Predictability of Bond Risk Premia and Affine Term Structure Models

Qiang Dai, Kenneth J. Singleton, and Wei Yang

This draft: June 6, 2004

1Dai is with the Stern School of Business, New York University, New York, NY, qdai@stern.nyu.edu. Singleton is with the Graduate School of Business, Stanford University, Stanford, CA 94305 and NBER, ken@future.stanford.edu. Yang is a Ph.D. student at the Graduate School of Business, Stanford University, wy29@stanford.edu. We are grateful for financial support from the Gifford Fong Associates Fund, at the Graduate School of Business, Stanford University.
Abstract

This paper explores the question of whether dynamic term structure models generate patterns of predictability in excess returns consistent with the recent historical evidence for U.S. Treasury bond yields. Our entire empirical analysis is carried out using four different data sets on zero-coupon bonds, in order to check robustness of key results in the literature to the spline methodology used to construct the zero coupon bond yields. Among the issues addressed are: (i) Is the degree of predictability documented in descriptive regression analyses largely the same as the degree of predictability implicit in multi-factor arbitrage-free term structure models?; (ii) Is there a single common “predictor factor” underlying the expected excess returns on bonds of all maturities?; (iii) To what extent is the evidence for predictability of excess returns a symptom of small-sample biases in estimated $R^2$? Our answer to the first question is yes: the degree of predictability documented in historical data is roughly the same as that implied by standard multi-factor dynamic term structure models. At the same time, treating a Gaussian three-factor model as the true data generating process, we find that the $R^2$ statistic for these regressions has a substantial upward small-sample bias of roughly 50%. Finally, we present strong evidence against the “common factor” hypothesis (ii). While most of our results are robust to the choice of spline, we argue that a portion of the predictability in excess returns in some data sets is induced by the choice of spline, and that patterns in the coefficients in prediction regressions are highly sensitive to the smoothness of the spline.


1 Introduction

Empirical evidence that expected excess returns on U.S. Treasury bonds exhibit predictable variation over time – equivalently, that there is predictable variation in bond risk premia – dates back at least to the compelling studies by Fama and Bliss [1987] and Campbell and Shiller [1991]. Subsequent studies have tended to reinforce these early findings through the inclusion of additional forecasting variables (see, e.g., Cochrane and Piazzesi [2002] and the references therein). The force of this evidence naturally gives rise to the question of whether dynamic term structure models (DTSMs) generate patterns of predictability in excess returns consistent with this historical evidence. This paper explores this issue in depth using historical data on U.S. Treasury bond yields.

In answering this question, we examine several auxiliary hypotheses that also have received considerable attention in the literature. Specifically, (i) Is the predictive power of forward rates largely the same as that of the first two or three principal components of bond yields? This inquiry is motivated by the high degree of correlation between the state variables in DTSMs and these principal components (PCs). (ii) Is there a single common “predictor factor” underlying the expected excess returns on bonds of all maturities? From their descriptive regressions of excess returns onto forward rates, Cochrane and Piazzesi [2002] (hereafter CP) conclude that a common “forward factor” underlies the predictability of holding period returns on U.S. Treasury bonds. We revisit their conclusion using forward rates and PCs, both descriptively and within the context of DTSMs. (iii) To what extent is the evidence for predictability of excess returns a symptom of small-sample biases in estimated $R^2$? We answer this question by assuming that the conditional first moments of excess returns are well described by a three-factor DTSM, and then examining where the historical $R^2$ lies within the small-sample distribution of the $R^2$ statistic implied by this DTSM.

A distinct, but complementary, objective of our analysis is an assessment of the robustness of the findings in both literatures – on DTSMs and on excess return predictability – to the construction of the zero-coupon bond yields used as inputs into past empirical analyses. Many of the recent studies of predictability of changes in bond yields or excess returns have used data on zero-coupon bond yields obtained, by various curve fitting methods, from market prices of coupon bond yields. Using this pre-processed data – effectively yields on hypothetical bonds that are not actually traded – has facilitated empirical analysis, because many of the expectational relations of interest take simple linear forms for zero-coupon yields. However, with this convenience comes a potential cost. Namely, both the zero-coupon bond yields and the associated forward rates are derived numbers. It follows that regressions of changes in yields or of excess returns on past yields or forward rates may give results that are influenced by the curve fitting methods used to extract zero yields from coupon bond yields. The literature has largely assumed (at least implicitly) that, for the purposes employed, the findings are robust across these alternative curve-fitting methods. We explore this robustness issue more systematically by reporting our major findings for four different spline methods for computing zero yields.

Throughout this paper we focus on the three-factor Gaussian $A_0(3)$ DTSM because of its prior success in forecasting levels of bond yields. Duffee [2002] showed that his extended
specification of the market prices of risk (MPR) in an $A_0(3)$ model provides better out-of-sample forecasts than the specification of MPRs in Dai and Singleton [2000]. Similarly, Dai and Singleton [2002] show that, using a similarly flexible specification of the MPR, Gaussian affine DTSMs are able to resolve the predictability puzzles associated with violations of the expectations theory of the term structure. Moreover, this choice is natural given our focus on the conditional means of bond yields. Our conclusions extend immediately to regime-switching Gaussian models (e.g., Dai, Singleton, and Yang [2003] and Ang and Bekaert [2003]) as these models effectively nest model $A_0(3)$.\footnote{We have confirmed this by fitting the model in Dai, Singleton, and Yang [2003] to the four data sets examined here.}

We conjecture that our qualitative findings will also carry over to quadratic-Gaussian models, because the market prices of risk in these models are of exactly the same form as in the affine model $A_0(3)$.\footnote{See Leippold and Wu [2002] and Kim [2004] for some related evidence on the predictability of excess returns in quadratic-Gaussian models.}

The remainder of this paper is organized as follows. Section 2 describes the spline methods used to derive the zero-coupon bond yields and shows that the pattern of coefficients from the projections of excess returns onto forward rates is not a robust feature of zero data. Section 3 presents descriptive evidence on the degree of predictability of excess returns using forward rates or the principal components of zero yields. The degrees of predictability based on forward rates or principal components are very similar, suggesting that these series embody essentially the same information about future excess returns. Two systematic features of these results are notable: the more flexible the spline (the more “choppy” is the underlying forward curve), the higher the $R^2$s in predicting excess returns with forward rates; and the lower is the percentage of this explained variation that is captured by the first few PCs.

The heart of our analysis is in Section 4 where we compare the predictability of excess returns implied by affine DTSMs to the sample predictability. We find that, for the most part, DTSMs replicate what we see in the data. Yet, within the family of DTSMs examined, the evidence also points towards an upward small-sample bias in the sample $R^2$ statistics. That is, our models suggest that excess returns are not as predictable as the historical data suggests, with the upward bias being as large as 50% for some model-dataset pairs.

Section 5 turns to the question of whether there is a single common factor underlying the predictability of excess returns. Formal (large-sample) tests provide strong evidence against this hypothesis. While this might seem surprising in the light of the evidence reported in $CP$, we show that the appearance of a common “forward factor” in their analysis is an artifact of the particular rotation of the bond yields used as predictors in their study.

Finally, we briefly reassess the evidence for predictability in Fama and Bliss [1987] through the lens of our DTSM and explore the nature of the predictability of excess returns over longer holding periods.


2 Splines and Zero Coupon Bond Yields

Four curve-fitting methods are used to derive zero-coupon bond yields from coupon bond yields, all of which have been examined in the recent empirical literature on term structure models: Unsmoothed Fama-Bliss (UFB), Fisher-Waggoner (FW), Smoothed Fama-Bliss (SFB), and Nelson-Siegel-Bliss (NSB). This section briefly describes these methods and explores the sensitivity of the coefficients in forward rate regressions to the spline method used to compute zero yields.

All four data sets – UFB, FW, SFB, and NSB – are generated using “The Bliss Term Structure Generating Programs”, and thus they are derived from the same set of underlying coupon bond prices. All data are monthly and cover the same period 1970 to 2000. This assures that any differences across data sets are due entirely to the spline methodology applied, and not to the use of different underlying coupon bond yields.

2.1 Spline Methods

The Unsmoothed Fama-Bliss method (Fama and Bliss [1987]) is an iterative method that extracts forward rates from coupon bond prices by building a piece-wise linear discount rate function. The forward rate is a (discontinuous) step function with respect to maturity, so the implied discount rates exhibit kinks at the maturities of the coupon bonds used.

The Fisher-Waggoner method uses a cubic spline to approximate the forward rate function itself. This method employs a large number of knots (as many as 50 to 60 knots) when minimizing the sum of squared fitted-price errors, and then a penalty is imposed on the excess variability of yields induced by the flexibility of the spline.

The Nelson-Siegel-Bliss method (based on Nelson and Siegel [1987]) approximates the discount rate with exponential functions of time to maturity. The Smoothed Fama-Bliss method starts with the piece-wise linear discount rates obtained from the Unsmoothed Fama-Bliss method, and resulting zero curves are “smoothed out” by fitting an NSB approximating function through the unsmoothed data. Since asymptotically flat exponential approximating functions are used, the resulting forward rate function is differentiable to infinite order.

