ESTIMATING YIELD CURVES IN TURKEY: FACTOR ANALYSIS APPROACH

C. Emre Alper * A

Aras Akdemir

Kazim Kazimov

14 July 2004

Abstract

In this paper, we perform factor analysis on yield curves estimated by Mc-Culloch and Nelson-Siegel methods. We estimate factors using nominal volumeweighted average monthly zero-coupon yields data from the Turkish Secondary Government Securities market. Our main aim is to characterize each monthly yield curve by three factors and forecast yield curves using time series properties of each factor. According to loadings of each factor, we label the factors as level, slope and curvature, respectively. We also examine their explanatory power in different sub-samples and explore their time series properties using an unrestricted VAR. We next forecast yield curves using AR-GARCH and random walk processes for the factors and compare their relative performance. We find encouraging results regarding explanatory power of three factor model and superior forecasting power of the AR-GARCH specification.

JEL classification: E43, C53, E47

Keywords: Term premium, Yield curve forecasting, Factor loadings

^{*}Corresponding author. Professor, Department of Economics, Boğaziçi University, Bebek 34342, Istanbul, Turkey. E-mail: alper@boun.edu.tr

1 Introduction

In this study, we characterize monthly yield curves in Turkey by estimating three factors which evolve through time. The data used in the study are the estimated one to fourteen month zero-coupon yields estimated by Alper *et al.* (2004) using McCulloch (1975) and Nelson-Siegel (1987) methods. We first explain time series properties of yield curves using factor analysis and then calculate out-of-sample forecasts by recursive estimations of the three factors.

Factor analysis is a widely used method to describe correlation relation between variables. Its use in empirical finance literature started with Litterman and Scheinkman (1991), who identified three factors explaining the variation in returns of fixed-income securities with various maturities. They labelled these factors as level, slope and curvature. Knez *et al.*(1994) extended the argument of Litterman and Scheinkman (1991) to analyze money market returns using the method developed by Joreskog (1967). Bliss (1997) emphasize hedging applications of factor analysis and included an extensive survey of literature.

Since factor analysis assumes that sufficient number of factors can characterize a yield curve, one can forecast yield curves by forecasting the factors. Duffee (2000) forecasted yields using affine term structure models. Diebold and Li (2003) forecasted yields by interpreting parameters of Nelson-Siegel model as level, slope and curvature factors.

We use a three-factor model and estimate monthly factors of yields between January 1992- March 2004. Rather than using the change in yields like Bliss (1997) or returns like Knez *et al* (1994), we use nominal yields because we aim to forecast nominal yields. We first analyze explanatory power of factors across time using a two year moving window. Next, we discuss time series properties of factors and estimate an unrestricted VAR to determine the persistency and exogeneity of each of the three factors. We then identify a data generating process underlying each factor and make forecasts using this specification. Finally we compare forecast results with the random walk model specification, as a benchmark case.

We proceed as follows. Section 2 discusses the basics of factor analysis. Section 3 explains the methodology and the dataset used in estimations. Section 4 provides estimation and forecast results. Section 5 concludes.

2 Factor Analysis

In this section, we introduce the basics of factor analysis. First, we broadly describe orthogonal factor model. Next, we discuss issues pertaining to factor loading and factor estimations.

2.1 Orthogonal factor model

Essentially factor analysis determines the covariance among many variables in terms of a few underlying, but unobservable random quantities called the *factors*. Within the context of term structure of interest rates, factors are a chosen in order to characterize the yield curve compactly.

