

## **The Information in the Term Structure: A Non-parametric Investigation**

BRUCE MIZRACH  
*Rutgers University, USA*

### ABSTRACT

This paper proposes a non-linear term structure model that nests the discrete and continuous time models as special cases. I estimate the model non-parametrically using nearest-neighbours regression. In sample, the non-linear model matches the standard theories, but out of sample, it offers substantial improvement. Linear models fail to track future interest rates: a random walk dominates the forward rate as a predictor for 3-month Treasury bills. A non-linear forecast based on the spread is shown statistically to be the best forecast.

KEY WORDS Term structure; interest rates; non-parametric; non-linear; nearest neighbours

The expectations theory of the term structure links long-term interest rates to unobserved sequences of future short-term rates. In the standard risk-neutral framework, the  $\tau$ -period rate equals a discounted sequence of  $\tau$  future one-period rates. Empirical evidence on the linear model, though voluminous, is largely discouraging. As Fama (1984) notes in an article from which I draw the title, forward rates, unadjusted for risk premia, are poor forecasts of spot rates. Shiller, Campbell and Schoenholtz (1983) observe: 'The simple theory that the slope of the term structure can be used to forecast the direction of future changes in the interest rate seem worthless.'

Subsequent research has been less pessimistic. Fama and Bliss (1987) find predictive power at the long end of the maturity spectrum. Fama (1990) and Mishkin (1990) are able to predict future inflation and real rates with the yield curve. Still, one is led to question the validity of the expectations hypothesis when forecasts of short-term rates even 3 months ahead are so poor. I show below that a random walk dominates the forward rate as a predictor of the 3-month ahead, 3-month bill rate.

This paper contends that non-linearity is an important component of the empirical failure of the expectations hypothesis. The standard models impose linearity so as to make the bond pricing relation tractable empirically. In the discrete time framework developed by Shiller (1979) and Shiller, Campbell and Schoenholtz (1983), long-term rates are equated to expected short-term rates through a linearization of the holding period yield. Higher-order non-linear terms are omitted that I find to be important empirically. In the continuous time framework of

Address for correspondence

CCC 0277-6693/96/030137-17  
© 1996 by John Wiley & Sons, Ltd.

*Received January 1995*

Cox, Ingersoll and Ross (1985), the expected future spot rates are linear functions of the instantaneous rate of interest. This excludes interesting non-linearities like the time-varying risk premia found by Fama and Bliss (1987) and Stambaugh (1988).

The approach taken here is to allow future spot rates to vary non-linearly with underlying factors, which may include current spot rates. Standard estimation procedures are inappropriate in this setting, leading me to pursue a non-parametric approach. In so doing, I can examine a highly flexible functional relationship for the yield curve that nests the two popular term structure models as special cases.

Recent work by Dunn and Singleton (1986) and Lee (1989) has tried to incorporate non-linearities into the term structure as well. Dunn and Singleton look at a model with non-separable utility and durable goods. Lee allows for a covariance between the marginal utility of consumption in adjacent periods. Both papers, however, use only a portion of the information at hand, limiting their analysis to a method of moments estimation of the first-order conditions. While Euler equation methods provide tests of particular restrictions imposed by the model, they are generally unsuitable for the forecasting exercises I consider below.

The innovation of this paper is to test the non-linear model non-parametrically. A non-parametric approach provides great latitude in specification and distributional assumptions. The only restriction imposed on the data-generating mechanism for bond prices is that the expected future spot rate be a smooth function of the spread or forward rate differential. The disturbance terms can be from a general class of densities, not just the normal. Given the leptokurtosis ('fat-tailed' distributions) in bonds and bills, first uncovered by Fama and Roll (1968), the usual higher-moment assumptions may be counterfactual. These minimal restrictions on the model will enable me to test a number of competing theories.

I employ a non-parametric nearest-neighbours approach developed by Cleveland (1979). Nearest-neighbours regression uses the data scatter to weight more heavily observations near (in the mean square sense) to the dependent variables. Unlike ordinary least squares, which fits the regression surface globally, the nearest-neighbours estimator provides a local approximation. A consistent estimate of the conditional mean is achieved without parametric specification assumptions. Below, I use the non-parametric procedure successfully both in and out of sample.

I argue that the appropriate criterion by which to judge term structure models is in terms of forecasting. The paper therefore focuses on determining whether a non-linear model can produce superior predictions of future interest rates. I also employ formal testing procedures for determining what constitutes a significant forecast improvement. Relying on work in Mizrahi (1995), the sample correlation coefficient between forecast errors is adapted as a test statistic for forecast comparison. This procedure will enable me to verify statistically that a forecast improvement of a certain number of basis points is significant or not. Previous efforts have only reported Thiel's mean-squared error ratio or relied on Efron's (1983) bootstrap methodology.

The organization of the paper is as follows. The next section extends the continuous and discrete time frameworks for the term structure to a non-linear setting. The third section reports least squares estimates of the standard model. Looking first at forward rates, I confirm existing results on the non-risk adjusted term structure. The 91-day-ahead, 91-day forward rate is a consistently biased predictor of the future spot rate for 3-month Treasury bills. The 3- and 6-month interest differential fits badly in-sample. The poor results serve as motivation for the non-parametric procedure in the fourth section.

In the fifth section I look at a number of non-parametric models in-sample. A non-linear autoregressive model outperforms its linear counterpart, and the same is true of the spread equation. Interestingly, a squared neighbour term proves particularly useful for both estimation and forecasting.

The sixth section is devoted to forecasting. I construct the test statistic used to evaluate forecast improvement and compare the out-of-sample performance of several non-linear models. Non-linear versions of the spread proved to be the best predictors. They dominate the forward rate at a high level of significance.

In the seventh section, I pool some of the forecasts from Section 5 to extract more information out of the data. A simple weighted average of the two best non-parametric forecasts has a mean squared error less than half that of the forward rate. I am also able to show that a constant risk premium cannot salvage the linear expectations hypothesis. The final section presents some conclusions.

### THE EXPECTATIONS HYPOTHESIS

This section develops the standard term structure framework in both discrete and continuous time. I show that the non-linear model nests both of the linear models as special cases. I also point out the principal areas in which non-linearity is likely to enter the term structure.

