NEW YORK UNIVERSITY
STERN SCHOOL OF BUSINESS
FINANCE DEPARTMENT

Working Paper Series, 1994

The Conditional Distribution of Excess Returns: An Empirical Analysis

Silverio Foresi and Franco Peracchi

FD-94-33

Forthcoming: Journal of American Statistical Association

The Conditional Distribution of Excess Returns: An Empirical Analysis

Silverio Foresi

New York University

and

Franco Peracchi

University of Udine

October 1992

Revised: November 1994

We thank workshop participants at New York University, two referees, the Editor and an Associate Editor for helpful comments. Financial support from the C.V. Starr Center for Applied Economics at New York University is gratefully acknowledged.

Abstract

In this paper we describe the cumulative distribution function of excess returns conditional on a broad set of predictors that summarize the state of the economy. We do so by estimating a sequence of conditional logit models over a grid of values of the response variable. Our method uncovers higher-order multidimensional structure that cannot be found by modeling only the first two moments of the distribution.

We compare two approaches to modeling: one based on a conventional linear logit model, the other on an additive logit. The second approach avoids the "curse of dimensionality" problem of fully nonparametric methods while retaining both interpretability and the ability to let the data determine the shape of the relationship between the response variable and the predictors.

We find that the additive logit fits better and reveals aspects of the data that remain undetected by the linear logit. The additive model retains its superiority even in out-of-sample prediction and portfolio selection performance, suggesting that this model captures genuine features of the data which seem to be important to guides investors' optimal portfolio choices.

Keywords: Asset pricing, generalized additive models, nonparametric methods.

1 Introduction

This paper is an empirical investigation of the distribution of excess returns on a wide stock market portfolio. We define the rate of return on this portfolio as the ratio of its liquidation value at the end of a month (price plus dividends) to the purchasing value of the asset at the end of the previous month. Figure 1 shows the monthly rates of return on the equally-weighted portfolio of New York Stock Exchange common stocks and on the one-month U.S. Treasury bill and their difference – the excess returns – from April 1960 to July 1992.

Empirical analysis of excess returns is important to guide investors' choices in the face of uncertainty about future returns. For an investor with a one-month horizon, the uncertainty about future excess returns comes entirely from the uncertainty about the rate of return on stocks since the rate of return on the one-month Treasury bill is known in advance. This is because the Treasury bill is purchased at a discount from the stated value which the government pays in one month to the holder of the bill. Figure 1 shows the difficulty of identifying patterns in excess returns: there is no evidence of trends and hardly any of autocorrelation. In fact, traditional views of market efficiency held that past excess returns should not help explain their future values.

Indeed, there is now evidence that both the mean and the variance of excess returns change through time in a way that can be predicted using information which goes beyond past excess returns [see Fama (1991) and Bollerslev, Chou, and Kroner (1992) for a review of the evidence on conditional first and second moments respectively]. This evidence is understood to be consistent with efficient markets when productivity and risk evolve in a predictable manner.

Two assumptions have played a significant role in shaping empirical analysis to date. One is that conditional first and second moments of asset returns are sufficient to inform investors' choices.

The other is that these moments have a known parametric form. Besides their appeal as measures of location and dispersion, mean and variance are central to the intuition of modern portfolio theory and the Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965). The higher the risk, as measured by variance, the higher the average returns investors demand to hold stocks rather than riskless Treasury bills. Unfortunately, mean-variance analysis is adequate to guide investors' choices only under special assumptions, such as multivariate normality of asset returns or quadratic utility function of investors. In general, a precise definition of risk and an unambiguous ranking of portfolio strategies requires the entire distribution of future returns [Rothschild and Stiglitz (1971)]. Our study suggests that modelling the entire distribution may improve predictions for portfolio choice.

Several recent studies have recognized the importance of avoiding rigid assumptions about the data. Examples include the autoregressive conditional heteroskedasticity (ARCH) models reviewed in Bollerslev, Chou, and Kroner (1992), and the nonparametric approaches of Pagan and Hong (1991). The main difference between these studies and ours is their emphasis on moments, as opposed to the entire distribution.

In this paper, we develop and implement two distinct ideas. First, instead of limiting ourselves to the first two moments, we describe in detail the cumulative distribution function of excess returns conditional on a broad set of predictors, including dividend-price ratios, interest rates, and rates of growth of employment and money. Second, we propose a method of estimating the cumulative distribution function that avoids the need of strong parametric assumptions.

A fully nonparametric estimation of the conditional distribution is out of the question with more than two or three predictors, because of the well known "curse of dimensionality" problem. We attempt to capture the complex relation between future returns and a broad array of predictors with a representation based on methods developed for generalized additive models. This representation avoids the curse of dimensionality problem that makes it difficult to extend classical scatterplot smoothers, such as kernel regression, to high dimensions. The goal is similar to that of the density estimation approach of Gallant and Tauchen (1992), who use a quadratic polynomial Hermite expansion of the conditional density. One advantage of our approach is interpretability, as the use of an additive specification makes it easy to think of the effects of different predictors on the excess returns in terms of "derivatives."

2 The statistical model

We interpret the data on excess returns as the realization of a time-series $\{Y_t\}_{t=0}^{\infty}$, which is itself a component of a strictly stationary k-dimensional time-series $\{X_t\}_{t=0}^{\infty}$. Because of the strict stationarity assumption, the cumulative distribution function of Y_{t+1} given X_t is well defined and is time invariant, that is, $F(y|x) = \Pr(Y_{t+1} \leq y | X_t = x)$ does not depend on the time index t.

Under the strict stationarity assumption, knowledge of $F(y|x)$ represents the complete solution to the problem of predicting Y_{t+1} given X_t, for it contains all the information relevant for prediction, irrespective of the particular loss function that the analyst may be using. If X_t is suitably defined, the conditional distribution function $F(y|x)$ may also be related to the transition probability function implied by the class of Markov models of asset prices studied, among others, by Brock (1980) and Duffie and Singleton (1993).

In this Section we outline a flexible method for estimating $F(y|x)$.

2.1 Approaches to modeling

Suppose that we want to estimate the conditional probability of future excess returns falling below some threshold y. For example, we may be interested in the conditional probability of negative excess returns. The natural approach is to introduce a binary random variable Z_{t+1}, taking value one if excess returns are below y and value zero otherwise. Estimating $F(y|x)$, viewed as a smooth function of x for fixed y, is then equivalent to fitting a model for the conditional mean of Z_{t+1} given X_t. If we reparameterize the problem in terms of the log odds-ratio

$$\eta(x) = \ln \frac{F(y|x)}{1 - F(y|x)},$$

the estimation problem reduces to fitting a logistic regression model to Z_{t+1}.

We propose to extend this approach and estimate not just one but J distinct functions $F_1(x)$, ..., $F_J(x)$, where $F_j(x) = F(y_j|x)$, and $-\infty < y_1 < \ldots < y_J < \infty$ are distinct points in the support of Y_{t+1}. This entails fitting a separate logistic regression to each binary random variable $Z_{j,t+1} = I_{(-\infty,y_j]}(Y_{t+1})$, $j = 1, \ldots, J$, where $I_A(\cdot)$ denotes the indicator function of the event A. After selecting a model for the log odds-ratios $\eta_j(x) = \ln[F_j(x)/(1 - F_j(x))]$, estimation may be carried by maximum likelihood, with the conditional log-likelihood function for a single observation (z, x) on $(Z_{j,t+1}, X_t)$ taking the simple form

$$l(\eta_j(x)) = z\, \eta_j(x) - \ln[1 + \exp \eta_j(x)].$$

There may be as many thresholds y_j as one wishes. By suitably choosing their number and position, one may hope to get a reasonably accurate description of $F(y|x)$.

If the conditional distribution of Y_{t+1} is continuous with support on the whole real line, then the sequence of functions $\{F_j(x)\}_{j=1}^J$ must satisfy

$$0 < F_j(x) < 1, \qquad j = 1, \ldots, J, \tag{1}$$

$$0 < F_1(x) < F_2(x) < \cdots < F_J(x) < 1, \tag{2}$$

for all x in the support of X. Modeling the log odds-ratio ensures among other things that condition (1) is automatically satisfied.

The monotonicity condition (2) is harder to impose. Since $\eta_j(x)$ is strictly increasing in $F_j(x)$, (2) is equivalent to the condition that

$$-\infty < \eta_1(x) < \eta_2(x) < \ldots < \eta_J(x) < \infty$$

for all x in the support of X. If we write $\eta_j(x) = \gamma_j + \mu_j(x)$, then sufficient conditions for monotonicity are that $\gamma_j > \gamma_{j-1}$ and $\mu_j(x) \geq \mu_{j-1}(x)$. One case where these conditions are satisfied is the ordered logit model, where $\gamma_j = y_j$ and $\mu_j(x) = \mu(x)$ for all j. This model is too restrictive, however, for it implies that changes in the predictors affect the distribution of future excess returns only through a location shift.

An alternative would be to model instead $F_1(x)$ and the probability increments

$$\pi_j(x) = F_j(x) - F_{j-1}(x) = P(y_{j-1} < Y \leq y_j | X = x), \qquad j = 2, \ldots, J.$$

If one can guarantee that $0 < \pi_j(x) < 1$, for example by modeling the log odds-ratio of $\pi_j(x)$, then monotonicity is clearly satisfied. We run, however, into difficulties of a different type, for the model does not automatically imply that $F_J(x) = F_1(x) + \sum_{h=1}^J \pi_h(x) < 1$.

Finally, one could model $F_1(x)$ and the conditional probabilities (or discrete hazards)

$$\lambda_j(x) = P(Y \leq y_j | Y > y_{j-1}, X = x) = \frac{S_{j-1}(x) - S_j(x)}{S_{j-1}(x)}, \qquad j = 1, 2, \ldots, J,$$

where $S_j(x) = 1 - F_j(x) = P(Y > y_j|X = x)$ is the survivor function evaluated at y_j. Using the recursion $S_j(x) = [1 - \lambda_j(x)]S_{j-1}, j = 2, \ldots, J$, one can write $F_j(x)$ as

$$F_j(x) = 1 - S_1(x) \prod_{h=2}^{j} [1 - \lambda_h(x)]. \tag{3}$$

If $\lambda_j(x)$ is modelled to guarantee that $0 < \lambda_j(x) < 1$, then both monotonicity and the constraint that $0 < F_j(x) < 1$ for all j are automatically satisfied.

The resulting model is difficult to interpret, however, because the link between its parameters and the conditional distribution function is quite involved. To see this, consider the impact of the predictors on $F_j(x)$. If $\theta_1(x)$ is the log odds-ratio of $F_1(x)$ and $\theta_j(x)$ is the log odds-ratio of $\lambda_j(x)$, then $F_j(x) = 1 - \prod_{h=1}^{j}(1 + \exp\theta_h(x))^{-1}, \quad j = 2, \ldots, J$. Hence, denoting by a $'$ differentiation with respect to x, we get

$$F_j'(x) = S_j(x) \sum_{h=1}^{j} \frac{\exp\theta_h(x)}{1 + \exp\theta_h(x)}\, \theta_h'(x) = S_j(x)\, [F_1(x)\, \theta_1'(x) + \sum_{h=2}^{j} \lambda_h(x)\, \theta_h'(x)],$$

which is a complicated weighted average of the gradients of the first j log odds-ratios. This may be contrasted with our choice of modeling $\eta_j(x)$, the log odds-ratio of $F_j(x)$. In this case

$$F_j'(x) = \eta_j'(x)\, S_j(x)\, F_j(x).$$

Thus, the gradient of $\eta_j(x)$ is immediately interpretable in terms of the gradient of $F_j(x)$.