3Bliss [1997] provides a more detailed description of these methods. The Nelson-Siegel-Bliss method was originally labeled the extended Nelson-Siegel method in Bliss [1997]. The Fisher-Waggoner method (Waggoner [1997]) is a modified version of the Fisher-Nychka-Zervos method (Fisher, Nychka, and Zervos [1995]).

4In particular, we use the filtered “long data set”, which filters out bonds with option features and liquidity problems, but otherwise contains all eligible issues (Bliss [1997]).

5CP checked the robustness of their findings on two data sets, the UFB data from CRSP, and the McCulloch-Kwon (MK) data. The latter data set, which we do not examine here, is based on a modified version of the McCulloch procedure (McCulloch [1975]). A cubic spline is used to approximate the discount function, and the spline is estimated with ordinary least squares. Since the estimated discount function is twice differentiable, this results in an instantaneous forward rate with a continuous first derivative.

A potential limitation of cubic splines is that they may lead to discount rates that approach positive or negative infinity as maturity is increased without bound.

7The original Nelson-Siegel method has only four free parameters. The Nelson-Siegel-Bliss method has five parameters free so as to provide a better fit for longer maturities (Bliss [1997]).
Table 1: Unsmoothed vs. Smoothed Fama-Bliss yields. $\hat{a}$ and $\hat{b}$ are the constant and slope coefficients from the following regression of the $UFB$ yields on the $SFB$ yields (monthly, 1970–2000): $y_{UFB}^{n,t} = a + b y_{SFB}^{n,t} + \epsilon_{n,t}$, where $y_{n,t}$ is the yield on an $n$-year zero coupon bond issued at date $t$.

| Maturity (year) | $\hat{a}$ (bp) | $\hat{b}$ | $R^2$ | SE($\epsilon$) (bp) | max $|y_{UFB} - y_{SFB}|$ (bp) |
|-----------------|----------------|---------|-------|---------------------|-------------------------------|
| 1               | -0.95          | 0.9994  | 0.9987 | 9.17                | 41.4                          |
| 2               | -4.90          | 1.0049  | 0.9990 | 7.81                | 40.3                          |
| 3               | -0.93          | 0.9999  | 0.9991 | 6.91                | 51.7                          |
| 4               | -1.03          | 1.0015  | 0.9986 | 8.58                | 47.6                          |
| 5               | -6.75          | 1.0057  | 0.9987 | 7.97                | 38.5                          |

Absent a formal economic model, a definitive choice among these methods seems infeasible. Which method one chooses in practice will likely depend on one’s priors about the smoothness of the zero and forward rate curves implicit in the market-generated coupon bond data. If one’s prior is that zero curves are relatively choppy (and forward rate curves have multiple discontinuities), then the $UFB$ method may be a reasonable choice. Alternatively, if one prefers smoother discount curves and differentiable forward rate curves, then one among the $FW$, $SFB$, and $NSB$ methods may be preferred.\(^8\)

All four splined zero curves are based on the same set of coupon bond prices. Hence, we expect high correlations among the resulting zero yields. Indeed this is the case, as is illustrated in Table 1 where we display the results from projecting $UFB$ yields onto the $SFB$ yields with matching maturities (plus a constant term) for the 1970–2000 sample period. The slopes and the $R^2$s are very close to unity, and the fitted constant terms are only a few basis points. The standard deviations of the residuals are about 7 to 9 basis points, thus confirming that the two data sets are almost perfectly correlated.\(^9\)

Nevertheless, as illustrated in Figure 1, these methods do generate different zero coupon bond yield curves. The $UFB$ yield curve, constructed using discontinuous forward rate functions is relatively choppy compared with all of the other curves. The Fisher-Wagner ($FW$) data uses the flexibility of the 50 to 60 knot points to smooth out the underlying coupon bond yields, without foregoing the many small dips and humps in the latter data. The curves from the $SFB$ and $NSB$ curves are much smoother. Even at the major yearly maturities (indicated by the symbols) there are differences in the spline-implied zero yields.

---

\(^8\)It is our impression that financial institutions typically employ relatively smooth, differentiable forward curves in their pricing models.

\(^9\)We repeated this analysis on other pairs of data sets, such as $FW$ and $SFB$, $UFB$ and $NSB$, etc. Again, despite the very different methods used to build zeros in these cases, the pairs were also nearly perfectly correlated.
2.2 A Tent-Shaped Forward Factor?

One of the notable findings by CP is that the coefficients in the projections of excess returns over one-year hold periods on the one-year forward rates, $f_{0\rightarrow1}$, $f_{1\rightarrow2}$, $f_{2\rightarrow3}$, $f_{3\rightarrow4}$, and $f_{4\rightarrow5}$, out to five years exhibit a tent-like pattern. This pattern is replicated in Figure 2(a) for the UFB data over the sample period 1970 – 2000, where it is seen that the tent shape is very similar across the maturities of the underlying bonds.

However, as illustrated by Figures 2(b) and 2(c), for the NSB and SFB data sets, respectively, this is not a robust feature of zero coupon bond yields. Rather, two of the four estimation methods, UFB and FW, give rise to (approximately) tent-shaped patterns for the projection coefficients, while the other two, SFB and NSB, produce wave patterns. In the case of the SFB data, for example, the wave pattern loads positively on the two- and four-year forward rates and negatively on the three- and five-year forward rates.

Checking sub-periods using the UFB data, the tent shape is (more roughly) sustained. In contrast, for most of the other data sets there is considerable variation over sub-periods in the coefficient patterns. For the SFB data, the post-1983 sub-sample gives a similar pattern to the entire sample. On the other hand, for the pre-1979 period, the wave shape is inverted, with negative loadings at two and four years and positive loadings at three and five years.

These differences across data are striking in the light of the near perfect correlations.
among the spline-implied zero yields of the same maturity. Our suspicion was that even small differences in the zero yields due to differing degrees of smoothing from the spline methodologies could explain these very different patterns. Pursuing this idea, using the regression results in Table 1, we estimated first-order autoregressive models for each of the residuals from the five regressions.\textsuperscript{10} Using the estimated parameters, we then simulated shocks from a conditional normal distribution with the same means, variances, and autocorrelations and added these shocks to the corresponding $SF_B$ yields. Finally, we re-ran the excess return regressions for the resulting “noisy” $SF_B$ yields. The pattern obtained is the distorted tent shape in Figure 2(d).\textsuperscript{11}

\textsuperscript{10}We fit univariate processes, because when we estimated a vector-autoregression for all five residual series, only the coefficients on the “own lags” were statistically significant.

\textsuperscript{11}The autoregressive component of the errors is not critical to replicating the tent shape in the error-perturbed $SF_B$ yields. When we perturb the $SF_B$ yields with i.i.d. errors that match the standard deviation of the residuals from Table 1, we obtain essentially the same results.

Figure 2: Slope coefficients from the projections of one-year excess returns on the one-year forward rates over the sample period 1970–2000. The legend refers to the maturity of the zero-coupon bond used to compute excess returns.
These calculations illustrate a more general principle. Namely, in projections of excess returns onto forward rates, the data based on the relatively choppy forward curves will give rise to projection coefficients that are “biased” in the direction of having a tent-like shape. We develop this idea formally in Appendix A.

Whether the wave-like patterns in Figures 2(b) and 2(c) are more representative of the “true” pattern is unclear. Surely the forward curves used in the construction of all four data sets are mismeasured, at least to some degree. We have shown that even small measurement errors, in the presence of highly correlated forward rates, can lead to large differences in fitted projection coefficients. Accordingly, to assign economic significance to any patterns obtained, it seems essential to start with an economic theory that identifies the forward curves directly from the market prices of coupon bonds. The adoption of an arbitrage-free DTSIM is one means of introducing such economic structure.

Of course one might reasonably argue that it is not the shape of the projection coefficients per se (tent-shaped or otherwise) that ultimately is of interest, but rather the fact that a single common forward factor appears to underlie the predictability of the holding period returns for bonds of all maturities. As in CP’s analysis, all of our results so far are suggestive of this conclusion in that the sample $R^2$s and the patterns of the projection coefficients are roughly the same across the maturities of the underlying bonds. We explore the statistical and economic underpinnings of the single-factor view in more depth in Section 5.

3 Predictability Regressions Using Forward Rates and Principal Components

Perhaps the most striking feature of recent studies of predictability is the high levels of $R^2$s obtained. In this section we first examine the robustness of this finding to the method of constructing the zero yields. Then we ask whether the first three PCs of zero yields, what are often referred to as the “level”, “slope”, and “curvature” factors, give rise to similar levels of predictability.

3.1 Prediction Using Forward Rates

Following CP, we regress excess returns over the one-year holding period onto five one-year forward rates (known today for one-year loans commencing in the future at annual intervals). Table 2 confirms that these regressions give relatively large $R^2$, though there are differences both across data sets and sample periods. For the entire sample period, the values of the $R^2$s obtained using the data sets FW, SFB, and NSB are largely insensitive to both the maturity of the bond underlying the construction of the excess return or the spline method used to compute zero yields. For UFB, the $R^2$ are somewhat higher (by about 4-6%), but they are also largely the same across bond maturities.