#### The discrete time framework

The linear discrete time framework assumes that bonds are priced using the present value formula.<sup>1</sup> The price at time  $t$  of a 1.00 dollar,  $\tau$ -period bond paying coupon  $\bar{i}$ ,  $p_t^{(\tau)}$ , is given by

$$p_t^{(\tau)} = \sum_{t=1}^{\tau} \frac{\bar{i}}{(1+i_t^{(1)})^t} + \frac{1.00}{(1+i_t^{(1)})^{\tau}} \quad (1)$$

where  $i_t^{(1)}$  is the one-period spot rate. The holding period yield is the change in price plus the coupon,

$$H_t^{(\tau)} = (p_{t+1}^{(\tau)} - p_t^{(\tau)} + \bar{i})/p_t^{(\tau)} \quad (2)$$

The Shiller (1979) linearization proceeds by assuming that the capital asset pricing model holds, that is,

$$E[H_t^{(\tau)}] = i_t^{(1)} + \phi^{(\tau)} \quad (3)$$

where  $\phi$  is a constant term premium. Substituting equation (2) into (3) and linearizing around the coupon rate yields the following expression for the long rate:

$$i_t^{(\tau)} = \sum_{j=0}^{\tau-1} W_j E_t(i_{t+j}^{(1)}) + \Phi^{(\tau)} \quad (4)$$

where  $W_j = g_j(1-g)/(1-g^{\tau})$ ,  $g = 1/(1+\bar{i})$ , and  $\Phi^{(\tau)}$  is a weighted sequence of risk terms  $\phi^{(\tau)}$

In Shiller, Campbell and Schoenholtz (1983, SCS), a compact expression for the holding period yield is derived using the concept of *duration*. The duration of a  $\tau$ -period bond with coupon  $\bar{i}$  is given by

$$D_{\tau} = (1-g^{\tau})/(1-g) \quad (5)$$

The resulting linear approximation for the holding period yield is

$$H_t^{(\tau)} = \frac{D_{\tau} i_t^{(\tau)} - (D_{\tau} - D_1)(i_{t+1}^{(\tau-1)})}{D_1} \quad (6)$$

Note that since the coupon effects are second order, they drop out of equation (6).

<sup>1</sup> This formula can be derived in, say, the Lucas (1978) framework with linear utility.

Denote the  $j$ -period-ahead forward rate on a  $\tau - j$  period bill by  $F_t^{(j, \tau-j)}$ . The forward rate is also conveniently written in terms of duration,

$$F_t^{(j, \tau-j)} = \frac{D_\tau i_t^{(\tau)} - D_j i_t^{(j)}}{D_\tau - D_j} \quad (7)$$

For concreteness, let  $j = 1$  and  $\tau = 2$ ,

$$F_t^{(1,1)} = \frac{D_2 i_t^{(2)} - D_1 i_t^{(1)}}{D_2 - D_1}$$

For bills, the duration is simply their time to maturity. Hence,

$$F_t^{(1,1)} = 2i_t^{(2)} - i_t^{(1)}$$

The expectations hypothesis assumes that forward rates equal expected future spot rates:

$$(\Delta i_{t+1}^{(1)})^e = F_t^{(1,1)} - i_t^{(1)} = 2(i_t^{(2)} - i_t^{(1)}) \quad (8)$$

One can also allow for a forward term premium

$$\theta^{(j, \tau-j)} \equiv F_t^{(j, \tau-j)} - (i_{t+j}^{(\tau-j)})^e \quad (9)$$

Equation (8) then becomes

$$(\Delta i_{t+1}^{(1)})^e = -2\theta^{(1,1)} + 2(i_t^{(2)} - i_t^{(1)})$$

Non-linearity is implicit in the discrete framework. Rewriting the holding period yield in terms of yield to maturity as in Shiller,

$$H_t^{(\tau)} = \frac{\bar{i} + (\bar{i}/i_{t+1}^{(\tau-1)}) + (i_{t+1}^{(\tau-1)} - \bar{i})/i_{t+1}^{(\tau-1)}[1 + i_{t+1}^{(\tau-1)}]^{\tau-1}}{(\bar{i}/i_t^{(\tau)}) + (i_t^{(\tau)} - \bar{i})/i_t^{(\tau)}[1 + i_t^{(\tau)}]^\tau} - 1$$

For  $\tau = 2$  and  $\bar{i} = 0$ , taking expectations yields

$$\frac{(i_{t+1}^{(1)})^e / (i_{t+1}^{(1)})^e [1 + (i_{t+1}^{(1)})^e]}{i_t^{(2)} / i_t^{(2)} [1 + i_t^{(2)}]^2} = (1 + (i_t^{(1)})) + \phi^{(2)} \quad (10)$$

As Shiller notes, equation (10) is a non-linear rational expectations difference equation in the one-period expected spot rate. A conjectural solution takes the form

$$i_2^{(1)} = E_t[f(i_t^{(1)}, i_{t+1}^{(1)}, \phi^{(2)})] \quad (11)$$

Assuming that equation (11) is invertible, this is an equation for the expected future spot rate,

$$(\Delta i_{t+1}^{(1)})^e = g(i_t^{(1)}, i_t^{(2)}, \phi^{(2)}) \quad (12)$$

There is much empirical support that  $\phi$  in equation (3) is time varying (see e.g. Startz, 1982). Equation (12) can also then be interpreted as the SCS model in which  $\phi$  has a time subscript.

### The continuous time framework

Alternatively, Cox, Ingersoll and Ross (1985, CIR) consider a continuous time representative agent economy. Agents maximize

$$E \int_t^T e^{-\rho s} \left[ \frac{c(s)^\gamma - 1}{\gamma} \right] ds \quad (13)$$

where  $\rho$  is a constant discount factor and  $\gamma$  is a risk aversion parameter. Agents must choose optimal consumption,  $c^*$ , and proportions,  $\alpha^*$ , to invest in a vector of production technologies,  $Y$ .

Consider the special case in which  $Y$  is a single state variable. Assume also that the mean and variance of the rates of return,  $r(Y)$ , are proportional to  $Y$ . Let  $Y$  evolve according to the following stochastic differential equation:

$$dY(t) = [\xi Y + \zeta] dt + \nu \sqrt{Y} dw(t) \quad (14)$$

where  $\xi$  and  $\zeta$  are constants, with  $\xi > 0$ ,  $\nu$  is a vector whose first argument is a constant  $\nu_0$ , and  $w(t)$  is a one-dimensional Weiner process. The equilibrium interest rate, CIR show, is a diffusion process

$$dr = \kappa(\theta - r) dt + \sigma \sqrt{r} \nu \sqrt{Y} dw(t) \quad (15)$$

with  $\kappa$ ,  $\theta$ ,  $\sigma^2$  constant, and  $\kappa\theta > 0$ ,  $\sigma^2 > 0$ . The first term is the drift of the diffusion, and the second term is the variance.

