ECONOMIC RESEARCH REPORTS

THE CONDITIONAL DISTRIBUTION OF EXCESS RETURNS: AN EMPIRICAL ANALYSIS

BY

Silverio Foresi
and
Franco Peracchi

RR # 92-49 *October, 1992*

C. V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

The Conditional Distribution of Excess Returns: An Empirical Analysis

Silverio Foresi
Department of Finance
New York University
New York, NY 10012

and

Franco Peracchi
Department of Economics
New York University
New York, NY 10003

October 1992

Abstract

This paper presents a new method for nonparametric estimation of the conditional distribution function of excess returns. The method, based on techniques recently developed for generalized additive models, avoids the "curse of dimensionality" problem while retaining interpretability and may be viewed as dual to the one based on modelling the quantiles of the conditional distribution. Our results indicate the presence of nonlinearities in the impact of a monetary aggregate on the conditional probability of future excess returns that do not arise from structural breaks and constitute a robust feature of the data.

JEL classification: C13, C14, G12.

Keywords: Additive models, asset pricing, nonparametric methods, regression quantiles.

We thank workshop participants at New York University for helpful comments, and the C.V. Starr Center for Applied Economics at New York University for financial support.

1 Introduction

Empirical work in asset pricing has been devoted almost exclusively to the study of the first two moments of the probability distribution of asset returns. Such an emphasis, largely due to the simplicity of classical Capital Asset Pricing Models, is too narrow.

More generally, any equilibrium model of asset pricing can be thought of as imposing restrictions on the conditional probability distribution of future excess returns Y_{t+1}, defined as the difference between the return on the stock market and a risk-free asset, conditional upon the current state X_t. Such a model would have as primitives a law of motion for the state of the economy X_t, a Markov process defined by its transition function $P(X_{t+1} \leq x|X_t)$, the preferences of the investors and a production technology. Given the structure of the model, one can characterize the functional relation linking current returns to the state of the economy [see *e.g.* Lucas (1978)]. From this solution, one can derive the probability of future returns Y_{t+1} conditional upon the current state X_t, $P(Y_{t+1} \leq y|X_t)$. It is clear that while the equilibrium solution implies restrictions on the conditional first and second moments of excess returns, the dependence of these moments on the conditioning variables need not be of a simple parametric form.

To go beyond the limitations of moments and simple parametric specifications, in this paper we try to accomplish two things. First, we look at the entire distribution function of excess returns conditional on a set of predictors including dividend yields, interest rates, and rates of growth of money and employment. Second, we model the conditional distribution function of excess returns as a flexible function of the predictors. Our estimates are based on techniques recently developed for generalized additive models [Hastie and Tibshirani (1990)].

In an effort to model flexibly the functional relation between excess returns and predictors, nonparametric estimation of conditional means has been advocated and implemented. For example, Boudhouk, Richardson, and Whitelaw (1992) use kernel regression and flexible Fourier transforms to estimate the functional relation linking expected excess returns to a single predictor, an indicator of the term structure of interest rates.

One advantage of our approach is that it avoids the "curse of dimensionality" problem that makes it difficult to extend classical scatterplot smoothers, such as kernel regression, to high dimensions. Another advantage is interpretability, due to our use of an additive specification for certain functionals of the conditional distribution of excess returns.

As a by-product we obtain measures of the conditional location, spread, symmetry, and tail weight of the distribution of excess returns, based on the quantiles of the estimated conditional distribution function. This offers an appealing alternative to the use of conditional moments. While estimates of first and second moments may be sensitive to the presence of outliers, conditional distribution functions and conditional quantiles are robust functionals by definition. This may be quite important in the case of the distribution of excess returns which may comprise extreme events, such as market crashes.

Our approach may be viewed as dual to the one based on modelling the quantiles of the conditional distribution. Unlike linear regression quantiles [Koenker and Bassett (1978)], however, we do not restrict the quantiles to be of a specific parametric form. Further, our approach is easier to implement using existing software than quantile smoothing splines [Koenker and Ng (1992)].

The remaining of this paper is organized as follows. In section 2 we describe our method. Section 3 describes the data and motivates the choice of predictors. Section 4 presents our results and shows that a standard parametric specification fails to capture important characteristics of the data. We document the presence of nonlinearities in the impact of a monetary aggregate, the rate of growth of M2, on the conditional probability of future excess returns. The shape of the nonlinear relation suggests a nonmonotonicity which does not arise from structural breaks and constitutes a robust feature of the data examined.