Factor analysis applications to the term structure generally utilizes the orthogonal factor model. The orthogonal factor model assumes a linear relationship between the observable random vector X, unobservable random variables F_1, F_2, \dots, F_k , called common factors, and k sources of variation, called errors, at each period. In matrix notation,

$$X_{n \times 1} - \mu = L_{n \times k} \cdot F_{k \times 1} + \varepsilon_{n \times 1} \tag{1}$$

where $E(X) = \mu$. The matrix L is the matrix of factor loadings. Each element in the factor loading matrix gives information about the effect of a unit change in a factor on the observed random vector X. For each period there are nobservations and $(n + 1) \cdot k$ unknowns, hence direct estimation is not possible. Restrictions on the factors F, and the errors, ε , are needed to make estimation feasible. Orthogonal factor model assumes

$$E(F) = 0; \ E(\epsilon) = 0; \ Cov(F) = I; \ Cov(\varepsilon) = \Psi; \ Cov(\varepsilon, F) = 0$$
(2)

where Ψ is a diagonal matrix. These assumptions of the orthogonal factor model implies the following variance-covariance matrix for X

$$\Sigma = Cov(X) = LL' + \Psi.$$
(3)

Since the factors, F, and the errors, ε , do not appear in the variance-covariance matrix, Σ , the number of parameters is reduced and estimation of factor loadings, L, and idiosyncratic variances, Ψ , becomes feasible.

The share of the variance of each element of X explained by the common factors is called the *communality*, and the remaining portion is called *idiosyncratic variance*. Let σ_{ii} denote the variance of the i^{th} variable of X. Then σ_{ii} can be written as

$$\sigma_{ii} = \underbrace{l_{i1}^2 + \dots + l_{ik}^2}_{communality} + \underbrace{\Psi_i}_{idiosyncratic \ variance} \tag{4}$$

where l_i is the i^{th} row of factor loading matrix L and Ψ_i is the i^{th} element of the diagonal matrix Ψ .

When k > 1, there is always an ambiguity concerning the factor loadings. This is because factor loadings L can be multiplied by any orthogonal matrix Tsuch that the resultant loading matrix $L^* = LT$ and L both give same communalities and factors with identical statistical properties. This indeterminacy can be utilized to "rotate" the original solution until the loadings have meaningful interpretation.

2.2 Estimation of loadings and factors

There are two popular methods of estimating factor loadings, namely, the *Princi*pal Component Method and the Maximum Likelihood Method. We use Maximum Likelihood in our analysis since it is the only method for factor extraction that provides us with basis for statistical testing procedures. We next explain the maximum likelihood factor extraction method by sketching Joreskog's (1967) iterative procedure.

Given that X and F come from a joint normal distribution and the assumptions of orthogonal factor model are satisfied, the log likelihood function can be written as

$$logL(L,\Psi) = -\frac{1}{2}N[ln|LL' + \Psi| + tr(S(LL' + \Psi))]$$
(5)

where S is the sample variance-covariance matrix¹. Joreskog (1967) shows that

¹One may also use sample correlation matrix with standardized values of X

when $L'\Psi^{-1}L = \Delta$ is diagonal, it is possible to find an \hat{L} that maximizes the log-likelihood function for a given Ψ . \hat{L} is given by

$$\hat{L} = \hat{\Psi}^{1/2} \hat{E} (\hat{\Lambda} - I)^{1/2} \tag{6}$$

where \hat{E} is the $n \times k$ matrix of first k normalized eigenvectors of $\hat{S} = \hat{\Psi}^{-1/2} S \hat{\Psi}^{-1/2}$ and $\hat{\Lambda}$ is the $k \times k$ diagonal matrix of corresponding eigenvalues.

The estimation process starts by choosing an initial value for $\hat{\Psi}$ and using equation (6) calculates \hat{L} . Then, plugging this \hat{L} and starting value of $\hat{\Psi}$ into equation (5), finds iteratively $\hat{\Psi}$ that maximizes the log-likelihood function. With the new value of $\hat{\Psi}$ calculates a new \hat{L} and continues this process until the convergence is achieved, when the difference between successive values of \hat{L} and $\hat{\Psi}$ are negligible.

Once the estimated loadings are obtained, factors can be computed by

$$F = (L'\Psi^{-1}L)^{-1}L'\Psi^{-1}(X-\mu).$$
(7)

3 Data and Methodology

The data set used in this study include the following two series: volume weighted monthly average yield series and number of days to maturity for zero-coupon bonds and bills traded in the Turkish government secondary securities market. The zero-coupon yields used in this study are the one to fourteen month yields estimated by McCulloch and Nelson-Siegel methods. The maximum maturity of fourteen month is selected since scarcity of observations beyond this maturity decreases the reliability of yield estimations.