Interestingly, this difference between UFB and the other data sets seems to largely disappear when the sample is split into two subperiods by omitting the period between 1979 and 1982 when the Federal Reserve changed to a monetary target for setting monetary
Table 2: \( R^2 \) for predictability regressions using one-year forward rates, for different data sets and sample periods.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UFB FW SFB NSB</td>
<td>UFB FW SFB NSB</td>
<td>UFB FW SFB NSB</td>
</tr>
<tr>
<td>2</td>
<td>0.36 0.32 0.30 0.30</td>
<td>0.28 0.30 0.28 0.25</td>
<td>0.44 0.49 0.40 0.43</td>
</tr>
<tr>
<td>3</td>
<td>0.37 0.33 0.31 0.31</td>
<td>0.29 0.30 0.28 0.25</td>
<td>0.41 0.46 0.39 0.42</td>
</tr>
<tr>
<td>4</td>
<td>0.39 0.35 0.31 0.32</td>
<td>0.31 0.32 0.27 0.25</td>
<td>0.43 0.47 0.38 0.42</td>
</tr>
<tr>
<td>5</td>
<td>0.36 0.34 0.32 0.32</td>
<td>0.30 0.31 0.26 0.24</td>
<td>0.40 0.45 0.37 0.40</td>
</tr>
</tbody>
</table>

3.2 Predictability Regressions Using PCs

It is instructive to recast this evidence of predictability in terms of the PCs of zero coupon bond yields. Taking account of all five PCs, these PCs span the same linear space as the five forward rates and, therefore, they have the same predictive content. Our interest is in the degree of predictive content of the first three PCs, since they have a close link to the three factors studied in many DTSMs (e.g., Dai and Singleton [2000] and Duffee [2002]). PCs are also orthogonal, by construction, and thus using them avoids the high degree of multi-collinearity inherent in the use of forward rates.

To address this issue, we constructed principal components from each of four data sets and then, for each data set, regressed the excess returns on the first three PCs. Table 3 shows that the first three PCs from data sets FW, SFB, and NSB account for well over 90% of the predictive power of forward rates for excess returns. The notable exception is data set UFB, where the percentages range between 76% and 83%. Notice also (from Table 3 Panel B) that by far the majority of the explanatory power for excess returns of these PCs is due to PC2, the “slope” of the yield curve. The smallest percentage contribution for PC2 is again for data set UFB.

These findings suggest that the fourth and fifth PCs, and their relationships to excess returns, in data set UFB are different than for the other three data sets. One way to see that this is indeed the case is to examine the coefficients from the projections of excess returns onto the five PCs. Table 4 shows that the coefficients on the first two PCs, the “level” and the “slope” factors, are very similar across all four data sets. As we move out the list of PCs, the magnitudes of the coefficients become increasingly different across data sets. For PC5, the differences are large with the magnitudes being positive for the choppiest data (UFB) and then declining monotonically to large negative numbers as the zero data

\[\text{12} \]The loadings of PCs on yields are indeterminate up to a positive/minus sign, and the exact choice depends on the computer algorithm. Whenever the sign choice has an effect on the results (for example, the sign of the regression coefficients on PCs), we follow the following convention. The loadings for PC1 on the yields are all positive. The loadings for PC2 are upward sloping with respect to maturity. The loadings for PC3 are negative on the short and long ends, and positive on the intermediate maturities, and so on.
Table 3: Contributions of the PCs to the $R^2$ from the regressions of excess returns on all five PCs.

becomes increasingly smooth.

That the variation in yields associated with the fifth PC in data set UFB is “excess” relative to the variation in the yields from other data sets is seen from Table 5. The volatilities of the first three PCs are quite similar across data sets. However, the volatilities of PC4 and PC5 are larger in data set UFB than in the other data sets. It is this extra variation in data set UFB that is not explained by the usual “level”, “slope,” and “curvature” factors, and that underlies the differences in Table 3.A.

These differences, that largely show up on the properties of the fourth and fifth PCs, are entirely attributable, of course, to the choice of spline methodology used to construct the zero coupon yields. What seems striking is how much even small differences in the smoothness of the zero curves affect the properties of the PCs. For instance, comparing FW and UFB, the former appears to merely smooth out some of the roughness retained in the latter (see Figure 1), yet the proportion of the predictability in excess returns captured by the first three PCs increases from between 75%-83% in the UFB data to over 90% in the FW data. We say more about nature of the predictability in the UFB data set that is not captured by the first three PCs in Section 7.

4 Excess Return Predictability in Affine DTSMs

The predictability regressions studied by Fama and Bliss [1987] and Cochrane and Piazzesi [2002], among others, are linear projections of the excess holding period returns on linear combinations of yields. These linear projections have an exact counterpart in affine DTSMs, as zero-coupon bond yields, and their implied forward rates and PCs, are all linear functions of the state vector which itself has a linear conditional mean. In this section we explore the degree of predictability of excess returns implied by $A_0(3)$ DTSMs, with the goal of assessing whether the descriptive evidence is matched by arbitrage free DTSMs.

The $A_0(3)$ model assumes that, under the historical measure $\mathbb{P}$, the $N \times 1$ state factor $Y$
### Table 4: Coefficients from the projections of excess returns to holding an $n$-year bond for one year onto PCs for the sample period 1970-2000.

<table>
<thead>
<tr>
<th>Data</th>
<th>const.</th>
<th>$PC1$</th>
<th>$PC2$</th>
<th>$PC3$</th>
<th>$PC4$</th>
<th>$PC5$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: 2-year bond</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$UFB$</td>
<td>0.0055</td>
<td>0.1003</td>
<td>1.3492</td>
<td>3.3983</td>
<td>6.4428</td>
<td>0.9943</td>
</tr>
<tr>
<td>$FW$</td>
<td>0.0054</td>
<td>0.0931</td>
<td>1.3809</td>
<td>3.8050</td>
<td>6.1916</td>
<td>-1.3166</td>
</tr>
<tr>
<td>$SFB$</td>
<td>0.0055</td>
<td>0.0931</td>
<td>1.3550</td>
<td>4.6544</td>
<td>8.0302</td>
<td>-83.3516</td>
</tr>
<tr>
<td>$NSB$</td>
<td>0.0056</td>
<td>0.0926</td>
<td>1.3489</td>
<td>4.5167</td>
<td>7.3497</td>
<td>-118.0631</td>
</tr>
<tr>
<td><strong>Panel B: 3-year bond</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$UFB$</td>
<td>0.0085</td>
<td>0.1486</td>
<td>2.5633</td>
<td>5.6305</td>
<td>13.3937</td>
<td>5.5403</td>
</tr>
<tr>
<td>$FW$</td>
<td>0.0082</td>
<td>0.1436</td>
<td>2.5689</td>
<td>6.7347</td>
<td>14.1288</td>
<td>-0.1123</td>
</tr>
<tr>
<td>$SFB$</td>
<td>0.0084</td>
<td>0.1469</td>
<td>2.5729</td>
<td>8.1283</td>
<td>16.9284</td>
<td>-225.3335</td>
</tr>
<tr>
<td>$NSB$</td>
<td>0.0084</td>
<td>0.1467</td>
<td>2.5624</td>
<td>7.7860</td>
<td>15.7404</td>
<td>-290.0392</td>
</tr>
<tr>
<td><strong>Panel C: 4-year bond</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$UFB$</td>
<td>0.0111</td>
<td>0.1814</td>
<td>3.6952</td>
<td>7.2509</td>
<td>20.4826</td>
<td>2.5542</td>
</tr>
<tr>
<td>$FW$</td>
<td>0.0109</td>
<td>0.1857</td>
<td>3.7430</td>
<td>8.6906</td>
<td>20.9955</td>
<td>-7.5994</td>
</tr>
<tr>
<td>$SFB$</td>
<td>0.0105</td>
<td>0.1810</td>
<td>3.6964</td>
<td>10.8716</td>
<td>18.6808</td>
<td>-368.8545</td>
</tr>
<tr>
<td>$NSB$</td>
<td>0.0105</td>
<td>0.1807</td>
<td>3.6793</td>
<td>10.3041</td>
<td>18.2554</td>
<td>-453.2284</td>
</tr>
<tr>
<td><strong>Panel D: 5-year bond</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$UFB$</td>
<td>0.0111</td>
<td>0.2169</td>
<td>4.6044</td>
<td>8.7377</td>
<td>20.4875</td>
<td>4.2047</td>
</tr>
<tr>
<td>$FW$</td>
<td>0.0120</td>
<td>0.2063</td>
<td>4.7527</td>
<td>10.6847</td>
<td>22.0768</td>
<td>-7.8984</td>
</tr>
<tr>
<td>$SFB$</td>
<td>0.0120</td>
<td>0.2073</td>
<td>4.7466</td>
<td>13.3017</td>
<td>14.5760</td>
<td>-490.0547</td>
</tr>
<tr>
<td>$NSB$</td>
<td>0.0119</td>
<td>0.2062</td>
<td>4.7310</td>
<td>12.5427</td>
<td>15.4478</td>
<td>-589.5576</td>
</tr>
</tbody>
</table>