2 The statistical model

Let the continuous random variable Y_{t+1} denote future excess returns, and let X_t be a random k-dimensional vector that helps predict Y_{t+1}. We are interested in studying $F(y|x) = P(Y_{t+1} \leq y|X_t = x)$, the conditional distribution function of Y_{t+1} given $X_t = x$, which is assumed to be time-invariant. Knowledge of $F(y|x)$ represents the complete solution to the problem of predicting Y_{t+1} given X_t, in the sense that it contains all the information relevant for prediction, irrespective of the particular loss function that the analyst may be using.

2.1 Model specification

The problem of estimating the conditional distribution function $F(y|x)$ from the data is simplified considerably if only a finite set $-\infty < y_1 < \cdots < y_J < \infty$ of evaluation points is considered. Define a sequence of binary random variables $Z_{j,t+1} = 1(Y_{t+1} \leq y_j)$, $j = 1, \ldots, J$, where $1(\cdot)$ is the usual indicator function. For each j, the conditional distribution of $Z_{j,t+1}$ given $X_t = x$ is a Bernoulli with parameter $F(y_j|x) = E(Z_{j,t+1}|X_t = x)$. This reduces the problem to estimation of the finite sequence $\{F(y_j|x)\}_{j=1}^J$.

Although the sequence $\{F(y_j|x)\}$ must satisfy the monotonicity constraint

$$0 < F(y_1|x) < \cdots < F(y_J|x) < 1 \tag{1}$$

for all x in the support of X_t, the estimation problem is much easier if (1) is not imposed.

The Bernoulli is a one-parameter linear exponential family of distributions with canonical parameter

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

the log odds-ratio. The conditional log-likelihood function for a single observation (z, x) on $(Z_{j,t+1}, X_t)$, expressed in terms of the canonical parameter, is of the form

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

Among other things, reparameterizing the model in terms of the canonical parameter ensures that the constraint $0 < F(y_j|x) < 1$ is automatically satisfied.

The conventional parametric approach to estimation proceeds by restricting $\eta(y_j, \cdot)$ to a family of functions indexed by a finite-dimensional real parameter. If

$$\eta(y_j, x) = \gamma_{0j} + x\delta_j, \tag{2}$$

where x is a row vector, we have the classical logit model. Beside ease of estimation, one important feature of the linear specification (2) is interpretability, namely the fact that the h-th component of the vector δ_j can be interpreted as the partial effect of the h-th variable in X_t on the log odds-ratio. One problem with this approach is that the linear specification may be too restrictive. A more flexible parametric specification may be obtained by specifying $\eta(y_j, x)$ as a polynomial function of x, such as a quadratic or a cubic. Even in this case, however, the global nature of the approximation may be too restrictive.

The alternative of a fully nonparametric specification is out of the question, in the case of small to moderate sample sizes, if one wants to control for a sufficiently rich set of predictors. One way of attacking the curse of dimensionality problem would be to consider projection-pursuit logistic regression, with the canonical parameter specified as

$$\eta(y_j, x) = \gamma_{0j} + \sum_{h=1}^{m} g_{hj}(x\delta_{hj}), \tag{3}$$

where $\{x\delta_{hj}\}_{h=1}^{m}$ are linear combinations or projections of the elements of x, and $\{g_{hj}\}_{h=1}^{m}$ are arbitrary univariate smooth functions. As in standard projection pursuit regression

[Friedman and Stuetzle (1981)], the number m of linear combinations, the vectors $\{\delta_{hj}\}$ and the functions $\{g_{hj}\}$ would be determined from the data by minimizing a measure of predictive power.

Although projection pursuit may be a promising approach, in this paper we choose for simplicity to work with a specification that is intermediate between (2) and (3), namely

$$\eta(y_j, x) = \gamma_{0j} + \sum_{h=1}^{k} g_{hj}(x_h), \tag{4}$$

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 (4) makes it possible to interpret g_{hj} as the derivative of the log odds-ratio with respect to the h-th component of X_t, thereby retaining interpretability, which is one of the important features of the linear specification (2).

