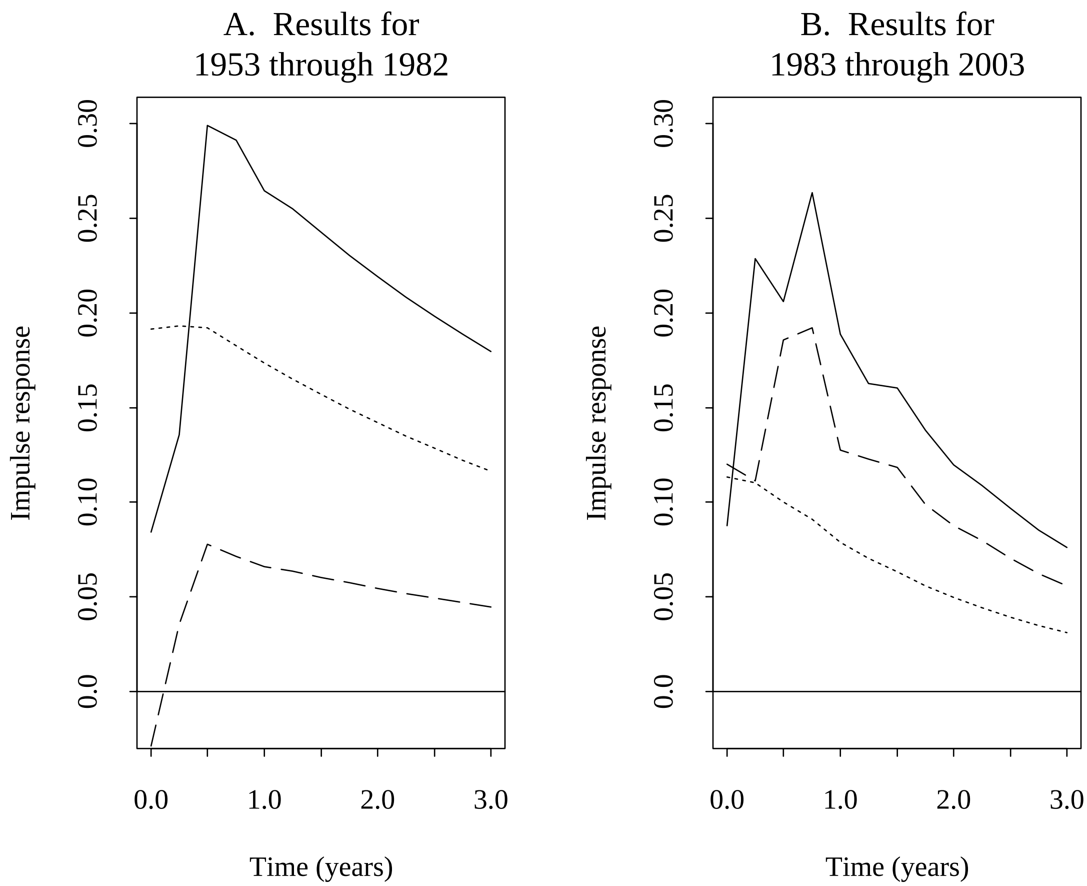

Figure 1: The reaction of bond yields to inflation

Quarterly inflation is fit to an AR(4) and three-month Treasury bill yields are regressed on lags zero through three of inflation. The solid line is the implied impulse response of the three-month bill yield to a one percentage point shock to quarter-t inflation. This response is used to compute the theoretical response of a five-year bond yield with the assumption that risk premia are constant (the dotted line). The actual response of the five-year Treasury yield is given by the dashed line.