Because we see no simple way of ensuring monotonicity while preserving interpretability and ease of estimation, (2) will not be imposed here.

2.2 Specification of the log odds-ratios

The conventional parametric approach proceeds by restricting the log odds-ratio $\eta_j(x)$ to a family of functions indexed by a finite-dimensional real parameter. If

$$\eta_j(x) = \gamma_j + \sum_{h=1}^{k} x_h \delta_{hj} = \gamma_j + \delta_j^\top x, \tag{4}$$

then we have the classical logit model. Beside ease of estimation, one important feature of the linear specification (4) is interpretability, for δ_{hj} may be interpreted as the constant partial derivative of the log odds-ratio of $F_j(x)$ with respect to the h-th variable in X_t.

While the linear specification is probably restrictive, the alternative of a fully nonparametric specification is out of the question for small to moderate sample sizes, if one wants to allow for a broad enough set of predictors. One way of attacking the curse of dimensionality problem would be to consider projection-pursuit logistic regression. We choose for simplicity to work with a specification that is intermediate between projection pursuit and the classical linear logit, namely

$$\eta_j(x) = \gamma_j + \sum_{h=1}^{k} g_{hj}(x_h), \tag{5}$$

where $\{g_{hj}\}$ are arbitrary univariate smooth functions, one for each component of the vector X_t, to be estimated nonparametrically. The additivity imposed by (5) makes it possible to interpret the gradient of g_{hj}, now a function of x_h, as the partial derivative of the log odds-ratio of $F_j(x)$ with respect to the h-th component of X_t, thereby retaining interpretability, which is one of the attractive features of the linear specification (4).

We shall also compare (5) with the semi-additive model

$$\eta_j(x) = \gamma_j + \delta_j^\top x^{(1)} + \sum_{h=1}^{k-k_1} g_{hj}(x_h^{(2)}), \tag{6}$$

where the k_1 variables in $x^{(1)}$ enter linearly and the remaining $k - k_1$ variables in $x^{(2)}$ enter nonparametrically. A specification similar to (6) was used by Engle *et al.* (1986) in the context of modeling a conditional expectation function.

2.3 Model estimation

We estimate models (5) and (6) using the local scoring algorithm proposed by Hastie and Tibshirani (1990) for generalized additive models. Their method modifies the scoring algorithm for the linear logit model by constructing at each iteration nonparametric estimates of the univariate functions $\{g_{hj}\}$, obtained by smoothing the transformed residuals from the previous iteration. Smoothing may be based on nearest neighborhood methods, such as Cleveland's (1979) loess, or on cubic smoothing splines. The latter have certain advantages, including convergence of the local scoring algorithm. The degree of smoothing may be chosen subjectively or an automatic selection criterion, such as cross-validation, may be employed. For a discussion of the properties of these estimators in the case of independently and identically distributed observations see Stone (1986).

The linear, semi-additive and additive specifications form a nested sequence. In the empirical section below we compare them using the logarithm of the likelihood ratio. Although the distribution theory for this statistic is not yet developed, simulation results in Hastie and Tibshirani (1990) show that the χ^2 distribution provides a useful approximation. Following their suggestion, we compute the number of degrees of freedom of the asymptotic χ^2 approximation as the value of a quadratic approximation to the expectation of the likelihood ratio statistic under the truth of the restricted model.

3 Data

Our raw data are monthly time-series from March 1959 to July 1992. Details on their sources are provided in the Appendix. The time-series has been split in two. The data prior to April 1989 are used for model selection and estimation. At the model selection stage we choose the predictors, their transformations, the type and degree of smoothing for the additive model, and the best semi-additive model. The data from April 1989 to July 1992 are left aside for the out-of-sample prediction exercise of Section 4.4.

3.1 Dependent variable

The dependent variable is the excess return, defined as the difference between the return on the equally-weighted portfolio of New York Stock Exchange common stocks and the one-month U.S. Treasury bill. The time-series plot of these data is shown in Figure 1. The dotted vertical bar indicates April 1989. It separates the data used for model selection and estimation from those left aside for out-of-sample prediction. The choice of excess returns rather than returns on stocks follows previous theoretical and empirical analysis. Because the return on a one-month Treasury bill is known at the beginning of the month, one needs to explain only the excess returns. Another reason for using excess returns is inflation. In periods of inflation, stocks and Treasury bills offer higher returns to compensate for the loss of purchasing power. Working with excess returns eliminates, to a first approximation, the scale effect of inflation. This simplifies the statistical model because excess returns do not display trends in the level.

3.2 Choice of predictors and data transformations

Previous empirical evidence guides our choice of predictors. Our general frame of reference is an equilibrium model of the economy in which future excess returns are determined by the current state of the financial, real, and monetary sectors of the economy.

Current excess returns, the short term interest rate, and the dividend yield (dividend-price ratio) summarize the state of the financial sector. Theoretical models suggest that the short term interest rate tracks movements of the investment opportunity set over time [Merton (1973)] and its ability to predict asset returns is well documented [Fama and Schwert (1977)]. The use of current excess returns as a predictor captures possible persistence in excess returns [Conrad and Kaul (1988)]. The dividend yield contains information on expected stock returns, as documented by several authors [Rozeff (1984), Campbell and Shiller (1988)]. In theory, however, the dividend yield encompasses both expected stock returns and expected dividend growth. To control for dividend growth, we consider measures of growth of the economy based on industrial production and employment.

Industrial production and employment also contain information on the state of the real sector, as they capture the level of business activity and the labor market conditions. Industrial production [Chen, Roll, and Ross (1986)] and unemployment [Gertler and Grinols (1982)] have been considered as proxies for macroeconomic risk "factors" in the literature on the cross sectional variation in asset returns. In as much as current values of these factors can help forecast future ones, they contain information on future returns. We consider also two other financial variables that contain information on the state of the economy: a default spread – the monthly average yield to maturity of corporate bonds rated Baa by Moody's Investor Services less the Aaa corporate bond yield – and a term structure variable – the one-month rate of return of a 3-month Treasury bill less the one-month return on a one-month bill. It is usually argued that the default spread contains information useful to forecast future business conditions, since the return on private debt instruments reflects the near-term risk in the economy. Term structure variables may capture the relative availability of credit with respect to the demand. The predictive power of default and term structure variables for excess returns is also well documented in the empirical literature [see for example Keim and Stambaugh (1986) for the default spread, and Campbell (1987) for term structure variables].

Finally, we consider a broad monetary aggregate, M2, as an indicator of the state of the monetary sector. Monetary variables have received considerable attention in studies of the relation between stock returns and inflation such as Fama (1981), Geske and Roll (1983), and Kaul (1987). These studies have mainly emphasized the contemporaneous reaction of money and stock prices to current and anticipated economic activity. Recent evidence [Chan, Foresi, and Lang (1992)] suggests however that monetary aggregates such as M2 and M3 have incremental explanatory power for future excess returns over and above that of other business-cycle predictors such as the term spread and the default spread. This is not surprising in light of the well recognized role of money as a leading indicator of future business conditions. The default spread and the term structure spread, two predictors we are using to forecast excess returns, have a similar leading indicator role.

We work with the continuously compounded rates of growth of employment, industrial production, and M2. We consider both month-to-month and year-to-year growth rates. The latter eliminate seasonal components up to the monthly frequency. Year-to-year growth rates of industrial production, employment, and M2 display a smoother profile than the month-to-month ones and show no recognizable seasonal pattern. The use of year-to-year growth rates has a natural interpretation, as stock market returns are related not only to short-run changes, but also to changes in industrial activity over longer horizons.

Industrial production and employment contain similar information for predicting excess returns. Thus we drop industrial production and use only employment as a proxy for the business cycle. M2 and employment for a given month are made available only the following month. Thus we use M2 and employment growth of month $t-1$ to predict excess returns in month $t+1$. We do not transform the other predictors other than by rescaling the rates of returns to express them all in percentage form.

To summarize, the predictors that we consider are: the dividend yield, the excess returns, the term-structure spread, the default spread, the one-month Treasury bill yield, the twelve-month differences of the log of M2, and the twelve-month differences of log employment. Ferson and Harvey (1991) use the first five to explain excess returns. To their list, we have added two standard macroeconomic variables – the year-to-year rates of growth of employment and M2.

After transforming the predictors, our estimation sample covers the period from April 1960 through March 1989, for a total of 348 data points. Figure 2 presents some features of the data. In plotting the data we symmetrically trim 2.5 percent of the data at the top and the bottom to eliminate the visual effect of a few extreme outliers. The top-left panel shows the density of excess returns (solid line) estimated using a kernel smoother with a triangular kernel and a fixed bandwidth chosen subjectively. For comparison, we overlay a Gaussian density (broken line) with mean and variance equal to the sample mean and variance of excess returns. The two densities differ mainly in kurtosis: an overall test based on the empirical third and fourth moments rejects the assumption of normality of excess returns.

The remaining panels are bivariate scatterplots of excess returns against each of the predictors.

To help visualize the central tendency of the data, we add a locally linear scatterplot smooth estimated by Cleveland's (1979) loess method using neighborhoods covering 40 percent of the data. On the right side of each scatterplot, we reproduce the univariate density of excess returns and, on the top, we present the univariate density of the predictor. These scatterplots show some structure in the data and suggest that the location of the response variable changes with the predictors. However, it remains difficult to see how other aspects of the distribution vary with the predictors.

4 Findings

We now present the results of estimating the conditional distribution function of excess returns by a sequence of logit models, where the binary response variable takes value one if future excess returns do not exceed a threshold y_j and value zero otherwise. The predictors are the ones described in Section 3.2. As thresholds, we select the percentiles of the empirical distribution of excess returns from the 10-th to the 90-th with increments of five percent. Because the financial literature on market timing is often interested in the probability of positive excess returns, we also consider zero as an additional threshold. This gives a total of eighteen evaluation points for $F(y|x)$. We use S-PLUS as a computation environment.

Section 4.1 presents the results for the linear logit model. Section 4.2 compares these results with the ones obtained by fitting the more flexible additive logit model. Section 4.3 presents the results of our search for a semi-additive specification intermediate between the linear and the additive logit. Section 4.4 compares the three models in terms of out-of-sample prediction.

4.1 Linear logit

We begin with log odds-ratios that are specified as linear in the predictors, $\lambda_j(x) = \gamma_j + \delta_j^\top x$. In order to keep the dimensionality of the model to a manageable dimension we exclude interactions among the predictors. Notice, however, that absence of interactions in the logit space does not imply absence of interactions in the probability space.

Figure 3 plots the estimated intercepts $\hat{\gamma}_j$ and slopes $\hat{\delta}_j$ for the 18 evaluation points. The vertical bars at the bottom of each plot correspond to the evaluation points. The estimates are connected by a solid line to help visualizing the differential impact of a predictor at different values of y in $F(y|x)$. The dotted lines are pointwise two-standard error bands. In reading these pictures one should keep in mind that the smoothness of the lines is partly an artifact of the high correlation between estimated coefficients for adjacent threshold values.

The coefficients on the dividend yield are negative and significant for all threshold values. This implies that an increase in this variable, keeping all other predictors constant, shifts the conditional distribution function of excess returns to the right. This result agrees with the well-known finding of a positive relationship between dividend yields and mean excess returns. For the one-month Treasury bill yield the effect is just the opposite. Since the coefficients are now always positive and significant, an increase in the Treasury bill yield shifts the conditional distribution to the left. Observe that in both cases the coefficients do not change much across threshold values. If the distribution of excess returns was logistic, constancy of the logit coefficients across thresholds would correspond to a pure location shift in the distribution.