We shall also compare the additive specification (4) with the semi-additive model

$$\eta(y_j, x) = \gamma_{0j} + x^{(1)}\delta_j + \sum_{h=1}^{k_2} g_{hj}(x_h^{(2)}), \tag{5}$$

where the k_1 variables in $x^{(1)}$ enter linearly, and the $k_2 = k - k_1$ variables in $x^{(2)}$ enter nonlinearly. A specification similar to (5) was used by Engle *et al.* (1986) in the context of modelling a conditional expectation function. Clearly, the linear, semi-additive and additive specifications form a sequence of nested models.

Estimation of models (4) and (5) may be based on the local scoring algorithm proposed by Hastie and Tibshirani (1990) for generalized additive models. Their method modifies the scoring algorithm for the logit model by constructing at each step smooth estimates of the univariate functions $\{g_{hj}\}$, obtained by fitting an additive model to the adjusted dependent variable from the previous step. Smoothing may be based on nearest neighborhood methods, such as Cleveland's (1979) loess, or on cubic smoothing splines. The latter choice has certain advantages, including convergence of the local scoring algorithm. The smoothing parameter may be chosen subjectively, or an automatic selection criterion, such as cross-validation, may be employed.

Stone (1986) showed that under appropriate regularity conditions, the estimates based on model (4) are consistent for the best additive approximation, in the Kullback-Leibler sense, to the true population log odds-ratio. Similarly, the estimates based on model (5) are consistent for the best semi-additive approximation to the true population log odds-ratio.

By analogy with standard logit models, we compare nested models using the ratio of the log-likelihoods. Although the distribution theory for these statistics is not yet developed, simulation results in Hastie and Tibshirani (1990) show that the χ^2 distribution provides a useful approximation. Of course, a computationally expensive alternative would be to use the bootstrap. Following Hastie and Tibshirani (1990), 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.

2.2 Relation with regression quantiles

In this section we relate our method to the regression quantiles introduced by Koenker and Bassett (1978). The original motivation for regression quantiles was robustness with respect to outliers in the y-space. This property is also shared by our method, because $Z = 1(Y \leq y)$ is a bounded random variable.

A π-th quantile of the conditional distribution of Y given $X = x$, or regression quantile for short, is a root $\theta(\pi, x)$ of the implicit equation in y

$$F(y|x) = P(Y \leq y|X = x) = \pi,$$

for given x and $\pi \in (0, 1)$ (we dropped time subscripts for simplicity). Knowledge of the whole family of regression quantiles is equivalent to knowledge of the conditional distribution of Y given X. A regression quantile is linear in x if it is of the form $\theta(\pi, x) = \alpha(\pi) + x\beta(\pi)$.

The assumption of linear regression quantiles is restrictive. For example, suppose that Y is generated by a location-scale transformation of a standardized continuous random variable U, that is, $Y = \mu(x) + \sigma(x)U$, where $0 < \sigma(x) < \infty$. The π-th regression quantile of Y is then equal to

$$\theta(\pi, x) = \mu(x) + \sigma(x)\theta(\pi),$$

where $\theta(\pi)$ denotes the π-th quantile of the distribution of U. Clearly, even when $\mu(x)$ is linear in x, $\theta(\pi, x)$ is generally nonlinear.

In what follows we consider the restrictions on the family of regression quantiles implied by a specific assumption on the log odds-ratios and, viceversa, the restrictions on the family of log odds-ratios implied by a specific assumption on the regression quantiles.

Suppose first that $\eta(y, x)$ is a known continuously differentiable function of (y, x). Also suppose that $\eta_y(y, x)$ is strictly positive for every (y, x), where subscripts denote partial derivatives. These assumptions are equivalent to the assumption that $F(y|x)$ is continuously differentiable in (y,x), with a density $f(y|x)$ that is strictly positive for every (y,x). For a fixed $\pi \in (0,1)$, the π-th regression quantile $\theta(\pi,x)$ is a solution with respect to y of the implicit equation

$$\eta(y,x) = \ln \frac{\pi}{1-\pi}.$$

By the implicit function theorem, $\theta(\pi,x)$ is unique and continuously differentiable in x, with

$$\theta_x(\pi,x) = -\frac{\eta_x(y,x)}{\eta_y(y,x)}\bigg|_{y=\theta(\pi,x)} = -\frac{F_x(y|x)}{f(y|x)}\bigg|_{y=\theta(\pi,x)}. \tag{6}$$

Notice that, $\theta_x(\pi,x)$ and $\eta_x(\theta(\pi,x),x)$ have opposite sign. Clearly, linearity of regression quantiles obtains if and only if the right-hand-side of (6) does not depend on x. In the special case when the log odds-ratio is linear in x, that is $\eta(y,x) = \gamma(y) + x\delta(y)$, linearity of the regression quantiles obtains if and only if $\partial\delta/\partial y = 0$, that is, $\delta(y)$ is a constant function, as in the ordered logit model.