The advantage of process (15) is that the conditional moments are described solely by past observations. The conditional mean and variance are given by CIR as

$$E[r(s) | r(t)] = r(t)e^{-\kappa(s-t)} + \theta(1 - e^{-\kappa(s-t)}) \quad (16)$$

$$\text{var}[r(s) | r(t)] = r(t)(\sigma^2/\kappa)(e^{-\kappa(s-t)} - e^{-2\kappa(s-t)}) + (\theta\sigma^2/\kappa)(1 - e^{-\kappa(s-t)})^2 \quad (17)$$

A closed-form solution for bond prices is then obtained. Let  $P(\tau, t)$  be the price at time  $t$  of a default-free discount bond with time to maturity  $\tau$ , and let  $p(\tau, t)$  be  $\log(P(\tau, t))$ . CIR then show

$$p(\tau, t) = a(\tau, t) + b(\tau, t)r(t) \quad (18)$$

where

$$a(\tau, t) = \log \left[ \frac{2\psi e^{[(\kappa + \lambda + \psi)(\tau - t)]/2}}{(\kappa + \lambda + \psi)(e^{\psi(\tau - t)} - 1) + 2\psi} \right]^{2\kappa\theta/\sigma^2} \quad (19)$$

$$b(\tau, t) = \left[ \frac{2(e^{\psi(\tau - t)} - 1)}{(\kappa + \lambda + \psi)(e^{\psi(\tau - t)} - 1) + 2\psi} \right] \quad (20)$$

with

$$\psi = ((\kappa + \lambda)^2 + 2\sigma^2)^{1/2} \quad (21)$$

$\lambda$  is a factor risk premium.

Assumptions (13)–(15) required to linearize the continuous time framework give it many of the same empirical implications as the discrete time model. To see this, define the yield to maturity,  $e^{[-(\tau - t)i(\tau, t)]} \equiv P(\tau, t)$ . One then has the following simple expression for the spot rate:

$$i(\tau, t) = [r(t)b(\tau, t) - a(\tau, t)]/(\tau - t) \quad (22)$$

The spread becomes a linear function of the instantaneous rate of interest,

$$i(\tau, t) - i(\tau - 1, t) = \alpha + \beta r(t) \quad (23)$$

where  $\alpha = a(\tau - 1, t)/(\tau - 1 - t) - a(\tau, t)/(\tau - t)$ , and  $\beta = b(\tau, t)/(\tau - t) - b(\tau - 1, t)/(\tau - 1 - t)$ . The forward rate is given by

$$F(\tau - 1, t) = p(\tau - 1, t + 1) - p(\tau, t) \quad (24)$$

Consider for concreteness  $\tau = 1, 2$ . The forward rate reduces to

$$F(1, t) = (2 - t)i(2, t) - (1 - t)i(1, t)$$

which is identical to equation (7) with  $t = 0$ . As for the holding period yield,

$$H(\tau, t + 1) = p(\tau - 1, t + 1) - p(\tau, t) \quad (25)$$

For two-period bonds,

$$H(2, t + 1) = p(1, t + 1) - p(2, t)$$

The classic implication in the discrete time framework is that expected holding period yields should be equal across maturities. Taking expectations on both sides of equation (25),

$$E_t H(2, t + 1) = p(1, t + 1)^e - p(2, t) = H(1, t + 1) \quad (26)$$

Substituting from above, it follows that

$$i(1, t + 1)^e = F(1, t) \quad (27)$$

The expected one-period spot rate equals the forward rate, just as in equation (8).

The CIR one-factor model does not appear to be supported empirically. Brown and Dybvig (1986) estimated the underlying parameters of equation (15) using a cross-section of interest rates. They found considerable time series variation in these parameters across sub-periods. Gibbons and Ramaswamy (1993) derive non-linear restrictions from the CIR model, but they look only at the unconditional moments of bonds returns. Gibbons and Ferson (1985) and Stambaugh (1988) have tried to estimate the dimension of the factor matrix. Both reject the one-factor model, though Stambaugh cannot reject a model with three factors.

Langtieg (1980) and Oldfeld and Rogalski (1987) have extended the arbitrage approach to multifactor models.<sup>2</sup> The pricing expression (18) becomes

$$p(\tau, t) = a(\tau, t) + b(\tau, t)x(t) \quad (28)$$

where  $x(t)$  is now a  $k$ -dimensional vector of factors.

In the empirical work that follows, I utilize a non-linear version of equation (28), allowing the spot rates to vary smoothly with the factors,

$$i(\tau, t) = f(x(t)) \quad (29)$$

This functional form can capture the time variation in the parameters of equation (15) found by Brown and Dybvig. Similarly, if the factors followed more complicated stochastic processes, the conditional expectations (16) and (17) might be non-linear functions as well. Forecasting exercises below will use equation (29) to predict out of sample,

$$i(1, t + 1) = f(x(t + 1) | x(t)) \quad (30)$$

This is identical to the discrete model (12) with  $x(t) = [i(1, t), i(2, t), \phi]$ . Therefore, the non-linear model nests the two popular models as special cases.

<sup>2</sup>These models have a much more tenuous equilibrium foundation. There is no guarantee, for arbitrary stochastic processes for the factors, that an equilibrium will even exist.

## THE LINEAR MODEL REVISITED

I begin the empirical investigation where previous work left off by estimating the standard linear model for 3- and 6-month treasury bills. Its many inadequacies leads me to the more general models that follow.

**In-sample estimation of the linear model**

In the simplest version of the expectations hypothesis, the term premia are ignored. The current two-period rate equilibrates the current one-period rate and the expected future one-period spot rate. A straightforward test of the model would be that the forward rate be an accurate predictor of the realized change in the spot rate. The standard test is to regress the realized  $\Delta i_{t+1}^{(1)}$  on the forward spread:

$$\Delta i_{t+1}^{(1)} = \alpha + \beta(F_{t+1}^{(1,1)} - i_t^{(1)}) + \mu_{t+1} \quad (31)$$

The linear model with rational expectations implies that  $\alpha = 0$ , and  $\beta = 1$ , and that  $\mu_{t+1}$  should be serially uncorrelated. These restrictions imply that the forward rate is the only variable necessary to predict changes in the spot rate.