For current excess returns and M2 growth we observe a different pattern. Coefficients change sign from negative to positive, and lose statistical significance, as we move from negative to positive threshold values. In this case, an increase in either variable, keeping all other predictors constant, lowers the probability of negative excess returns and may also lower the probability of very high excess returns. This corresponds to a decrease in the spread of the distribution – a reduced volatility of excess returns, which is consistent with the findings of Black (1976) and others. However, because the magnitude of the effect is different at different thresholds, not only the spread but also other aspects of the distribution may be affected.

Figure 4 gives an alternative way of summarizing the information provided by the linear logit fit. It shows how the estimated conditional distribution changes when a single predictor changes and all the others remain constant. Each panel in the figure is constructed by evaluating the fitted probabilities $\hat{F}_j(x)$, $j = 1, \dots, J$, over a grid of 100 equally spaced points between the 2.5-th and the 97.5-th percentile of one of the predictors, with all the others kept constant at their average value. We then plot the contours of iso-probability using an interpolating algorithm.

These iso-probability contours have a nice interpretation as conditional quantiles of excess returns. For example, projecting the curve indexed by .5 onto the y-axis traces out the behavior of the conditional median as one of the predictors changes and all the others remain constant. A positive (negative) slope of an iso-probability contour indicates a positive (negative) impact of the predictor on the corresponding conditional quantile. Horizontal contours correspond to a predictor having no impact. One can detect violation of the monotonicity constraint (2) by looking at the iso-probability maps, as lines parallel to the y-axis should never cross twice the same contour. Although we see some violations in the case of the dividend yield, current excess returns, and the one-month Treasury bill yield, these occur in regions of the predictors' space where the sample information is too low to afford reliable conclusions.

The figure gives a convenient way of representing the effect of a predictor on various aspects of the distribution of excess returns. Iso-probability contours that are parallel to each other indicate that the variable affects only the location of the distribution, not its shape. This appears to be the case with the dividend yield and the term spread. The figure suggests that while other predictors may have limited information about location, they carry information about spread, symmetry, and tail weight. For example, higher current excess returns or M2 growth decrease dispersion of future returns as measured by the interquartile range (IQR). However, they also affect the degree of symmetry of the distribution, which tends to become more skewed to the right. Another example is employment growth: an increase in employment growth lowers the median and the IQR of future excess returns but also increases the weight in the tail of the distribution, as measured by the ratio of the distance between the 10-th and the 90-th percentile and the IQR, and changes the distribution from left- to right-skewed.

We conclude that although the significance of a predictor in a model of mean excess returns may be limited, this does not imply that the predictor contains no information, because its role may only be appreciated by looking at other aspects of the distribution.

4.2 Additive logit

We now consider what may be gained by allowing for a more flexible relation between predictors and excess returns.

Fitting the additive logit model (5) requires making a number of choices. After some experiments with both cubic splines and loess, we decided to use cubic splines to smooth the data. The overall shape of the estimated relations is similar for the two smoothers, but the visual appearance of splines is more regular. Loess is also computationally more cumbersome: the local scoring algorithm requires more iterations and in a few cases failed to converge. The degree of smoothing, which corresponds to roughly five degrees of freedom for each nonparametric term, is the same for all predictors and was chosen subjectively, after an informal comparison of Akaike criteria.

Figure 5 has been constructed in exactly the same way as Figure 4 and shows the iso-probability contours corresponding to the additive fit. Although some broad features are common to the two figures, the additive contour maps are much less regular: they are far from linear and often change orientation. Further, condition (2) is violated more frequently than for the linear case, although in most cases this occurs in regions that correspond to infrequent values of the predictors. In the absence of measures of statistical accuracy, it is hard to draw firm conclusions from a comparison of these two figures. We note, however, that the conditional medians from the additive logit are in much closer agreement with the loess smooths in Figure 1. This is somewhat reassuring, especially if one considers that our models are not explicitly designed to track location.

In Figures 6a and 6b we compare in more detail the linear and the additive logit fit for the probability of negative excess returns, $y = 0$. Each figure consists of seven panels, one for each predictor. The plots in Figure 6a are on the scale of the log odds-ratio and present the contribution of each predictor to the fit. The vertical bars at the basis of each plot show the frequency distribution of that predictor. The thicker solid curves represents the nonparametric components $\hat{g}_{hj}(x_h)$ from the additive logit, while the lighter solid lines represent the corresponding linear components $x_h\hat{\delta}_{hj}$ from the linear logit. To facilitate the comparison, all components are normalized to be equal to zero at the average value of the predictor. The dotted lines are two standard-error bands associated with the linear component. If the linear logit model was correct, these standard-error bands could be interpreted as approximate 95-percent confidence intervals for the variation in the slope of the linear component.

Figure 6a helps visualizing the differences between the fit the nonparametric and linear components. These differences are not confined to regions of low density of the data. The inversions in the slope of the additive components are the most apparent departures from the linear logit. Increases in M2 growth, for example, may be good or bad news for the stock market. If M2 growth is below-average, an increase in M2 growth lowers the probability of negative excess returns. If M2 growth is above-average, an increase in M2 growth increases instead the probability of negative excess returns.

Even when the nonparametric components look approximately linear, we still observe interesting departures from the linear logit. Consider for example the dividend yield. Although the nonparametric component is monotone over most of its range, its slope is steeper than the linear component, and the difference becomes more pronounced for above-average values of the predictor. The lower slope of the linear logit is clearly a compromise to get a good fit at extreme values of the predictor.

The plots in Figure 6b are on the probability scale and show how estimated probabilities from the two models vary as functions of a single predictor, keeping all the others constant at their average value. Estimated probabilities from the additive logit model are represented by a thicker solid curve, while those from the linear model are represented by a lighter solid curve. The broken curves are estimates from a nonparametric logit model fitted to the response variable $Z_{j,t+1}$ using just a single predictor. They give useful benchmarks for the multivariate models. The vertical bars at zero and one indicate the occurrence of zeros and ones in the response variable $Z_{j,t+1}$. They help detect local concentrations of ones and zeros which, by inflating the nonparametric slopes, may produce extreme fitted probabilities.

A comparison of the univariate smooth with the additive fit again suggests that the nonlinearities captured by the additive model are genuine aspects of the data and not the result of overfitting or spillover effects from other terms of the model. Interestingly, the additive and the univariate nonparametric fit are in reasonable agreement with each other except for two predictors, the dividend yield and the one-month Treasury bill yield. For either predictor the effect on the probability of negative excess returns is enhanced by the presence of the other predictor.

An overall comparison of the linear and the additive model is presented in Table 1. This table reports, for each threshold value of excess returns and for each model, the residual deviance (minus twice the maximized log likelihood) and the effective degrees of freedom. It also reports the likelihood ratio statistic and the p-value of a likelihood ratio test based on the χ^2 distribution with number of degrees of freedom equal to the difference in degrees of freedom of the two models. Although one may question the use of conventional significance levels, we notice that the likelihood ratio test rejects the linear logit at the 10 percent level in 15 out of 18 cases, and at the 5 percent level in 10 out of 18 cases. To interpret this pattern one should keep in mind that all the results in Table 1 are correlated, some highly; however, the small p-values in the middle of the table provide some evidence that the additive model is fitting better. Interestingly, we observe very few rejections of linearity in the upper part of the distribution of excess returns.

4.3 Semi-additive logit

The additive logit fits the data better than the linear one but uses many more degrees of freedom. This suggests that by selecting only a few nonparametric terms one may attain a better balance between goodness-of-fit and model parsimony. Hopefully, this would give a model that, while still flexible, is better able to predict future excess returns.

One measure of predictive accuracy is the Akaike Information Criterion (AIC)

$$\text{AIC} = -2\ln\hat{L} + 2\,df,$$

where $\hat{L}$ denotes the maximized value of the likelihood and df denotes the number of degrees of freedom. We use this criterion to select, for each threshold value of excess returns, the best semi-additive model among all possible combinations of linear and nonparametric terms. For simplicity, we keep the amount of smoothing the same for all nonparametric components.

Figure 7 plots, for six threshold values of excess returns, the residual deviance and the degrees of freedom of semi-additive models combining linear and nonparametric terms: each "+" mark represents one of the resulting 128 ($= 2^7$) semi-additive models, including the fully additive and the linear logit ones. The six threshold values are roughly evenly spaced in the range from the 10-th and 90-th percentile of the unconditional distribution. The top broken line represents all combinations of deviance and degrees of freedom that give the same AIC as the additive logit, while the bottom line corresponds to the linear logit. Any line parallel to the two broken lines describes a collection of different models with the same value of AIC. The lower the line, the better the AIC associated with the model. The linear logit always dominates the additive on the basis of AIC. There are, however, semi-additive models that do better than the linear logit. The best of these models are indicated by a black square, while black diamonds denote the ones that do worst in terms of AIC.

Tables 2 shows which nonparametric components are included in the best and worst models according to AIC. The best models may be interpreted as the directions of departure from the linear logit that result in the highest decrease in AIC. As an example, the best AIC model of the probability of negative excess returns is the one where the one-month Treasury-bill yield and the growth rates of M2 and employment are allowed to enter nonparametrically. This model results in a decrease of the deviance from 447.1 of the linear logit to 422.8, which more than offsets the increase in degrees of freedom from 8 to roughly 19.

Few of the "+" marks are to the lower left of the AIC line corresponding to the linear logit model, indicating that the linear logit is nearly best for all the quantiles of y displayed. If we replace the AIC with a more general criterion of the form $C(\alpha) = -2\ln \hat{L} + \alpha\, df$, we can compute the critical value of α above which the linear logit model would be chosen. In the case of the probability of negative excess returns, this critical value is equal to 2.19. On the other hand, the additive model would be chosen only if α falls below .98. Thus, a small departure from the Akaike criterion ($\alpha = 2$) towards a more conservative one such as the Schwarz criterion ($\alpha = \ln 348 = 5.85$), which corresponds to a steepening of the broken lines in Figure 7, would be enough for the linear model to be chosen. Choosing the additive model requires instead a large reduction in the trade-off between goodness-of-fit and parsimony.

4.4 Out-of-sample prediction

A reader may wonder to what extent our results depend on the particular data set used for model specification and estimation. After all, smoothing operations are very much data-specific and, at least to some extent, the choice of the predictors and their transformation is also data-based.

We address this issue by considering how the three models predict the data after March 1989. Recall that these data have been left aside at the model specification and estimation stages. Our out of sample prediction exercise covers the period from May 1989 through July 1992, a total of thirty-nine data points. We look at four summaries of predictive accuracy for the different models of the probability of negative excess returns. The first three are based on the fitted model for the probability of negative excess returns. The last statistic uses also the probabilities associated to the other 17 evaluation points of $F(y|x)$.

To simplify notation, let the response variable Z_t be a binary indicator taking value zero if future excess returns are positive and value one otherwise, with F_t as a short-hand for $P(Y_t \leq 0|X_{t-1})$. We denote the predicted value of F_t by $\hat{F}_t$. The top-left panel of Figure 8 shows the time-series plot of the estimated probabilities for the additive, linear, and best (AIC) semi-additive models of negative excess returns.

