Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models*

Glenn D. Rudebusch† Tao Wu‡

November 2004
Revised June 2005

Abstract

This paper examines a recent shift in the dynamics of the term structure and interest rate risk. We first use standard yield-spread regressions to document such a shift in the U.S. in the mid-1980s. Then, for the pre- and post-shift subsamples, we estimate dynamic, affine, no-arbitrage models, which exhibit a significant difference in behavior that largely can be attributed to changes over time in the pricing of risk associated with a "level" factor. Finally, we suggest a link between the shift in term structure behavior and changes in the risk and dynamics of the inflation target as perceived by investors.

*Golnaz Motiey provided excellent research assistance. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of San Francisco. The previous title of an earlier version of this paper was "The Recent Shift in Term Structure Behavior from a No-Arbitrage Macro-Finance Perspective."

†Federal Reserve Bank of San Francisco; www.frbsf.org/economists/grudebusch; Glenn.Rudebusch@sf.frb.org.

‡Federal Reserve Bank of San Francisco; Tao.Wu@sf.frb.org.

1. Introduction

During the past few decades, the U.S. economy has undergone an important transformation that has likely altered the nature of uncertainty and risk in the economy as well as investors' attitudes and pricing of that risk. A key aspect of this transformation is the precipitous decline in overall macroeconomic volatility: Since the middle of the 1980s, the volatility of real GDP growth has been about 35 percent lower than earlier in the postwar period (as noted by Kim and Nelson 1999 and McConnell and Perez-Quiros 2000). Several factors may underlie this "Great Moderation" in economic fluctuations.[1] For example, improved economic policies may have helped stabilize the economy; indeed, many have argued that the conduct of U.S. monetary policy improved dramatically during the mid-1980s, helping to usher in this period of diminished output volatility as well as remarkably low and stable inflation. Alternatively, the recent quiescence in real activity and inflation may largely reflect good luck—that is, a temporary run of smaller economic shocks. Other potential factors include non-policy changes in the dynamics of the economy arising from, for example, improved inventory management or a greater share in aggregate output accounted for by the relatively stable service sector. Finally, the development of deeper and more integrated financial markets may also have played an important role both in damping the magnitude of economic fluctuations and in mitigating their effects on investors. Given such dramatic shifts in the economic environment, a change in the behavior of the term structure of interest rates, and especially in the size and dynamics of risk premiums, would hardly be surprising.

This paper examines how the dynamics of the term structure and interest rate risk may have changed over time. We use affine, no-arbitrage, asset pricing models of the type popular in the finance literature to investigate the recent shift in the behavior of the term structure; however, our investigation is also informed by the above literature on the recent transformation of the U.S. economy and by consideration of the macroeconomic underpinnings of the term structure factors in finance models.[2] The payoff from this joint analysis is bi-directional as well. The macro-finance perspective helps illuminate the nature of the shift in the behavior of the term structure, highlighting in particular the importance of a shift in investors' views regarding the risk associated with the inflation goals of the monetary authority. In addition, the

[1] For references to the quickly growing literature on this topic, see Blanchard and Simon (2001) and Stock and Watson (2003).

[2] The connection between the macroeconomic and finance views of the term structure has been a very fertile area for recent research, including, for example, Piazzesi (2003), Diebold, Rudebusch, and Aruoba (2004), Hördahl, Tristani, and Vestin (2004), Rudebusch and Wu (2004), Wu (2001), Dewachter and Lyrio (2002), Duffee (2004), and Kozicki and Tinsley (2001).

shift in term structure behavior, as viewed using a no-arbitrage finance model, sheds light on the nature of recent macroeconomic changes. Specifically, if one assumes that the factors underlying recent changes in the *macroeconomy* also have left their imprint on the *yield curve*, the finance models suggest that more than just good luck was responsible for the recent macroeconomic transformation. Instead, a favorable change in economic dynamics, likely linked to a shift in the monetary policy environment, appears to have been an important element of the Great Moderation.

We begin our analysis in Section 2 with a simple empirical characterization of the recent shift in the term structure in the U.S.. For this purpose, we use regressions of the change in a long-term interest rate on the lagged spread between long and short rates. Following Campbell and Shiller (1991), such regressions have been widely used to test the expectations hypothesis of the term structure, which assumes that the risk or term premiums embedded in long rates are constant. We find—as have many others—that these tests often reject the expectations hypothesis; however, of more interest for our purposes is the apparent significant shift in the estimated coefficients from these regressions. Indeed, since the mid-1980s, there is much less evidence against the expectations hypothesis than before, which suggests a shift in risk pricing and in the properties of risk premiums.

We use these term structure regression results as a summary statistic for characterizing the changing empirical behavior of the term structure. Accordingly, the regression evidence is a useful first step to a more formal modeling perspective on the term structure change, which is provided in Section 3 using an estimated dynamic, affine, no-arbitrage model of bond pricing. The no-arbitrage model provides an obvious setting in which to examine changes in interest rate behavior and time-varying term premiums. Indeed, as demonstrated by Backus et al. (2001), Duffee (2002), and Dai and Singleton (2002), affine, no-arbitrage models with a rich specification of the dynamics of risk premiums are broadly consistent with the usual full-sample term structure regression results of the type obtained in Section 2. We conduct a similar consistency check between models and regression results, though from a somewhat different perspective. Namely, given our evidence of a significant shift in the term structure regression results, we estimate affine, no-arbitrage models for each of the two subsamples that are associated with the different regression results. We find a statistically significant difference between the two estimated bond pricing models. In addition, the subsample models are able to account for much of the disparity between the subsample term structure regression results, thus supporting the empirical characterization of structural change in Section 2.

Beyond merely documenting the recent change in term structure behavior through regression analysis and model estimates, we also consider the more difficult task of understanding and accounting for such time variation. In Sections 4 and 5, we illuminate the economic changes that may account for the shift in term structure behavior. We first use the estimated subsample no-arbitrage models to parse out whether a shift in the *dynamics* of risk or a shift in the *pricing* of that risk is likely the most important factor accounting for the shift in term structure behavior. We find that changes in the pricing of risk associated with a "level" factor are crucial determinants of the change in term structure behavior. We then try to provide an interpretation of this shift in terms of possible recent macroeconomic changes. For this purpose, we employ the macro-finance model of Rudebusch and Wu (2004) and link the recent shift in term structure behavior to changes in the risk and dynamics of the central bank's inflation target as perceived by investors.

At this point, it is perhaps useful to discuss other recent related research. There has been little analysis of the effects on asset pricing induced by the important structural shifts in the economy documented in the macroeconomics literature cited above. Indeed, the finance literature often treats the entire postwar period as a long homogenous sample. An exception to this practice is the literature on regime-switching models of interest rates, including, for example, Hamilton (1988), Gray (1996), Ang and Bekaert (2002), Bansal and Zhou (2002), and Dai, Singleton, and Yang (2003). These papers attempt to capture the postwar dynamics of interest rates with models that contain a succession of alternating regimes that are often linked informally to business cycles or interest rate policies. In contrast, we are interested in a single break in the behavior of the term structure, with our attention focused by the macroeconomic evidence that suggests the shift occurred during the middle to late 1980s. Also, following the macroeconomic evidence, we have no expectation that this change will be reversed (and we incorporate no pricing of further regime change risk). Of course, regime switching at a cyclical frequency could coexist with a single large shift in risk pricing as well, but our interest here is in the latter. Accordingly, our analysis is related to other work, including Watson (1999), who examined a shift in the unconditional volatility of interest rates, and Lange, Sack, and Whitesell (2003), who considered a change in the forecastability of short-term interest rates. However, in contrast to these analyses, we examine a shift in behavior of risk pricing using both simple regression indicators as well as formal dynamic bond pricing models.

2. Regression Evidence of a Term Structure Shift

In this section, to help guide our subsequent model-based analysis, we provide a simple empirical characterization of the recent shift in the behavior of the term structure. This characterization, which also provides a metric to assess the *extent* of any such shift, is based on a regression test of the expectations hypothesis that was popularized by Campbell and Shiller (1991).

To derive this regression test, consider the following decomposition of the yield of a pure discount bond into average expected future yields and a term premium $\mathrm{E}_t\theta_{m,t}$:

$$i_{m,t} = (1/m)\sum_{j=0}^{m-1} \mathrm{E}_t(i_{t+j}) + \mathrm{E}_t\theta_{m,t}, \tag{2.1}$$

where $i_{m,t}$ is the continuously compounded yield to maturity at time t of an m-month nominal zero-coupon bond with the notational simplification for the one-month rate of $i_t \equiv i_{1,t}$. Bekaert, Hodrick, and Marshall (1997a) derive equation (2.1) from a modern asset pricing equation and show that the term premium is a function of second- and higher-order conditional moments of the stochastic discount factor (or pricing kernel). If these moments vary over time, then so will the term premium. If not, then term premiums will be constant, the expectations hypothesis will hold, and changes in long-term rates will result only from changes in expected future short-term rates. In this case, from equation (1) we can obtain

$$mi_{m,t} - (m-1)i_{m-1,t+1} = i_t + \text{const.} + \sum_{j=1}^{m-1}(\mathrm{E}_t(i_{t+j}) - \mathrm{E}_{t+1}(i_{t+j})), \tag{2.2}$$

where the left-hand side is the one-month holding period return of a bond of maturity m and the right-hand side is the one-month short rate plus a constant premium plus an expectational term.[3] This expectational term represents the capital gains or losses resulting from revisions to expected future short rates made between periods t and $t+1$. With rational expectations, these revisions are unpredictable at time t, so they can be interpreted as a white noise error term. Equation (2.2) then leads naturally to the "long-rate regression" form of Campbell and Shiller

[3] The holding period return is the profit or loss from buying an m-period bond at time t and selling the same (aged) $(m-1)$-period bond at time $t+1$. If $b_{m,t}$ is the price of this m-period nominal bond, then the return is $b_{m-1,t+1}/b_{m,t}$, the log of which is the left-hand side of the equation.

(1991):

$$i_{m-1,t+1} - i_{m,t} = \alpha_m + \beta_m(i_{m,t} - i_t)/(m-1) + \varepsilon_{m,t}, \tag{2.3}$$

where α_m and β_m are maturity-specific regression intercept and slope coefficients, and $\varepsilon_{m,t}$ is the white noise expectational term (scaled by $1-m$). Under the expectations hypothesis, the estimated slope coefficient β_m will equal unity; that is, the term spread will be an optimal forecast of future change in the long rate (adjusted for a constant risk premium), so when the spread between long and short rates widens (narrows), the long rate should rise (fall) in the following period.