Equation (31) is one of the most widely studied relationships in economics. The strict version has been widely tested and rejected. Shiller, Campbell and Schoenholtz (1983), Fama (1984) and Froot (1989) reject the model, finding that  $\nu < 1$ , and  $\alpha > 0$ . For the very short rates,  $\beta$  is often insignificantly different from zero. Mankiw and Miron (1986), Shiller (1979) and Mankiw and Summers (1984) also reject variants of equation (31). Froot (1989) also contains an excellent survey of these results.

While equation (31) is soundly rejected, authors such as Fama (1984) find the yield curve does contain information. His innovation was to ask the alternative question as to whether the slope of the yield curve does provide any guide to changes in future short rates. Consider the linear version of equation (12):

$$\Delta i_{t+1}^{(1)} = \alpha + \beta(i_t^{(2)} - i_t^{(1)}) + \mu_{t+1} \quad (32)$$

As Fama shows, even when tests of equation (31) fail, one often finds that  $\beta$  in equation (32) is significantly positive.

**Data and testing**

I have a sample of almost 3000 daily interest rate observations from the Bank of America for the period January 1980 to December 1987 for 3- and 6-month (13 and 26 weeks respectively) Treasury bills. This data set was split into two blocks for analysis. The period from May 1980 to December 1984 was used for estimation purposes to allow consistent comparisons throughout the paper. January 1985 to December 1987 is then reserved for forecasting exercises. Wednesday rates were chosen from the daily data to form a weekly data set of 413 observations.<sup>3</sup>

OLS estimation of equation (31) is reported in line one in Table I. While significantly positive, the slope term is significantly different from one. The constant term is strongly negative. This reflects in part the drop in interest rates in the first part of the sample. This will show up again as a strong negative bias. The  $\bar{R}^2$  is above 10%, and the forward rate proved to be the best in-sample predictor.

<sup>3</sup>There were five observations missing from the sample. In each case, the Thursday rates were used instead.

Table I. OLS estimation of the linear model<sup>a</sup>

Model	Constant	Slope	$\bar{R}^2$	DW
Forward rate (31)	-4.248 (-5.52)	0.366 (-5.30)	0.101	0.095
Spread (32)	-0.336 (-2.06)	0.886 (-2.55)	0.022	0.087

<sup>a</sup>The dependent variable is the 13-week difference in the 91-day Treasury bill rate. The estimation interval is 5/14/80 to 12/26/84. *T*-statistics are in parentheses. DW is the Durbin-Watson statistic.

Line two of Table I reports the OLS estimation of equation (32). The spread is the difference between current 3- and 6-month rates. The slope coefficient is significantly different from zero, but also significantly different from two as would be implied by equation (8). The constant is significantly negative which is compatible with the existence of a term premium. Clearly, additional work needs to be done. The next section develops the non-parametric procedures I will need to estimate the more complicated term structure models.

### NEAREST-NEIGHBOURS ESTIMATION

#### Motivation

Non-parametric statistical procedures are ideal tools when there is little *a priori* information as to the structural model. Minimal assumptions are required for purposes of inference. Consider a function for the conditional mean of  $y$  given observations on exogenous variables  $x$ :  $E[y | x] = f(x)$ . Suppose that the functional form is unknown and the realized values of  $y$  differ from  $f(x)$  by disturbances drawn from a general class of densities. Parametric assumptions will almost surely result in specification error. Even when the functional form is appropriately specified, the disturbances may not be normally distributed, and least squares estimates may be highly inefficient.

In practice, most data analysts would assume the conditional mean to be smooth, take a Taylor expansion, drop the higher-order terms, and assume that the disturbance term is normal. Much of the empirical work on the term structure proceeds in one of these two ways.

However, neglecting the higher-order terms of a Taylor expansion can lead to misleading inference. A truncated Taylor expansion estimated by least squares does not provide any information *per se* on the linear terms of the expansion. As White (1980) has stressed, least squares is a global approximation to a regression function. Least squares solves the problem

$$\min_{\beta \in B} \sum_{i=1}^n [y_i - f(\beta)]^2 \quad (33)$$

If I were to replace an analytic<sup>4</sup>  $f(\beta)$  with the first two terms of a Taylor series approximation, the solution to equation (33), call it  $\beta^*$ , will generally *not* resemble in any way the coefficients of the Taylor expansion. White shows that  $\beta^*$  is the solution to a weighted least squares approximation to

<sup>4</sup>This implies that  $f$  is a smooth function and that all its partial derivatives exist.



the true model. Only in the case where the approximation differs by an independent additive error term will  $\beta^*$  provide correct inference. In addition, if the disturbances to equation (33) are not normally distributed, least squares will not be the maximum likelihood estimator.

The nearest-neighbours estimator is motivated by the desire to achieve the appropriate local weights for a Taylor expansion. To see this, index the sample points into pairs  $(y_i, x_i)$ ,  $i = 1, \dots, n$ . Now search across the entire sample of  $\xi_i$ 's,  $i \neq t$ , near to the current  $x, x_t$ . A Taylor expansion of the conditional mean  $x_t$  around any  $x_i$ ,

$$y_t = f(x_i)(x_t - x_i) + f'(x_i)(x_t - x_i)^2 + \dots \quad (34)$$

can be fit more accurately if I choose points near to  $x_t$ . For example, suppose  $y$  is thought to be the first-order Markov process

$$y_t = \rho x_t + \varepsilon_t \quad (35)$$

A regression of the  $y$ 's on the  $x$ 's nearest to  $x_t$  will provide an accurate estimate of  $\rho$ . Suppose that equation (35) is misspecified and the true model includes a squared term,

$$y_t = \rho x_t + \rho x_t^2 + \varepsilon_t \quad (36)$$

This falls into the case considered by White, and least squares would not accurately estimate  $\rho$  using the specification (35).