Now suppose that $\theta(\pi,x)$ is a known continuously differentiable function of (π,x). This implies that $f(\theta(\pi,x)|x)$ is strictly positive for all x. If $\theta(\pi,x) = y$, where y is a fixed number, then by definition

$$\pi = F(\theta(\pi,x)|x) = F(y|x).$$

Hence, by the chain rule,

$$F_x(y|x) = -\theta_x(\pi,x)\, f(y|x)|_{\pi=F(y|x)}$$

or, equivalently,

$$\eta_x(y,x) = -\theta_x(\pi,x)\,\eta_y(y,x)|_{\pi=F(y|x)}\,. \tag{7}$$

Clearly, linearity of the log odds-ratio obtains if and only if the right-hand-side of (7) does not depend on x. If the regression quantiles are linear, that is $\theta(\pi,x) = \alpha(\pi) + x\beta(\pi)$, the log odds-ratio is linear if and only if

$$\eta_x(y,x) = -\beta(\pi)\,\eta_y(y,x)|_{\pi=F(y|x)}$$

does not depend on x. A sufficient condition is that $\beta(\pi)$ is a constant function and $\eta_y(y,x)$ does not depend on x.

As an example, suppose that $Y = x\beta + U$, where U has a logistic distribution with mean zero and unit variance. Then the π-th regression quantile of Y is linear

$$\theta(\pi,x) = \alpha(\pi) + x\beta,$$

where $\alpha(\pi)$ is equal to the π-th quantile of the distribution of U. On the other hand,

$$F(y|x) = \frac{\exp(y - x\beta)}{1 + \exp(y - x\beta)},$$

and so $\eta(y, x) = \gamma(y) + x\delta$, where $\gamma(y) = y$ and $\delta = -\beta$.

One motivation for regression quantiles is their robustness with respect to outliers in the y-space. This property is also shared by our method, because $Z = 1(Y \leq y)$ is a bounded random variable.

3 Empirical Results

3.1 Choice of predictors

Our general frame of reference is an equilibrium model in which interest rates and future stock returns are determined by the current state of the financial, real, and monetary sectors of an economy.

To summarize the state of the financial sector we use the short term interest rate and the dividend yield. The short term rate is a proxy for the location of the investment opportunity set over time [Merton (1973)]. The dividend-price ratio, or dividend yield for short, is a proxy for expected future returns [Rozeff (1984)]. Since the dividend yield encompasses expected dividend growth, it is a noisy proxy for future returns. To control for dividend growth, we also include measures of growth of the economy, based on industrial production and employment. Industrial production and employment are also proxies for the state of the real sector, as they capture the business cycle and the labor market conditions. Similar cyclical variables have been used in the studies of Gertler and Grinols (1982) and Chen, Roll, and Ross (1986) as macrofactors in an Arbitrage Pricing Theory context.

Finally, as an indicator of the monetary sector, we include the rate of growth of M2. Monetary variables have received considerable attention in studies of the relation between stock returns and inflation [Fama (1981)]. More recently, the use of an endogenous monetary aggregate such as M2 has been advocated as a proxy for consumption in a consumption-oriented Capital Asset Pricing Model [Chan, Foresi, and Lang (1992)].

3.2 Data sources

Our data are monthly from January 1960 to December 1990.

The macroeconomic variables include industrial production, employment, inflation, and M2. The financial variables include the dividend yield on the Standard & Poor Common

Stock Composite, the one month yield on Treasury bills, a term spread equal to the difference between the six-month and the three-month Treasury bill yields, and a default spread equal to the difference of the yields on BAA and AAA corporate bonds. Excess returns are the difference between the returns on the equally weighted portfolio of New York Stock Exchange common stocks and the yield on a one-month Treasury bill.

