Short-Term Predictability of the Term Structure

HAIM REISMAN AND GADY ZOHAR

HAIM REISMAN
is an associate professor of finance in the Department of IE & Management at Technion-Israel Institute of Technology in Haifa, Israel.
reisman@ie.technion.ac.il

GADY ZOHAR
is a lecturer in finance in the Department of IE & Management at Technion-Israel Institute of Technology in Haifa.
gady@tx.technion.ac.il

Despite the crucial implications of term structure predictability for bond portfolio management as well as risk management, this issue has received relatively little attention. We show that a seemingly insignificant short-term predictability of the term structure may allow substantially higher returns through frequent rebalancing of the bond portfolio.

Recent attempts to forecast the term structure include Duffee [2002] and Diebold and Li [2003]. Duffee [2002] shows that *essentially affine* models may account for predictability, but the extent to which predictions are significant is not clear. Diebold and Li [2003] decompose the term structure into three factors using Nelson and Siegel [1987] exponential components, and attempt to predict the factors themselves. While they show surprising success in predicting the term structure six months to one year ahead, their model performs quite poorly in short-term predictions (one month ahead).

Similar to Diebold and Li [2003], we decompose the term structure into factors $z_i(t)$. We write

$$y(t,\tau) = \sum_{i=1}^{F} u_i(\tau) z_i(t)$$

where $y(t, \tau)$ is the spot rate for τ years. Unlike Diebold and Li, we use principal components analysis (PCA), and then the factors are pairwise uncorrelated in-sample, thus providing some justification for an independent treatment of each factor.

Indeed, using our approach on monthly data, our one-month ahead prediction agrees with Diebold and Li [2003]. Autoregressive moving average (ARMA) modeling does not improve the results obtained by assuming a simple random walk. This also turns out to be the case in application of standard ARMA modeling to weekly data, and testing the predictions for one week ahead.

The results using weekly data improve significantly when instead of modeling the factor dynamics, $z_i(t)$, we model the series of the differences, $dz_i(t)$. This is consistent with Diebold and Li's [2003] observation that the first two factors test positive for a unit root.

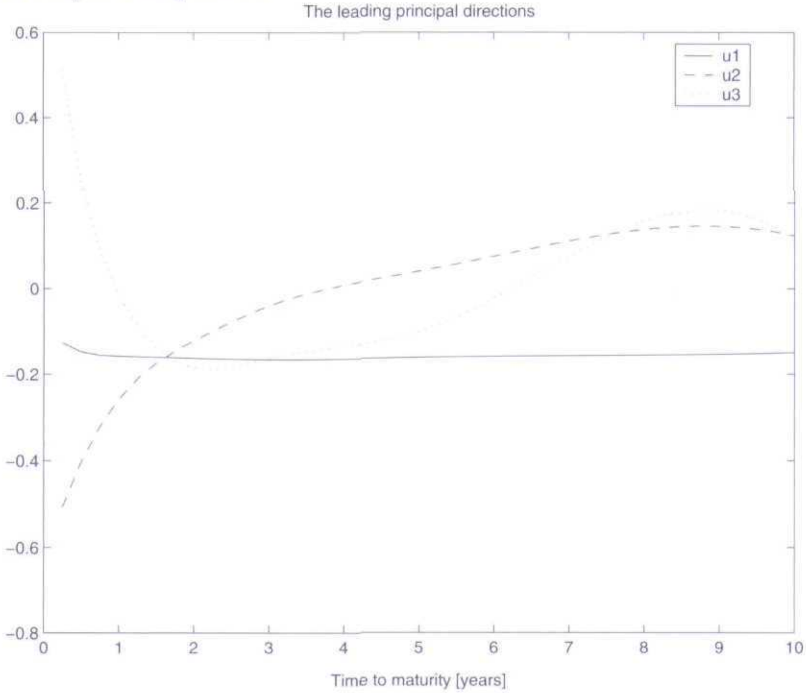
In other words, while it may be natural to assume that $z_i(t)$ evolves like a possibly mean-reverting Itô process, we observe a consistent non-trivial autocorrelation pattern for the process of differences $x_i(t) \equiv dz_i(t)$. For these processes we find a strictly positive short-term correlation, which implies that there is short-term momentum in Treasury yields.

While our prediction improves the root mean squared error only slightly, it turns out that taking into account the prediction results to allocate funds among bonds may lead to substantial excess returns by frequent rebalancing of the portfolio.