Given our assumptions on  $f$ , though, there exists a smooth inverse function relating  $x$  to  $y$ . The nearest-neighbours procedure chooses the  $q$ -nearest  $y_i$ 's for which  $\sqrt{(x_i - x_t)^2}$  is small. I can then express the conditional expectation as a function of values for the *dependent* variable,

$$E[y | x] = \sum_{k=1}^q w_k y_k \quad (37)$$

where the  $w_k$  are weights that vary with the size of the distance of  $x_i$  from  $x_t$ . One can provide a greater weighting to nearby points with exponential functions of the distance.

Since  $x_i$ 's near to  $x_t$  will be associated with  $y_i$  values near to  $y_t$  for any functional form smoothly relating the variables, the nearest-neighbour's estimator is robust to non-linearities that invalidate the least squares estimator. I can more closely approximate the Taylor expansion by fitting the regression surface locally. Regularity conditions are quite minimal. Intuitively, all that is required is that the functional form possess a Taylor expansion. This intuition is formalized in the next subsection.

Non-parametric approaches have begun to filter into financial economics. Diebold and Nason (1990) have analysed ten foreign exchange rates using the locally weighted regression procedure introduced by Cleveland (1979). Mizrach (1992) extends Diebold and Nason to a multivariate setting. Frank and Stengos (1989) have used the nearest-neighbours estimator to look at precious metal prices.

Non-parametric statistical techniques are ideal for estimating models that routinely appear in finance. Non-linearities seem inherent in market relationships. In addition, non-normality seems to be a key stylized fact about the unconditional distributions of financial data. These facts lead me to expect that non-parametric approaches will prove fruitful for the term structure.

### Neighbour selection and consistency

This section develops the uniformly weighted nearest-neighbours estimator. Consider a regression model

$$y_t = f(x_t) + \varepsilon_t \quad (38)$$

where  $x_t = (x_{t1}, \dots, x_{tp})$  is a  $1 \times p$  vector of explanatory variables,  $f$  is a smooth function mapping  $R_p \rightarrow R$ , and  $\varepsilon_t$  is an independent and identically distributed mean zero disturbance term. The nearest-neighbour regression function estimate at time  $t$  is given by

$$\widehat{f}(x_t) = \sum_{i=1}^n w_{it} I[\|x_i - x_t\| < \eta] y_i \quad i \neq t \quad (39)$$

where  $I$  is the indicator function,  $\| \cdot \|$  is the Euclidean norm

$$\|x_i - x_t\| = \left[ \sum_{j=1}^p (x_{ji} - x_{jt})^2 \right]^{1/2} \quad (40)$$

$\eta$  is some constant, and  $w_{it} = (w_{i1t}, \dots, w_{int})$  is a sequence of probability weights.

Stone (1977) cleverly formulated the consistency problem as being one of obtaining a consistent sequence of weights for the neighbours. For example, let  $q$  be the cardinality of the vector of nearest neighbours less than  $\eta$  from  $x_t$ ,  $\# \{k_t^q\} = (k_{t1}, \dots, k_{tq})$ . A simple average of the  $q$  neighbours would set  $w_{it} = 1/q$  for all  $i$  such  $\|x_i - x_t\| < \eta$ , and 0 otherwise.

Stone proved consistency for probability weights satisfying the following necessary and sufficient conditions:

$$E \sum_{i=1}^n w_{it} f(x_i) \leq C E f(x_t) \quad \forall t \geq 1, \infty > C > 0 \quad (41)$$

$$\sum_{i=1}^n w_{it} I[\|x_i - x_t\| > a] \rightarrow 0 \text{ in probability } \forall a > 0 \quad (42)$$

$$\max_i w_{it} \rightarrow 0 \text{ in probability} \quad (43)$$

If the weights are uniform, quadratic or triangular, these revert to the familiar conditions that as  $n \rightarrow \infty$ ,  $q \rightarrow \infty$ , but  $q/n \rightarrow \infty$ . Intuitively, this requires that, as the sample grows large, the number of neighbours must go off to infinity but at a slower rate than the sample size increases. Consistency then becomes a matter of imposing a selection rule involving  $\eta$ . As a practical matter, I will look over a range of  $q$ 's.

Having selected the neighbours, the next step is to construct estimates of the weights using least squares. Denote the  $n \times (q+1)$  matrix of neighbouring interest rates by

$$K_n = \begin{bmatrix} 1 & k_{11} & \dots & k_{q1} \\ 1 & k_{12} & \dots & k_{q2} \\ \vdots & \vdots & & \vdots \\ 1 & k_{1n} & \dots & k_{qn} \end{bmatrix} \quad (44)$$

which includes a constant term in the first column. The locally weighted regression estimate is

$$\widehat{f}(x_t) = k_t (K_n' K_n)^{-1} K_n' Y_n \quad (45)$$

where  $k_t$  is the  $t$ th row of  $K_n$ , and  $Y_n = [y_1, \dots, y_n]'$ . The least squares estimates of the weights will correspond to the simple average only in the unlikely event that the constant is the only significant regressor in equation (44). This procedure allows the data to weight the neighbours according to mean-square error loss for the entire regression function.

## THE NON-PARAMETRIC TEST OF THE EXPECTATIONS HYPOTHESIS

As noted earlier I can write a very large class of term structure models in the following manner:

$$E[\Delta i_{t+1}^{(1)} | x_t] = f(x_t) \quad (46)$$

where  $f: R_p \rightarrow R$  is a smooth function relating the  $p$ -vector  $x_t$  of state variables to the future spot rate. In general,  $f$  is unknown and the class of disturbances to equation (46) is not restricted to the normal density. Nearest neighbours can handle these difficulties, requiring only two inputs to the estimation, apart from the choice of state variables. We must choose  $q$ , the number of neighbours, and the weights  $w_k$  in equation (37).

The regularity conditions (41)–(43) require an increase in the number of neighbours as the sample grows large. Since the sample size is fixed in practice, I have chosen to report the results for a variety of values for  $q$ . In preliminary work, I looked at values of  $q$  from 1 to 100 (about 25% of the estimation sample). To remind the reader, this means that I regressed the spot rate on the  $q$  nearest values of the spot rate, where distance is measured as in equation (38). The temporal ordering of the data is irrelevant as the nearby spot rate may come from anywhere in the estimation sample.<sup>5</sup> While a large number of neighbours improved the fit in-sample, the best forecasting models had 10 or fewer neighbours.