### Table 5: Standard deviations of PCs in basis points.

<table>
<thead>
<tr>
<th>Data</th>
<th>$PC1$</th>
<th>$PC2$</th>
<th>$PC3$</th>
<th>$PC4$</th>
<th>$PC5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$UFB$</td>
<td>528</td>
<td>67.3</td>
<td>10.6</td>
<td>7.39</td>
<td>6.73</td>
</tr>
<tr>
<td>$FW$</td>
<td>528</td>
<td>66.2</td>
<td>9.9</td>
<td>4.99</td>
<td>2.47</td>
</tr>
<tr>
<td>$SFB$</td>
<td>527</td>
<td>64.2</td>
<td>9.0</td>
<td>1.47</td>
<td>0.16</td>
</tr>
<tr>
<td>$NSB$</td>
<td>528</td>
<td>64.7</td>
<td>8.9</td>
<td>1.45</td>
<td>0.16</td>
</tr>
</tbody>
</table>
follows the process
\[ Y_{t+1} = Y_t + \kappa^p (\theta^p - Y_t) + \Sigma \epsilon_{t+1}, \] (1)
where \( \Sigma \) is a constant volatility matrix and \( \epsilon_{t+1} \sim N(0, I) \). The market price of factor risk is assumed to be
\[ \Lambda_t = \Sigma^{-1}(\lambda_0 + \lambda_Y Y_t), \] (2)
so the risk-neutral \( \kappa^Q = \kappa^p + \lambda_Y \). Finally, the short rate follows \( r_t = \delta_0 + \delta'_Y Y_t \). These assumptions imply that the log of price of a zero coupon bond with a maturity of \( n \) periods is given by
\[ p_{n,t} = A_n + B'_n Y_t, \] (3)
for known scalar \( A_n \) and \( N \times 1 \) vector \( B_n \).

The excess return of holding a \( n \)-period bond for \( m \) periods is
\[ r_{n,t+m}^m = p_{n-m,t+m} - p_{n,t} + p_{m,t} = A_{n-m} - A_n + A_m + B'_{n-m} Y_{t+m} - B'_n Y_t + B'_m Y_t. \] (4)
Since \( Y \) is a Markov process, the expected excess return– the conditional mean of \( r_{n,t+m}^m \) given \( Y_t \)– is linear in \( Y_t \):
\[ r_{n,t+m}^m = C_{n,m} + L'_{n,m} Y_t + \nu_{t+1,t+m}, \] (5)
where
\[ L'_{n,m} = -\delta'_Y \kappa^{Q,-1} (I - \kappa^Q)^n [I - (I - \kappa^Q)^{-m} (I - \kappa^F)^m] - \delta'_Y \kappa^{Q,-1} [(I - \kappa^F)^m - (I - \kappa^Q)^m]. \] (6)
If the mean reversion matrices \( \kappa^Q \) and \( \kappa^p \) are the same– equivalently, the market price of factor risk is constant– then the \( L_{n,m} \) are zero. Excess returns are predictable only when investors’ market prices of risk are state dependent.

These results, as well as the associated \( R^2 \)s derived in Appendix B, can be re-expressed in terms of observed predictor variables, because of the linear relationship between the state \( Y \) and zero yields. In particular, the yield on an \( n \)-period zero is
\[ R_{n,t} = -\frac{p_{n,t}}{n} = -\frac{A_n}{n} - \frac{B'_n}{n} Y_t = -a_n - b'_n Y_t, \] (7)
the \( PC \)s of a set of bond yields \((R_{n_1,t}, \ldots, R_{n_K,t})\) are simply linear combinations of these yields, and the associated forward rates are
\[ f_{t}^{m-n} = p_{m,t} - p_{n,t} = A_m - A_n + (B_m - B_n)' Y_t. \] (8)
Fama and Bliss [1987] focus on a single forward spread
\[ f_{t}^{n-1-n} - R_{1,t} = p_{n-1,t} - p_{n,t} - R_{1,t} = A_{n-1} - A_n + A_1 + (B_{n-1} - B_n + B_1)' Y_t \] (9)
as a predictor of excess returns. Their explanatory variable therefore spans only a subspace of the information spanned by the forward rates or \( PC \)s.

\[ ^{13} \text{Details of the subsequent derivations in this subsection are given in Appendix B.} \]
4.1 How Much Predictability of Excess Returns is Implicit in Affine DTSMs?

For each of the four data sets, we estimate discrete-time, three-factor, Gaussian DTSMs by the method of maximum likelihood (ML), using monthly data over the common period of 1970–2000. Then, treating the ML estimates as the true population parameters, we undertake two complementary exercises. First, from each model, we simulate time series of bond yields of length $10^5$ and then, using the simulated data, we estimate the projections of excess returns onto the five forward rates. We interpret the resulting projection coefficients and $R^2$ as the population values implied by the models.\(^{14}\)

Second, to explore the small-sample properties of the model-implied $R^2$ in these projections, we simulated $10^4$ time series of yields, each of length $N$, estimated the projections of excess returns onto forward rates for each simulation, and then computed the sample mean and standard deviation of the $R^2$ across simulations. In other words, we computed the means and standard deviations of the small-sample distributions of the $R^2$ statistic implied by the $A_0(3)$ model. This exercise was undertaken for $N = 360$ and $N = 600$, corresponding to thirty and fifty years of data, respectively.

Given the linear dependence of excess returns on $Y$ in affine DTSMs, the number of latent factors limits the number of non-collinear forward rates that can be used in these regressions. For our three-factor models, we choose to use the forward rates $f^{0\rightarrow 1}$, $f^{2\rightarrow 3}$, and $f^{4\rightarrow 5}$, in addition to a constant term, as the regressors. To be consistent, we compare with the sample $R^2$s of the excess return regressions on the same three forward rates.\(^{15}\)

<table>
<thead>
<tr>
<th>Maturity</th>
<th>$UFB$</th>
<th>$FW$</th>
<th>$SFB$</th>
<th>$NSB$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Samp</td>
<td>360</td>
<td>600</td>
<td>Pop</td>
</tr>
<tr>
<td></td>
<td>Samp</td>
<td>360</td>
<td>600</td>
<td>Pop</td>
</tr>
<tr>
<td></td>
<td>Samp</td>
<td>360</td>
<td>600</td>
<td>Pop</td>
</tr>
<tr>
<td></td>
<td>Samp</td>
<td>360</td>
<td>600</td>
<td>Pop</td>
</tr>
<tr>
<td></td>
<td>Samp</td>
<td>360</td>
<td>600</td>
<td>Pop</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.33</td>
<td>0.26</td>
<td>0.23</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
<td>0.27</td>
<td>0.23</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.35</td>
<td>0.29</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
<td>0.31</td>
<td>0.28</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.08)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Monte Carlo simulations for model $A_0(3)$. For each dataset and model, the first column reports the $R^2$s for regressions estimated from the observed data, the second and third columns report the small-sample means and standard errors (in parenthesis) of the $R^2$s estimated from simulated samples of length 360 and 600 months, and the fourth column reports those in the population.

The model-implied population $R^2$s (Pop in Table 6) are very similar across all four data

\(^{14}\) We confirmed the reliability of the the analytical formulas in Section 4 and Appendix B using these simulations.

\(^{15}\) Similar conclusions can be drawn even if we compare with the sample $R^2$s obtained using all five forward rates, which, as reported in Table 2, are mostly only 1 or 2% larger.
sets. This is surely a consequence of the fact that the fitted yields from the three-factor models share similar smoothness properties, even though the underlying data sets are not equally smooth. Furthermore, the sample $R^2$s are consistently larger than their population counterparts (compare columns “Samp” and “Pop”). Additionally, the differences between these $R^2$s tend to be smaller for the data generated by the smoother splines.

The general tendency for population $R^2$s to be below their sample counterparts suggests that finite-sample $R^2$s are upward biased. To explore this possibility more systematically, we conducted a Monte Carlo analysis under the presumption that each $DTSM$, evaluated at the $ML$ estimates, is the true data generating model for the term structure data. The results confirm that, under the null hypotheses that our affine models accurately describe conditional first-moment properties of these bond yields, the actual (population) degree of predictability in excess returns is much less than what is indicated by the sample $R^2$s presented in $CP$ and Table 2. The small-sample biases are large, on the order of 50%.

Moreover, the means of the small-sample distributions of the $R^2$s are quite close to the realized values in the data sets (compare the column labeled “360” with the column “Samp”). The largest difference occurs for dataset $UFB$. Thus, if one believes that $UFB$ most accurately measures the zero and forward rates, then this finding constitutes mild evidence that model $A_0(3)$ does not fully generate the degree of predictability in excess returns. However, the differences between the $R^2$s in the historical sample and the mean of the small-sample distribution of the model-implied $R^2$s are all within one standard deviation (of the small-sample distribution of the $R^2$) of each other.