I. FACTOR DECOMPOSITION

To specify our model for a market with N discount bonds of times to maturity $\tau_1, \ldots, \tau_N$, we assume the economy is driven by F factors

EXHIBIT 1
Principal Components



of uncertainty z_i, $i = 1, \ldots, F$, and that the spot rate for τ years admits the specification $y(t, \tau) = \sum_{i=1}^{F} u_i(\tau)z_i(t) + \epsilon(t, \tau)$, which immediately implies that:[1]

$$dy(t,\tau) = \sum_{i=1}^{F} u_i(\tau)dz_i(t) + d\epsilon(t,\tau) \tag{1}$$

We can choose $u_i(\cdot)$ arbitrarily, but such a representation becomes useful only if there is little residual noise $d\epsilon$. Consequently, we set $u_i(\tau)$ as the (orthonormal) principal directions determined by applying standard principal components analysis (PCA) to a sequence of vectors $(dy(t, \tau_1), \ldots, dy(t, \tau_N)) \in \mathbb{R}^N$ (for discrete times).[2] For more details on PCA, see Jolliffe [1986].

By taking u_i as the principal directions, and then defining dz_i, $i = 1, \ldots, F$ as the projections of $dy(t, \cdot)$ onto $u_i(\cdot)$:

$$dz_i(t) = \sum_{j=1}^{N} dy(t,\tau_j)u_i(\tau_j)$$

we minimize the in-sample mean squared error $\sum_t \sum_{i=1}^{N} d\epsilon^2(t, \tau_i)$.

As Equation (1) is the only assumption we make on dy, we can say nothing regarding the statistical properties of either $z_i(t)$, $i \in \{1, \ldots, F\}$ or $\epsilon(t, \tau)$. If the term structure evolves in a complex manner, so will the factors $z_i(t)$. They are not assumed to be martingales, and may have non-zero correlations. Nonetheless, for a given F, the factors obtained will explain most of the in-sample variance in the sense of minimizing the squared error.

Using PCA, we obtain the affine decomposition that best explains the observed variance of dy among all decompositions with F factors. PCA has the significant advantage that it requires no ad hoc assumptions on the dynamics of processes. That no assumptions are needed for PCA to recover the most significant principal directions *in-sample* does not guarantee that the principal vectors will perform well *out of sample*. Stationarity is a sufficient (but not a necessary) condition for PCA to perform well also on out-of-sample data.

Finally, since the in-sample mean of $dy(\cdot, \tau)$ is close to zero, it can be shown that at least in-sample the dz_i are approximately (unconditionally) mutually uncorrelated. This motivates modeling of the dz_i as *independent* processes.