Deviations from the expectation hypothesis will push the slope coefficient away from one. In particular, as noted early on by Mankiw and Miron (1986), a time-varying term premium can drive the estimated β_m to zero or even to negative values as the resulting term spread reflects variation in expected risk premiums rather than in future rates. In our analysis below, we construct models in which the time variation in the term premium (or equivalently the conditional heteroskedasticity of the discount factor) is sufficient to generate the regression coefficients found in the data, which are often significantly less than one. However, we are not primarily interested in the slope coefficients as indicators of the expectations hypothesis; instead, we use them as simple summary statistics of term structure behavior, and we interpret shifts in these coefficients as indications that the term structure behavior has changed. Of course, the fact that so many researchers have focused so much effort on estimating these slope coefficients makes them of particular interest, but other simple metrics of term structure change could also be considered (as in Watson 1999 and Lange, Sack, and Whitesell 2003).

Table 1 collects estimates of the slope coefficient β_m in equation (2.3) over various samples for eight different long-rate maturities—each column uses a different maturity m. In each case, the underlying interest rate data are from end-of-month, zero-coupon U.S. Treasury securities. The original full-sample (1952-1987) estimates from Campbell and Shiller (1991) are shown at the top along with coefficient standard errors in parentheses. Estimates and standard errors from a more recent sample (1970-1995) from Dai and Singleton (2002) are shown directly below. These two sets of estimates are similar and representative of the literature. In particular, both sets of estimates are uniformly negative and decrease steadily as the maturity of the long rate increases—falling from about -0.3 for $m = 3$ to less than -4.0 at a long-rate maturity of 10 years.

The long-rate regression slope estimates from our full data sample, which runs from 1970 to 2002, are shown in the middle rows of Table 1. Despite differences in the sample range, our full-sample estimates match the earlier results of Campbell and Shiller and Dai and Singleton quite closely.[4] In particular, our full-sample estimates of the slope coefficients are uniformly negative and decline with maturity to almost -4.0 at the long end. The numbers in brackets below the standard errors are the p-values of the null expectations hypothesis that $\beta_m = 1$. These p-values indicate that for each of the nine regressions over our full sample the expectations null hypothesis can be rejected at the 5 percent significance level and often at the 1 percent level. It should be noted that—as in the remainder of this section—the reported standard errors and p-values are based on the usual asymptotic distributions (with a standard correction for heteroskedasticity). Questions have been raised in the literature about the appropriateness of such asymptotic distributions for inference in small samples; therefore, in the Appendix, we report monte carlo simulations that indicate that in this application the small-sample biases are not leading us astray.

We are primarily interested in regression results from shorter samples, and our strong prior—based on the shifts in the economy described in the introduction—is that the most likely potential breakdate for term structure behavior would occur around the middle or late 1980s. In particular, econometric evidence (e.g., Kim and Nelson 1999 and McConnell and Perez-Quiros 2000) suggests that a likely date for the start of reduced volatility in economic activity is 1984. In addition, there appears to have been an important shift in the conduct of monetary policy during the 1980s, perhaps triggered or reinforced by the appointment of Fed Chairman Alan Greenspan in late 1987. Of course, for pricing risk in real time, investors may have needed some time to learn about and assess the importance of these changes, which makes the choice of a breakdate somewhat indeterminate. We will examine a variety of potential breakdates below; however, for an initial look at the data with an a priori choice of a breakdate, the lower half of Table 1 provides estimates when the sample is split into an earlier "subsample A" that runs from 1970 through 1987 and a later "subsample B" that goes from 1988 to 2002. (This is the split suggested by the change in Fed Chairmen and conveniently supplies two subsamples of nearly

[4] A second difference across these estimates is the exact methodology used for creating the yields data—particularly in interpolating missing maturities and smoothing out idiosyncratic observations (e.g., Bliss 1997). Our data are unsmoothed Fama-Bliss yields data, kindly supplied by Robert Bliss, but we obtained qualitatively similar breakpoint results with smoothed Fama-Bliss data (the type of data used in Dai and Singleton 2002). Another difference is in approximating $i_{m-1,t+1}$ by $i_{m,t+1}$. We have the entire maturity set of yields, so we do *not* employ this approximation, but the other authors in Table 1 do apply it.

equal size.)

The long-rate regression results in the lower half of Table 1 show an interesting difference across the two subsamples. The slope estimates from the nine long-rate regressions are all negative in subsample A, as in the full sample, while they are predominately positive in the later subsample B. Furthermore, the expectations hypothesis is rejected in every subsample A regression, while it is rejected in only one subsample B regression (at the 3-month horizon). Note that this lack of rejection does not reflect inflated standard errors from a short sample. In fact, for each maturity, the standard errors from the subsample B regressions are smaller than the full-sample ones.

Evidence from a formal break test is given in the bottom line in Table 1, which shows the p-value at each maturity for a Chow-type F-test that the slope coefficient has not shifted between subsamples A and B.[5] Taken one maturity at a time, the evidence of a shift in the slope coefficient is decidedly mixed. For the three regressions using 6-, 9-, and 12-month long rates, the evidence suggests a clear break, while at other maturities, the p-values are typically in the 15 to 20 percent range. The Table 1 coefficients and standard errors from the A and B subsamples are also displayed in Figure 1. It is clear that the ± 2 standard error bands overlap considerably except at fairly short horizons, which is consistent with the predominance of insignificant individual breakdate p-values.

Still, the fact that *all* of the slope coefficients, taken as a group, have shifted in the same direction in the later subsample is highly suggestive of a structural break in the behavior of the term structure. Rigorous statistical evidence on this point requires the formulation of a joint test. The next section will develop closely related evidence in the context of an empirical no-arbitrage model of the entire term structure. However, in the spirit of the regression analysis of this section, we also examine evidence on the joint significance of simultaneous changes in several of the slope coefficients by stacking several long-rate regressions for different maturities into one system regression. Although none of these long-rate regressions share a common regressor or regressand, it is highly likely that their error terms are correlated, so the system Seemingly Unrelated Regression (SUR) technique will generate more precise estimates.[6] Specifically, we stack

[5] The specific test used adds two variables to the long-rate regression: a dummy variable that is non-zero only during subsample B and a spread times that dummy. The break test is an F-test of the significance of the latter.

[6] As the term structure literature has stressed (e.g., Litterman and Scheinkman 1991, Duffie and Kan 1996), almost all movements in the yield curve can be captured by a few factors; thus, the errors in individual long-rate regressions are likely correlated across the regressions. On the other hand, the term spreads used in the regressions at different maturities are also likely correlated for the same reason. The efficiency gains from running SUR will depend on which correlation dominates, and the Appendix provides some evidence on this issue.

the individual long-rate regressions for the 3-, 24-, and 60-month maturities, which are three representative maturities for which the stability null hypotheses of unchanged slope coefficients were *not* rejected in the individual regressions. The system regression for these three maturities is

$$\begin{bmatrix} i_{2,t+1} - i_{3,t} \\ i_{23,t+1} - i_{24,t} \\ i_{59,t+1} - i_{60,t} \end{bmatrix} = \begin{bmatrix} \alpha_3 \\ \alpha_{24} \\ \alpha_{60} \end{bmatrix} + \begin{bmatrix} \beta_3 & 0 & 0 \\ 0 & \beta_{24} & 0 \\ 0 & 0 & \beta_{60} \end{bmatrix} \begin{bmatrix} (i_{3,t} - i_t)/2 \\ (i_{24,t} - i_t)/23 \\ (i_{60,t} - i_t)/59 \end{bmatrix} + \begin{bmatrix} \varepsilon_{3,t} \\ \varepsilon_{24,t} \\ \varepsilon_{60,t} \end{bmatrix}. \quad (2.4)$$

The estimation results for this SUR regression are shown in Table 2 for the full sample and for subsamples A and B. The slope coefficient estimates in subsamples A and B continue to show the same stark quantitative differences apparent in the individual regressions in Table 1; however, the coefficient standard errors are, on average, about half as large in magnitude. This greater precision sharpens inference, and for these three maturities (which again were chosen for their individual non-rejection of stability null), the p-value of .007 clearly rejects the joint null hypothesis of no change in the three slope coefficients between the A and B subsamples. These system break test results are representative of other combinations of three or more yields.[7]

Finally, while we have considered a specific breakdate, based on a prior view of the timing of changes in the behavior of aggregate output, inflation, and monetary policy, it is also useful to consider testing more generally the null of parameter stability without such a prior. To do this, we consider all possible breakdates in the middle 70 percent of the full sample for the system regression, and calculate a Chow-type test statistic at each of these breakdates. Figure 2 shows this set of test statistics as well as two 10 percent critical values. The less stringent one—the lower dashed line—is the usual χ^2 critical value (6.25) for the hypothesis that a specific (a priori) known breakdate is significant. The more stringent one—the upper dashed line (12.27)—is based on a test that does not assume any prior knowledge about potential breakdates. It tests the significance of the maximum value of all Chow-type test statistics calculated at all possible breakdates in the middle 70 percent of the sample, as given in Andrews (1993).[8] Applied to all possible breakdates for the system regression, the break test statistic does exceed the Andrews critical value during the late 1980s. This evidence supports our earlier selection of a breakdate, though, not surprisingly, the test is not sensitive enough to single out just one date.

[7] The expectations hypothesis, namely, that all three slope coefficients equal unity, is also rejected in each system regression in Table 2. For subsample B, this rejection reflects the low value of β_3.

[8] For our application, in which the variables are not highly persistence, it appears from various small-sample simulation studies that this asymptotic distribution is appropriate (see Diebold and Chen 1996 and O'Reilly and Whelan 2004).

In summary, we take the regression results as indicative of a break in term structure behavior in the 1980s. Determining the nature of that break in terms of changes in the dynamics of the short rate, of risk, or of the pricing of that risk is the subject of the rest of our analysis.