We follow maximum likelihood iterative procedure developed by Joreskog (1967) to estimate factor loadings. We use correlation matrix to estimate loadings and standardized values of yields to estimate factors²(Figures 1-3). In order to determine the number of factors likelihood ratio (LR) statistic as proposed in

 $^{^{2}}$ Factors estimated using correlation matrix or covariance matrix is essentially the same. Loadings are different however transformation of loadings is possible, see Johnson and Wichern (2002) for details.

Joreskog (1969) is used. Factors are rotated so that loadings are identical on the first factor³. As a result, a change in first factor shifts the whole yield curve. Thus one may interpret the first factor as determining the level of the yield curve. One may also interpret the remaining two factors as slope and curvature without further rotation. Interpretation of factors will be discussed in section 4 in detail.

In order to assess the explanatory power of each factor across time, we use a two year moving window as proposed by Bliss (1997). Starting with January 1992, we compute proportion of total sample variance due to each factor using two year data windows, then move the window until the end of the dataset.

In order to explore the persistency of each factor, we estimate an unrestricted VAR and interpret impulse-response functions.

We obtain one to twelve months ahead yield forecasts in the following manner. First, we identify each factors' time series property through Box-Jenkins methodology. We find that each of the three factors for the two methods are stationary at level and follow ARIMA(3,0,0)-GARCH(1,1) process. We assume that this is the underlying data generating process for each factor and proceed with accordingly henceforth. Secondly, we calculate out-of-sample forecasting based on assumed time series properties for each factor. Out-of-sample forecasts are calculated using AR(3)-GARCH(1,1). We also calculated out-of-sample forecasts for the random walk process, as a benchmark case, and compare them. Finally, from forecasted factors, we calculate yields for maturities for one to fourteen months, and compare the forecasting performance of random walk process to the AR(3)-GARCH(1,1), across time as well as belonging to different maturities.

Out-of-sample forecasting is based on recursive estimations of factor loadings and AR-GARCH terms. First we obtain one to twelve months ahead forecasts for the sample January 1992- January 1997. Next we add one observation and forecast using the updated loadings and AR-GARCH terms. We continue to add observations and calculate one to twelve-month-ahead forecast errors until March 2003.

³See Appendix for details on rotation of factors.

4 Results

4.1 Estimation Results

Using both Nelson-Siegel and McCulloch methods, we estimate loadings for yields. Based on the LR test statistics as suggested by Joreskog (1969), we conclude that three factor model is sufficient for the whole sample for both methods. The p-values are large for the hypothesis that number of factors is three.

The paths of rotated factor loadings with respect to maturity given in Figures 4-5, provide strong evidence for labelling them as level, slope and curvature. The first factor has same loadings on yields of all maturities, which implies that increase in this factor will increase yields for all maturities equally. Hence we interpret the first factor as the level of the yield curve. Loadings on the second factor are small in magnitude for short maturities and large for magnitude in long maturities, which implies that any increase in the second factor increases yields for longer maturities relatively more. Hence, one can interpret the second factor are smaller in magnitude for middle maturities; increase in the third factor will increase short and long maturities more than middle maturities. Hence the third factor can be interpreted as the curvature of the yield curve.

When we consider different sub-samples, we note that loadings on slope factor are increasing with maturity and plots of loadings on curvature factor make inverted humps in the middle maturities (Figures 6 and 7).