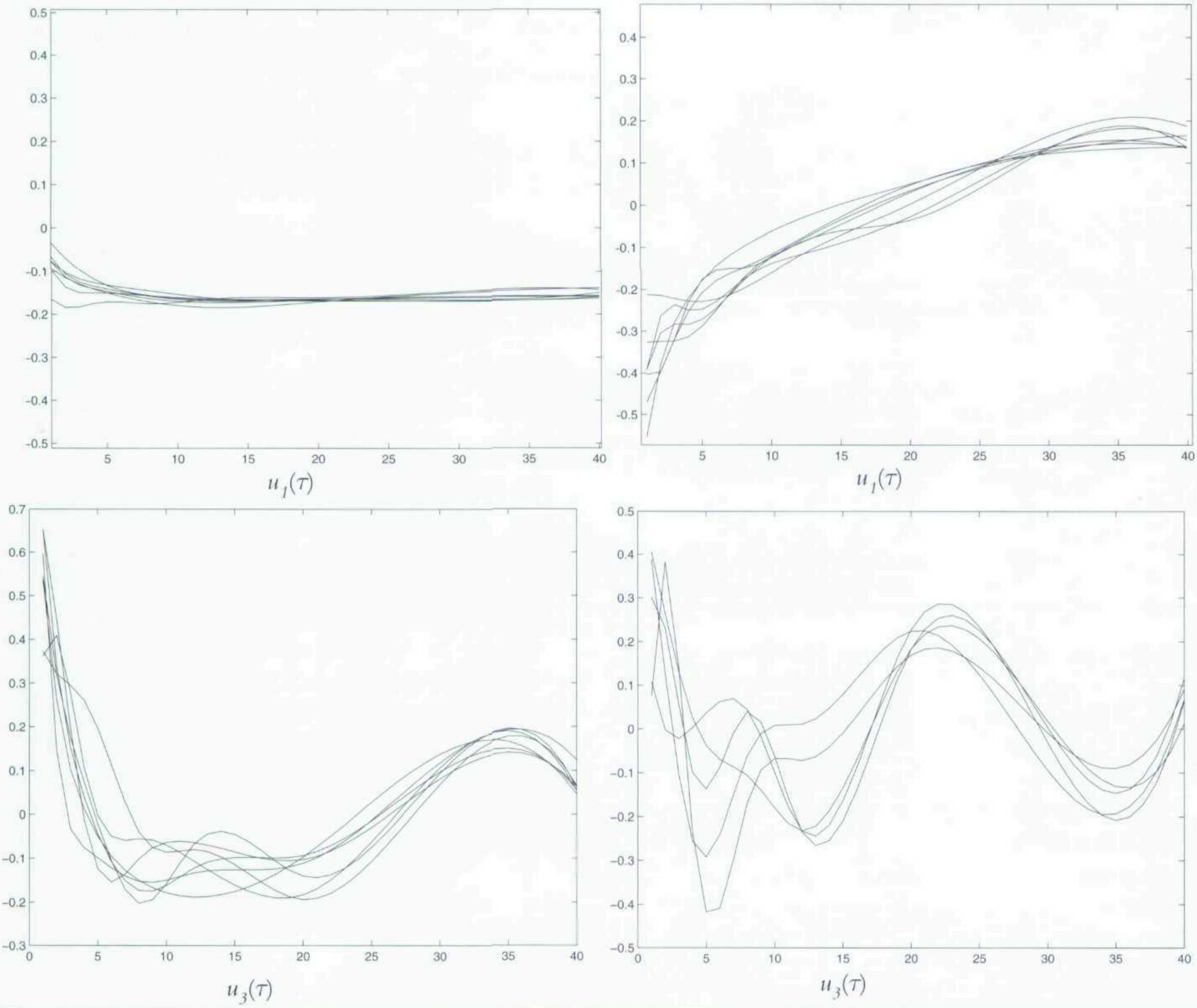
Our data include yields to maturity (YTM) of U.S. discount bonds that mature in three and six months and one, two, three, five, seven, and ten years, dating back to the beginning of 1982. The numbers are interpolated from the YTM of U.S. bonds and notes provided by the Federal Reserve weekly. We use PCA to estimate the principal directions $u_i(\cdot)$ of the data set. The three leading components that are drawn in Exhibit 1 seem to agree with those found in other works (see, e.g., Litterman and Scheinkmann [1991]).

The three leading principal directions account for about 99% of the variance in bond yields. The first principal component explains about 87% of the variance in the term structure. From the plot in Exhibit 1 it is seen to be almost constant as a function of maturity, and hence it is responsible for parallel shifts of the yield curve. The second factor accounts for an additional 10.5% of the variance, and mainly tilts the curve. The third factor contributes an additional 1.5%, and is responsible for changes in curvature.

Note that the $u_i(\cdot)$ are assumed to be constant over time. In Exhibit 2 we examine their stability. Here the principal directions are estimated over disjoint time periods

EXHIBIT 2

Factor Loadings at Various Time Intervals



of two years each. It is evident that u_1 and u_2 are quite stable, but u_3 is marginally stable, while u_4 (as well as the rest of the factor loadings not drawn here) changes rapidly over time.

Typically only two to three leading factors are used for modeling because of the obvious instability.