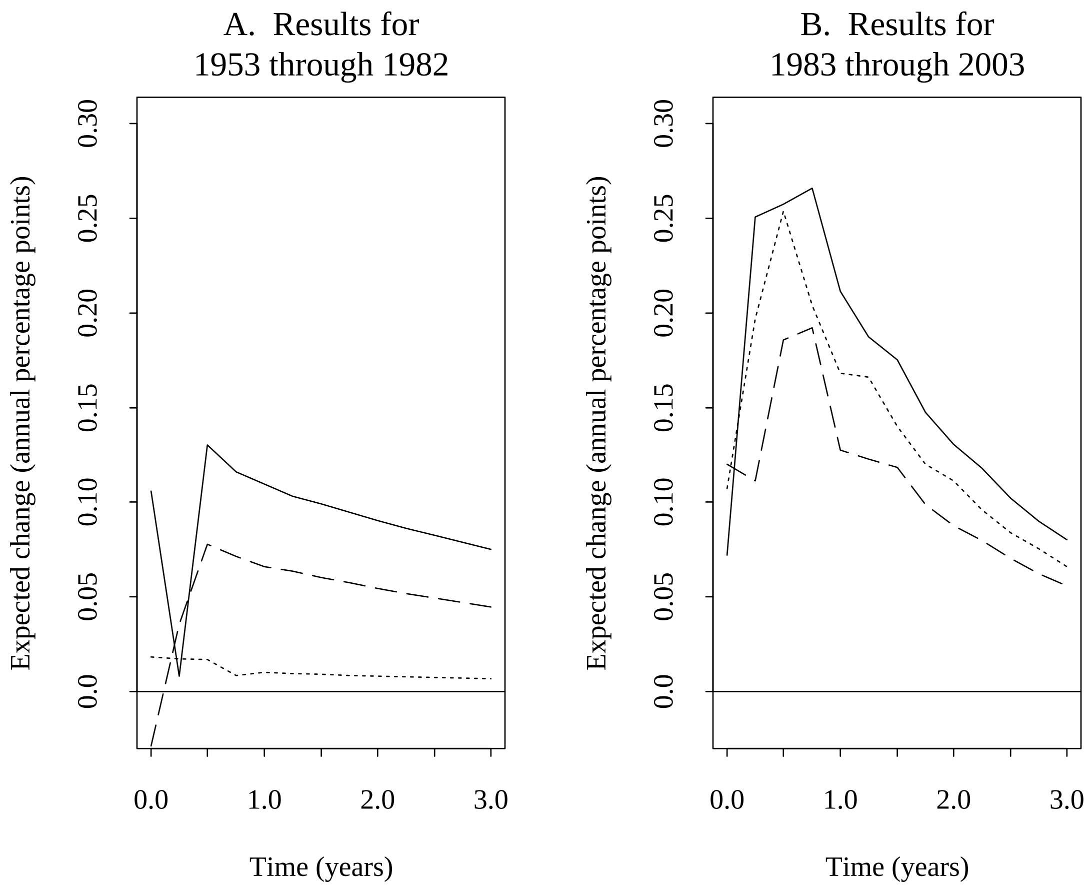

Figure 2: The reaction of bond yields to inflation, based on estimates of a partial term structure model

The panels display the time-paths of changes in expected future bond yields, given a percentage point shock to quarter-t inflation. The solid line is the expected change in the three-month yield. The dotted line is the expected change in the five-year yield. Both are computed from the point estimates of the partial term structure estimated in Section 5. The dashed line is taken from Figure 1.

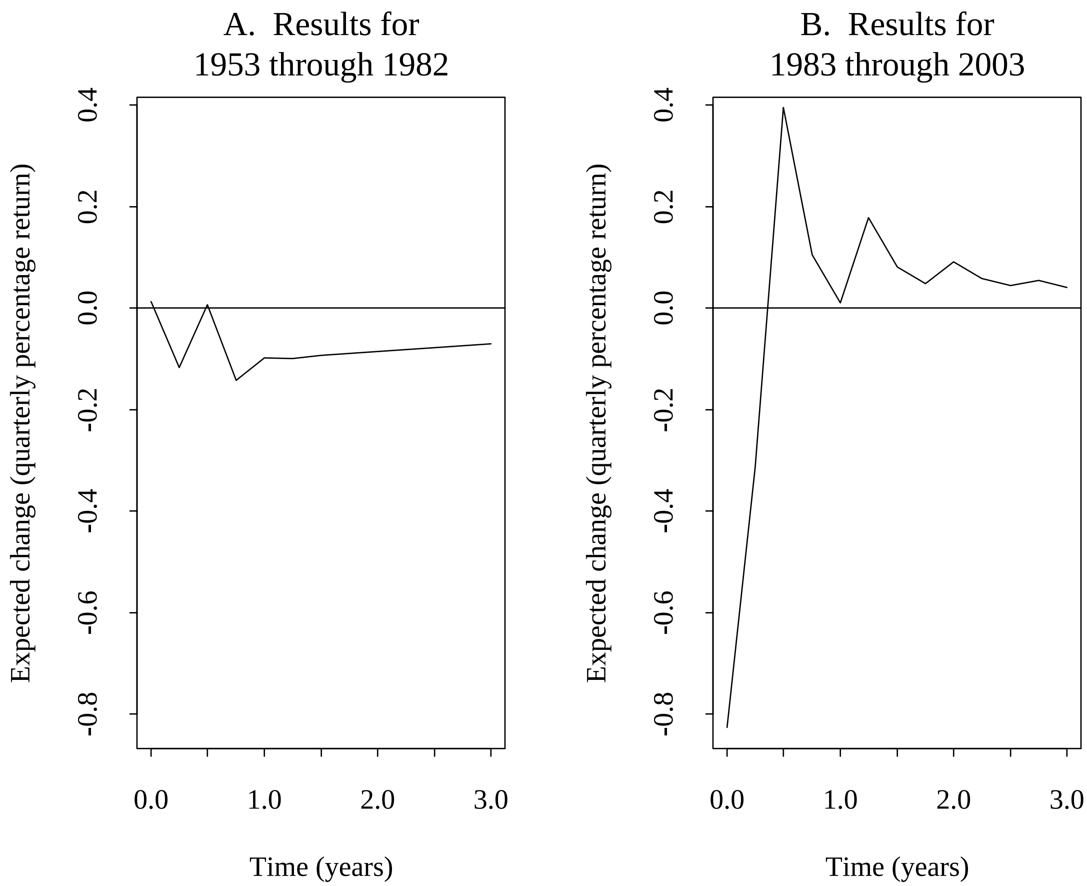

Figure 3: The reaction of expected excess bond returns to inflation, based on estimates of a partial term structure model

The panels display the time-path of the change in the expected quarterly excess log return to a five-year bond, given a one percentage point shock to quarter-t inflation. They are computed from the point estimates of the partial term structure model estimated in Section 5.