The first statistic is the sum of the squared prediction errors

$$S_1 = \sum_{t=1}^{39}(Z_t - \hat{F}_t)^2.$$

The additive logit does best with $S_1 = 9.06$, the linear logit gives $S_1 = 9.23$, and the semi-additive does worst with $S_1 = 9.38$. The second measure is the sum of weighted squared prediction errors,

$$S_2 = \sum_{t=1}^{39}\frac{(Z_t - \hat{F}_t)^2}{w_t},$$

where the weights w_t are equal to the reciprocal of the variance of Z_t, as estimated by the additive model. The ordering of the three models remain the same, with $S_2 = 43.9$ for the additive, $S_2 = 44.9$ for the linear, and $S_2 = 45.6$ for the semi-additive. The ranking does not change if the variance estimates from the linear logit are used instead as weights.

The top-right panel of Figure 8 shows the time-series plot of the squared prediction errors (the plot of the weighted squared error, not shown, is quite similar). Being a less parsimonious model, it is remarkable how well the additive does until the end of the prediction period. We expected its predictive accuracy to deteriorate over time. The additive model does display some large errors at the beginning, but there is no indication that it is performing worse than the linear logit in the later part of the prediction period.

To spice up the comparison, the third statistic is the overall return from a simple market-timing strategy. Starting with one dollar in April 1989, each month we decide to invest in Treasury bills or stocks depending on whether the predicted probability of negative excess returns is above or below .5.

The total return for the market-timing strategy based on the predictions of the linear logit (Strategy L) is $1.80, or 80 per cent, the additive (Strategy A) gives $1.85, while the semi-additive (Strategy SA) gives $1.79. The count of wrong timing predictions is 14 for L, 15 for A, and 16 for SA. To see that our market-timing strategies fare quite well, consider that an investor holding Treasury bills over the same thirty-nine month period would get a total return of $1.23, while one holding stocks would get $1.32. Investors flipping a coin to time the market (the unconditional probability of negative excess returns is .46) would average somewhere between the overall return on Treasury bills and that on stocks.

Reshuffling one's portfolio to time the market is costly because it requires paying commission fees, for example. Strategy L is more aggressive and switches portfolio ten times, A and SA are more conservative and require three and four switches respectively. Assuming a rather crude proportional adjustment cost of .5 percent per each portfolio switch, Strategy A comes clearly ahead with a net return of $1.82 against $1.72 for L and $1.76 for SA. By comparing the cumulative rates of return on the lower-left panel of Figure 8 with the estimated probabilities on the top-left panel, we see that all three strategies outperform the buy-and-hold all-stocks strategy because they switch to Treasury bills after August 1989, as the stock market was just starting on a long negative trend. Strategies

A and SA are "bearish" until October 1990, when they turn "bullish". Being conservative turns out to be a plus for A and SA: by timing the market more aggressively, L starts better but it takes a false step when it turns bullish one month too early in September 1990. This, and the lower adjustment costs, explain why A and SA end up above L in overall returns. On the other hand, the difference between A and SA is entirely due to the last month, when SA shifts to Treasury bills missing a stock market increase.

The three summaries of predictive accuracy for the models of negative excess returns indicate that the additive logit model performs better than the linear and the semi-additive ones. In particular, the use of monetary weights to assess performance strongly favors the additive over the linear. Contrary to our expectations, the semi-additive does not come out so well from the exercise.

A portfolio strategy which invests entirely in Treasury bills or entirely in stocks is only a coarse approximation to an optimal strategy. To hedge his bets, an investor may prefer investing a fraction ω_t in stocks, and $1-\omega_t$ in Treasury bills. The return on such a composite portfolio is the weighted average of the rate of return on Treasury bills, b_{t+1}, and the rate of return on stocks s_{t+1}, that is $\omega_t s_{t+1} + (1-\omega_t)b_{t+1}$ (these are gross rates of return, so that 1.03 means a 3% net rate of return). For simplicity, we assume that the investor cannot borrow ($\omega_t \leq 1$) and cannot sell stocks they do not own (no "short" selling, $\omega_t \geq 0$). The investor chooses a portfolio weight to maximize the expected utility of next period portfolio wealth W_{t+1}, where the utility function is a standard power utility function. More precisely, in each period t the investor chooses a portfolio weight ω_t^* by solving the problem

$$\begin{aligned} &\max_{\omega_t \in [0,1]} \; E_t \left(\frac{W_{t+1}^{1-a}}{1-a} \right), \\ &\text{s.t.} \;\; W_{t+1} = W_t \; [b_{t+1} + \omega_t(s_{t+1} - b_{t+1})], \end{aligned} \tag{7}$$

where the expectation is taken with respect to the estimated conditional distribution of future excess returns $s_{t+1} - b_{t+1}$ given current information. The real valued parameter a regulates the degree of risk aversion. For risk averse investors a is positive and there is consensus in the literature that it should exceed one.

We approximate the expectation in (7) using the predicted probabilities from the linear and additive logit fits associated with the 18 evaluation points of $F(y|x)$, and choose ω by a grid-search over the interval $[0, 1]$ with step size equal to .005.

The end-of-period wealth from the strategy $\{\omega_t^*\}$ is the last statistic we consider. For any value of the risk-aversion parameter a, the ranking of the end-of-period wealth across different probability models is the same as the ranking of the end-of-period utilities. This statistic is interesting for two reasons. First, it provides a comprehensive measure of predictive accuracy, as it involves not only the probability of negative excess returns, but also the probability corresponding to the other evaluation points of $F(y|x)$. Second, it allows to relate our model to a standard Gaussian model of log-stock prices. When log-stock prices are conditionally normal, the choice of the portfolio weights depends only on the mean and variance of log-stock returns

$$\omega_t^* = \frac{E_t(\ln s_{t+1}) + .5\,\mathrm{Var}_t(\ln s_{t+1}) - \ln b_{t+1}}{a\,\mathrm{Var}_t(\ln s_{t+1})}, \tag{8}$$

a result which is exact in the continuous-time optimal portfolio model of Merton (1973). The weights (8) can be computed by estimating the conditional mean and variance of log-stock returns. For the mean, we fit an additive model to the logarithm of stock returns at time $t + 1$, with the same covariate vector X_t used in the linear and additive logit models. For the variance, we fit an additive model to the logarithm of the squared residuals from the previous regression, with X_t as the covariate vector. Both regressions are fitted to the data before April 1989 using a degree of smoothing corresponding roughly to five degrees of freedom for each nonparametric term, which is the same used for the additive logit models.

The lower-right panel of Figure 8 plots, for different values of the risk-aversion parameter, the end-of-period wealth from the Gaussian model and the linear and additive logits. Because there is no simple choice of which semiparametric terms to include, since different semi-additive models do best at different quantiles, we leave the semiadditive model out of this last comparison. The additive logit model does best for investors with risk-aversion above two, while the Gaussian does best for less risk-averse investors; the linear model is the worst performer. The benefits (and costs) of using the entire distribution depend on the risk aversion of the investors. Investors with low risk aversion are interested mainly in average performances: a model of mean log-stock returns makes the best use of information to this end. As risk aversion increases, higher-order moments of the distribution become important. In this case, the Gaussianity assumption appears to be too restrictive, as the additive logit model outperforms the Gaussian one for more risk-averse investors.

5 Conclusions

In this paper we describe the cumulative distribution function of excess returns conditional on a broad set of predictors that summarize the state of the economy. We do so by estimating a sequence of conditional logit models over a grid of values of the response variable. Our method uncovers highly multidimensional structure that could not be found by modeling only the first two moments of the distribution. Our analysis suggests also that focusing on location, for example on the conditional mean or median, may lead one to overlook the role of certain predictors, such as M2 and employment growth, whose impact is mainly on higher order aspects of the distribution.

We compare two approaches to modeling: a conventional linear logit and an additive logit model. The second model avoids the "curse of dimensionality" problem of fully nonparametric methods while retaining both interpretability and the ability to let the data determine the shape of the relationship between the response variable and the predictors.

We find that the additive logit fits better and reveals aspects of the data that remain undetected by the linear logit. It uses many more degrees of freedom, however. Surprisingly, the additive model retains its superiority even in out-of-sample prediction, where it also outperforms a Gaussian model in choosing optimal portfolios for more risk-averse investors. This suggests that the within-sample performance is not due to overfitting and confirms our view that this model captures genuine features of the data which seem to be important for optimal portfolio selection.

References

Black, F. (1976), "Studies of stock prices volatility changes," *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Bollerslev, T., Chou, R., and Kroner, K. (1992), "ARCH modeling in finance: A selective review of the theory and empirical evidence, with suggestions for future research", *Journal of Econometrics*, 52, 5–61.

Brock, W. (1980), "Asset prices in a production economy", in *The Economics of Uncertainty*, ed. J.J. McCall, Chicago: University of Chicago Press, 1-46.

Chan, K.C., Foresi, S., and Lang, L. (1992), "Does money explain asset returns? Theory and empirical evidence", New York University, Dept. of Finance.

Chen, N., Roll, R., and Ross, S. A. (1986), "Economic forces and the stock market", *Journal of Business*, 59, 383–403.

Cleveland, W.S. (1979), "Robust locally weighted regression and smoothing scatterplots", *Journal of the American Statistical Association*, 74, 829–836.

Conrad, J., and Kaul, G. (1988), "Time variation in expected returns", *Journal of Business*, 61, 409–425.

Duffie, D., and Singleton, K.J. (1993), "Simulated moments estimation of Markov models of asset prices", *Econometrica*, 61, 929–952.

Engle, R.F., Granger, C.W.J., Rice, J.A., and Weiss, A. (1986), "Semiparametric estimates of the relationship between weather and electricity sales", *Journal of the American Statistical*

Association, 81, 310–320.

Fama, E.F. (1981), "Stock returns, real activity and money", *American Economic Review*, 71, 545–565.

Fama, E.F. (1991), "Efficient capital markets: II", *Journal of Finance*, 46, 5, 1575–1617.

Fama, E.F., and Schwert, G.W. (1977), "Asset returns and inflation", *Journal of Financial Economics*, 5, 115–146.

Ferson, W.E., and Harvey, C.R. (1991), "The variation of economic risk premiums", *Journal of Political Economy*, 99, 2, 385–415.

Gallant, R.A., and Tauchen, G. (1992), "A nonparametric approach to nonlinear time series analysis: Estimation and simulation", in *New Dimensions in Time Series Analysis*, eds. D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M. Taqqu, New York: Springer-Verlag, 71-92.

Gertler, M., and Grinols E.L. (1982), "Unemployment, inflation, and common stock returns", *Journal of Money, Credit, and Banking*, 14, 216–233.

Geske, R., and Roll, R. (1983), "The fiscal and monetary linkage between stock returns and inflation", *Journal of Finance*, 38, 1–33.

Hastie, T.J., and Tibshirani, R.J. (1990), *Generalized Additive Models*, New York: Chapman and Hall.

Kaul, G. (1987), "Stock returns and inflation: the role of the monetary sector", *Journal of Financial Economics*, 18, 253–276.

Lintner, J. (1965), "The Valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets", *Review of Economics and Statistics*, 47, 13–37.

Merton, R.C. (1971), "Optimum consumption and portfolio rules in a continuous time model", *Journal of Economic Theory*, 3, 373–413.

Merton, R.C. (1973), "An intertemporal capital asset pricing model", *Econometrica*, 41, 867–887.

Pagan, A., and Hong, Y. (1991), "Nonparametric estimation and the risk premium", in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, eds. W.A.Barnett, J.Powell, and G.Tauchen, Cambridge: Cambridge University Press, 51-75.