The dependence of the small-sample distribution of the $R^2$ on the size of the sample is illustrated in Figure 3. The mean decreases substantially as sample length is increased from ten to twenty years. Beyond about two thousand observations, the means are essentially constant while the standard deviation bands narrow in towards the mean.

Finally, in Table 7, we report the projection coefficients implied by the $A_0(3)$ model. Since comparable results are obtained for all four data sets, we present results only for $UFB$ and $SFB$. Further, since we are fitting three-factor models, we choose to use the PCs of the 1-, 3-, and 5-year yields in our calculation of the model-implied regression coefficients. The means of the model-implied small-sample distributions ($T = 360$) and the coefficients obtained from historical regressions match quite closely.

5 A “Single-Factor” Model for Excess Returns?

The patterns of projection coefficients illustrated in Figure 2 is suggestive of there being a single “forward factor” underlying the predictability of excess returns. A priori, such a finding would be striking, because it is now well documented that multi-factor models describe the distribution of bond yields much better than single-factor models, and excess returns are linear functions of the factors in affine $DTSM$s. In this section we examine the “common-factor” hypothesis for excess returns with both formal empirical tests and the

16 These results assure us of sufficient accuracy in obtaining the population results for the excess return regressions at our simulation length of $10^5$. 

13
<table>
<thead>
<tr>
<th>Maturity</th>
<th>const.</th>
<th>$PC_1$</th>
<th>$PC_2$</th>
<th>$PC_3$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UFB Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0055</td>
<td>0.1249</td>
<td>1.4819</td>
<td>5.4664</td>
<td>0.3157</td>
</tr>
<tr>
<td>3</td>
<td>0.0085</td>
<td>0.1837</td>
<td>2.8032</td>
<td>11.3469</td>
<td>0.3227</td>
</tr>
<tr>
<td>4</td>
<td>0.0111</td>
<td>0.2220</td>
<td>4.0193</td>
<td>13.7289</td>
<td>0.3152</td>
</tr>
<tr>
<td>5</td>
<td>0.0111</td>
<td>0.2655</td>
<td>5.0375</td>
<td>15.7518</td>
<td>0.3144</td>
</tr>
<tr>
<td><strong>UFB Population</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0033</td>
<td>0.1366</td>
<td>0.8492</td>
<td>7.3953</td>
<td>0.1863</td>
</tr>
<tr>
<td>3</td>
<td>0.0053</td>
<td>0.2248</td>
<td>1.7171</td>
<td>12.3235</td>
<td>0.1889</td>
</tr>
<tr>
<td>4</td>
<td>0.0066</td>
<td>0.3132</td>
<td>2.7333</td>
<td>15.5353</td>
<td>0.2098</td>
</tr>
<tr>
<td>5</td>
<td>0.0073</td>
<td>0.4050</td>
<td>3.8880</td>
<td>17.7156</td>
<td>0.2320</td>
</tr>
<tr>
<td><strong>UFB Small Sample Simulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0033</td>
<td>0.2000</td>
<td>0.8795</td>
<td>7.0743</td>
<td>0.2618</td>
</tr>
<tr>
<td>3</td>
<td>0.0053</td>
<td>0.3359</td>
<td>1.8459</td>
<td>11.8752</td>
<td>0.2647</td>
</tr>
<tr>
<td>4</td>
<td>0.0066</td>
<td>0.5825</td>
<td>4.3240</td>
<td>16.8452</td>
<td>0.3045</td>
</tr>
<tr>
<td>5</td>
<td>0.0073</td>
<td>0.7825</td>
<td>6.1370</td>
<td>20.7564</td>
<td>0.3978</td>
</tr>
<tr>
<td><strong>SFB Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0055</td>
<td>0.1162</td>
<td>1.5213</td>
<td>6.4509</td>
<td>0.2897</td>
</tr>
<tr>
<td>3</td>
<td>0.0084</td>
<td>0.1823</td>
<td>2.8823</td>
<td>11.3020</td>
<td>0.2957</td>
</tr>
<tr>
<td>4</td>
<td>0.0105</td>
<td>0.2234</td>
<td>3.1357</td>
<td>14.9208</td>
<td>0.2995</td>
</tr>
<tr>
<td>5</td>
<td>0.0120</td>
<td>0.2545</td>
<td>4.3072</td>
<td>17.9297</td>
<td>0.3013</td>
</tr>
<tr>
<td><strong>SFB Population</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0033</td>
<td>0.1512</td>
<td>0.9139</td>
<td>6.4533</td>
<td>0.2216</td>
</tr>
<tr>
<td>3</td>
<td>0.0052</td>
<td>0.2402</td>
<td>1.7109</td>
<td>10.8798</td>
<td>0.2071</td>
</tr>
<tr>
<td>4</td>
<td>0.0064</td>
<td>0.3110</td>
<td>2.5106</td>
<td>13.8443</td>
<td>0.2050</td>
</tr>
<tr>
<td>5</td>
<td>0.0071</td>
<td>0.3728</td>
<td>3.3183</td>
<td>15.9725</td>
<td>0.2048</td>
</tr>
<tr>
<td><strong>SFB Small Sample Simulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0033</td>
<td>0.2094</td>
<td>0.9625</td>
<td>6.1082</td>
<td>0.2901</td>
</tr>
<tr>
<td>3</td>
<td>0.0052</td>
<td>0.3439</td>
<td>1.8812</td>
<td>10.3430</td>
<td>0.2800</td>
</tr>
<tr>
<td>4</td>
<td>0.0063</td>
<td>0.3452</td>
<td>2.8385</td>
<td>12.9962</td>
<td>0.2800</td>
</tr>
<tr>
<td>5</td>
<td>0.0071</td>
<td>0.5444</td>
<td>3.8245</td>
<td>14.7322</td>
<td>0.2813</td>
</tr>
</tbody>
</table>

Table 7: Coefficients from the projections of one-year excess returns on $PC$s of the 1-, 3-, and 5-year yields for data sets $UFB$ and $SFB$. The population results are obtained through analytical calculations described in Section 4. The length of the small sample simulations is 360 months.
theoretical results from affine DTSMs. We argue that there is not a common factor (be it a single factor based on forward rates, PCs, or simply yields) across bond maturities underlying the predictability of excess returns.

We first use a GMM-based test to examine the common factor hypothesis. In previous sections we have shown that, in excess returns regressions, using all five forward rates is susceptible to multicollinearity, and the coefficients on higher order PCs are unlikely to be reliably estimated. Hence, in our test of the common factor restrictions we use only three forward rates ($f^{0-1}$, $f^{2-3}$, and $f^{4-5}$) and the first three PCs (in addition to a constant term) as regressors. The test results we obtained using all five forward rates or PCs (not reported) indicate even stronger rejections of the common-factor hypothesis.

Let $\beta_n$ ($n = 2$ to $5$) denote the $4 \times 1$ vector of projection coefficients in the excess returns regression of a $n$-year bond. The common-factor null hypothesis states that there exist a vector $\gamma$ and scalars $a_n$, such that $\beta_n = a_n \gamma$. Hence, while there are 16 free projection coefficients in the unrestricted model, there are only 7 free parameters under the common-

---

17 Details about the tests used can be found in Appendix C.
18 Another consideration is that using only three regressors substantially reduces the number of covariance matrix parameters estimated in constructing the GMM-based test statistics. We suspect that this may improve the small-sample properties of our test statistics.
Table 8: $GMM$ test of common factor restrictions. Under the null hypothesis, the test statistic follows a $\chi^2(df = 9)$ distribution. The 95% quantile is 16.919. The Newey-West heteroskedasticity and autocorrelation adjusted covariance matrix is estimated with 15 lags.

To interpret these rejections, we explore in more depth the projection coefficients implied by the $A_0(3)$ $DTSM$. The linear dependence of excess returns $r_{m,t+m}$ on $Y_t$ is given by (5). Fixing the holding period length $m$, the dependence of the projection coefficients on bond maturity $n$ is reflected in the term

$$-\delta Y \kappa^Q^{-1} (I - \kappa^Q)^n [I - (I - \kappa^Q)^{-m}(I - \kappa^P)^m]Y_t = -\xi'(I - \kappa^Q)^n Z_t,$$

where $\xi$ is a column vector, and $Z_t$ is a linear combination of $Y_t$. Hence $(I - \kappa^Q)^n$ governs the dependence of the loadings on $n$. The loading on $Z_{t,j}$ is $D_{n,j} = \Sigma_i \xi_i [(I - \kappa^Q)^n]_{ij}$. A single factor exists if and only if the ratios of the $D_{n,j}$, $j = 1, \ldots, N$, are invariant to changes in $n$.