3. Estimating Subsample No-Arbitrage Models

In the preceding section, we provided regression evidence of a significant shift in the behavior of the term structure during the 1980s. In this section, we estimate dynamic term structure models that can capture that shift in behavior. The framework we use is a standard no-arbitrage representation from the empirical bond pricing literature that assumes no opportunities for financial arbitrage across bonds of different maturities.[9]

We focus on a two-factor, Gaussian, affine, no-arbitrage term structure model, or an $A_0(2)$ model as defined in Dai and Singleton (2000). The model features a constant volatility of term structure factors but the risk pricing is state-dependent, which implies conditionally heteroskedastic risk premiums. Dai and Singleton (2002) compare the performance of different dynamic term structure models and find that this type of specification performs the best in matching the full-sample long-rate regression coefficients.[10]

The model is formulated in discrete time. The state vector relevant for pricing bonds is assumed to be summarized by two latent term structure factors, L_t and S_t. These are stacked in the vector $F_t = (L_t, S_t)'$, which follows a Gaussian VAR(1) process:

$$F_t = \rho F_{t-1} + \Sigma \varepsilon_t, \tag{3.1}$$

where ε_t is i.i.d. $N(0, I_2)$, Σ is diagonal, and ρ is a 2×2 lower triangular matrix. The short (one-month) rate is defined to be a linear function of the latent factors:

$$i_t = \delta_0 + L_t + S_t = \delta_0 + \delta_1' F_t. \tag{3.2}$$

Without loss of generality, this implicit definition of δ_1 implies unitary loadings on the two factors of the short rate because of the normalization of the unobservable factors. Finally, following Constantinides (1992), Dai and Singleton (2000, 2002), Duffee (2002), and others, the prices of risk associated with the conditional volatility in the L_t factor, denoted $\Lambda_{L,t}$, and in the S_t factor,

[9] See Dai and Singleton (2002) and Rudebusch and Wu (2004) for references and discussion.

[10] Dai and Singleton (2002) use a three-factor model, but following Rudebusch and Wu (2004) we obtain an adequate fit, especially in subsample B, with two factors.

denoted $\Lambda_{S,t}$, are defined to be linear functions of the factors:

$$\Lambda_t = \begin{bmatrix} \Lambda_L \\ \Lambda_S \end{bmatrix}_t = \lambda_0 + \lambda_1 F_t. \tag{3.3}$$

Note that if all of the elements of λ_1 are zero, then the price of risk and the risk premium are constant, and, in this special case, the expectations hypothesis holds.

Under the no-arbitrage assumption, the logarithm of the price of a j-period nominal bond is a linear function of the factors

$$\ln(b_{j,t}) = \overline{A}_j + \overline{B}_j' F_t, \tag{3.4}$$

where the coefficients $\overline{A}_j$ and $\overline{B}_j$ are recursively defined by

$$\begin{aligned} \overline{A}_1 &= -\delta_0; \qquad \overline{B}_1 = -\delta_1 \\ \overline{A}_{j+1} - \overline{A}_j &= \overline{B}_j'(-\Sigma\lambda_0) + \frac{1}{2}\overline{B}_j'\Sigma\Sigma'\overline{B}_j + \overline{A}_1 \\ \overline{B}_{j+1} &= \overline{B}_j'(\rho - \Sigma\lambda_1) + \overline{B}_1; \qquad\qquad j = 1, 2, ..., J. \end{aligned} \tag{3.5}$$

Given this bond-pricing formula, the continuously compounded yield to maturity $i_{j,t}$ of a j-period nominal zero-coupon bond is given by the linear function

$$i_{j,t} = -\ln(b_{j,t})/j = A_j + B_j' F_t, \tag{3.6}$$

where $A_j = -\overline{A}_j/j$ and $B_j = -\overline{B}_j/j$.

The above model is estimated by maximum likelihood using end-of-month data on U.S. Treasury zero-coupon bond yields of maturities 1, 3, 12, 36, and 60 months (the yields are expressed at an annual rate in percent.) In estimating the model, the mean of the short rate δ_0 is set to the unconditional mean of the short rate in each subsample period (and λ_L^0 is normalized to zero). Therefore, the estimated model parameters for factor dynamics, risk pricing, and factor shocks are

$$\rho = \begin{bmatrix} \rho_L & 0 \\ \rho_{SL} & \rho_S \end{bmatrix}, \ \lambda^0 = \begin{bmatrix} 0 \\ \lambda_S^0 \end{bmatrix}, \ \lambda^1 = \begin{bmatrix} \lambda_{LL}^1 & \lambda_{LS}^1 \\ \lambda_{SL}^1 & \lambda_{SS}^1 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} \sigma_L & 0 \\ 0 & \sigma_S \end{bmatrix}.$$

In addition, following standard practice in the literature, the 1-month and 60-month bond yields are assumed to be measured without error, while bond yields of the other three maturities are measured with i.i.d. shocks with mean zero. The standard errors of these measurement errors are denoted σ_3, σ_{12}, σ_{36}. Finally, out of concern that the model may be over-parameterized, we impose certain zero restrictions on λ_0 and λ_1 on the entries with insignificant estimates in a preliminary round of model estimation. This procedure is common in the finance literature (e.g., Dai and Singleton 2002 and Ang and Piazzesi 2003) and introduces little change to the value of the likelihood function.

We estimate the model separately on the full sample of data (1970-2002) and on subsamples A (1970-1987) and B (1988-2002) as suggested by the results in Section 2. The maximum likelihood estimates of the model in different sample periods, along with their estimated standard errors and the value of the log-likelihood function, are displayed in Table 3. The most important result in Table 3 is that the hypothesis of a single unchanged data-generating process during the full sample is rejected at any significance level—the likelihood ratio test statistic, which follows a $\chi^2(12)$ under the hypothesis, is 481.88. This evidence provides another strong rejection of the joint stability hypothesis, consistent with the SUR test results of Table 2, and it helps validate the splitting of the sample.

The two subsample models exhibit interesting similarities and differences in parameter estimates. As is typically found, both the subsample A and subsample B models have a very persistent L_t factor ($\rho_L \approx .99$) and a less persistent S_t factor ($\rho_S \approx .95$). These two factors are often given the labels "level" and "slope," respectively, since a positive shock to L_t pushes up yields at all maturities while a positive shock to S_t predominantly pushes up yields at short maturities. Indeed, the factor loadings of both of our subsample estimated models are consistent with such a designation. Although both level and slope are a bit more persistent during the later subsample, a more striking difference is found in the factor shock volatilities in the two subsample periods. In particular, the volatilities of both factor shocks are significantly larger in the earlier subsample than in the later subsample.[11] The estimates of the standard deviations of the level and slope factor shocks are 41 and 60 basis points during subsample A, but only 15 and 42 basis points in subsample B. This finding is consistent with the view that 1970 to 1987 is a turbulent period for financial markets and the macroeconomy, while the more recent period has lower financial risks and a more tranquil economy. Finally, there is also a clear difference in risk pricing in the two subsamples. The subsample A estimates of the elements of λ_1 are uniformly larger than in subsample B. Time variation in risk premiums in this model solely reflects the time variation in the price of risk (as the volatilities of risks are constant), which is

[11] Interestingly, past regime-switching studies (Gray 1996, Ang and Bekaert 2002, and Dai, Singleton and Yang 2003) also find that the term structure factors exhibit less mean reversion (i.e., more persistence) in regimes with low volatilities of term structure risks.

a linear function of the factors. Therefore, the larger λ_1 in the earlier subsample will generate more variation in the price of risk and the risk premiums for a given level of factor volatility.

Overall, the subsample model estimates appear consistent with the notion of a shift in term structure behavior as suggested by the regression evidence. In the next section, we will link the differences in model parameter estimates to the different regression results.

4. Accounting for the Shift in Term Structure Behavior

Section 2 provided evidence of a significant break in the estimated coefficients in various long-rate term structure regressions, and Section 3 provided evidence of a significant break in a no-arbitrage dynamic term structure model. In this section, we link these two results together by investigating the ability of the two subsample no-arbitrage models, which were estimated in Section 3, to account for the long-rate regression results through changing factor and risk price dynamics .

Our examination focuses on long-rate regression coefficients implied by a particular no-arbitrage term structure representation. For a given no-arbitrage model of the form described in Section 3, the population value of the long-rate regression coefficient for maturity m is given by

$$\begin{aligned}
\beta_m &\equiv \frac{\mathrm{cov}[(i_{m-1,t+1} - i_{m,t}), (i_{m,t} - i_{1,t})/(m-1)]}{\mathrm{var}[(i_{m,t} - i_{1,t})/(m-1)]} \\
&= \frac{\mathrm{cov}[(B'_{m-1}F_{t+1} - B'_m F_t), (B'_m F_t - B'_1 F_t)]}{\mathrm{var}[B'_m F_t - B'_1 F_t]}(m-1) \\
&= \frac{(B'_{m-1}\rho - B'_m)\Omega(B'_m - B'_1)'}{(B'_m - B'_1)\Omega(B'_m - B'_1)'}(m-1) \qquad (4.1)
\end{aligned}$$

where the B_m's are the factor loadings defined in Section 3 and Ω denotes the unconditional variance-covariance matrix of the two factors in F_t. From equation (3.5), note that the B_m's are determined by Σ, λ_1, and ρ—that is, by the covariance of the factor shocks, the sensitivity of the price of risk to the factors, and the parameters of the autoregressive dynamics of the factors, respectively. From equation (3.1), note that Ω depends on the parameters Σ and ρ. Therefore, the population regression coefficients associated with different no-arbitrage model structural estimates are straightforward to compute.

The implied long-rate regression coefficients associated with the subsample model estimates shown in Table 3 are given in Figures 3 and 4, for subsamples A and B, respectively. The thick solid lines in these figures plot the model-implied population β_m coefficients at all bond maturities while the thin solid lines give the actual historical regression estimates from subsamples A and B. These implied model population estimates match the historical regression results fairly well. For subsample A (Figure 3), the model-implied regression coefficients decrease quite rapidly as the maturity of the long rate increases. Although the population coefficients are not quite as low as the historical estimates for maturities less than 48 months, there is a fairly close match at longer maturities. For subsample B (Figure 4), the model-implied projection coefficients are all positive and quite close to the empirical regression estimates.

In order to account for possible small-sample biases, we also simulate data from each model and calculate the regression coefficients in repeated finite samples. Specifically, we take random draws of ε_t (with the number determined by the particular sample period under investigation), simulate data from the no-arbitrage model, and compute the long-rate projection coefficients. This procedure is repeated 1000 times, and Figures 3 and 4 also plot the medians and 90 percent frequency or confidence bands from these simulations. In both figures, the median estimates from the simulations lie very close to the population estimates, indicating that the small-sample biases are fairly modest in this application (which is consistent with the results reported in the Appendix). In addition, the empirical estimates typically lie inside the 90 percent confidence bands of the model simulations.