Rozeff, M. (1984), "Dividend yields and equity risk premiums", *Journal of Portfolio Management*, Fall, 68–75.

Sharpe, W.F. (1964), "Capital asset prices: a theory of market equilibrium under conditions of risk", *Journal of Finance*, 19, 425–442.

Stone, C.J. (1986), "The dimensionality reduction principle for generalized additive models", *Annals of Statistics*, 14, 590–606.

Table 1

For each threshold value of excess returns and for each model, we report the residual deviance (minus twice the maximized log likelihood) and the effective degrees of freedom (*df*), along with the likelihood ratio statistic and the *p*-value of a likelihood ratio test based on the χ^2 distribution with number of degrees of freedom equal to the difference in degrees of freedom of the two models. The number of degrees of freedom are computed as in Hastie and Tibshirani (1990).

	Linear logit		Additive logit		Test of linear *vs* additive		
Excess returns	Deviance	*df* (1)	Deviance	*df* (2)	Log likelihood ratio	(1)-(2)	*p*-value
$\leq$ -5.52	175.28	8	129.95	35.09	45.33	27.09	.015
$\leq$ -4.38	260.40	8	219.99	35.01	40.42	27.01	.047
$\leq$ -3.06	314.50	8	269.20	35.31	45.29	27.31	.017
$\leq$ -2.24	356.67	8	317.08	35.46	39.59	27.46	.063
$\leq$ -1.69	397.03	8	355.46	35.52	41.57	27.53	.042
$\leq$ -1.25	413.81	8	371.15	35.77	42.66	27.77	.035
$\leq$ -.78	435.93	8	396.81	35.29	39.12	27.29	.067
$\leq$ 0.00	447.08	8	403.71	35.38	43.37	27.38	.027
$\leq$ 0.12	447.29	8	402.24	35.36	45.05	27.36	.018
$\leq$ 0.81	450.87	8	408.93	34.95	41.94	26.95	.033
$\leq$ 1.33	444.21	8	403.64	35.40	40.56	27.40	.051
$\leq$ 1.99	433.13	8	390.76	35.68	42.37	27.68	.037
$\leq$ 2.79	421.90	8	383.80	35.77	38.10	27.77	.091
$\leq$ 3.35	397.43	8	360.08	35.81	37.35	27.81	.107
$\leq$ 4.00	371.54	8	335.90	35.74	35.64	27.74	.144
$\leq$ 4.83	327.01	8	288.00	35.54	39.01	27.54	.072
$\leq$ 5.74	269.42	8	237.97	34.95	31.45	26.95	.251
$\leq$ 6.81	207.75	8	160.78	35.36	46.98	27.36	.011

Table 2

The + sign denotes whether a nonparametric component is included in the best and the worst semi-additive model according to AIC. A blank indicates that a predictor enters linearly.

Excess returns	Dividend yield	Excess return at t	Term spread	Default spread	1-month Treasury-bill yield	M2 growth	Employment growth
Best AIC model							
$\leq$ -5.52		+				+	
$\leq$ -4.38		+	+				+
$\leq$ -3.06			+		+		
$\leq$ -2.24					+		
$\leq$ -1.69						+	
$\leq$ -1.25			+		+	+	
$\leq$ -.78			+		+		+
$\leq$ 0.00					+	+	+
$\leq$ 0.12			+			+	+
$\leq$ 0.81					+		
$\leq$ 1.33							
$\leq$ 1.99					+		
$\leq$ 2.79	+						
$\leq$ 3.35	+		+				
$\leq$ 4.00	+						
$\leq$ 4.83				+	+		
$\leq$ 5.74				+			
$\leq$ 6.81	+			+			+
Worst AIC model							
$\leq$ -5.52	+		+	+	+	+	
$\leq$ -4.38	+	+		+	+	+	+
$\leq$ -3.06	+	+		+	+	+	
$\leq$ -2.24	+	+	+	+	+	+	+
$\leq$ -1.69	+	+	+	+	+	+	+
$\leq$ -1.25	+	+		+	+	+	+
$\leq$ -.78	+	+		+	+	+	
$\leq$ 0.00	+	+	+	+	+	+	
$\leq$ 0.12	+	+		+	+	+	
$\leq$ 0.81	+	+	+	+	+	+	
$\leq$ 1.33	+	+	+	+	+	+	+
$\leq$ 1.99	+	+	+	+	+	+	+
$\leq$ 2.79		+	+	+	+	+	+
$\leq$ 3.35		+	+	+	+	+	+
$\leq$ 4.00		+	+	+	+	+	+
$\leq$ 4.83		+	+	+		+	+
$\leq$ 5.74	+	+	+		+	+	+
$\leq$ 6.81	+	+	+		+	+	+

Figure 1

Monthly rates of return on US stocks and one-month Treasury bills and their difference – the excess returns – from April 1960 to July 1992. The dotted vertical bar indicates April 1989.

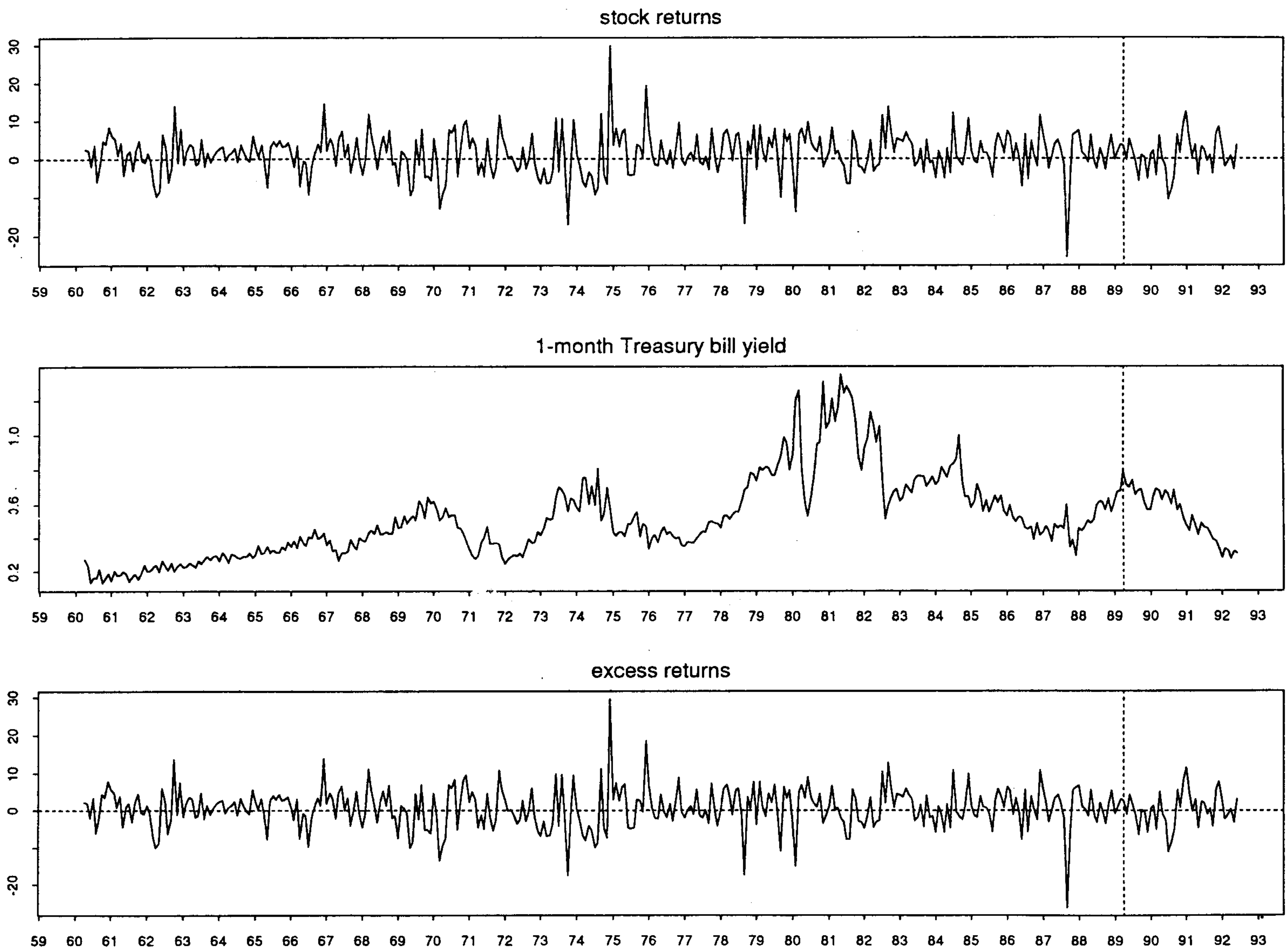


Figure 2

The data used in estimation are from April 1960 through March 1989. The top-left panel shows the density of excess returns (solid line), estimated using a kernel smoother with a triangular kernel and a fixed bandwidth chosen subjectively. As a reference point, we have overlaid the density of a normal distribution (broken line) with mean and variance equal to those of the excess returns. The remaining panels are bivariate scatterplots of future excess returns against each of the predictors. The solid line is a loess scatterplot smooth. On the right and the top of each scatterplot are the univariate densities of excess returns and the predictor. In plotting the data, we symmetrically trimmed 2.5 percent at the top and the bottom to eliminate the visual effect of a few extreme outliers.