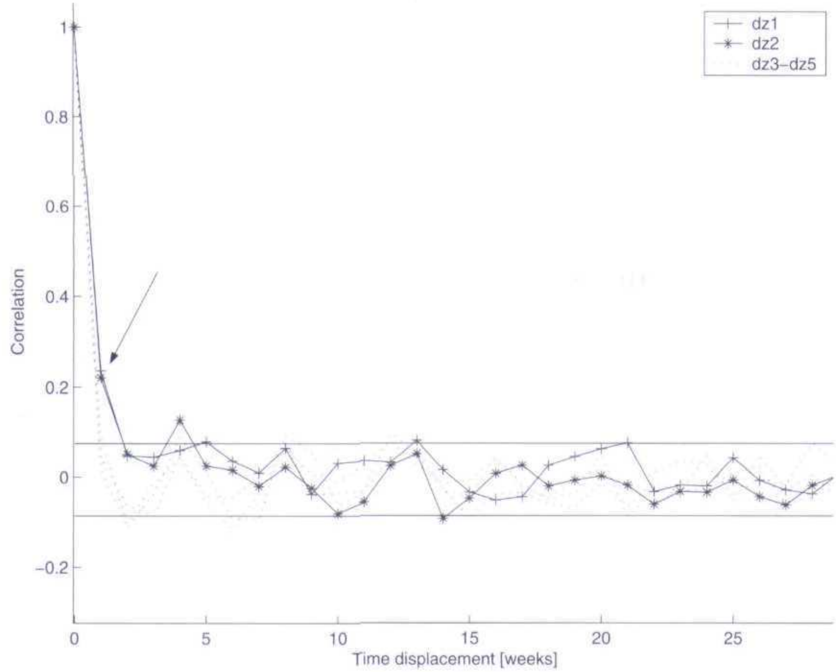
II. ARMA MODELING

We model the series of differences $dz_i(t) \equiv z_i(t + dt) - z_i(t)$, where dt is one week, as independent ARMA processes. In Exhibit 3 we plot the whole sample autocorrelations of these processes for $i = 1, \ldots, 5$.

All autocorrelation functions except dz_1 and dz_2 are essentially zero, and therefore indicate that those factors are not predictable. The non-vanishing correlation on the one-week lag in dz_1 and dz_2 reaffirms the observation of Diebold and Li [2003] regarding a unit root in the series $z_1(t)$ and $z_2(t)$. The correlograms imply that both dz_1 and dz_2 should be modeled as either AR(1) or MA(1), which implies that the original z_i processes for $i = 1, 2$ are best modeled as ARIMA(1, 1, 0) or ARIMA(0, 1, 1).

Concentrating only on factors z_1 and z_2, in Exhibit 4 we plot the autocorrelation functions for different time intervals, to demonstrate their stability. The dotted horizontal lines bound the 95% confidence interval

EXHIBIT 3
Autocorrelations of dz_i, i = 1, . . ., 5



for each dashed plot, and the solid horizontal lines do the same for the average (solid) plot. It is evident that either AR(1) or MA(1) is suitable for the modeling of dz_1 and dz_2. AR(1) also seems better than any other model if we compare the final prediction errors (FPE).

Modeling dz_i, $i = 1, 2$ as AR(1) (performance differs little if we use MA(1) instead), we estimate the processes of dz_1 and dz_2 using the entire sample (1982-2002):

$$dz_1(t) - 0.26dz_1(t - dt) = \epsilon_1(t)$$

$$dz_2(t) - 0.2dz_2(t - dt) = \epsilon_2(t)$$

For a general AR(1) model, $dz(t) = \alpha dz(t - dt) + \epsilon(t)$, we estimate α in different time intervals in order to demonstrate the robustness of the fit. These results are presented in Exhibit 5.

In Exhibit 6 we provide the autocorrelations of the residuals of dz_1 and dz_2. Once again the horizontal lines stand for Bartlett's 95% confidence band. The low autocorrelation values attained for almost all displacements show that our models reliably describe dz_1 and dz_2. Once the dz_i are predicted, the prediction of dy follows directly from Equation (1).

For out-of-sample testing we estimate the AR(1) models for dz_i on the first half of the data (1982-1992), yielding $dz_1(t) - 0.237dz_1(t - dt) = \epsilon_1(t)$, and $dz_2(t) - 0.233dz_2(t - dt) = \epsilon_2(t)$. For in-sample testing, we learn on the second half (1992-2002), and then we obtain slightly