I found that the weights had virtually no impact on the analysis. Cleveland (1979) used a cubic weighting structure. I tried several weights and found little difference from the local regressions reported in Table II. Down weighting the outlying neighbours only had the effect of making models with a large number of neighbours perform similarly to those with only a few ( $q \leq 10$ ).

In Table II, I contrast the nearest-neighbours estimator with the two reduced forms prevalent in the literature. I first considered autoregressive models of order  $p$ ,

$$E[\Delta i_{t+1}^{(1)} | x_t] = f(\Delta i_t^{(1)}, \dots, \Delta i_{t-p+1}^{(1)}) \quad (47)$$

In line three of Table II an OLS estimate of the linear AR(1) performed dismally. The  $\bar{R}^2$  is negative. Higher-order autoregressive models, though not reported, fared little better.

Based on Fama's suggestion, I set the factor matrix equal to current and lagged values of the spread between 3- and 6-month bills:

$$E[\Delta i_{t+1}^{(1)} | x_t] = f(i_t^{(2)} - i_t^{(1)}, \dots, i_{t-p+1}^{(2)} - i_{t-p+1}^{(1)}) \quad (48)$$

As noted earlier this model with  $p = 1$  showed some merit, but it was still dominated by the forward rate in sample.

Non-parametric formulations of these two equations fill out most of the rest of Table II. In column one the first information provided is the choice of state vector,  $x$ . For instance, AR(4) indicates equation (47) with  $p = 4$ , and Spread AR(4) is equation (48) with  $p = 4$ .  $q - NN$  denotes the number of neighbours in the locally weighted regression; thus, 5 - NN indicates 5 neighbours.

I also developed non-linear formulations of the neighbour regressions, allowing the spot rate to vary with the cross products of the neighbours,

$$E[\Delta i_{t+1}^{(1)} | x_t] = g(k_t) = \sum_{r=1}^q \sum_{s=1}^q k_{rs} k_{rs} \quad (49)$$

<sup>5</sup>In the forecasting exercises, the sample is dynamically updated.

Table II. Term structure models: in-sample estimation<sup>a</sup>

Model	$\bar{R}^2$	Rank
Forward rate	0.101	1
Spread	0.022	4
AR(1)	-0.004	13
Spread 1 — NN	0.022	6
Spread 5 — NN	0.015	9
Spread 10 — NN	0.022	5
Spread AR(4) — NN	-0.002	10
AR(1) 1 — NN	-0.004	12
AR(1) 5 — NN	0.060	3
AR(1) 10 — NN	0.100	2
AR(4) 1 — NN	-0.002	11
Non-linear Spread 1 — NN	0.018	8
Non-linear AR(1) 1 — NN	0.020	7

<sup>a</sup>The estimation interval is 5/14/80 to 12/26/84, a total of 242 observations. The dependent variable in each cases is the change in the 3-month Treasury bill rate.

I found that squared own neighbour terms worked best. This restricts  $s$  to equal  $r$  in equation (49). In Tables II and III the models labelled as non-linear include squares of the neighbours as explanatory variables. These models performed poorly in-sample but were quite helpful in the out-of-sample analysis below.

The in-sample nearest-neighbour estimates fared no better than their linear counterparts. For the spread equation, models of 1, 5 and 10 neighbours provided less than 5% of the explained variation in the 3-month rate. A fourth-order autoregressive spread model fared little better. For

Table III. Term structure models: forecasting<sup>a</sup>

Model	Bias	MSE	Rank	$K$ -stat.
Random walk	-15.32	34.17	11	1.780
Forward rate	-35.47	45.21	14	2.967
Spread	9.74	31.31	4	0.612
AR(1)	9.37	32.94	10	1.594
Spread 1 — NN	9.74	31.29	3	0.365
Spread 5 — NN	8.03	32.17	7	0.966
Spread 10 — NN	5.36	32.10	6	0.911
Spread AR(4) 1 — NN	3.44	31.97	5	1.296
AR(1) 1 — NN	9.40	32.92	9	1.584
AR(1) 5 — NN	4.88	40.38	13	3.433
AR(1) 10 — NN	4.86	76.29	15	5.634
AR(4) 1 — NN	2.61	34.36	12	1.789
Non-linear Spread 1 — NN	6.24	30.83	2	0.365
Non-linear AR(1) 1 — NN	9.40	32.91	8	1.554
Combination	0.18	30.38	1	

<sup>a</sup>The estimation interval is 5/14/80 to 12/26/84, a total of 242 observations. Fifty-two observations from 3/27/85 to 3/26/86 are used for forecast combination. The forecast interval is 4/2/86 to 12/30/87, a total of 92 observations. Bias and mean squared error (MSE) are  $\times 10^{-2}$ .  $K$ -stat. is given in the text by equation (53). It is distributed as a standard normal variable, with a 90% two-sided critical value of 1.65.

the non-parametric autoregressive models, things seem a little brighter. While the AR(1) and AR(4) 1-neighbour equations perform badly, the 5- and 10-neighbour equations seem promising. The 10-neighbour equation is the only non-linear model to explain as much as the forward rate.<sup>6</sup>

Based on  $\bar{R}^2$ , the non-linear models do not distinguish themselves from their linear analogs in sample. However, they do offer significant improvement in their out-of-sample predictive power. Forecasting is arguably the best way to compare these models, so I now turn to the analysis out of sample.

## A COMPARISON OF FORECASTS

### Evaluating forecast improvement

While finding correlation in the term structure with future state variables provides useful information, the expectations hypothesis is, above all, a bond pricing theory. The true test of such a model is its ability to predict the path of future interest rates, and I need a precise way to discriminate between alternatives. Following Mizrach (1995), I propose a formal test for evaluating improvement in forecast performance. Surprisingly, there is very little literature on this subject.

Consider two forecasts,  $z_1$  and  $z_2$ , and let the respective forecast errors be  $e_i \equiv y - z_i$ ,  $i = 1, 2$ . Let  $MSE_i$  be the mean squared error of forecast  $i$ :

$$MSE_i = (1/n) \sum_{t=1}^n e_{it}^2 \quad (50)$$

This application requires some way to determine if two mean squared errors are different from one another. The standard  $F$ -test is tempting but not appropriate here. The two MSEs are not draws from independent random samples.