The source for the differences in the term structure dynamics between the two subsamples can be illuminated with model perturbations. Specifically, we look at the effect on a long-rate regression coefficient from changing a subset of the model parameters from their subsample A estimated values to their subsample B estimated values. This model variation can uncover the specific factors driving the different subsample regression results. However, because the long-rate regression coefficients are nonlinear functions of the model parameters, the effect of changing a particular model parameter depends on the exact constellation of the other parameters. To reduce the number of model permutations to a more manageable size, we focus on three blocks of parameters—in Σ, λ_1, and ρ—as sets that contain either all subsample A estimates or all subsample B estimates. For example, the ρ_{LL}, ρ_{SS}, and ρ_{SL} in ρ are all either from subsample A or subsample B. Thus, there are only eight possible combinations of the two subsample estimates of Σ, λ_1, and ρ to consider. Each of each these eight cases is identified by a parameter triple, with ρ_A, Σ_A, λ_A, ρ_B, Σ_B, and λ_B representing the estimates of ρ, Σ, and λ_1 in subsamples

A and B, respectively.

The model perturbation results obtained from varying these three sets of parameters are given in Table 4. For conciseness, the table only displays the effect on the β_{120} coefficient, but our conclusions generalize to the other long-rate maturities as well. The top line shows the change in the population estimate of β_{120} resulting from a shift from all subsample A no-arbitrage parameter estimates (denoted as the ρ_A, Σ_A, λ_A model) to all subsample B parameter estimates (denoted as ρ_B, Σ_B, λ_B). This change, which is 2.88, is also the difference between the right-hand-side endpoints of the thick solid lines in Figures 3 and 4. The rest of Table 4 provides a quantitative accounting of the source of this change. Specifically, the next block of lines investigates a change in just the autoregressive parameters from ρ_A to ρ_B, holding fixed the other parameters across the four possible permutations of Σ and λ_1 (namely, Σ_A, λ_A; Σ_A, λ_B; Σ_B, λ_A; Σ_B, λ_B). The average effect of such a change in factor dynamics would cause β_{120} to decrease by 1.04—that is, β_{120} is pushed in the opposite direction from what was observed. In contrast, as shown in the middle lines, the shift in the factor shock volatility parameters from Σ_A to Σ_B induces, on average, a 0.46 increase in β_{120}, which is a modest step in the observed direction of change. Finally, as shown in the bottom panel of Table 4, the change in the risk pricing parameters from λ_A to λ_B more than accounts for the total observed change in the population β_{120}.

The risk pricing parameters therefore appear crucial in generating the changing profile of the long-rate regression coefficients across the two subsamples. The more factor-sensitive risk pricing in subsample A—since the subsample A estimates of λ_1 are larger than in subsample B—generates greater time variation in the risk premiums for a given level of factor volatilities. These more variable subsample A term premiums induce greater deviations from the expectations hypothesis and push the β_m estimates below those in subsample B. This effect is reinforced to a limited extent by the higher variances of the factor shocks in the first subsample (since the elements of Σ_A are larger than those of Σ_B). These higher factor shock variances induce higher factor volatilities and hence greater time variation in the price of risk and risk premiums. However, a partial offset to the above two factors comes from the higher autoregressive parameters in the later subsample. Specifically, because the elements of ρ_B are higher than those of ρ_A, these work to boost the volatility of the factors and risk premiums in subsample B and lower the regression coefficients.[12]

Table 5 reports on a model permutation procedure that considers individual parameters instead of blocks of parameters. There are eight key individual model parameters in ρ, Σ, and λ_1: ρ_{LL}, ρ_{SS}, ρ_{SL}, σ_L, σ_S, λ^1_{LL}, λ^1_{LS}, and λ^1_{SS}. Table 5 provides the average effect on β_{120} of changing each one of these coefficients from its subsample A estimate to its subsample B estimate, holding the other coefficients fixed.[13] These results further narrow the source of the upward shift in the long-rate regression coefficients in the later subsample to just a few model parameters, all of which are related to the level factor. In particular, the two most influential parameters are λ^1_{LL} and λ^1_{LS}, which control the way in which the price of risk that is attached to fluctuations in the level factor varies with the magnitude of level and slope. The reduced size of these risk pricing parameters in subsample B can account on their own for the shift in the long-rate regression coefficients across the two subsamples. The reduction in σ_L, the variance of shocks to level, also plays some role by reducing level factor volatility (and the associated risk premium variability), but this effect is offset by the increase in the level factor autoregressive parameter ρ_{LL}, which tends to boost the level factor variability.

To summarize, the standard no-arbitrage bond pricing model suggests that the recent historical shift in term structure behavior predominantly reflects a change in the way investors price risk associated with the level factor. Changes in factor dynamics and factor shock volatility appear to have played a relatively modest role. These term structure results may also help illuminate the nature of the moderation and transformation of the U.S. economy that occurred in the 1980s. As noted in the introduction, one hypothesis is that there was a run of less volatile economic shocks in the more recent period. Our estimates support the presence of less volatile factor shocks in the recent subsample; however, the effect of this change on the behavior of the term structure appears modest. Instead, our estimates indicate that there was a important change in the dynamics of the economy that effected risk pricing. The next section elaborates on this interpretation using a no-arbitrage macro-finance model that links movements in the level factor to observable variables in the economy.

[12] We have also found that more persistent factors, even holding the volatility of the factors constant (as opposed to the holding constant the volatility of the factor shocks), leads to lower regression coefficients.

[13] For investigating the effects of a change in any given parameter, there are 128 possible mixed sample A and B permutations for the other seven parameters. (Note that λ^1_{SL} is zero in both samples.) We do not investigate all of these permutations; instead, Table 5 provides the average change in β_{120} using a representative sample of eight of these configurations using the same blocks of parameters in Table 4.

5. A Macro-Finance Perspective on the Term Structure Shift

The analysis so far suggests that an important transformation occurred in the U.S. economy in the 1980s regarding the behavior of the level factor and, in particular, the pricing of risk associated with that factor. A natural next step is to provide an economic interpretation of these changes. We pursue this task in the structural macro-finance model of Rudebusch and Wu (2004), which we describe briefly before considering some model perturbations.

The Rudebusch-Wu macro-finance model combines the above canonical no-arbitrage term structure representation with elements from a standard macroeconomic model. A key point of intersection between the finance and macroeconomic specifications is the short-term interest rate. The short rate remains a linear function of two latent term structure factors as in the finance model, so

$$i_t = \delta_0 + L_t + S_t. \tag{5.1}$$

As demonstrated in Rudebusch and Wu (2004), however, there is a close connection among these level and slope factors and a simple Taylor (1993) rule for monetary policy:

$$i_t = r^* + \pi_t^* + g_\pi(\pi_t - \pi_t^*) + g_y y_t, \tag{5.2}$$

where r^* is the equilibrium real rate, π_t^* is the central bank's inflation target, π_t is the annual inflation rate, and y_t is a measure of the output gap. This link reflects the fact that the Federal Reserve sets the short rate in response to macroeconomic data in an attempt to achieve its goals of output and inflation stabilization. Therefore, level and slope are not simply modeled as purely autoregressive time series; instead, they form elements of a monetary policy reaction function. In particular, L_t is interpreted the medium-term inflation target of the central bank as perceived by private investors. Investors are assumed to modify their views of this underlying rate of inflation slowly, as actual inflation, π_t, changes, so L_t is linearly updated by news about inflation:[14]

$$L_t = \rho_L L_{t-1} + (1 - \rho_L)\pi_t + \varepsilon_{L,t}. \tag{5.3}$$

The slope factor S_t captures the Fed's dual mandate to stabilize the real economy and keep inflation close to its medium-term target level. Specifically, S_t is modeled as the Fed's cyclical response to deviations of inflation from its target, $\pi_t - L_t$, and to deviations of output from its

[14] As shown in Rudebusch and Wu (2004), L_t is primarily associated with yields of maturities from 2 to 5 years, which is an important indication of the relevant horizon for the associated inflation expectations.

potential, y_t:

$$S_t = \rho_S S_{t-1} + (1-\rho_S)[g_y y_t + g_\pi(\pi_t - L_t)] + u_{S,t} \tag{5.4}$$

$$u_{S,t} = \rho_u u_{S,t-1} + \varepsilon_{S,t}. \tag{5.5}$$

In addition, a very general specification of the dynamics of S_t is adopted that allows for both policy inertia and serially correlated elements not included in the basic Taylor rule.[15]

The dynamics of the macroeconomic determinants of the short rate are then specified with fairly standard New Keyesian equations for inflation and output (adjusted for monthly data):

$$\pi_t = \mu_\pi L_t + (1-\mu_\pi)[\alpha_{\pi_1}\pi_{t-1} + \alpha_{\pi_2}\pi_{t-2}] + \alpha_y y_{t-1} + \varepsilon_{\pi,t} \tag{5.6}$$

$$y_t = \mu_y E_t y_{t+1} + (1-\mu_y)[\beta_{y1} y_{t-1} + \beta_{y2} y_{t-2}] - \beta_r(i_{t-1} - L_{t-1}) + \varepsilon_{y,t}\ . \tag{5.7}$$

That is, inflation responds to the public's expectation of the medium-term inflation goal (L_t), two lags of inflation, and the output gap. Output depends on expected output, lags of output, and a real interest rate.

The specification of long-term yields in the macro-finance model follows the standard no-arbitrage formulation described in Section 3. Accordingly, the state space of the combined macro-finance model can be expressed by equation (3.1) with the state vector F_t redefined to include output and inflation. The dynamic structure of this transition equation is determined by equations (5.3) through (5.5). There are four structural shocks, $\varepsilon_{\pi,t}$, $\varepsilon_{y,t}$, $\varepsilon_{L,t}$, and $\varepsilon_{S,t}$, which are assumed to be independently and normally distributed. The short rate is determined by (5.1). For pricing longer-term bonds, the risk price associated with the structural shocks is assumed to be a linear function of only L_t and S_t, which matches the formulation in Section 3 and allows for easy comparison.[16] However, it should be noted that the macroeconomic shocks $\varepsilon_{\pi,t}$ and $\varepsilon_{y,t}$ *are* able to affect the price of risk through their influence on π_t and y_t and, therefore, on the latent factors, L_t and S_t.