EXHIBIT 4
Autocorrelations of dz_i at Various Time Intervals

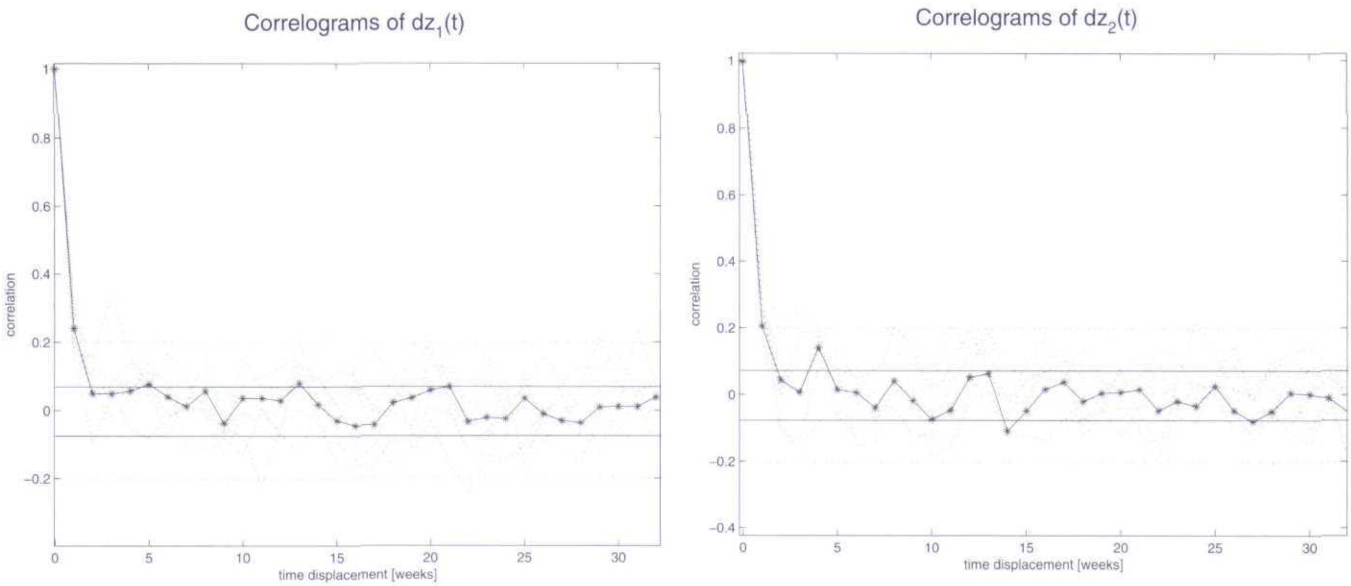


EXHIBIT 5
Models for dz_1 and dz_2 at Various Periods

Years:	1982-1986	1986-1990	1990-1994	1994-1998	1998-2002
α in dz_1:	0.19	0.36	0.27	0.19	0.34
α in dz_2:	0.3	0.26	0.21	0.09	0.26

different results: $dz_1(t) - 0.261dz_1(t-dt) = \epsilon_1(t)$, and $dz_2(t) - 0.247dz_2(t-dt) = \epsilon_2(t)$. In both cases, we test on the second half of the data (1992-2002).

Exhibit 7 compares the root mean squared error (RMSE) of the prediction of our model with the random walk model where the optimal prediction of $z_i(t+dt)$ is simply $z_i(t)$, and with the fitted AR(1) model for z_1 and z_2. We show the results for the six-month and two- and seven-year maturities. Both in-sample and out-of-sample results are provided.

Our model is clearly superior to the others. It achieves an average RMSE improvement of approximately 2.8%, both in-sample and out. Note that the RMSE does not change by much over different time intervals. Thus the improvement may look slight and might cause one to think there is no reason to use our model for prediction.

Nonetheless, we argue that since the prediction is for a short horizon, and since it is used for intertemporal portfolio adjustments, small improvement in the RMSE may produce greater improvements in portfolio performance.

This can happen because by holding a long-term (dynamic) portfolio, a high percentage of the noise is diversified away by averaging over time. This irrelevant noise is responsible for the level of the RMSE, while for an investor with long-term objectives only the non-diversifiable RMSE should be relevant. On the other hand, the improvement to the RMSE is measured in absolute terms. Relative to the non-diversifiable RMSE, the percentage improvement is therefore much greater than 2.8%.

Accounting for this predictability by dynamic rebalancing of bond weights in a portfolio can produce substantially higher returns.

III. APPLICATIONS IN PORTFOLIO SELECTION

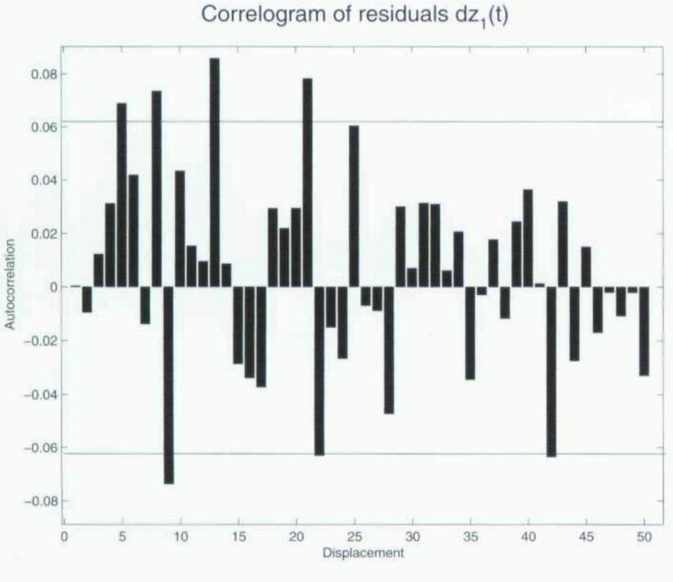
We have been able to fit a model to the data, and obtained a robust although minor improvement to the root mean squared error. Next we demonstrate the substantial improvement that can be achieved by applying the prediction results to bond portfolio selection.