Following Granger and Newbold (1986), assume that  $(e_1, e_2)$  is a bivariate normal population with zero mean and finite variances  $\sigma_1^2$  and  $\sigma_2^2$ . Each forecast is assumed to be unbiased and serially uncorrelated. Consider now the normally distributed random variables  $(e_1 + e_2)$  and  $(e_1 - e_2)$ :

$$E[(e_1 + e_2)(e_1 - e_2)] = \sigma_1^2 - \sigma_2^2 \quad (51)$$

The two error variances will be equal if and only if this pair of random variables is uncorrelated, i.e.  $\rho = 0$ . A simple test for equality of the MSEs is then based on the sample correlation coefficient:

$$r \equiv \frac{\sum_{t=1}^n (e_{1t} + e_{2t})(e_{1t} - e_{2t})}{\left[ \sum_{t=1}^n (e_{1t} + e_{2t})^2 \sum_{t=1}^n (e_{1t} - e_{2t})^2 \right]^{1/2}} \quad (52)$$

<sup>6</sup>Regressions with 25 neighbours or more were able to push the  $\bar{R}^2$  well above the forward rate.

a combined forecast which is a weighted average of  $z_1$  and  $z_2$ . The forecast error variance is:

$$\sigma_c^2 = \omega^2 \sigma_1^2 + (1 - \omega)^2 \sigma_2^2 + 2\omega(1 - \omega)\rho\sigma_1\sigma_2 \quad (55)$$

I can minimize this expression by setting

$$\omega = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (56)$$

In implementing equation (55), I use sample analogs, using the first year of forecasts. I then use the combined forecast for the remainder of the sample. In a number of combinations not reported below, the weights

$$\omega = \frac{\sum_{t=1}^n e_{1t}^2}{\sum_{t=1}^n (e_{1t}^2 + e_{2t}^2)} \quad (57)$$

which ignore the correlation between forecasts, worked best. A combined forecast of the non-linear AR(1) and non-linear spread models has the lowest MSE in Table III.

To ascertain the statistical significance of these improvements, the statistic  $K$  in equation (53) was computed for all the forecasts relative to the combined forecast. The best forecast is statistically better than the random walk, forward rate or simple AR(1) model. The only parametric model that compares with the best forecast is the simple spread equation, (32). Accounting for non-linearities, though, reduces the MSE by 3%, a large amount for weekly returns. The pooling also eliminates nearly all the bias.

#### CONCLUSION

The empirical tests of traditional term structure models have generally been discouraging. This paper argues that this is due to the linear approximations and the failure to account for non-normality in the error structure. This paper has developed a framework that allows for a non-linear specification of the spot rate and a general class of disturbances. By using a non-parametric procedure, nearest-neighbours regression, I was able to test generalizations of the two standard theories.

After looking both in and out of sample, I chose a forecast competition as the metric. I was able to statistically validate an improvement in forecast performance over the traditional models. In-sample fit proved to be a highly unreliable guide as to how a model would perform out of sample. This makes a strong case for forecast comparisons as model selection criteria.

My results indicate that there is some fundamental information in the spread for the path of future spot rates. None of the autoregressive models, including the non-parametric estimates, predicted as well as the spread. I believe that this is because the purely autoregressive models are misspecified. The forecasts only remind us that sophisticated estimation and forecasting procedures cannot rescue a misspecified model. However, non-parametric methods greatly expand the class of models that can be considered for empirical analysis, and I hope that this paper inspires new theoretical as well as empirical work on the term structure.

In all, the results indicate that non-linearities are abundant. Even at very short horizons and at the short end of the maturity spectrum, non-linear models improve point prediction. A power function of the neighbour terms proved to be the real breakthrough in the forecast competitions.

Pooling information from competing non-linear models enabled me to further improve the forecasts.

I conclude that a great deal of information can be gained from a non-parametric approach to the term structure. In future work, I intend to look at bonds, where the non-linearities are likely to be even more pronounced.

#### ACKNOWLEDGEMENTS

I would like to thank Tony Santomero and members of the Wharton Macro Lunch Group for helpful comments.

#### REFERENCES

- Brown, S. and Dybvig, 'The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates', *Journal of Finance*, **41** (1986), 617-30.
- Cleveland, W. S., 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association*, **74** (1979), 829-36.
- Cox, J., Ingersoll, J. Jr. and Ross, S. A., 'A theory of the term structure of interest rates', *Econometrica*, **53** (1983), 385-407.
- Diebold, F. and Nason, J., 'Nonparametric exchange rate prediction?' *Journal of International Economics*, **28** (1990), 315-32.
- Dunn, K. and Singleton, K., 'Modelling the term structure of interest rates under non-separable utility and durability of goods', *Journal of Financial Economics*, **17** (1986), 27-45.
- Efron, B., 'Estimating the error rate of a prediction rule: improvement on cross-validation', *Journal of the American Statistical Association*, **78** (1984), 316-331.
- Fama, E. F., 'The information in the term structure', *Journal of Financial Economics*, **13** (1984), 509-28.
- Fama, E. F., 'Term structure forecasts of interest rates, inflation and real returns', *Journal of Monetary Economics*, **25** (1990), 59-76.
- Fama, E. F. and Bliss, R. R., 'The information in long maturity forward rates', *American Economic Review*, **77** (1987), 680-92.
- Fama, E. F. and Roll, R., 'Some properties of symmetric stable distributions', *Journal of the American Statistical Association*, **63** (1968), 817-35.
- Frank, M. and Stengos, T., 'Measuring the strangeness of gold and silver rates of return' *Review of Economic Studies*, **56** (1989), 553-67.
- Froot, K. A., 'New hope for the expectations hypothesis of the term structure of interest rates', *Journal of Finance*, **44** (1989), 283-305.
- Gibbons, M. R. and Ferson, W., 'Testing asset pricing models with changing expectations and an unobservable market portfolio', *Journal of Financial Economics*, **14** (1989), 217-36.
- Gibbons, M. R. and Ramaswamy, K., 'The term structure of interest rates: empirical evidence', *Review of Financial Studies*, **6** (1993), 619-58.
- Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, San Diego, CA: Academic Press.
- Langtieg, T., 'A Multivariate model of the term structure', *Journal of Finance*, **35** (1980), 71-97.
- Lee, B., 'A nonlinear expectations model of the term structure of interest rates with time-varying risk premia', *Journal of Money, Credit and Banking*, **21** (1989), 339-67.
- Lucas, R., 'Asset prices in an exchange economy', *Econometrica*, **46** (1978), 1429-45.
- Mankiw, N. G., 'The term structure of interest rates revisited', *Brookings Papers on Economic Activity*, **1** (1986), 61-110.
- Mankiw, N. G. and Miron, J. A., 'The changing behavior of the term structure of interest rates', *Quarterly Journal of Economics*, **101** (1986), 211-28.
- Mankiw, N. G. and Summers, L. H., 'Do long-term interest rates overreact to short-term interest rates?' *Brookings Papers on Economic Activity*, **1** (1984), 211-28.
- Mishkin, F. S., 'What does the term structure tell us about future inflation?' *Journal of Monetary Economics*, **25** (1990), 77-95.
- Mizrach, B., 'Multi-variate nearest neighbor forecasts of EMS exchange rates', *Journal of Applied Econometrics*, **7** (1992), S151-63.