Given the geometric decay of $(I - \kappa^Q)^n$, a stable ratio of the $D_{n,j}$ is not likely. This is illustrated by the special case of a diagonal $\kappa^Q$, in which case the relevant terms are

$$\xi_1 (1 - \kappa^Q_{11})^n : \xi_2 (1 - \kappa^Q_{22})^n : \ldots : \xi_N (1 - \kappa^Q_{NN})^n.$$

As $\xi$ is non-zero, if, as suggested by historical estimates (see, e.g., Dai and Singleton [2003]), different factors exhibit different degrees of mean reversion, then the ratios cannot be invariant across different values of $n$.

---

19 Here the common factor constraint is imposed on the constant term as well as the slope coefficients. Unreported results indicate similar strong rejections of the common factor hypothesis if we only impose the constraint on slope coefficients. Moreover, in the minimal form of the common factor test, when we only impose the restriction on the slope coefficients of two forward rates or $PCs$ (leaving the constant term and the rest of regressors unconstrained), we still obtain strong rejection results.

20 For each data, the test statistics obtained using all five forward rates or $PCs$ exhibit the same values. Hence, the differences across Panels A and B in Table 8 are due to the fact that the three forward rates and $PCs$ chosen for the tests do not span exactly the same linear space.

21 The parameters estimated using $GMM$ under the common-factor restriction are similar to those estimated using restricted regressions as described in $CP$. Using the $CP$ restricted regression estimates to calculate the $GMM$ objective function only results in higher values of the test statistics, and thus even stronger rejection of the null hypothesis.
This can be seen numerically in Table 9 using the $A_0(3)$ estimates obtained with the UFB data. In panel (a), column headings $Y_i$, we calculate the loadings of the expected one-year excess returns on the state factors for bond maturities of $n = 2$ to 5. To determine whether there is a common factor across $n$, we scale the loadings for each bond maturity so that the coefficients on $Y_2$ are 1 (Panel B). A common factor exists if, after scaling, the coefficients are the same across $n$. This equality is not obtained and, in fact, the differences are substantial.

If $DTSM$s do not imply that a common factor underlies the predictability of excess returns across maturities, then why do the coefficients on forward rates so strikingly suggest a single factor? The answer to this dilemma is given in Table 9, columns $f^{0-1}$ through $f^{4-5}$. These loadings on the forward rates are simply linear transformations of the loadings in columns $Y_1$ through $Y_3$. In other words, what shows up in forward rate coefficients looking like a common ("forward") factor is simply an artifact of the space in which the projections are being represented. Small differences in the coefficients on the forward rates are in fact highly statistically significant differences. This finding is similar to that of Backus, Foresi, Mozumdar, and Wu [2001], who found that small departures from the null hypothesis in certain forward-rate based tests of the expectations theory of the term structure translated into very large departures from the null hypothesis in the Campbell-Shiller style yield regressions.

These observations are reinforced by the coefficients in the projections of excess returns onto the PCs which, as we have noted, are highly correlated with the factors in $DTSM$s. From Table 9, columns PC1 through PC3, it is seen that there are considerable differences between scaled coefficients across bond maturities. Hence, the absence of a single factor is more evident from using the orthogonal PCs.

---

22The $A_0(3)$ ML estimates obtained from other data sets produce very similar results.
6 Fama-Bliss Regressions

Fama and Bliss [1987] presented evidence of predictable excess returns using a single forward rate. The findings of $CP$ (and indeed all of the multi-factor DTSMs) imply that adding more forward rates will strengthen the evidence for forecastable excess returns. Table 10, columns “Sample” confirm that sample $R^2$s increase with the inclusion of more forward rates (compare these results to those in Table 2). Further, as with the results using more forward rates, the model $A_0(3)$-implied population $R^2$ are smaller (roughly one half) of the values obtained from the historical sample. However, absent the multi-collinearity of the multiple forward-rate regressions, the means of the model-implied small-sample distributions of the $R^2$ are close to the model-implied population values (compare the results under “Population” with those under “$A_0(3)$ Small Sample”). So, assuming that model $A_0(3)$ is a reasonable representation of the first-moment properties of excess returns, there does not appear to be substantial bias in the $R^2$ statistic. Finally, though the historical sample $R^2$ are above the population (and mean of the model-implied small-sample distribution), the former are typically within one-standard deviation of the latter.

<table>
<thead>
<tr>
<th>Dataset/ Maturities</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$UFB$</td>
<td>0.144</td>
<td>0.147</td>
<td>0.149</td>
<td>0.067</td>
<td>0.048</td>
<td>0.042</td>
<td>0.062</td>
<td>0.092</td>
<td>0.057</td>
<td>0.054</td>
<td>0.073</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$FW$</td>
<td>0.144</td>
<td>0.139</td>
<td>0.142</td>
<td>0.094</td>
<td>0.054</td>
<td>0.042</td>
<td>0.056</td>
<td>0.079</td>
<td>0.062</td>
<td>0.054</td>
<td>0.067</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$SFB$</td>
<td>0.141</td>
<td>0.125</td>
<td>0.114</td>
<td>0.106</td>
<td>0.086</td>
<td>0.064</td>
<td>0.064</td>
<td>0.070</td>
<td>0.084</td>
<td>0.071</td>
<td>0.074</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$NSB$</td>
<td>0.140</td>
<td>0.126</td>
<td>0.118</td>
<td>0.111</td>
<td>0.060</td>
<td>0.048</td>
<td>0.062</td>
<td>0.084</td>
<td>0.065</td>
<td>0.057</td>
<td>0.070</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
</tbody>
</table>

Table 10: Sample, $A_0(3)$ implied population, and $A_0(3)$ implied mean of the small-sample distribution of the $R^2$ in the Fama-Bliss regressions. The population $R^2$s are analytically calculated. The length of the small sample simulations is 360 months.

7 Concluding Remarks

This paper has addressed several questions related to the predictability of excess returns on bonds using forward rates, including:

1. Is the degree of predictability documented in descriptive regression analyses largely the same as the degree of predictability implicit in multi-factor DTSMs?

2. Is there a single common “predictor factor” underlying the expected excess returns on bonds of all maturities?
3. To what extent is the evidence for predictability of excess returns a symptom of small-sample biases in estimated $R^2$?

The answer to the first question is yes. Within samples, excess returns appear to be highly predictable and the degree of predictability is roughly the same whether one is using forward rates or PCs of bond yields. Furthermore, the mean of the small-sample distribution of the $R^2$ statistic (implied by our $A_0(3) \ DTSM$) is very similar to the $R^2$s obtained with the historical time series.

As was noted in the introduction, we conjecture that these findings carry over to other DTSMs with similar flexibility in their specifications of the market prices of factor risks. This is the case with quadratic-Gaussian models, for example. Additionally, our findings (implicitly) support those of Bansal, Tauchen, and Zhou [2003] that a regime-switching model can reproduce many of the findings in CP as our model is nested in a multiple regime model. However, our analysis also suggests that regime-switching is not central to replicating the sample evidence on predictability within a DTSM.

At the same time, we find substantial statistical evidence against the view that there is a single common factor underlying the predictability of excess returns on bonds of different maturities. This is perhaps not surprising, viewed in the light of observations that (i) affine term structure models imply that excess returns are linear functions of all of the risk factors, and (ii) there is substantial evidence that at least two, and more likely three, factors are needed to adequately describe the historical distributions of bond yields. The fact that the patterns of coefficients from forward rate projections “look like” results from a common factor is shown to be an artifact of the particular rotation of the risk factors giving rise to forward rates as predictors. When we rotate to orthogonal factors, e.g. PCs, the coefficients clearly reveal that there is not a common predictive factor.

Regarding question (3), Monte Carlo analyses of the small-sample distributions of the $R^2$ statistic suggest that the within sample degree of predictability documented by CP is likely to be an overstatement of the actual degree of predictability. Depending on the method used to construct the zero curve, this degree of upward bias in sample $R^2$ is as much as 50%. There was much less bias in the small-sample distribution of the $R^2$ statistic in Fama-Bliss regressions onto a single forward rate.

Throughout this analysis we maintained a parallel analysis of four different data sets, constructed with four distinct spline functions. While many of our findings are robust across these datasets, some notable differences were obtained. Specifically, the shapes of the coefficients in the projections of excess returns onto forward rates were highly sensitive to the choice of spline method. This appears to be a consequences of differences in the spline-induced smoothness of forward rates.

Differences in smoothness of the splines also underlie the degree to which affine DTSMs replicate the predictability in the historical data (question (1)). The smoother the spline, the more closely the model-generated predictability matches up with that in the data. Put differently, in the data sets based on splines that generate relatively choppy forward rates, there is predictability in excess returns associated with variation in the fourth and fifth principal components. Since the three-factor models examined are primarily capturing predictability
based on the first three principal components, this extra predictability is being missed.