Our prediction of yields naturally induces a prediction of bond prices via the bond price formula. Denoting by $B(t, T) = \exp[-(T-t)y(t, T-t)]$ the price of a discount bond maturing at time T, using Itô's lemma it is not hard to show that:

$$\frac{dB(t,T)}{B(t,T)} = y(t,\tau)dt + \tau\frac{\partial y(t,\tau)}{\partial \tau}dt - \frac{1}{2}\tau^2 d\langle y(t,\tau)\rangle - \tau dy(t,\tau) \tag{2}$$

where $\tau \equiv T - t$, and $\langle\cdot\rangle$ denotes quadratic variation (see Reisman and Zohar [2004]).

EXHIBIT 6
Autocorrelations of Residuals



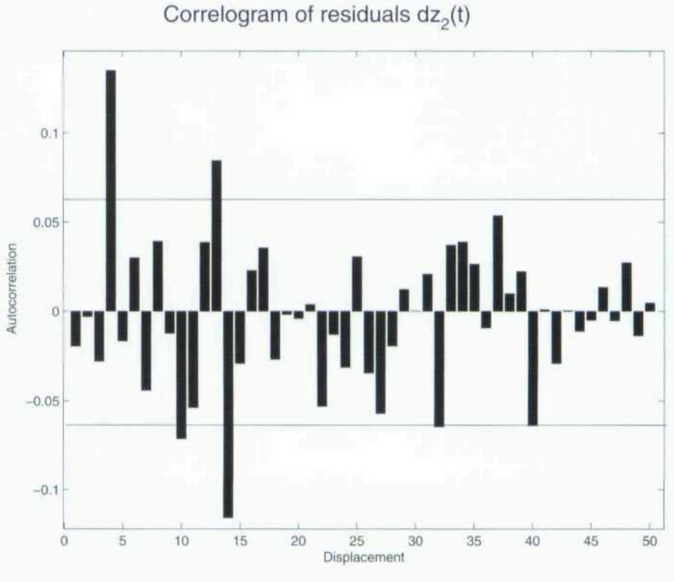


EXHIBIT 7
Comparison of RMSE in One Week-Ahead Prediction

Test:	in sample			out of sample		
Time to maturity:	6M	2Y	7Y	6M	2Y	7Y
Random walk:	0.1512	0.1396	0.1313	0.1512	0.1396	0.1313
AR(1):	0.1512	0.1395	0.1312	0.1515	0.1396	0.1312
Our model:	0.1482	0.1346	0.1273	0.1481	0.1346	0.1274

The prediction of bond returns is then obtained by substituting $dy(t, \tau)$ in Equation (2) by its prediction $d\hat{y}(t, \tau)$. Hence, we can devise portfolios that maximize instantaneous expected return under desired constraints. In accordance with the prediction results, we predict the dy in Equation (1) by its first two leading factors: $d\hat{y}(t, \tau) \approx u_1(\tau)d\hat{z}_1(t) + u_2(\tau)d\hat{z}_2(t)$.

We now construct a zero-cost portfolio that consists of investing \$100 in a single bond, and borrowing the same amount by selling short the shortest maturity, which is the three-month note in our data. The portfolio is not risk-free because of variance in the price of the bond, so instantaneous returns are not zero.

The profit plots in Exhibit 8 accumulate weekly gains without discounting. These gains are on a notional initialized to \$100 on a weekly basis. In this case, the slope of each graph is the instantaneous profit rate. Over the last ten years, the average excess yearly return on holding a constant maturity of seven years (first graph) has been approximately 3.5%, as is evident from the plot labeled *without prediction*. Holding a two-year note would return about 1.5% on average (second graph).

When we incorporate prediction, sometimes it turns out that the expected return on the long maturity is lower than the risk-free rate. In this case, we flip the position, i.e., sell the long maturity short, and buy the three-month note instead. Then the variance of the returns remains unchanged, thus facilitating a fair comparison to the case without prediction.