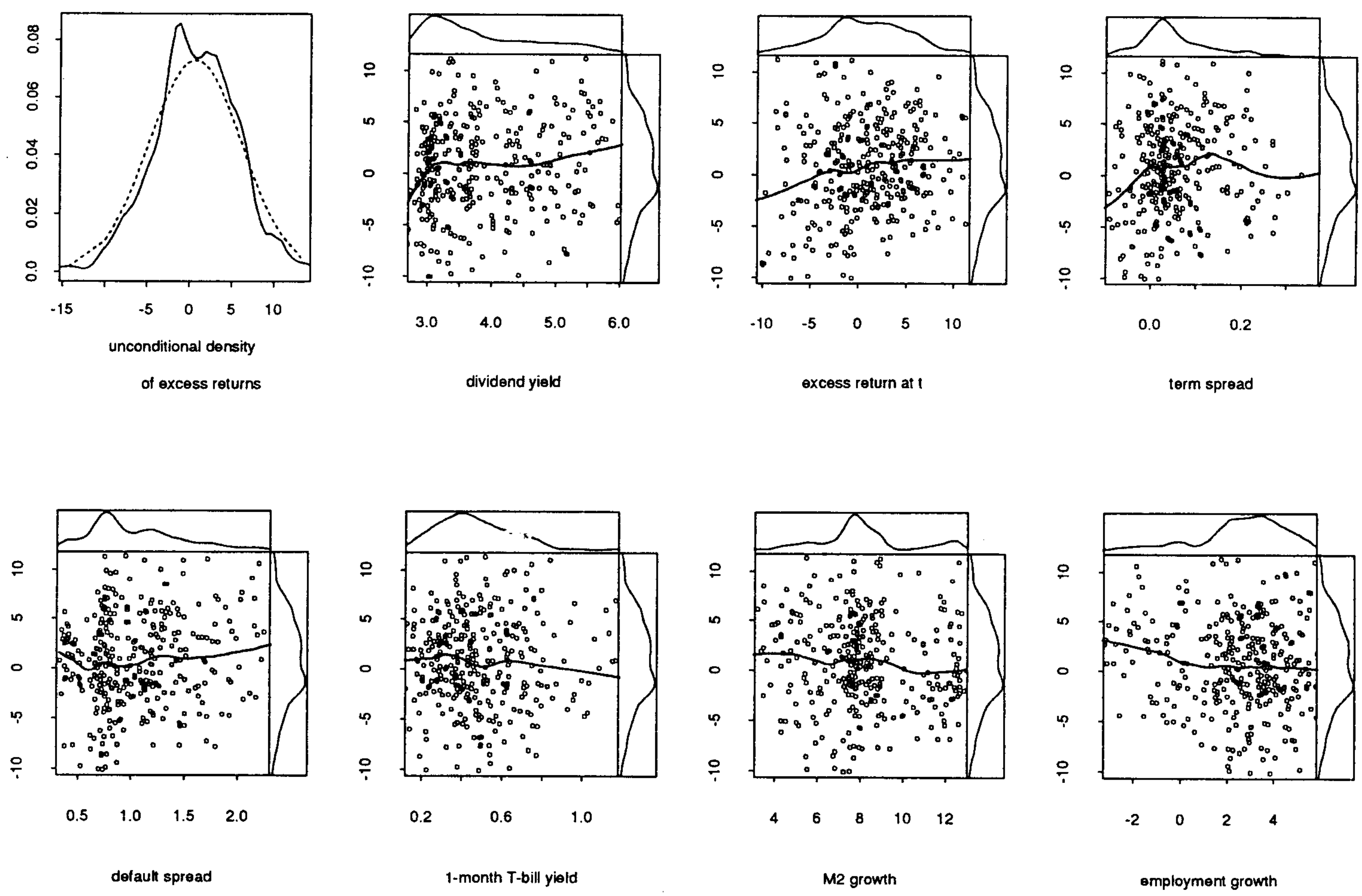


Figure 3

Estimated coefficients for the linear logit model (4). The vertical bars on the horizontal axis correspond to the 18 values of excess returns at which the conditional distribution function is estimated. The dotted lines are pointwise two-standard error bands.

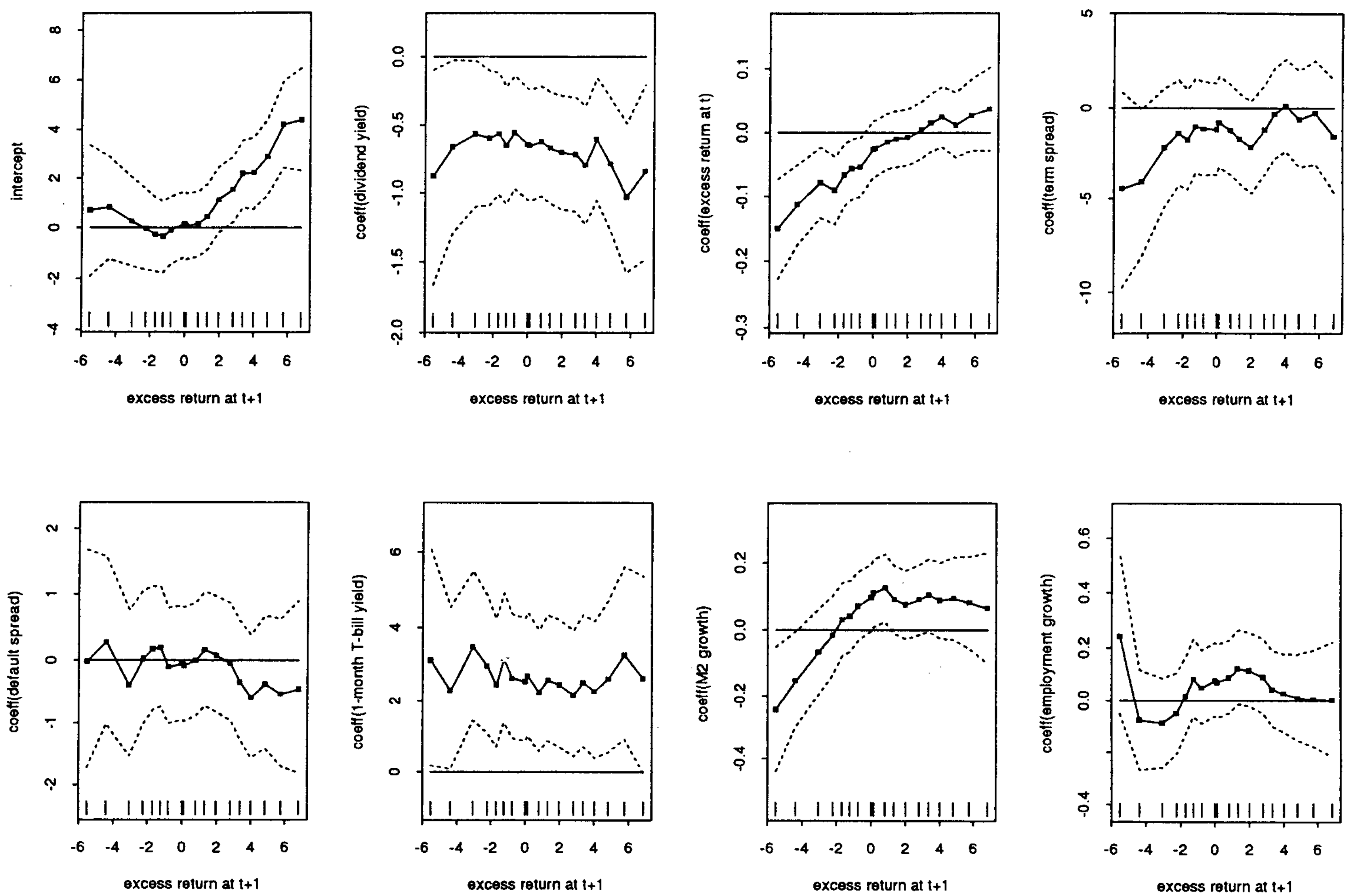


Figure 4

Iso-probability contours of the conditional distribution function estimated by the linear logit model (4). Each plot is constructed by evaluating the fitted probabilities $\hat{F}_j(x)$, $j = 1, \ldots, J$, over a grid of 100 equally spaced points between the 2.5-th and the 97.5-th percentile of one of the predictors, keeping all the others constant at their average value.

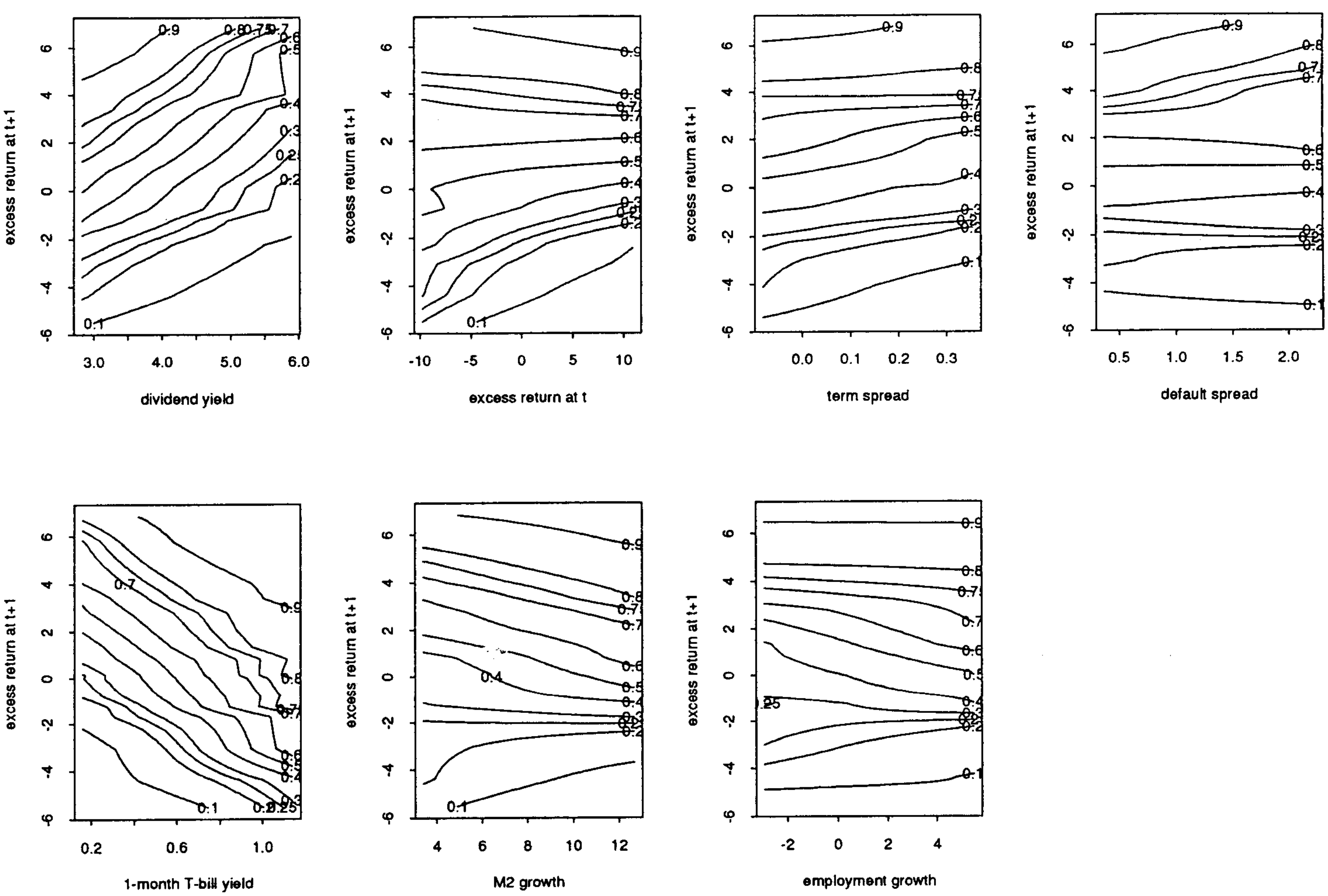


Figure 5

Iso-probability contours of the conditional distribution function estimated by the additive logit model (5). Each plot is constructed by evaluating the fitted probabilities $\hat{F}_j(x)$, $j = 1, \ldots, J$, over a grid of 100 equally spaced points between the 2.5-th and the 97.5-th percentile of one of the predictors, keeping all the others constant at their average value.