Pooling information from competing non-linear models enabled me to further improve the forecasts.

I conclude that a great deal of information can be gained from a non-parametric approach to the term structure. In future work, I intend to look at bonds, where the non-linearities are likely to be even more pronounced.

#### ACKNOWLEDGEMENTS

I would like to thank Tony Santomero and members of the Wharton Macro Lunch Group for helpful comments.

#### REFERENCES

- Brown, S. and Dybvig, 'The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates', *Journal of Finance*, **41** (1986), 617-30.
- Cleveland, W. S., 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association*, **74** (1979), 829-36.
- Cox, J., Ingersoll, J. Jr. and Ross, S. A., 'A theory of the term structure of interest rates', *Econometrica*, **53** (1983), 385-407.
- Diebold, F. and Nason, J., 'Nonparametric exchange rate prediction?' *Journal of International Economics*, **28** (1990), 315-32.
- Dunn, K. and Singleton, K., 'Modelling the term structure of interest rates under non-separable utility and durability of goods', *Journal of Financial Economics*, **17** (1986), 27-45.
- Efron, B., 'Estimating the error rate of a prediction rule: improvement on cross-validation', *Journal of the American Statistical Association*, **78** (1984), 316-331.
- Fama, E. F., 'The information in the term structure', *Journal of Financial Economics*, **13** (1984), 509-28.
- Fama, E. F., 'Term structure forecasts of interest rates, inflation and real returns', *Journal of Monetary Economics*, **25** (1990), 59-76.
- Fama, E. F. and Bliss, R. R., 'The information in long maturity forward rates', *American Economic Review*, **77** (1987), 680-92.
- Fama, E. F. and Roll, R., 'Some properties of symmetric stable distributions', *Journal of the American Statistical Association*, **63** (1968), 817-35.
- Frank, M. and Stengos, T., 'Measuring the strangeness of gold and silver rates of return' *Review of Economic Studies*, **56** (1989), 553-67.
- Froot, K. A., 'New hope for the expectations hypothesis of the term structure of interest rates', *Journal of Finance*, **44** (1989), 283-305.
- Gibbons, M. R. and Ferson, W., 'Testing asset pricing models with changing expectations and an unobservable market portfolio', *Journal of Financial Economics*, **14** (1989), 217-36.
- Gibbons, M. R. and Ramaswamy, K., 'The term structure of interest rates: empirical evidence', *Review of Financial Studies*, **6** (1993), 619-58.
- Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, San Diego, CA: Academic Press.
- Langtieg, T., 'A Multivariate model of the term structure', *Journal of Finance*, **35** (1980), 71-97.
- Lee, B., 'A nonlinear expectations model of the term structure of interest rates with time-varying risk premia', *Journal of Money, Credit and Banking*, **21** (1989), 339-67.
- Lucas, R., 'Asset prices in an exchange economy', *Econometrica*, **46** (1978), 1429-45.
- Mankiw, N. G., 'The term structure of interest rates revisited', *Brookings Papers on Economic Activity*, **1** (1986), 61-110.
- Mankiw, N. G. and Miron, J. A., 'The changing behavior of the term structure of interest rates', *Quarterly Journal of Economics*, **101** (1986), 211-28.
- Mankiw, N. G. and Summers, L. H., 'Do long-term interest rates overreact to short-term interest rates?' *Brookings Papers on Economic Activity*, **1** (1984), 211-28.
- Mishkin, F. S., 'What does the term structure tell us about future inflation?' *Journal of Monetary Economics*, **25** (1990), 77-95.
- Mizrach, B., 'Multi-variate nearest neighbor forecasts of EMS exchange rates', *Journal of Applied Econometrics*, **7** (1992), S151-63.



- Mizrach, B., 'Forecast comparison in  $L_2$ ', Working Paper 95-24, Rutgers University, <http://www-rci.rutgers.edu/~mizrach/>.
- Oldfeld, G. S. and Rogalski, R. J., 'The stochastic properties of term structure movements', *Journal of Monetary Economics*, **19** (1987), 229-45.
- Shiller, R. J., 'The volatility of long-term interest rates and expectations models of the term structure', *Journal of Political Economy*, **87** (1979), 1190-1219.
- Shiller, R., Campbell, J. and Schoenholtz, K., 'Forward rates and future policy: interpreting the term structure of interest rates', *Brookings Papers on Economic Activity*, **1** (1983), 173-217.
- Stambaugh, R., 'The innovation in forward rates: implications for models of the term structure', *Journal of Financial Economics*, **21** (1988), 41-70.
- Startz, R., 'Do forecast errors of term premia really make the difference between long and short rates', *Journal of Financial Economics*, **10** (1982), 323-9.
- Stone, C. J., 'Nonparametric regression and its applications', (with discussion), *Annals of Statistics*, **5** (1977), 595-645.
- White, H., 'Consequences and detection of misspecified nonlinear regression models', *Journal of the American Statistical Association*, **76** (1980), 419-33.

*Author's biography:*

**Bruce Mizrach** is an assistant professor at Rutgers University, having previously held appointments at Boston College, The Wharton School, and the Federal Reserve Bank of New York. He has published widely in the field of non-linear time series analysis and is serving as editor of the new MIT Press journal *Studies in Nonlinear Dynamics and Econometrics*.

*Author's address:*

**Bruce Mizrach**, Department of Economics, Rutgers University, 424 New Jersey Hall, New Brunswick, NJ 08903, USA.  
e-mail: [mizrach@rci.rutgers.edu](mailto:mizrach@rci.rutgers.edu).