The cumulative proportion of total sample variance due to each factor and for each window are presented in Figures 8 and 9. It can be observed that the explanatory power of the level factor is lower during periods of high volatility. The explanatory power of the level factor during 1994 and 2001 crises are lower than the other periods. This is due to low correlation between yields during periods of high volatility. During 1994 and 2001 crises, we observe high term premiums, steep yield curves and almost no change in total explanatory power of the three factors. The increase in the explanatory power of the slope and curvature factors compensate the decrease in the explanatory power of the level factor. This finding indicates the robustness of the three factor model. For the VAR analysis, we select four lags using the LR criteria. The impulseresponse functions given in Figures 10 and 11 exhibit persistence for the level and the slope factors but no persistence for the curvature factor. In addition, level factor has significant effects on slope and curvature factors. We observe that one standard deviation shock to level factor increases slope factor for one period and curvature factor up to four periods. We may interpret this observation as follows: given the Turkish Secondary Government Securities Market for the period 1992-2004, an adverse shocks to the economy that shift yield curve up in the current period, not only increases the overall level of the interest rates but also makes the yield curve steeper, affecting yields for longer maturities more in the next periods. We also note that shocks to the slope factors do not affect level and curvature and shocks to curvature factor do not affect any factor. These results are robust to the choice of method for constructing monthly yield curves. We conclude that level factor is the major determinant in forecasting future yield curves for Turkey.

4.2 Forecast results

We follow Box-Jenkins methodology to identify, estimate and diagnose the time series properties of the three factors. Following the identification stage, we discover that the data generating process for each of the three factors is ARIMA(3,0,0)-GARCH(1,1). Nevertheless, we estimate each factor by assuming two different data generating precesses, namely, the random walk, as a benchmark and AR(3)-GARCH(1,1).

In order to compare the relative performances of AR-GARCH and random walk models, we use the average RMSE (Root Mean Squared Error) of forecasts. Table 1 shows average RMSE statistics for one to twelve-month-ahead forecasts. For all forecasts horizons RMSE criterion is lower for the AR-GARCH specification. Tables 2 and 3 show average RMSE statistics for different maturities. Based on the RMSE criterion, we conclude that AR-GARCH specification is superior for all forecast horizons and across all maturities.

Tables 4 and 5 present descriptive test statistics of the forecast errors for selected forecast horizons and methods. Based on the Augmented Dickey-Fuller

(ADF) test statistics, all series are stationary at the level. When we consider autocorrelation LM test statistics for two lags, AR-GARCH forecast errors for one-month-ahead forecasts have no autocorrelation whereas, for one year forecast errors as well as for the forecast errors from random walk specification, we fail to reject the null hypothesis of no autocorrelation. Hence we conclude that forecasts from random walk specification as well as AR-GARCH specification for longer forecast horizons may be suboptimal.

Diebold and Li (2003) conduct a similar analysis for the US market. They compare the AR(1)-GARCH(1,1) specification and random walk and find inconsistent results concerning the forecasting performance of two models. Diebold an Li (2003) also find out that forecast errors are autocorrelated hence suboptimal for all forecast horizons. When compared to their findings, our AR-GARCH specification outperforms random walk for all maturities and forecast errors are serially uncorrelated in the one month horizon.

5 Conclusion

In this paper, we use Turkish Secondary Government Securities Market data and characterize monthly yield curves by three factors which evolve dynamically through time. We use one to fourteen months zero-coupon yields, estimated by Alper *et al* (2004) using the McCulloch and the Nelson-Siegel methods. We analyzed explanatory power of the three factor model, time series properties of the estimated factor loadings and factors and finally, relative performance of one to twelve-month-ahead forecasts of AR-GARCH and random walk models.

We conclude that three factor model is able to capture most of the variation of monthly yield curves in Turkey. Additionally, the level explanatory power of three factor model is robust to the choice of sub-samples in the dataset. During periods of high volatility, explanatory power of slope and curvature factors increase and the level factor decreases without affecting the sum total of the three factors. This indicates that term structure models utilizing a single factor may not perform sufficiently well even worse in periods of high volatility, whereas three factor models procedures will perform equally good in stable and volatile periods. The VAR analysis indicates persistent level and slope factors and significant effects of level on slope and curvature factors. These results imply that shocks that affects the level of the yield curve this period will affect both the level and the shape in following periods.

One to twelve-month-ahead forecasts are calculated for both the AR(3)-GARCH(1,1) and random walk specifications. The forecasts are based on recursive estimations of AR-GARCH terms and factor loadings. We found that AR-GARCH specification outperforms random walk for all periods and maturities. In contrast to the findings by Diebold and Li (2003), we usually obtain non serially correlated forecast errors for the one-month-ahead AR-GARCH specification.