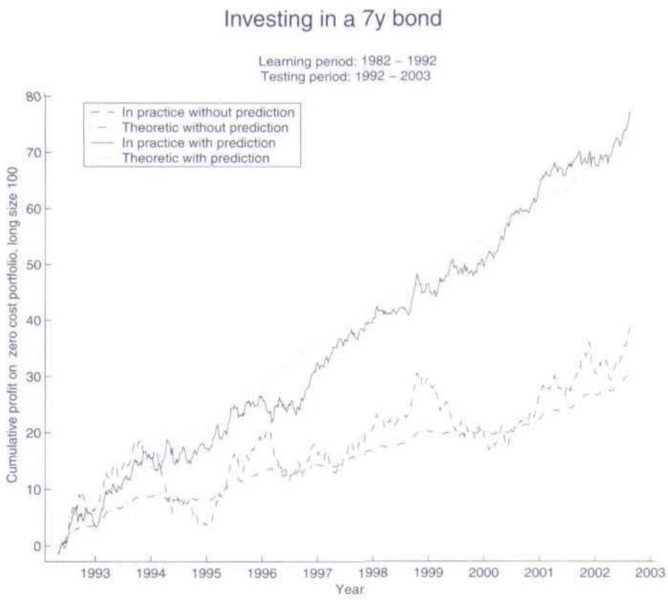
Over the same testing period of ten years, such an investment strategy with the seven-year note would have gained an average yearly return of about 7.5%, which more than doubles the return attained by the constant-maturity portfolio. The standard deviation of returns on both prediction-based and constant-maturity portfolios is approximately 5.2%. Consequently, the Sharpe ratio of the constant-maturity is approximately 0.65, and as high as 1.45 for the prediction-based allocation.

Similarly, we gain a yearly return of approximately 2.5% doing the same with the two-year note, which almost doubles the return without prediction. The standard deviation here is approximately 2%, resulting in a Sharpe ratio of 0.75 without prediction and 1.25 with prediction.

These are two arbitrary examples, but we obtain results of the same nature for all maturities.

Note that our investment strategy is not an optimal way to take advantage of predictability, but rather the simplest and the most intuitive way to demonstrate the sig-

EXHIBIT 8
Cumulative Profit from Prediction-Based Investment



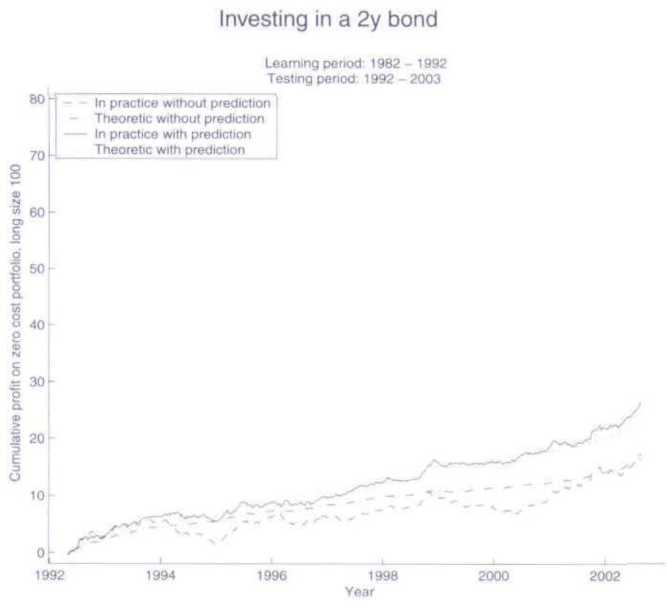


EXHIBIT 9
Time-Dependence of Optimal Allocation

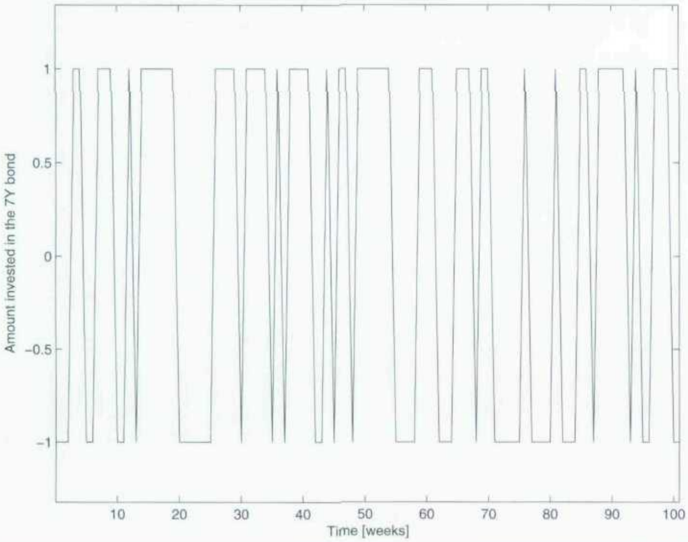
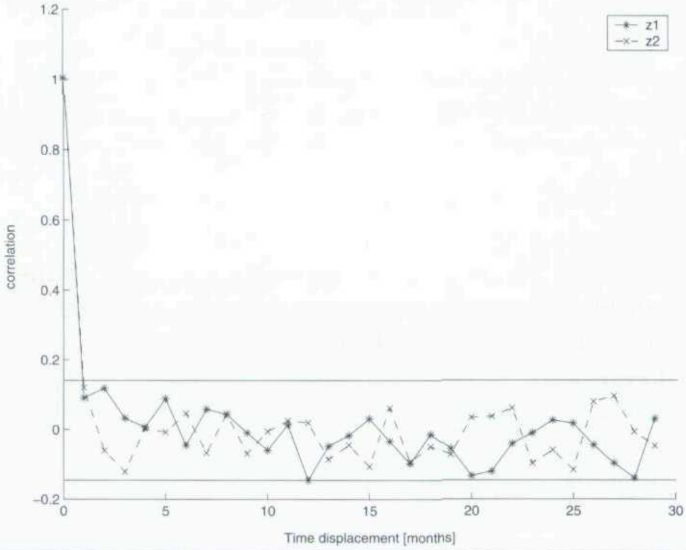