The estimates of this macro-finance model from Rudebusch and Wu (2004), which are based on U.S. term structure data that are essentially from subsample B (1988 to 2000), are shown in Table 6. As above, the factor L_t is very persistent, with a ρ_L estimate of 0.989, which implies a

[15] If $\rho_u = 0$, the dynamics of S_t arise from monetary policy partial adjustment; conversely, if $\rho_S = 0$, the dynamics reflect the Fed's reaction to serially correlated information or events not captured by output and inflation. Rudebusch (2002) shows that the latter is often confused with the former in empirical applications.

[16] Therefore, λ_1 continues to have just four potentially non-zero entries (λ^1_{LL}, λ^1_{LS}, λ^1_{SL}, and λ^1_{SS}), thus greatly reducing the number of parameters to be estimated.

small but significant response to actual inflation. The monetary policy interpretation of the slope factor is supported by the reasonable estimated inflation and output response coefficients, g_π and g_y, which are 1.25 and 0.20, respectively. These values, as well as the estimated parameters describing the inflation and output dynamics, appear to be in line with other estimates in the literature.

We next turn to the implied long-rate regression coefficients from this model.[17] As before, we conduct a model simulation exercise in which repeated samples of data are generated from the macro-finance model and used in the calculation of regression coefficients. Figure 5 shows median values of the regression coefficients obtained from the macro-finance model simulated data as a solid line. The coefficients are predominantly positive and decline from about 1 at a very short maturity to slightly negative at a 120-month maturity. These estimates are a bit closer to the actual historical estimates from the subsample B data (shown as the dotted line) than the coefficients implied by the estimated subsample B no-arbitrage model from Sections 3 and 4 (the dashed line).

The analysis in Sections 3 and 4 suggested that changes in the conditional volatility of the level factor and in the pricing of level factor risk were the most important factors in accounting for the shift in long-rate regression coefficients. This same issue can be examined in the macro-finance model. In particular, as noted above, the key parameters λ^1_{LL}, λ^1_{LS}, and σ_L play the same role in both models. The effect of changing these parameters in the macro-finance model is shown in Table 7, which, as in Tables 4 and 5, focuses on just the coefficient β_{120} for conciseness. The first three lines of Table 7 show the effect on β_{120} of changing λ^1_{LL}, λ^1_{LS}, and σ_L from their estimates in Table 6 (-0.0045, 0.0168, and 0.342, respectively) to their subsample A estimates in Table 3 (-0.0146, 0.0342, and 0.41, respectively).[18] Increasing (in absolute value) λ^1_{LS} and σ_L gives clearly lower estimates of β_{120}, while changing λ^1_{LL} has little effect on its own. However, the combination of all three changes—line 4—shifts β_{120} down by a substantial 2.05. That is, as above in the basic no-arbitrage model, the risk pricing and dynamics of the level factor appear crucial for accounting for the shift in term structure behavior.

More importantly, the macro-finance model provides an economic interpretation of this shift.

[17] Hördahl, Tristani, and Vestin (2004) also examine long-rate regression coefficients from a macro-finance model for German data.

[18] Another experiment that we are investigating in further work would be to estimate the macro-finance model for sample A and conduct a comparison as in Section 4. This may be problematic because the estimated policy rule of sample A often induces nonstationarity in forward-looking rational expectations macroeconomic models (see Rudebusch 2004).

Since the level factor reflects the perceived inflation target, the macro-finance explanation of the shift in term structure behavior is that during the 1970s and early 1980s investors had a very different view of the medium-term outlook for inflation than they did later on. Investors in the early period appear to have viewed the inflation goal as particularly uncertain, in the sense that it had a greater conditional volatility (higher σ_L) and that its price of risk was more sensitive to fluctuations in the economy (in particular, λ^1_{LS} is higher early on). This explanation is broadly consistent with the view that expectations of the underlying goals for inflation were less firmly anchored in investors' minds during the earlier subsample, which is a common interpretation of the historical evolution of U.S. monetary policy. Alternatively, it could also be that changes to financial markets or institutions allow investors to hedge risk better in the later subsample, so that the risk compensation is less sensitive to changes in the economy.

Other changes in the economy may also have played a role in the shifting term structure behavior. Many authors have noted that the volatilities of shocks to output and inflation are significantly larger in the 1970s than in the 1990s. To consider the possibility that the higher conditional macroeconomic volatility in the earlier period helped account for the lower regression coefficients, we increase the standard deviations of the output and inflation shocks, σ_π and σ_y, by 50 percent, which is the order of magnitude suggested by previous empirical work, including Stock and Watson (2003) and Moreno (2004). As shown in the second line from the bottom in Table 7, this model perturbation has little effect on the estimate of β_{120}. Another important economic change that many estimated models of Federal Reserve behavior have highlighted is the substantially lower responsiveness of monetary policy to inflation that occurred before the 1980s.[19] To consider the possibility that a lower inflation response parameter in the earlier subsample may account for the lower regression coefficients, we reduced the inflation response coefficient in the monetary policy rule, g_π, by one-half, which is broadly in line with various empirical estimates. The result, as shown in the bottom row of Table 7, is only a very modest effect on the estimate of β_{120}.

[19] See, for example, Fuhrer (1996), Judd and Rudebusch (1998), Clarida, Galí, and Gertler (2000), and Rudebusch (2004) for discussion. In contrast to the inflation response coefficient, the evidence on a significant change in the monetary policy output response coefficient is mixed.

6. Conclusion

As noted in the introduction, the existence of a shift in the behavior of the term structure would not be surprising, given the dramatic changes in the economy over the past few decades. We indeed document such a shift in the behavior of the term structure using a simple regression technique as well as structural models. Our key result is that the volatility of term premiums appears to have declined over time; furthermore, this decline appears to have been induced by changes in the conditional volatility and price of risk of the term structure level factor, which we suggest may be related to investors' perceptions of the Fed's inflation goals.

Of course, as many have noted, a shift in the conduct of monetary policy will likely lead to a change in the behavior of the term structure (for example, Rudebusch 1995, Fuhrer 1996, Kozicki and Tinsley 2001, 2005 and Cogley 2003). However, our results suggest that the linkage is perhaps more subtle than is commonly appreciated. For example, although the Fed's short rate response to changes in inflation during the 1970s has been found to be less vigorous than in the 1990s, such a change—on its own—appears to have small direct effects on the evolution of term premiums and appears unlikely to account for the shift apparent in our empirical results. This conclusion appears to mirror that of Stock and Watson (2003), who found small direct effects of monetary policy rule changes on macroeconomic volatility. However, our results do suggest that broader, but likely closely related, shifts in the monetary policy environment may have played an important role. In particular, a change in the perceptions of the inflation goals of the Fed could have altered the dynamic evolution of term premiums. Such a change may reflect a greater willingness to anchor the inflation rate or a greater transparency about such desires.

A. Appendix on Small-Sample Inference

In Section 2, we conduct inference on the expectations hypothesis and the hypothesis of stable parameters using asymptotic distributions that have been called into question in certain circumstances (e.g., in Bekaert, Hodrick, and Marshall 1997b). In this section, we report monte carlo simulations using the data generated from estimated models of Section 3 in order to explore the appropriateness of this inference in small samples for our models.

Table A1 displays the results of testing the expectations hypothesis based on 1000 simulated samples of size 396, 216, and 180 observations, respectively, for the full sample, subsample A, and subsample B. Using these simulated data, the change in the 3-month rate and the change in the 60-month rate are regressed on the on the 3- and 1-month spread and the 60- and 1-month spread, respectively. Each entry in the table reports the frequency with which an F-test statistic rejects the null expectations hypothesis, which is the hypothesis that the slope coefficient (or a pair of slope coefficients) is equal to one. These rejections are calculated using the standard 5 percent asymptotic critical values. In the top panel of Table A1, model simulations are based on parameter estimates from each sample of data (given in Table 3) except that the price of risk is set to be a constant (λ_1 is set equal to 0). Constant risk prices in this term structure model imply that risk premiums are constant; thus, the null expectations hypothesis is true and the population slope regression coefficients are indeed equal to one. In this case, the reported rejection frequency is the empirical size of the F-test.

The first row of the top panel reports the frequency of rejection using as a data-generating process the full-sample model estimates in Table 3 (again with λ_1 set equal to 0). For the individual long-rate regressions, the hypothesis that $\beta_3 = 1$ is rejected 7.7 percent of the time, and the hypothesis that $\beta_{60} = 1$ is rejected 6.7 percent of the time. These empirical sizes are quite close to the 5 percent nominal size. The third entry of 5.7 percent gives the frequency of simulated samples in which $\beta_3 = 1$ and $\beta_{60} = 1$ were both rejected in the individual long-rate regressions. This statistic provides a relevant comparison for the system SUR estimation, which tests the joint null expectations hypothesis that $\beta_3 = \beta_{60} = 1$. This joint test has an empirical size of 7.7 percent. The second and third rows of Table A1 show that the F-test is only slightly less well-sized when the data are simulated from the subsample A model estimates and the subsample B model estimates (which may reflect the smaller samples in these cases).

The lower panel of Table A1 displays the frequency of rejections when the data are simulated from the exact estimated models given in Table 3. In each of these models, $\lambda_1 \neq 0$, so the price of risk is time-varying and the expectations hypothesis does not hold; thus, the rejection frequencies in this panel indicate the empirical power of the F-test. The power of this test, given our data-generating mechanism and sample sizes, is low (particularly for the 3-month maturity) to moderate (for the 60-month maturity). The simulation results indicate some advantage to running system SUR when testing the joint expectations hypothesis. When the simulations are based on full-sample model estimates, the system regression correctly rejects the null in 24.2 percent of the draws, while the individual long-rate regressions reject the null on both 3- and 60-month maturities in only 6.2 percent of the draws. This pattern is similar when simulations are based on subsample A estimates. When simulation is performed based on subsample B estimates, however, the power in running individual long-rate regressions becomes quite small — even smaller than the corresponding empirical size in the upper panel. This puzzle reflects two offsetting effects on the regression coefficients: a downward pressure from the time-varying risk prices and term premiums, and an upward small-sample bias as discussed in Bekaert, Hodrick, and Marshall (1997b) which tends to push the coefficients back to unity. The effect of small-sample bias is overwhelmed in the full-sample and subsample A simulations when the risk price variability is large, but it is quite important in the subsample B simulations when the risk price movements are small. However, the SUR reports smaller standard errors, so the power in running the SUR is much higher than for individual long-rate regressions. This again underscores the efficiency gains from the SUR.