What should we conclude from the different degrees of predictability, particularly between \( UFB \) and the other data sets? The choppiness in forward rates in the dataset \( UFB \), and to a lesser extent \( FW \), is of course simply a manifestation of the data. The coupon-bond yields underlying the construction of the splines do not all line up along a smooth curve. Most likely, the reasons that actual market yields are not smooth along the maturity spectrum is not related to any fundamental macroeconomic phenomena. Evans and Marshall [2001] found that, while the first two PCs of bond yields are correlated with such macroeconomic time series as output growth, inflation, and monetary shocks, this was not the case for the third PC in their data. We regressed \( PC_4 \) and \( PC_5 \) (the sources of the incremental predictability in the \( UFB \) data) onto current and lagged values of the growth in industrial production, unemployment, and inflation, and also found very low \( R^2 \)s and largely statistically insignificant coefficients.

A more likely explanation is differences in liquidity, as reflected in differences in repo rates. To our knowledge the coupon bond data to which the splines are applied are not adjusted for any differences in repo rates and, as such, they are not “true” liquidity-adjusted and risk-free coupon bond yields. Viewed this way, it is interesting that the choppiness in zero-coupon bond yields induced by such differences in liquidity lead, in turn, to small incremental amounts of forecastability in excess returns over monthly sampling intervals. If it is this incremental explanatory power that is of interest, then the \( UFB \) dataset may well be the best one to study. However, most of the literature on DTSMs has been focused on the lower order PCs, and their relationships to macroeconomic variables. For the analysis of these links, it appears that little is lost by using the smoother zero curves produced by the continuous and differentiable splines.

Finally, an interesting question that the previous literature has not addressed in depth is whether or not the evidence for the predictability of excess returns gets stronger as the length of the holding period is increased. This question is difficult to answer in descriptive studies, because the effective number of degrees of freedom rapidly decreases with increases in the length of the holding periods. However, within the context of our estimated pricing models, we can answer this question at the population level.

Both the sample and the population\(^{23}\) \( R^2 \)s for regressions of excess returns on the five-year zero-coupon bond over various holding periods are displayed in Table 11. The \( R^2 \)s remain at the same levels as the length of the holding period increases. Furthermore, for the models examined, the upward biases in the small-sample \( R^2 \), relative to their population counterparts, seem to largely disappear over the longer holding periods.

\(^{23}\)The population numbers are based on simulations with a length of \( 10^5 \) months.
<table>
<thead>
<tr>
<th>Hold Period (years)</th>
<th>Model</th>
<th>$UFB$</th>
<th>$FW$</th>
<th>$SFB$</th>
<th>$NSB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sample $A_0(3)$</td>
<td>0.37</td>
<td>0.35</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23</td>
<td>0.22</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>2</td>
<td>Sample $A_0(3)$</td>
<td>0.36</td>
<td>0.36</td>
<td>0.34</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.30</td>
<td>0.30</td>
<td>0.26</td>
<td>0.30</td>
</tr>
<tr>
<td>3</td>
<td>Sample $A_0(3)$</td>
<td>0.31</td>
<td>0.30</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.35</td>
<td>0.35</td>
<td>0.32</td>
<td>0.36</td>
</tr>
<tr>
<td>4</td>
<td>Sample $A_0(3)$</td>
<td>0.33</td>
<td>0.31</td>
<td>0.32</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.38</td>
<td>0.39</td>
<td>0.36</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Table 11: $R^2$ for regressions for excess returns of holding a five-year bond for $n$ years (over the returns to holding an $n$-year zero coupon bond), for $n = 1$ to 4.
A Projection Coefficients in Forward-Rate Regressions

In Section 2.2 we presented evidence that small errors in measuring forward rates could have large effects on the shapes of the coefficients in projections of excess returns onto forward rates. In particular, we conjectured that the data based on the relatively choppy forward curves will give rise to projection coefficients that are “biased” in the direction of having a tent-like shape. This appendix develops this idea more formally.

Suppose that the true data generating process implies that the excess return on an \( n \)-period bond is predictable by forward rates in the following manner:

\[
h_{t+1}^n \equiv \log \frac{P_{t+1}^{n-1}}{P_t^n} - f_{t}^{0-1} = \alpha_n^0 + \alpha' f_t + \epsilon_{t+1}^n,
\]

for all \( 2 \leq n \leq N \), for some \( N \geq 2 \), where \( P_t^n \) is the price of a \( n \)-period zero-coupon bond and \( f_t = \{f_t^{(i-1)-i}\}_{i=1}^N \) is a \( N \times 1 \) vector of forward rates including the spot rate as the first element. Suppose further that the measured forward rates \( \hat{f}_t \) equal the true rates plus errors:

\[
\hat{f}_t = f_t + u_t,
\]

where \( u_t \) is a \( N \times 1 \) vector of measurement errors. For simplicity, let us assume that \( u_t \) is i.i.d. with diagonal covariance matrix and that \( \text{Cov}(f_t, u_t) = 0 \).

By definition, the holding-period return constructed from the mis-measured forward rates is given by

\[
\hat{h}_{t+1}^n = h_{t+1}^n + v_t^n,
\]

where \( v_t^n \equiv \sum_{i=2}^n u_i^t - \sum_{i=1}^{n-1} u_{t+i}^t \). The CP regression coefficients based on the forward rates contaminated by measurement errors are therefore given by

\[
\hat{\alpha}_n = \text{cov}^{-1} \left( \hat{f}_t \right) \times \text{cov} \left( \hat{h}_{t+1}^n, \hat{f}_t \right) = \left( \Omega_f + \Omega_u \right)^{-1} \times \Omega_f \times \alpha_n + \left( \Omega_f + \Omega_u \right)^{-1} \times \beta_n,
\]

\[
\beta_n \equiv \text{cov} \left( \sum_{i=2}^n u_i^t, u_t \right) = \begin{pmatrix} 0 & [\Omega_u]_{22} & [\Omega_u]_{33} & \ldots & [\Omega_u]_{nn} & 0 & \ldots & 0 \end{pmatrix}',
\]

where \( \Omega_f \) and \( \Omega_u \) are, respectively, the unconditional covariance matrix of the forward rates and the measurement errors.

To examine the difference between \( \hat{\alpha}_n \) and \( \alpha_n \), let us consider the extreme case where the forward rates are perfectly correlated in levels\(^{25}\) and have the same volatility; i.e. \( \Omega_f = \omega_f \times \zeta \), where \( \zeta \) is a \( N \times N \) matrix of 1’s. Suppose also that the measurement errors have the same volatility; i.e., \( \Omega_u = \omega_u \times I \), where \( I \) is a \( N \times N \) identity matrix. In this case,

\[
\left( \Omega_f + \Omega_u \right)^{-1} = \omega_u^{-1} \left( I - \frac{\zeta}{N + \frac{\omega_u}{\omega_f}} \right).
\]

\(^{24}\)Since the forward rates are highly correlated in levels, \( \Omega_f \) may be close to singular. However, \( \Omega_f + \Omega_u \) typically has full rank. Thus, the OLS regressions based on contaminated yields are feasible, even if the true forward rates are perfectly correlated.

\(^{25}\)This is the most relevant case. For instance, in the UFB data, seven of the ten distinct correlations among the forward rates are 0.94 or larger, and the smallest correlation is 0.85.
It follows that
\[ \hat{\alpha}_n = \bar{\alpha}_n + \omega_u^{-1} \left( \beta_n - \frac{N}{N + \omega_f} \times \bar{\beta}_n \right) \approx \bar{\alpha}_n + \omega_u^{-1} (\beta_n - \bar{\beta}_n), \tag{12} \]
where \( \bar{\alpha}_n \equiv \frac{\alpha_n}{N} \) and \( \bar{\beta}_n = \frac{\beta_n}{N} = \frac{n-1}{N} \times \omega_u \) are the averages of the means of \( \alpha_n \) and \( \beta_n \), respectively.\(^{27}\)

The reason that \( \hat{\alpha}_n \) tends to have a tent shape, regardless of the shape of the \( \alpha_n \), is that the first element and the last \( N - n \) elements of \( \beta_n - \bar{\beta}_n \) are negative, while the middle \( n - 1 \) elements are positive. In fact, letting \( c \equiv \frac{n-1}{N} \), then
\[ \omega_u^{-1} (\beta_n - \bar{\beta}_n) = \begin{pmatrix} -c & 1 - c & 1 - c & \ldots & 1 - c & -c & \ldots & -c \end{pmatrix}', \]
the elements of which sum to 0.

In applying this reasoning to the patterns in Figure 2, let us view \( f_t \) as a smooth forward rate curve and \( \hat{f}_t \) as a relatively noisy series. We have seen that the UFB data, generated with a discontinuous and piece-wise constant forward rate curve behaves more like a noisy \( \hat{f}_t \) than a smooth \( f_t \) such as is generated by the SFB or NSB methods (see, e.g., Figure 1). Thus, it may well be that the tent-shape of the projection coefficients in Figure 2 (a) is driven by the component \( \omega_u^{-1} (\beta_n - \bar{\beta}_n) \) of \( \hat{\alpha}_n \) in (12). Even a small amount of noise turns the wave-like patterns for the data set SFB into the tent shape in Figure 2(d). Moreover, given \( N \), our analysis implies that the tent shape should be more pronounced as \( n \) increases, which is exactly what we see in both Figures 2 (a) and 2 (d).