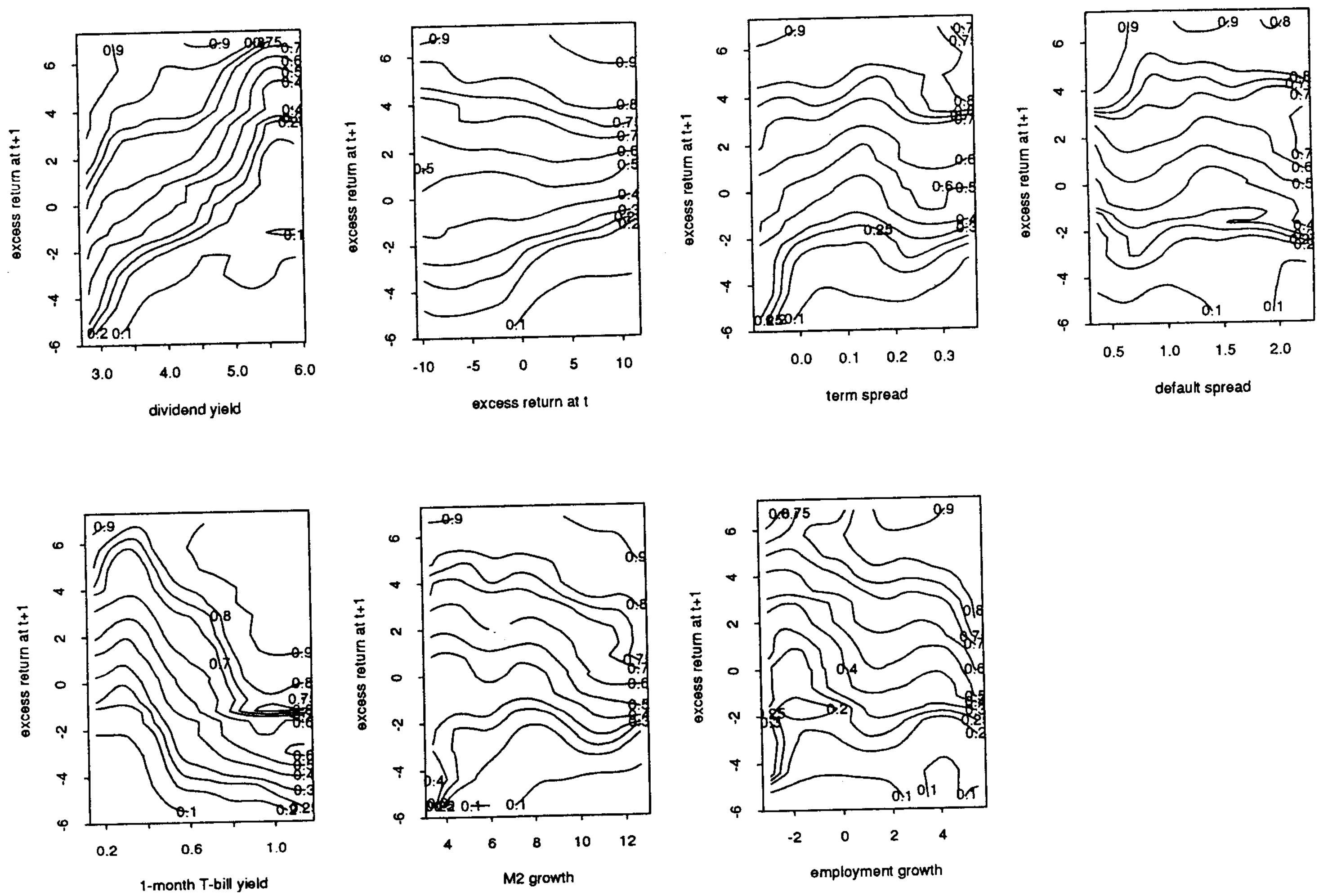


Figure 6a

Each plot is on the scale of the log odds-ratio of $P(Y_{t+1} \leq 0|X_t)$. The thicker solid curves represent the nonparametric components $\hat{g}_{hj}(x_h)$ from the additive logit, while the lighter solid lines represent the linear components $x_h\hat{\delta}_{hj}$ from the linear logit. All components are normalized to be equal to zero at the average value of the predictor. The dotted lines are two standard-error bands associated with the linear component. The vertical bars at the basis of each plot show the frequency distribution of the predictor.

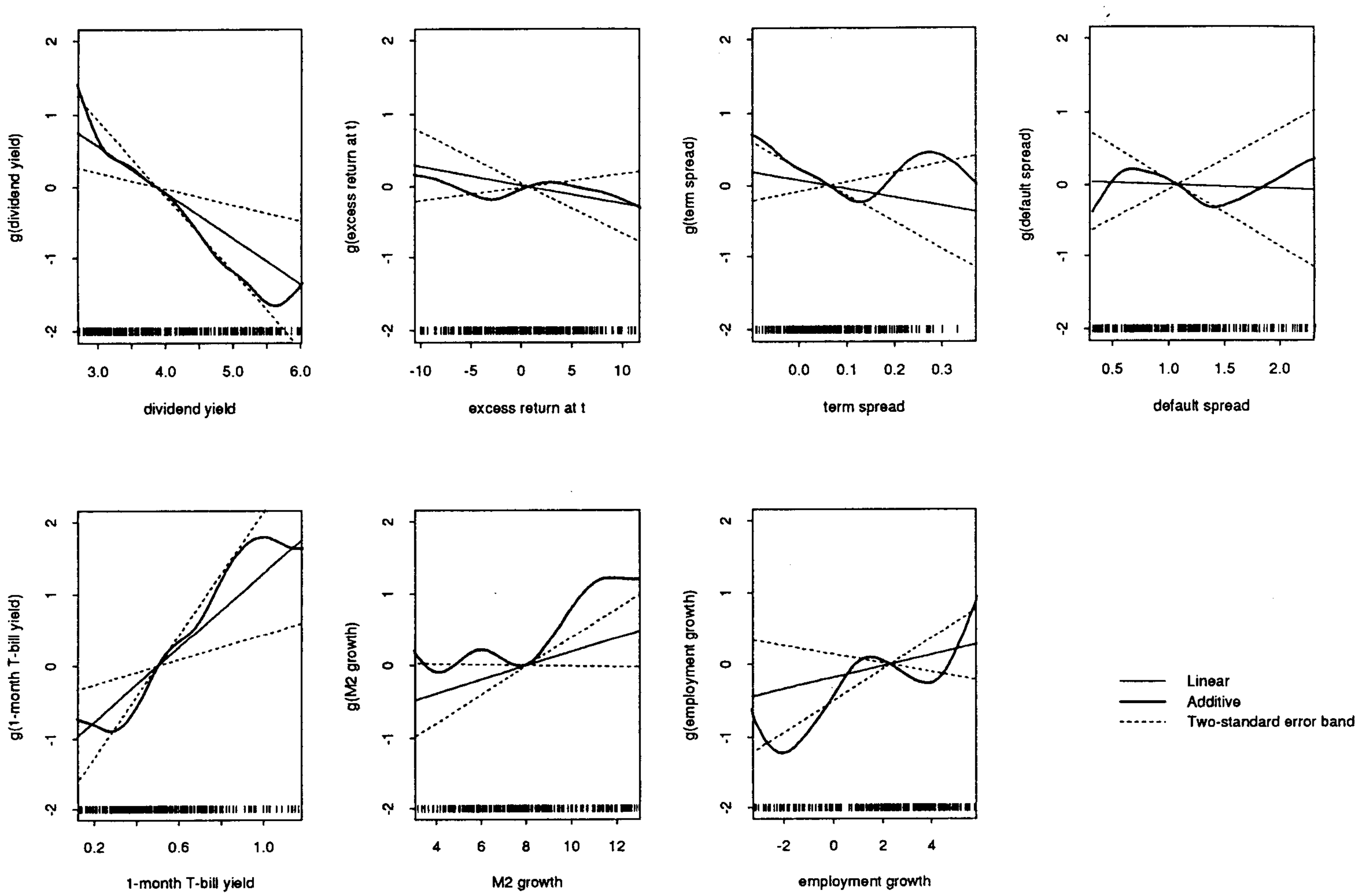


Figure 6b

Each plot shows the estimates of $P(Y_{t+1} \leq 0|X_t)$ as a function of a single predictor, keeping all the others constant at their average value. Estimates from the additive logit are represented by a thicker solid curve, those from the linear logit by a lighter solid curve. The broken curves are estimates from a nonparametric logit fitted using just a single predictor. The vertical bars at zero and one indicate the occurrence of zeros and ones in the response variable $Z = I_{(-\infty,0]}(Y_{t+1})$.

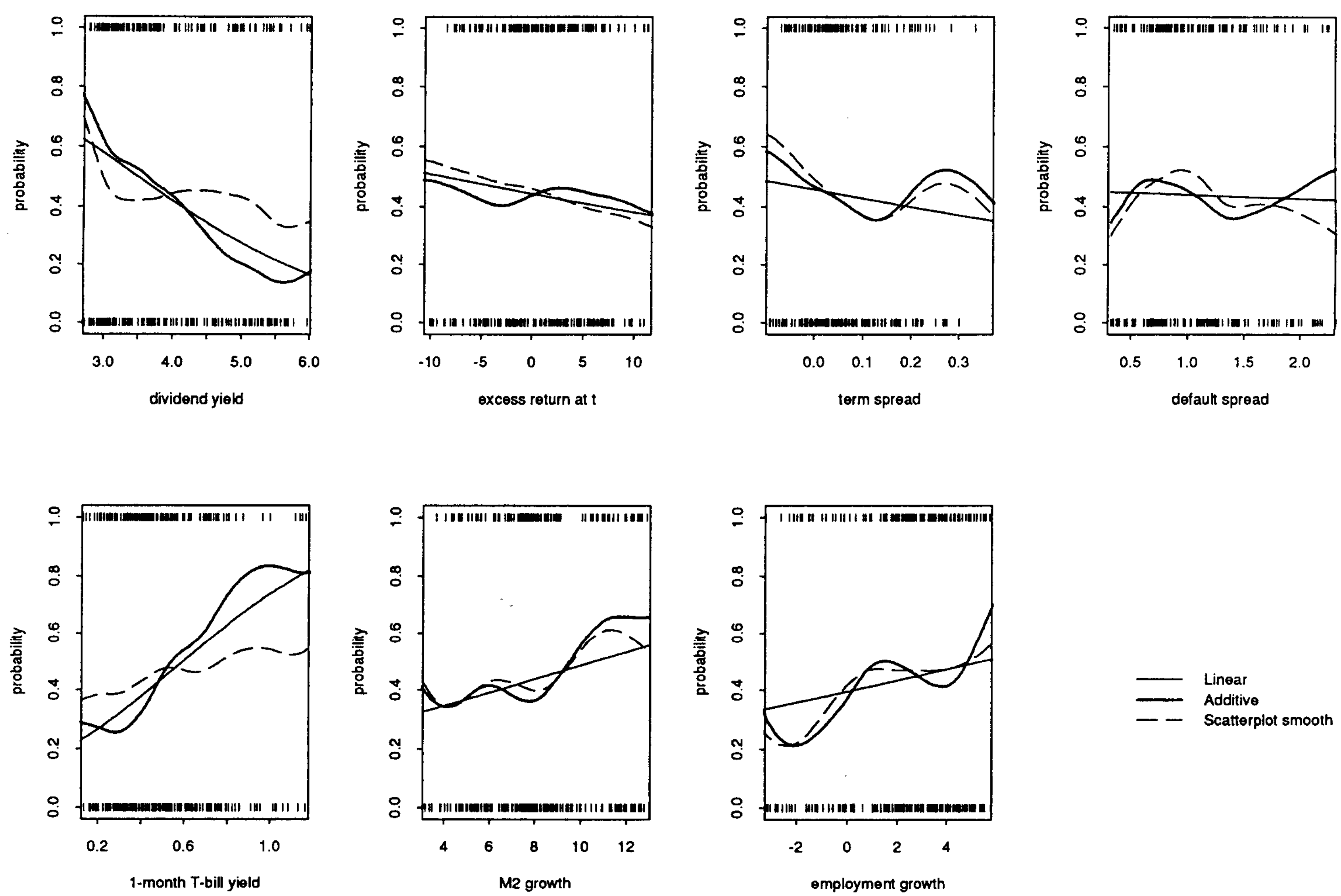


Figure 7

For six threshold values y, we plot the residual deviance and the degrees of freedom of the 128 models obtained by considering all possible combinations of linear and nonparametric terms. The upper dotted line represents all combinations of deviance and degrees of freedom that give the same AIC as the additive logit, while the lower line corresponds to the linear logit. Black squares and black diamonds denote the best and the worst semi-additive model, respectively.