Table A2 displays the frequency of rejection of a Chow-type test of the null hypothesis, namely, that there is no difference in the slope coefficient (or coefficients) between the earlier and later subsamples. Each entry is based on 1000 simulations of 396 observations and reports the frequency of the test statistic exceeding the 5 percent theoretical critical value, which indicates rejection of the null hypothesis of no breakpoint. First consider the empirical size of the test. In the top panel, the full-sample model estimates are used exclusively, so the data are generated under a single regime. The frequency of rejection is 5.1 percent for the 3-month regression, 3.8 percent for 60-month regression, and 2.7 percent for the system SUR, suggesting that the test is fairly well-sized though with some tendency to reject the null hypothesis less frequently than theory would predict.

The bottom panel of Table A2 provides results when the data-generating process contains a regime switch. In particular, the first 216 observations of each simulation are drawn from the model estimates in subsample A, and the remaining 180 observations are drawn from the model estimates in subsample B. Thus the proportions of rejections in this panel indicate the empirical power of the Chow test, which appears fairly high. The test correctly rejects the null 66 percent of the time for the 3-month rate regression and 64.8 percent of the time for the 60-month rate regression. For the joint hypothesis, the structural stability null is rejected 54.8 percent of the time with both individual regressions, which is lower than the 60.9 percent rejection rate obtained with SUR, suggesting some modest efficiency gains to system estimation.

References

[1] Andrews, Donald (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, 821-856.

[2] Ang, Andrew and Geert Bekaert (2002), "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics* 20(2), 163-182.

[3] Ang, A. and M. Piazzesi (2003), "No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics* 50, 745-787.

[4] Backus, D., S. Foresi, A. Mozumdar, and L. Wu (2001), "Predictable Changes in Yields and Forward Rates," *Journal of Financial Economics* 59(3), 281-311.

[5] Bansal, R. and H. Zhou (2002), "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance* 57, 1997-2043.

[6] Bekaert, Geert, Robert J. Hodrick, and David A. Marshall (1997a), "Peso Problem Explanations for Term Structure Anomalies," NBER working paper no. 6147.

[7] Bekaert, Geert, Robert J. Hodrick, and David A. Marshall (1997b), "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics* 44, 309-348.

[8] Blanchard, Olivier, and John Simon (2001), "The Long and Large Decline in U.S. Output Volatility," *Brookings Papers on Economic Activity* 1, 135-164.

[9] Bliss, Robert R. (1997), "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research* 9, 197-231, Greenwich, Conn. and London: JAI Press.

[10] 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.

[11] 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.

[12] Cogley, Timothy (2003), "An Exploration of Evolving Term-Structure Relations," manuscript, University of California-Davis.

[13] Constantinides, G.M. (1992), "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies* 5, 531-552.

[14] Dai, Q. and K.J. Singleton (2000), "Specification Analysis of Affine Term Structure Models," *Journal of Finance* 55, 1943-1978.

[15] Dai, Q. and K.J. Singleton (2002), "Expectations puzzles, time-varying risk premia, and affine models of the term structure," *Journal of Financial Economics* 63, 415-441.

[16] Dai, Q., K.J. Singleton, and Wei Yang (2003), "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields," Stanford University Working Paper.

[17] Dewachter, H. and M. Lyrio (2002), "Macro Factors and the Term Structure of Interest Rates," manuscript, Catholic University of Leuven.

[18] Diebold, Francis, and Celia Chen (1996), "Testing Structural Stability with Endogenous Break Point: A Size Comparison of Analytic and Bootstrap Procedures," *Journal of Econometrics* 70, 221-241.

[19] Diebold, Francis, Glenn D Rudebusch, and S. Boragan Aruoba (2004), "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," manuscript, Federal Reserve Bank of San Francisco, forthcoming in the *Journal of Econometrics*.

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

[21] Duffee, Gregory R. (2004), "A No-Arbitrage Term Structure Model Without Latent Factors," manuscript, University of California – Berkeley.

[22] Duffie, D. and R. Kan (1996), "A Yield-Factor Model of Interest Rates," *Mathematical Finance* 6, 379-406.

[23] Fuhrer, Jeffrey C. (1996), "Monetary Policy Shifts and Long-Term Interest Rates," *The Quarterly Journal of Economics* 111, 1183-1209.

[24] Gray, Stephen F. (1996), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics* 42, 27-62.

[25] Hamilton, J.D. (1988), "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control* 12, 385-423.

[26] Hördahl, Peter, Oreste Tristani, and David Vestin (2004), "A Joint Econometric Model of Macroeconomic and Term Structure Dynamics," manuscript, European Central Bank, forthcoming in the *Journal of Econometrics*.

[27] Judd, John, and Glenn Rudebusch (1998), "Taylor's Rule and the Fed: 1970-1997," *Economic Review*, Federal Reserve Bank of San Francisco, no. 3, 3-16.

[28] Kim, Chang-Jin, and Charles R. Nelson (1999), "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics* 81, 608-616.

[29] Kozicki, Sharon, and P.A. Tinsley (2001), "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics* 47, 613-652.

[30] Kozicki, Sharon, and P.A. Tinsley (2005), "What Do You Expect? Imperfect Policy Credibility and Tests of the Expectations Hypothesis," *Journal of Monetary Economics* 52, 421-447.

[31] Lange, Joe, Brian Sack, and William Whitesell (2003), "Anticipations of Monetary Policy in Financial Markets," *Journal of Money, Credit, and Banking* 35(6), 889-909.

[32] Litterman, Robert, and Jose A. Scheinkman (1991), "Common Factors Affecting Bond Returns," *Journal of Fixed Income* 1, 54-61.

[33] Mankiw, N.Gregory, and Jeff A. Miron (1986), "The Changing Behavior of the Term Structure of Interest Rates," *The Quarterly Journal of Economics* 101, 211-228.

[34] McConnell, Margaret M., and Gabriel Perez-Quiros (2000), "Output Fluctuations in the United States: What has Changed Since the Early 1980's," *American Economic Review* 90(5), 1464-1476.

[35] Moreno, Antonio (2004), "Reaching Inflation Stability," *Journal of Money, Credit, and Banking* 36, 801-825.

[36] O'Reilly, Gerard, and Karl Whelan (2004), "Has Euro-Area Inflation Persistence Changed Over Time?" Working Paper.

[37] Piazzesi, Monika (2003), "Bond Yields and the Federal Reserve," manuscript, *Journal of Political Economy*, forthcoming.

[38] Rudebusch, Glenn D. (1995), "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure," *Journal of Monetary Economics* 24, 245-274.

[39] Rudebusch, Glenn D. (2002), "Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia," *Journal of Monetary Economics* 49, 1161-1187.

[40] Rudebusch, Glenn D. (2004), "Assessing the Lucas Critique in Monetary Policy Models," manuscript, forthcoming in the *Journal of Money, Credit, and Banking.*

[41] Rudebusch, Glenn D., and Tao Wu (2004), "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," Working Paper.

[42] Stock, James H., and Mark W. Watson (2003), "Has the Business Cycle Changed? Evidence and Explanations," in *Monetary Policy and Uncertainty*, Federal Reserve Bank of Kansas City, 9-56.

[43] Taylor, John B. (1993), "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39, 195-214.

[44] Watson, Mark W. (1999), "Explaining the Increased Variability in Long-Term Interest Rates," Federal Reserve Bank of Richmond *Economic Quarterly* 85(4), 71-96.

[45] Wu, Tao (2001), "Macro Factors and the Affine Term Structure of Interest Rates," manuscript, Federal Reserve Bank of San Francisco.

Table 1

Slope Coefficients for Long-Rate Regressions

Maturity (months)								
3	6	9	12	24	36	48	60	120
Estimates from Campbell-Shiller (1991): 1952:01 to 1987:02								
-0.176	-1.029	-1.219	-1.381	-1.815	-2.239	-2.665	-3.099	-5.024
(0.362)	(0.537)	(0.598)	(0.683)	(1.151)	(1.444)	(1.634)	(1.749)	(2.316)
Estimates from Dai-Singleton (2002): 1970:02 to 1995:12								
-0.428	-0.883	-1.228	-1.425	-1.705	-1.190	-2.147	-2.433	-4.173
(0.481)	(0.640)	(0.738)	(0.825)	(1.120)	(1.295)	(1.418)	(1.519)	(1.985)
Full sample: 1970:01 to 2002:12								
-0.334	-0.636	-0.957	-1.249	-1.116	-1.615	-2.420	-2.042	-3.984
(0.370)	(0.470)	(0.560)	(0.650)	(1.007)	(1.279)	(1.424)	(1.541)	(1.920)
[0.00]	[0.00]	[0.00]	[0.001]	[0.036]	[0.041]	[0.016]	[0.048]	[0.009]
Subsample A: 1970:01 to 1987:12								
-0.519	-1.342	-2.058	-2.517	-1.844	-2.273	-3.256	-2.882	-5.297
(0.540)	(0.713)	(0.845)	(0.906)	(1.359)	(1.665)	(1.886)	(2.059)	(2.802)
[0.005]	[0.001]	[0.00]	[0.00]	[0.036]	[0.049]	[0.024]	[0.059]	[0.025]
Subsample B: 1988:01 to 2002:12								
0.289	0.812	1.227	0.962	0.788	0.358	-0.116	0.161	-1.123
(0.139)	(0.280)	(0.447)	(0.592)	(0.936)	(1.180)	(1.326)	(1.453)	(1.854)
[0.00]	[0.504]	[0.611]	[0.949]	[0.821]	[0.586]	[0.400]	[0.564]	[0.252]
Break Test: p-value for hypothesis of no change in slope coefficient								
0.147	0.005	0.001	0.001	0.111	0.197	0.173	0.227	0.214

Note: Except for the final row, the first number in each set is the estimated slope coefficient from regression (2.3). Asymptotic standard errors are shown in parentheses, and p-values of the expectations hypothesis that the coefficient equals unity are given in brackets. The final row gives Chow-type p-values testing the breakdate of 1988:01.