EXHIBIT 10
Autocorrelations of Monthly Differences of z_1, z_2



nificance of the prediction results. Results can be greatly improved by applying methods for optimal allocation and by investing in many maturities.

The striking success of the prediction-based allocation is not without cost. One needs to flip the portfolio and take a short position in a bond relatively frequently. In Exhibit 9 we plot the time-dependence of the position in the seven-year bond, which demonstrates the rate of switching between holding and shorting the asset. It is clear that enjoying the fruits of the prediction requires active portfolio management. Such frequent rebalancing may involve high transaction costs, and then one should apply a more sophisticated optimization scheme.

IV. CONCLUSION

Our analysis of the short-term predictability of U.S. Treasury yields finds significant prediction power in weekly data. While prediction improves the root mean squared error only slightly, we have shown that by applying the prediction results to bond portfolio selection, it is possible to increase returns dramatically.

It is natural to wonder what happens over periods other than one week. With monthly data, our results are consistent with those in Diebold and Li [2003]. The monthly data also seem to have a unit root, and the correlogram of the series of differences in Exhibit 10 indicates that these data should be regarded as random.

Indeed, when fitting an autoregressive moving average model to the data, using it for prediction, and rebalancing the portfolio on a monthly basis, we obtain approximately the same profit as we would without incorporating prediction.

Finally, it is interesting to note that while our prediction implies that in the short term there is a momentum in yield changes, Diebold and Li [2003] show there is mean reversion over horizons of six months to one year. In fact, this mean reversion is essential in order to keep the term structure from exploding.

A combination of our predictor and the one suggested by Diebold and Li [2003] should account for both short-term momentum and long-term mean reversion. This combination is modeled as

$$dz_i(t) = \alpha\, dz_i(t - dt) - \beta\, z_i(t)$$

which is simply an AR(2) model for $z_i(t)$. Fitting AR(2) to z_1 and z_2 over the whole sample results in $dz_1(t) = 0.259dz_1(t - dt) - 0.001z_1(t)$, and $dz_2(t) = 0.201dz_2(t - dt) - 0.0015z_2(t)$, so the autoregressive coefficient is very close to the one obtained by our model, and there is an additional slow mean-reverting term.

While this model may work well in capturing both short and long horizons, for short-term prediction we find our ARIMA(1, 1, 0) gives better results. This is not surprising, as an additional parameter for capturing long-term behavior makes the model overfitted for short-term purposes.

ENDNOTES

[1]The theory we develop can be fitted in a straightforward manner for a continuum of maturities also.

[2]There are many other reasonable choices of u_i. For example, Diebold and Li [2003] use the Nelson and Siegel [1987] factor loadings.

REFERENCES

Diebold, F.X., and C. Li. "Forecasting the Term Structure of Government Bond Yields." NBER Working Paper No. 10048, October 2003.

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

Jolliffe, I.T. *Principal Component Analysis.* New York: Springer-Verlag, 1986.

Litterman, R., and J. Scheinkman. "Common Factors Affecting Bond Returns." *The Journal of Fixed Income,* 1 (1) (1991), pp. 54-61.

Nelson, C.R., and A.F. Siegel. "Parsimonious Modeling of Yield Curves." *Journal of Business,* 60 (1987), pp. 473-489.

Reisman, H., and G. Zohar. "Excess Yields in Bond Hedging." *RISK*, 17 (12) (2004), pp. 93-97.