Directions for further research include an analysis of the efficacy of the monetary policy in Turkey by investigating the response of the term premium as well as the estimated factors of the yield curve.

References

- Alper, C. Emre, Aras Akdemir and Kazim Kazimov, 2004, "Estimating the Term Structure of Government Securities in Turkey," *Boğaziçi University Re*search Papers: ISS/EC-2004-03
- Bliss, Robert, 1997, "Movements in the Term Structure of Interest Rates," Federal Reserve Bank of Atlanta Economic Review, Vol. 82, 16-32.
- [3] Diebold, Francis X., and Canlin Li, 2003, "Forecasting the Term Structure of Government Bond Yields," NBER Working Paper no: 10009.
- [4] Duffee, Gregory, 2000, "Term premia and Interest Rate Forecasts in Affine Models," *The Journal of Finance*, Vol.57, 405-443.
- [5] Johnson Richard, and Dean Wichern, 2002, Applied Multivariate Statistical Analysis, Prentice Hall, Upper Saddle River, NJ.
- [6] Joreskog, Karl G., 1967, "Some contributions to Maximum Likelihood Factor Analysis," *Psysmetrika*, Vol. 32, 43-482.
- [7] Joreskog, Karl G, 1969, "A general Approach to Confirmatory Maximum Likelihood Factor Analysis," *Psysmetrika*, Vol. 34, 183:202.
- [8] Knez, Peter J., Robert Litterman, and Jose Scheinkman, 1994, "Explorations into Factors Explaining Money Market Returns," *The Journal of Finance*, Vol. 49, 1861-82.
- [9] Litterman, Robert, and Jose Scheinkman, 1991., "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, Vol.1 54-61.
- [10] McCulloch, Huston J, 1975, "The Tax-Adjusted Yield Curve." The Journal of Finance, Vol. 30, 811-830.
- [11] Nelson, Charles R., and Andrew F. Siegel, 1987, "Parsimonious Modeling of Yield Curves," *Journal of Business*, Vol.60, 473-489.

Appendix

Rotation of factor loadings

We perform orthogonal rotation to interpret factors as level, slope and curvature. To achieve this, we rotate the loadings so that the loadings on the first factor is approximately same. We follow Bliss (1997) to create the rotation matrix T. Since we have three factors, T is the product of three two-dimensional clockwise orthogonal rotation matrices. Each matrix leaves one column of the loading matrix L unchanged. The matrices we use for rotation are given as follows

$$\mathbf{T_1} = \begin{bmatrix} \cos \theta_1 & \sin \theta_1 & 0 \\ -\sin \theta_1 & \cos \theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{T_2} = \begin{bmatrix} \cos \theta_2 & 0 & \sin \theta_1 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{bmatrix}$$
$$\mathbf{T_3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_3 & \sin \theta_3 \\ 0 & -\sin \theta_3 & \cos \theta_3 \end{bmatrix}$$

 $T = T_1T_2T_3$ is the orthogonal rotation matrix. We minimize the variance of the first column of $L^* = LT$ with respect to $\theta = \{\theta_1, \theta_2, \theta_3\}$, subject to θ taking on values in the closed interval $[-\pi, \pi]$, using constrained optimization.

Tables

Method	Forecast Horizon (Months)								
	1	1 2 3 6 9 10 11 12							
AR-GARCH	0.384	0.396	0.397	0.420	0.429	0.431	0.434	0.436	
Random walk	0.418	0.467	0.442	0.508	0.537	0.543	0.499	0.500	

Table 2: RMSE for one-month-ahead forecasts from different maturities

Method	Maturity (Months)									
	1	1 2 3 6 9 10 12 14								
AR-GARCH	1.290	0.726	0.562	0.290	0.262	0.293	0.376	0.461		
Random walk	1.535	0.842	0.641	0.375	0.312	0.347	0.450	0.557		

Table 3: RMSE for one-year-ahead forecasts from different maturities

Method	Maturity (Months)							
	1 2 3 6 9 10 12 1							
AR-GARCH	2.005	0.624	0.356	0.098	0.081	0.103	0.168	0.247
Random walk	3.246	0.868	0.483	0.183	0.119	0.146	0.243	0.368