Table 2

Slope Coefficients for System SUR Long-Rate Regressions

Maturity (months)		
3	24	60
Full sample: 1970:01 to 2002:12		
-0.457	-1.267	-2.507
(0.148)	(0.399)	(0.642)
Subsample A: 1970:01 to 1987:12		
-0.462	-2.008	-4.053
(0.227)	(0.571)	(0.895)
Subsample B: 1988:01 to 2002:12		
0.101	0.403	1.051
(0.124)	(0.458)	(0.790)
P-value for hypothesis of no change		
0.007		

Note: These are estimated slope coefficients from system regressions with asymptotic standard errors in parentheses. The final row gives p-values for the null of no change in the three slope coefficients across subsamples A and B.

Table 3

Parameter Estimates of the No-Arbitrage Model

Parameter	Full sample		Sample A		Sample B	
δ_0	6.2030		7.3502		4.8263	
ρ_{11}	0.9930	(0.0039)	0.9899	(0.0094)	0.9950	(0.0063)
ρ_{22}	0.9594	(0.0011)	0.9444	(0.0167)	0.9616	(0.0030)
ρ_{21}	-0.0137	(0.0070)	0.0116	(0.0292)	0.0069	(0.0033)
λ_2^0	-0.0004	(0.0002)	0		-0.0003	(0.0004)
λ_{11}^1	-0.0174	(0.0041)	-0.0146	(0.0234)	-0.0090	(0.0058)
λ_{21}^1	-0.0163	(0.0089)	0.0175	(0.0118)	-0.0085	(0.0035)
λ_{12}^1	0		0.0342	(0.0341)	0	
λ_{22}^1	0		0		0	
Σ_{11}	0.2420	(0.0190)	0.3984	(0.1885)	0.1532	(0.0434)
Σ_{22}	0.5100	(0.0100)	0.6022	(0.1240)	0.4205	(0.0305)
σ_3	0.2440	(0.0030)	0.3000	(0.0112)	0.2200	(0.0135)
σ_{12}	0.4080	(0.0097)	0.4386	(0.0203)	0.3800	(0.0534)
σ_{36}	0.2380	(0.0103)	0.2700	(0.0223)	0.1900	(0.0229)
$\log L$	9587.86		5207.18		4621.62	

Note: These are ML estimates from three data sample periods of the no-arbitrage model with asymptotic standard errors in parentheses. The final row gives the value of the maximized log-likelihood function ($\log L$).

Table 4

The Effect of Model Changes (by Blocks of Parameters) on β_{120} Estimates

Model change	Effect on β_{120} estimate from change in model	Contribution to total effect on β_{120} estimate
Change in all model parameters from subsample A to subsample B estimates		
ρ_A, Σ_A, $\lambda_A \to \rho_B$, Σ_B, λ_B	2.88	100
Change in factor autoregressive parameters ($\rho_A \to \rho_B$)		
ρ_A, Σ_A, $\lambda_A \to \rho_B$, Σ_A, λ_A	-0.60	-21
ρ_A, Σ_B, $\lambda_A \to \rho_B$, Σ_B, λ_A	-0.61	-21
ρ_A, Σ_A, $\lambda_B \to \rho_B$, Σ_A, λ_B	-1.92	-67
ρ_A, Σ_B, $\lambda_B \to \rho_B$, Σ_B, λ_B	-1.05	-37
Average	-1.04	-36
Change in factor volatility parameters ($\Sigma_A \to \Sigma_B$)		
ρ_A, Σ_A, $\lambda_A \to \rho_A$, Σ_B, λ_A	0.23	8
ρ_A, Σ_A, $\lambda_B \to \rho_A$, Σ_B, λ_B	0.28	10
ρ_B, Σ_A, $\lambda_A \to \rho_B$, Σ_B, λ_A	0.21	7
ρ_B, Σ_A, $\lambda_B \to \rho_B$, Σ_B, λ_B	1.14	40
Average	0.46	16
Change in risk pricing parameters ($\lambda_A \to \lambda_B$)		
ρ_A, Σ_A, $\lambda_A \to \rho_A$, Σ_A, λ_B	3.66	127
ρ_A, Σ_B, $\lambda_A \to \rho_A$, Σ_B, λ_B	3.71	129
ρ_B, Σ_A, $\lambda_A \to \rho_B$, Σ_A, λ_B	2.34	81
ρ_B, Σ_B, $\lambda_A \to \rho_B$, Σ_B, λ_B	3.26	113
Average	3.24	112

Note: Each model is identified by a parameter triple, where ρ_A, Σ_A, λ_A, ρ_B, Σ_B, and λ_B represent the estimates of ρ, Σ, and λ_1 in subsamples A and B, respectively. The differences in the population estimates of β_{120} for each pair of models is reported in the middle column. The contribution to the total effect on β_{120} is calculated as the ratio (in percent) of the effect on β_{120} for any particular pair of models to the total effect given in the first row.

Table 5

The Effect of Model Changes (by Individual Parameters) on β_{120} Estimates

Parameter being changed (subsample A to B estimate)	Average effect on β_{120} from parameter change	Average contribution to total effect on β_{120} estimate
ρ_{LL}	-1.13	-39.29
ρ_{SS}	0.13	4.61
ρ_{SL}	-0.08	-2.78
λ^1_{LL}	1.24	43.20
λ^1_{LS}	2.23	77.41
λ^1_{SL}	0.02	0.76
σ_L	0.72	24.88
σ_S	-0.22	-7.70

Table 6

Parameter Estimates of the Macro-Finance Model

Factor dynamics					
ρ_L	0.989	(0.0068)	g_π	1.253	(0.0066)
ρ_S	0.026	(0.0111)	g_y	0.200	(0.0066)
ρ_u	0.975	(0.0062)			
Inflation dynamics					
μ_π	0.074	(0.0113)	$\alpha_{\pi 1}$	1.154	(0.0525)
α_y	0.014	(0.0074)	$\alpha_{\pi 2}$	-0.155	(0.0066)
Output dynamics					
μ_y	0.009	(0.0066)	β_{y1}	0.918	(0.0604)
β_r	0.089	(0.0067)	β_{y2}	0.078	(0.0066)
Risk price					
λ^1_{LL}	-0.0045	(0.0068)	λ^1_{LS}	0.0168	(0.0068)
λ^1_{SL}	-0.0223	(0.0064)	λ^1_{SS}	0.0083	(0.0067)
Standard deviations					
σ_L	0.342	(0.0089)	σ_π	0.238	(0.0110)
σ_S	0.559	(0.0313)	σ_y	0.603	(0.0128)
Standard deviations of measurement error					
σ_3	0.288	(0.0162)			
σ_{12}	0.334	(0.0194)			
σ_{36}	0.127	(0.0094)			

Note: Standard errors are in parentheses.

Table 7

The Effect of Macro-Finance Model Changes on β_{120} Estimates

Model permutation	Effect on β_{120} from parameter change
(1) $\sigma_L \rightarrow \sigma_L$ in Subsample A	-0.32
(2) $\lambda^1_{LL} \rightarrow \lambda^1_{LL}$ in Subsample A	0.02
(3) $\lambda^1_{LS} \rightarrow \lambda^1_{LS}$ in Subsample A	-0.54
(1) + (2) + (3)	-2.05
(4) $g_\pi \rightarrow 0.5 \times g_\pi$	-0.11
(5) $\sigma_\pi, \sigma_y \rightarrow 1.5 \times (\sigma_\pi, \sigma_y)$	0.12

Table A1

Simulated Frequency of Rejection of Expectations Hypothesis

DGP	Individual long-rate regressions 3-month	60-month	3- and 60-month	System SUR (3- and 60-month) 3- and 60-month
	Empirical size – Each DGP model modified so that $\lambda_1 = 0$			
Full Sample Model	0.077	0.067	0.057	0.077
Subsample A Model	0.080	0.085	0.052	0.121
Subsample B Model	0.113	0.090	0.076	0.127
	Power – Each DGP model as estimated with nonzero λ_1			
Full Sample Model	0.082	0.309	0.062	0.242
Subsample A Model	0.125	0.453	0.084	0.233
Subsample B Model	0.081	0.047	0.022	0.301

Note: Based on asymptotic 5 percent critical values, these numbers are the frequency of rejection over 1000 simulated data samples of the hypothesis that the slope coefficient is unity. The bottom panel uses the ML model estimates from Table 3 in the data-generating process (DGP), while the top panel sets $\lambda_1 = 0$ in each model, so the expectations hypothesis null is true. The data samples simulated from the full-sample model, the subsample A model, and the subsample B model have 396, 216, and 180 observations, respectively.

Table A2

Simulated Frequency of Rejection of No-Break Hypothesis

DGP	Individual long-rate regressions 3-month	60-month	3- and 60-month	System SUR (3- and 60-month) 3- and 60-month
	Empirical size – No shift in DGP coefficients			
Full-sample model	0.051	0.038	0.013	0.027
	Power – DGP Shift from subsample A to subsample B coefficients			
Split-sample model	0.660	0.648	0.548	0.609

Note: Based on asymptotic 5 percent critical values, these numbers are the frequency of rejection over 1000 simulated data samples of the hypothesis that the slope coefficient does not shift. Each data sample in the top panel has 396 observations from the full-sample ML model in Table 3. In the bottom panel, each data sample has 216 observations from the subsample A model and then 180 observations from the subsample B model.