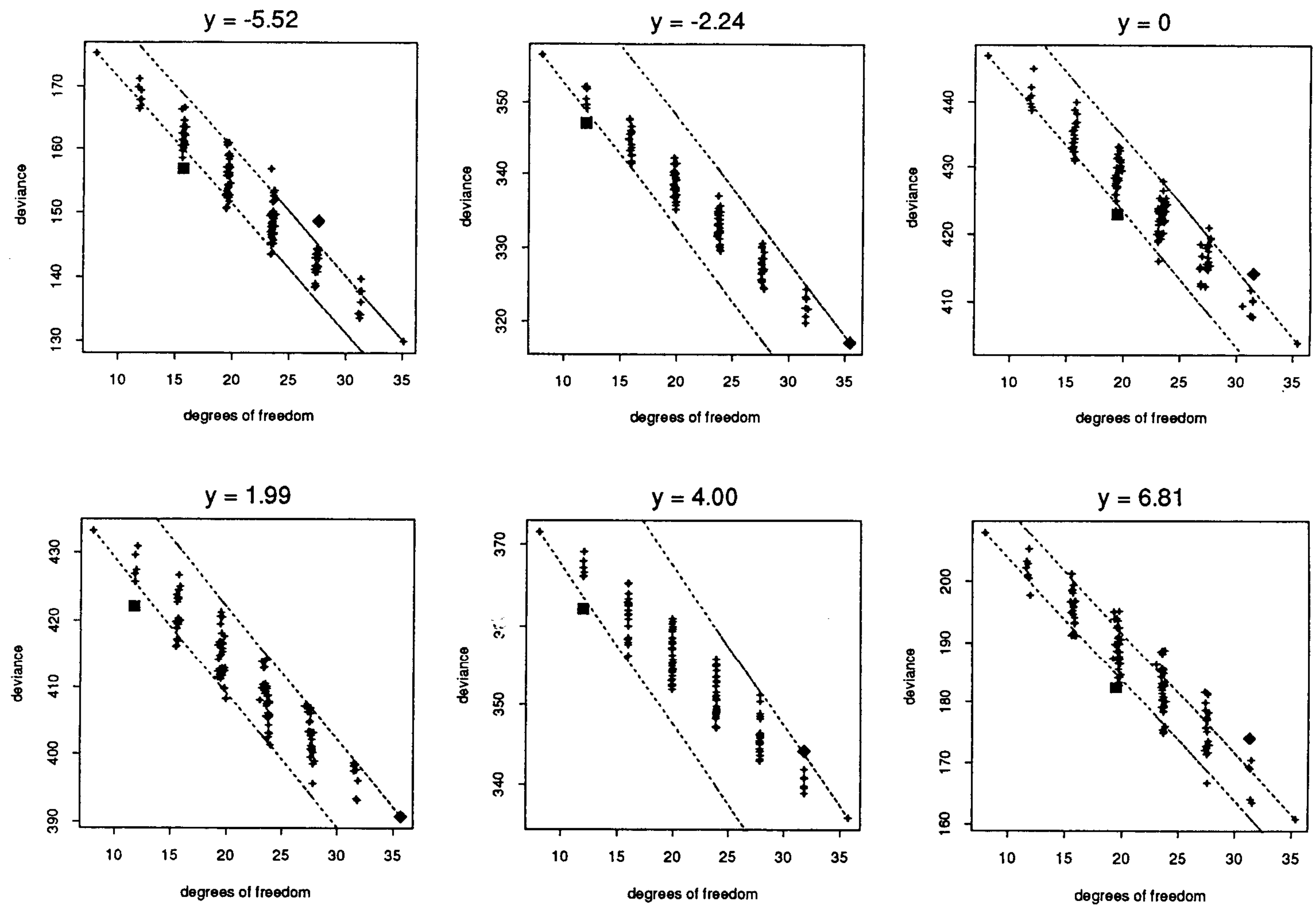
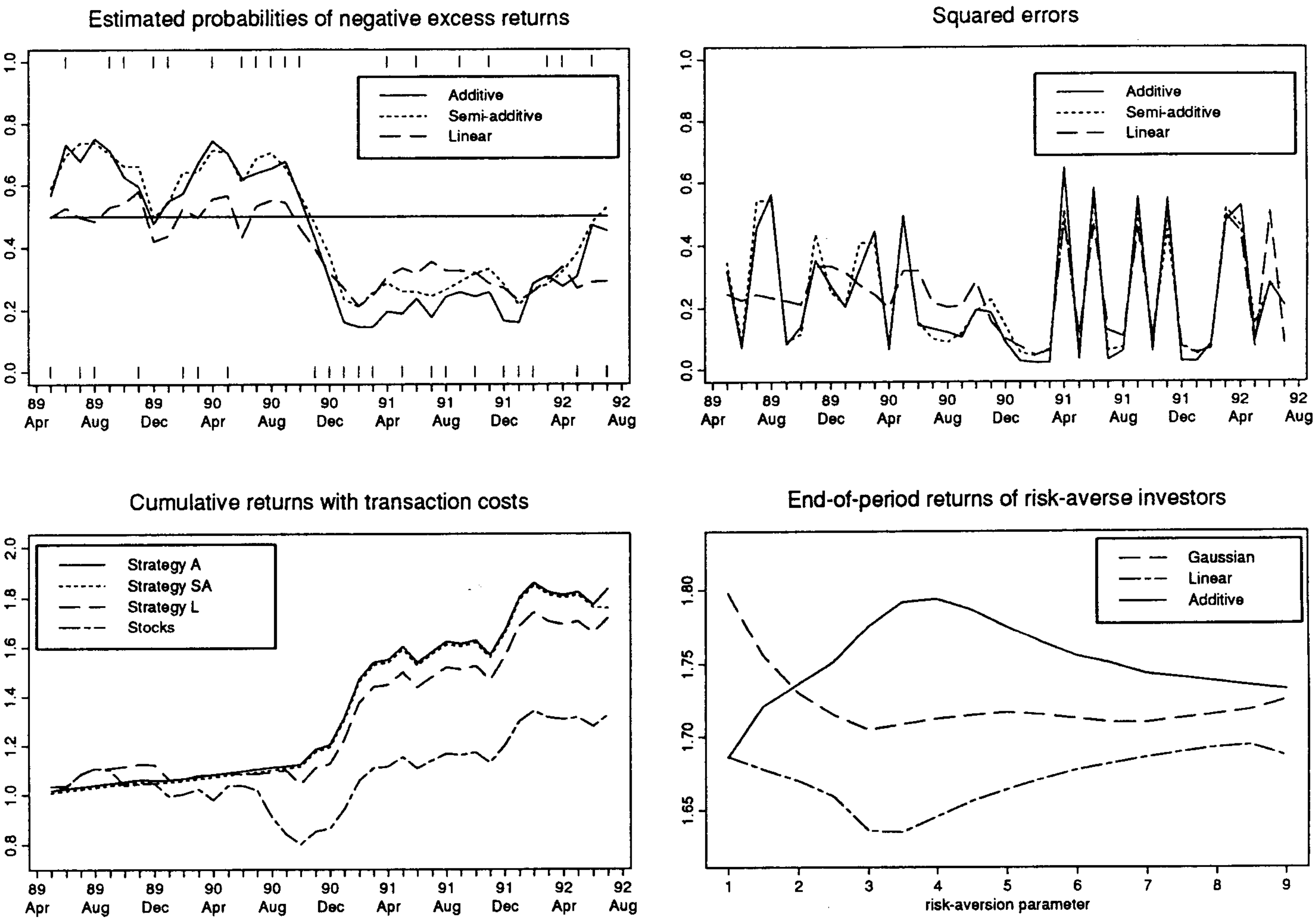


Figure 8

The top-left panel plots the estimates of the probability of negative excess returns. The vertical bars at one (zero) show when actual excess returns are negative (positive). The top-right panel plots the squared prediction errors of each model for the out-of-sample period May 1989–July 1992. The bottom-left panel plots the cumulated returns from $1 invested on April 1989. "Stocks" denotes the returns from holding the equally-weighted portfolio of New York Stock Exchange common stocks. Strategies A, L, and SA time the market by switching between stocks and a portfolio of 1-month Treasury bills depending on the probability of negative excess returns as predicted by the additive logit (A), the linear logit (L), and the semi-additive logit (SA). The bottom-right panel plots, for different levels of risk aversion, the end-of-period returns from $1 invested by risk-averse investors on April 1989. "Gaussian" denotes the return from a gaussian model of log-stock returns.



APPENDIX

Our raw data are monthly, from March 1959 to July 1992.

The following variables are from the Center for Research in Security Prices (CRSP) at the University of Chicago. The original return data in the CRSP data set are in decimal form. We express them in percentage, but we do not annualize them.

1. Returns on the equally-weighted portfolio of New York Stock Exchange common stocks

$$\mathbf{ewret}_t = \frac{1}{N}\sum_{n=1}^{N} r_{n,t},$$

where $r_{n,t} = \frac{p_{n,t}f_{n,t}+d_{n,t}}{p_{n,t-1}}$ and $r_{n,t}$ is the return on stock n purchased at $t-1$, $p_{n,t}$ is the stock price at time t, $d_{n,t}$ is the dividend, $f_{n,t}$ is the price adjustment factor to allow for stock splits. Monthly, not annualized.

2. **ustret**: return on U.S. Treasury bills with one month to maturity, from Ibbotson Associates. Monthly yield, not annualized. The yield for month $t+1$ is computed from the price quoted at the end of month t for the Treasury bill closest to one month maturity but with less than 30 days to maturity. The difference between **ewret** and **ustret** is the excess return.

3. Return on U.S. Treasury bills with one month to maturity, from the CRSP "Fama" Treasury bill term structure files. Monthly yield, not annualized, continuously compounded. The yield for month $t+1$ is computed from the price quoted at the end of month t for the Treasury bill closest to one month maturity. It may not coincide with **ustret**, as the maturity of the bill can exceed 30 days.

4. Monthly holding period returns on 3- and 1-month Treasury bills, from the CRSP "Fama" Treasury bill term structure files. Monthly rate, not annualized, continuously compounded.

The return for month $t + 1$ is computed from the price quoted at the end of the month t for the Treasury bill closest to 3- and 1-month maturities. The difference between the 3- and the 1-month Treasury-bill return is the term spread.

The following variables are from Citibase:

1. Industrial production, total index (1987=100). Seasonally adjusted. We use the year-to-year continuously compounded growth rate.

2. Employment, employees on nonagricultural payroll, total of private sector. Seasonally adjusted. We use the year-to-year continuously compounded growth rate.

3. M2 money stock: includes M1 (currency + traveler checks + bank demand deposits + other checkable deposits) + small-denomination time deposits + savings deposits + money market deposit accounts + money market mutual funds shares (noninstitutional) + overnight repurchase agreements + overnight Eurodollars. Daily averages, seasonally adjusted.

4. Dividend yield: dividend-price ratio of the Standard & Poor's 500 composite common stock portfolio. Per cent per annum. This is the 12-month moving sum of dividends divided by Standard & Poor's 500 index level for the current month.

5. Yield to maturity on speculative grade corporate bonds (Moody's Baa Corporate Bond Yield). Per cent per annum. Monthly averages of daily yields.

6. Yield to maturity on investment grade corporate bonds (Moody's Aaa Corporate Bond Yield). Per cent per annum. Monthly averages of daily yields. The difference of the Baa and the Aaa yields gives the "default spread".