When we re-estimated these projections for different sample periods, holding fixed the data set, we often found substantial changes in shapes. For instance, the peaks and troughs of the wave patterns were sometimes reversed across sample periods. The primary exception was the data set UFB, for which a tent shape was almost always obtained. This lends further support to the view that the choppiness induced by the discontinuous forward rate curve may be dominate in this data set.

### B Excess Returns in \( A_0(3) \) Models

This appendix derive analytical expressions for excess returns for the Gaussian \( A_0(N) \) DTSMs. Starting from (1) and (2), \( Y \) is also affine under the risk-neutral measure \( Q \) with \( \kappa^Q = \kappa^F + \lambda_Y \). The short rate follows \( r_t = \delta_0 + \delta'_Y Y_t \). It then follows (e.g., Dai, Singleton, and Yang [2003]) that the log price of a bond of maturity \( n \) is given by
\[ p_{n,t} = A_n + B'_n Y_t, \]
where
\[ B_{n+1} = -\delta_Y + (I - \kappa^Q')B_n, \quad B_0 = 0, \]

\(^{26}\)Under our presumption that the forward rates are perfectly correlated, \( \alpha_n \) is indeterminate. However \( \hat{\alpha}_n \) depends only on the average of the elements of \( \alpha_n \), \( \bar{\alpha}_n \).

\(^{27}\)The last step involves an highly accurate approximation of \( N - \frac{\omega_u}{\omega_f} \approx N \), because typically \( \omega_u << \omega_f \).
Here $I$ is an identity matrix. This further leads to

$$B_n = -[I + (I - \kappa^Q)^2 + \ldots + (I - \kappa^Q)^n - 1] \delta_Y$$

The excess return from holding a $n$-period bond for $m$ periods is defined as

$$r_{n,m,t+m} = p_{n-m,t+m} - p_{n,t} + p_{m,t} = A_{n-m} - A_n + A_m + B'_{n-m}Y_{t+m} - B'_n Y_t + B'_m Y_t.$$ 

To obtain the relation between the excess return $r_{n,m,t+m}$ and the state factor $Y_t$, we need to link $Y_{t+m}$ with $Y_t$. From the state factor process under the physical measure we obtain

$$Y_{t+m} - \theta^P = (I - \kappa^P)^m (Y_t - \theta^P) + \Sigma \epsilon_{t+m} + (I - \kappa^P) \Sigma \epsilon_{t+m-1} + \ldots + (I - \kappa^P)^{m-1} \Sigma \epsilon_{t+1},$$

which we reexpress as

$$Y_{t+m} = [I - (I - \kappa^P)^m] \theta^P + (I - \kappa^P)^m Y_t + \eta_{t+1,t+m}.$$ 

Here, $\eta_{t+1,t+m}$ represents the weighted cumulative effect of the shocks, and is independent of $Y_t$. Its variance matrix is

$$\text{var}(\eta_{t+1,t+m}) = \Sigma \Sigma' + (I - \kappa^P) \Sigma \Sigma' (I - \kappa^P) + \ldots + (I - \kappa^P)^{m-1} \Sigma \Sigma' (I - \kappa^P)^{m-1}.$$ 

Under stationarity conditions, $||(I - \kappa^P)^m|| \to 0$ as $m \to \infty$. Hence, we obtain the unconditional variance matrix of the state factor as

$$\text{var}(Y_t) = \lim_{m \to \infty} \text{var}(\eta_{t+1,t+m}).$$

This allows us to compute $\text{var}(Y_t)$.

Now

$$r_{n,m,t+m} = A_{n-m} - A_n + A_m + B'_{n-m}Y_{t+m} - B'_n Y_t + B'_m Y_t = C_{n,m} + L'_{n,m} Y_t + \nu_{t+1,t+m},$$

where,

$$L'_{n,m} = [(I - \kappa^P)^m B_{n-m} - B_n + B_m]'$$

$$= -\delta'_Y \kappa^Q - 1 (I - \kappa^Q)^n [I - (I - \kappa^Q)^{-m} (I - \kappa^P)^m]$$

$$- \delta'_Y \kappa^Q - 1 [(I - \kappa^P)^m - (I - \kappa^Q)^m].$$

If we regress excess returns on state factors, the model implied population coefficients will be $C_{n,m}$ and $L_{n,m}$, and the residual has a variance

$$\text{var}(\nu_{t+1,t+m}) = B'_{n-m} \text{var}(\eta_{t+1,t+m}) B_{n-m}.$$ 

24
Hence, we obtain a population $R^2$ of

$$\frac{L'_{n,m} \text{var}(Y_t)L_{n,m}}{L'_{n,m} \text{var}(Y_t)L_{n,m} + \text{var}(\nu_{t+1,t+m})}. $$

In general, if we project the excess returns onto a constant and the regressor $G + HY_t$,

$$r_{n,m,t+m} = \alpha + \beta'(G + HY_t) + u_{t+1,t+m},$$

where $H$ is an $N \times N$ matrix, then the population projection coefficients are given by

$$\beta = (H \text{var}(Y_t)H')^{-1}(H \text{var}(Y_t)L_{n,m}),$$

$$\alpha = C_{n,m} + L'_{n,m}\theta^p - \beta'(G + H\theta^p),$$

since $E(Y_t) = \theta^p$. The $R^2$ is

$$\frac{\beta' H \text{var}(Y_t) H' \beta}{L'_{n,m} \text{var}(Y_t)L_{n,m} + \text{var}(\nu_{t+1,t+m})}. $$

C Testing Constraints Under GMM

Suppose the multi-equation regression model for $M$ dependent variables is

$$y_{1,t+12} = x_{1t}'\beta_1 + \epsilon_{1,t+12},$$

$$\vdots$$

$$y_{M,t+12} = x_{Mt}'\beta_M + \epsilon_{M,t+12},$$

where $x_t = [x_{1t} \ldots x_{Kt}]'$ is a $K \times 1$ vector of predictors and the twelve-period forecast horizon arises because we are examining one-year holding periods. Under this alternative model, we assume that $E[\epsilon_{i,t+12}|x_{t-s}, y_{t-s}; s = 0, 1, \ldots] = 0$, $i = 1, \ldots, M$.

Each $\beta_i$ is a $K \times 1$ vector of coefficients, and all together they comprise the $MK \times 1$ vector $\beta = [\beta_1 \ldots \beta_M]'$.

There are $M \times K$ moment conditions

$$Eg_t = E \begin{bmatrix} \epsilon_{1,t+12}x_t \\ \vdots \\ \epsilon_{M,t+12}x_t \end{bmatrix} = \begin{bmatrix} (y_{1,t+12} - x_{1t}'\beta_1)x_t \\ \vdots \\ (y_{M,t+12} - x_{Mt}'\beta_M)x_t \end{bmatrix} = 0.$$

The sample analog for $Eg_t(b)$ is

$$\bar{g}(b) = \frac{1}{T} \sum_{t=1}^{T} g_t(b) = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} (y_{1,t+12} - x_{1t}'b_1)x_t \\ \vdots \\ (y_{M,t+12} - x_{Mt}'b_M)x_t \end{bmatrix}.$$
Let $W$ be a symmetric, positive definite weighting matrix, which has $MK$ columns by $MK$ rows. The GMM estimator is obtained by minimization of the objective function

$$\hat{\beta} = \arg\min_b J(b) = \arg\min_b T\bar{g}(b)'W\bar{g}(b)$$

Since the forecast horizon is twelve months, the errors $(y_{i,t+12} - \beta'_i x_t)x_t$ potentially have non-zero autocorrelations out to lag 11. Therefore, we construct the optimal distance matrix $W$ using the Newey and West [1987] method with 15 lags.

Under the unrestricted model, there are $MK$ moment conditions to identify $MK$ coefficients. We obtain the unrestricted estimates $\hat{\beta}$ from OLS, and calculate $\hat{S}$. Then we use $\hat{S}$ as the weighting matrix to obtain the coefficients estimates $\hat{\beta}^r$ under the common factor restriction. Under the null hypothesis that there is a common factor, the difference between the GMM objective function values, $J(\hat{\beta}^r) - J(\hat{\beta})$, follows a $\chi^2$ distribution with the degree of freedom equal to the number of parameter restrictions (Eichenbaum, Hansen, and Singleton [1988]).
References


Bansal, R., G. Tauchen, and H. Zhou (2003). Regime-Shifts in Term Structure, Expecta-
tions Hypothesis Puzzle, and the Real Business Cycle. forthcoming, *Journal of Busi-
ness and Economic Statistics*.

Options Research* 9, 197–231.


Dai, Q. and K. Singleton (2002). Expectations Puzzles, Time-varying Risk Premia, and


of Finance* 57, 405–443.

tative Agent Models of Consumption and Leisure Choice Under Uncertainty. *Quarterly


with Smoothing Splines. Working Paper 95-1, Finance and Economics Discussion Series
Federal Reserve Board.

Kim, D. (2004). Time-Varying Risk and Return in the Quadratic-Gaussian Model of the