Figure 1
Estimated Coefficients from Long-Rate Regressions

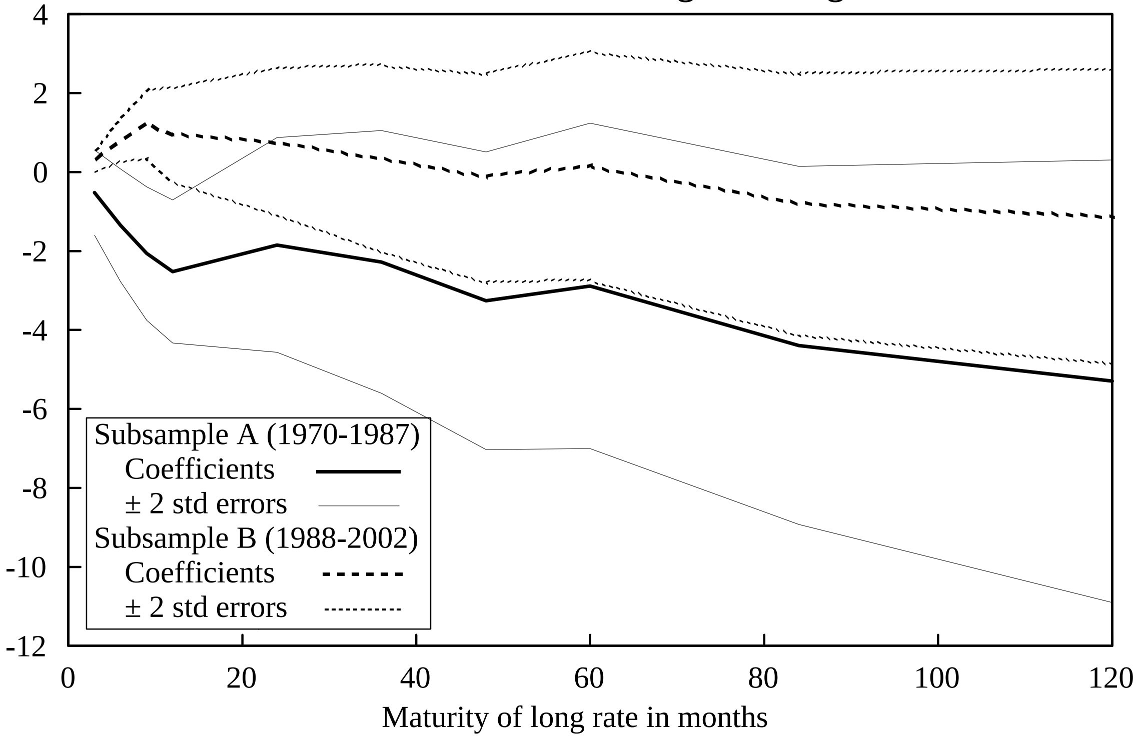

Figure 2
Test Statistics for a System Regression Shift at Various Breakdates

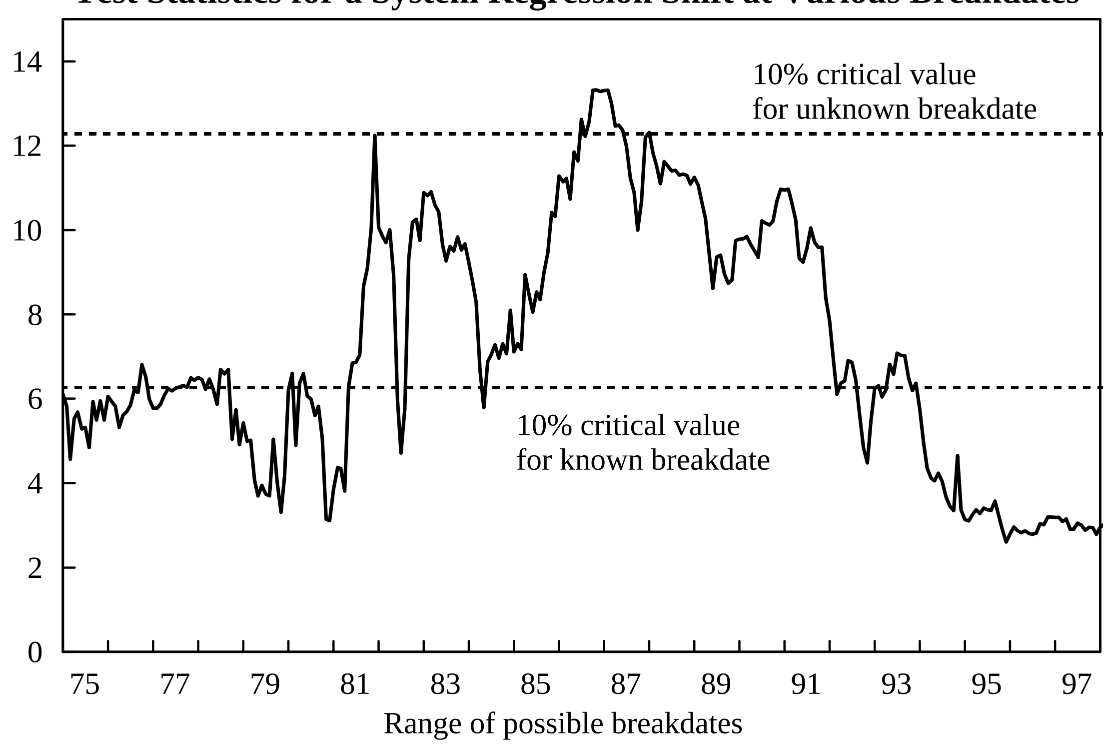

Figure 3
Regression Coefficients Implied by Subsample A No-Arbitrage Model

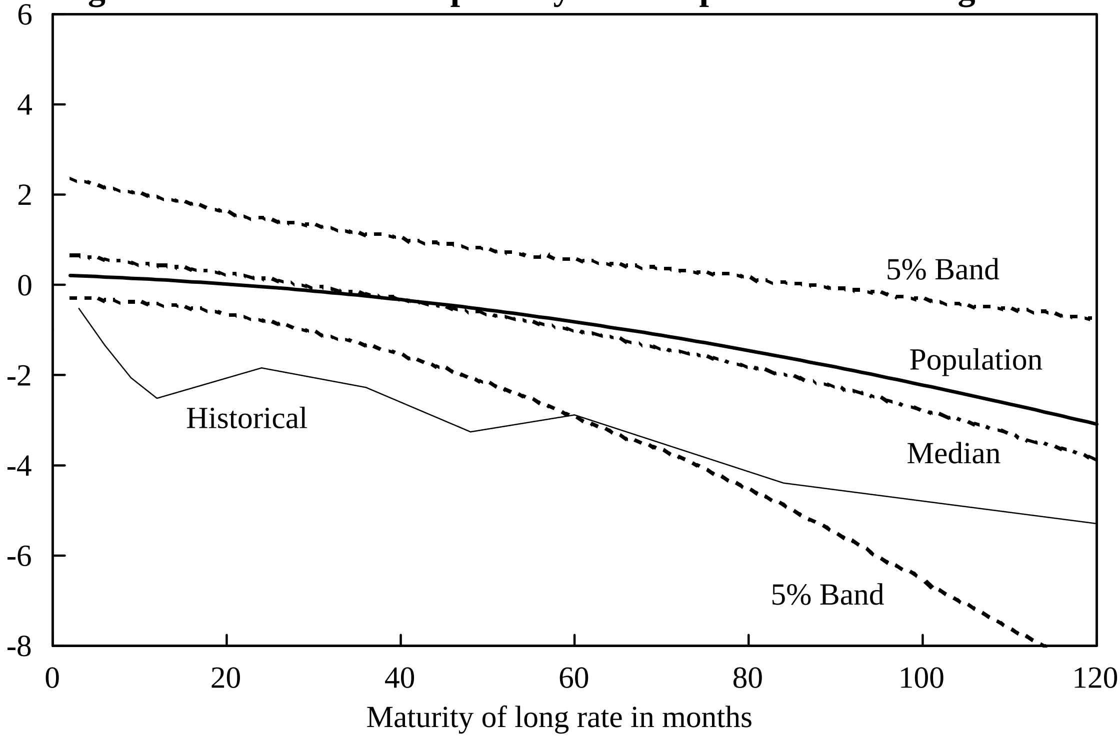

Figure 4
Regression Coefficients Implied by Subsample B No-Arbitrage Model

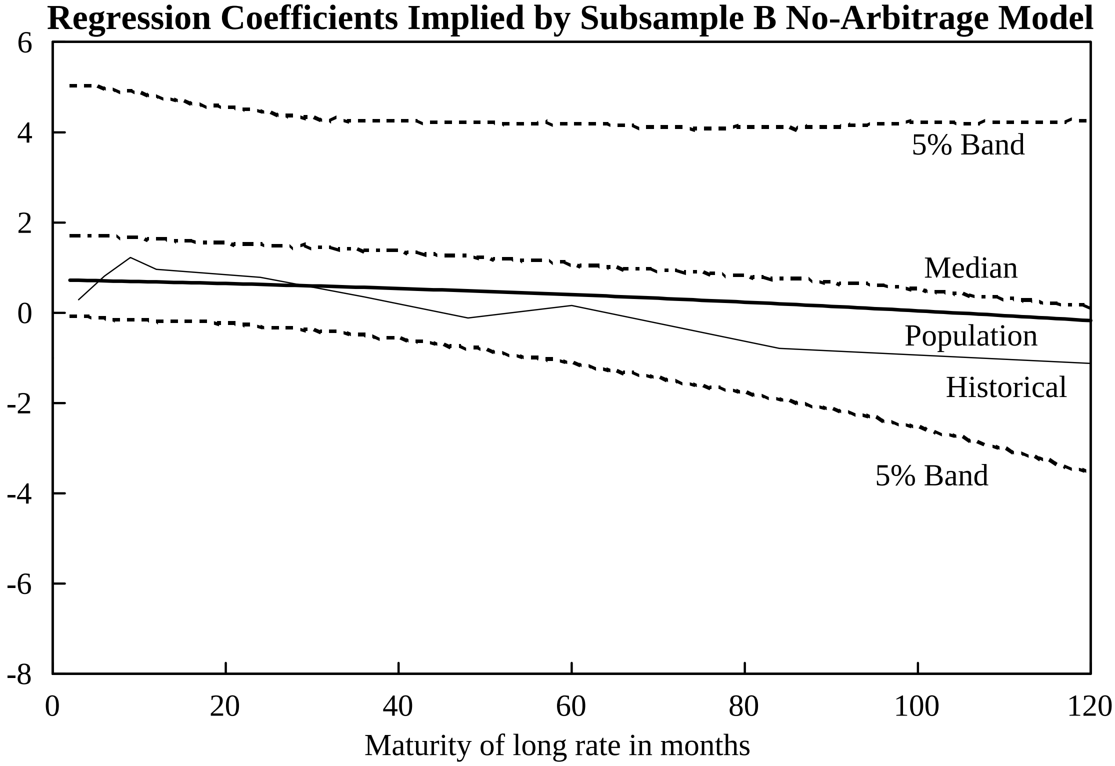

Figure 5
Regression Coefficients Implied by Macro-Finance Model