Table 4: Error statistics for one-month-ahead forecasts

Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.137	0.725	0.002	0.063	0.093	0.000			
6 month	0.132	0.144	0.002	0.271	0.078	0.000			
9 month	-0.104	0.126	0.000	0.148	0.000	0.000			
1 year	-0.051	0.156	0.000	0.229	0.256	0.000			
·	I	AR-GAR	CH (Ne	lson-Siegel)					
Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.068	1.290	0.000	0.767	0.896	0.000			
6 month	0.008	0.290	0.000	0.532	0.238	0.054			
9 month	-0.026	0.262	0.000	0.799	0.411	0.000			
1 year	0.001	0.376	0.000	0.076	0.056	0.000			
Random walk (McCulloch)									
Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.288	1.164	0.000	0.062	0.000	0.000			
6 month	0.129	0.152	0.030	0.000	0.000	0.000			
9 month	-0.075	0.129	0.000	0.003	0.000	0.000			
1 year	-0.001	0.228	0.000	0.024	0.000	0.000			
Random walk (Nelson-Siegel)									
Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.088	1.535	0.000	0.000	0.000	0.000			
6 month	0.009	0.375	0.000	0.000	0.000	0.000			
9 month	-0.014	0.312	0.000	0.001	0.000	0.000			
1 year	0.019	0.450	0.000	0.111	0.000	0.000			

AR-GARCH (McCulloch)

J-B stands for p-value of Jarque-Bera test statistics.

Table 5: Error statistics for one-year-ahead forecasts

Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.169	0.792	0.001	0.006	0.023	0.000			
6 month	0.153	0.098	0.015	0.071	0.176	0.000			
9 month	-0.058	0.117	0.022	0.000	0.000	0.000			
1 year	-0.071	0.187	0.000	0.001	0.002	0.000			
	ŀ	AR-GAR	CH (Ne	lson-Siegel)					
Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.089	1.222	0.012	0.002	0.000	0.000			
6 month	0.057	0.176	0.000	0.002	0.001	0.098			
9 month	-0.018	0.138	0.008	0.000	0.000	0.000			
1 year	-0.026	0.274	0.003	0.001	0.000	0.000			
Random walk (McCulloch)									
Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.349	1.222	0.010	0.004	0.011	0.000			
6 month	0.178	0.176	0.000	0.314	0.086	0.001			
9 month	-0.010	0.138	0.007	0.002	0.003	0.000			
1 year	-0.019	0.274	0.000	0.008	0.029	0.000			
Random walk (Nelson-Siegel)									
Maturity	Mean	RMSE	ADF	LM 1 lag	LM 2 lag	J-B			
1 month	-0.114	1.802	0.005	0.009	0.003	0.239			
6 month	0.037	0.428	0.000	0.098	0.229	0.982			
9 month	-0.005	0.345	0.006	0.000	0.000	0.000			
1 year	0.005	0.493	0.002	0.000	0.001	0.000			

AR-GARCH (McCulloch)

J-B stands for p-value of Jarque-Bera test statistics.

Graphs



Figure 1: Rotated values of first factor for the yields estimated by two methods



Figure 2: Rotated values of second factor for the yields estimated by two methods



Figure 3: Rotated values of third factor for the yields estimated by two methods



Figure 4: Rotated factor loadings for cont. compounded yields (McCulloch)



Figure 5: Rotated factor loadings for cont. compounded yields (Nelson-Siegel)



Figure 6: Rotated loadings on factor 2 on January for two year windows



Figure 7: Rotated loadings on factor 3 on January for two year windows



Figure 8: Percentage of variations in yields explained by three factors (McCulloch)



Figure 9: Percentage of variations in yields explained by three factors (Nelson-Siegel)



Figure 10: Impulse response graphs for rotated factors (McCulloch)



Figure 11: Impulse response graphs for rotated factors (Nelson-Siegel)



Figure 12: Variance decomposition graphs for rotated factors (McCulloch)



Figure 13: Variance decomposition graphs for rotated factors (Nelson-Siegel)