The macroeconomic data and the dividend yield are seasonally adjusted from Citibase. The other financial data are from the Center for Research in Security Prices at the University of Chicago.

Interest rates, dividend yields, and rates of returns are continuously compounded.

3.3 Data transformations and model selection

We take the natural logarithm of the macro variables. All variables are then standardized by subtracting the sample mean and dividing by the sample standard deviation.

Dividend yield, interest rates, industrial production, employment, and money display long memory. While we could include several lags of the predictors, for reasons of parsimony we restrict attention to levels and first differences. We consider both one-month first differences of a variable X, $dX_t = X_t - X_{t-1}$, and twelve-month differences $ddX_t = X_t - X_{t-12}$. Twelve-month differencing eliminates seasonal components up to the monthly frequency and is interpretable as a smoothing filter based on the first difference of an equally weighted MA(11). Thus, it has the flavor of adding more lags while retaining a tractable dimension of the vector X_t.

We use the linear logit specification as a benchmark when selecting the predictors. Twelve-month differences of the macroeconomic variables display a smoother profile than monthly first differences and no recognizable seasonal pattern. For these variables, twelve-month differences perform better in estimation than one-month differences. In contrast, one-month differences of the financial variables, such as the dividend yield and the interest rate, display no recognizable seasonal pattern and perform better in estimation than twelve-month differences.

While both levels and one-month differences of the dividend yield are strongly significant, neither one- nor twelve-month differences in interest rates add significant predictive power after the inclusion of the level.

Following the existing literature on asset returns [see *e.g.* Gertler and Grinols (1982) and Chen, Roll, and Ross (1986)], we also tried additional predictors such as the default spread, the term spread and inflation. 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. The term spread may capture the relative availability of credit with respect to the demand. After the inclusion of money, however, the default spread and the term spread are no longer significant. This is in line with the evidence in Stock and Watson (1990) that money, term spread, and default spread convey similar information as to the state of the business cycle and thus to the state of future returns. Similarly, we find that inflation does not enter as significant after the inclusion of our predictors.

More importantly, we find no evidence that past excess returns help predict future excess returns after conditioning on our set of predictors.

Finally, 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.

In conclusion, the best results obtain with the following five predictors: the dividend yield (**fsdxp**), one-month differences of dividend yield (**dfsdxp**), the one-month Treasury bill rate (**fygm1**), twelve-month differences of the log of M2 (**ddfm2**), and twelve-month differences of log employment (**ddlp**).

As for the dependent variable, the excess returns Y, we evaluate its conditional distribution function at the deciles of the unconditional distribution. The financial literature on market timing is often interested in the probability of positive excess returns. Thus we also consider $y = 0$ as an additional evaluation point.

4 Findings

We use S-PLUS and STATA as computation environments.

Figure 1 presents the scatter plot matrix of the data. The first row shows the relation between the predictors and future excess returns. Here, no clear patterns emerge. More recognizable patterns emerge in the relation between certain pairs of predictors, such as a positive relation between the one-month Treasury bill rate and the dividend yield, and a negative relation between the one-month bill rate and the dividend yield.

Figure 2 plots the estimates of the logit slope coefficients corresponding to the ten evaluation points and the associated two-standard error intervals. The vertical line corresponds to zero excess returns.

To interpret this figure consider the dividend yield and the rate of money growth. The dividend yield has a negative and statistically significant effect at all evaluation points of the conditional distribution. This implies that an increase in the dividend yield, keeping all other predictors constant, shifts the conditional distribution of excess returns to the right. Thus, our results are in line with previous findings of Rozeff (1984), who documents a positive relation between dividend yields and expected excess returns.

On the other hand, the impact of money on the probability of excess returns changes sign at different evaluation points and is significantly different from zero only at evaluation points corresponding to positive excess returns.

Figure 3.a and 3.c show the three-dimensional (3D) surface of the conditional distribution of excess returns as a function of the dividend yield and the one-month bill rate keeping all other predictors constant at their mean level (zero, because of standardization). Figure 3.b and 3.d show the corresponding iso-probability contours.

Notice that, although the monotonicity constraint (1) is not imposed in the estimation, no violation is found when fitting both parametric and nonparametric versions of our model.

A positive slope of the iso-probability contours indicates a negative impact of the conditioning variable on the probability that excess returns exceed any given level. For example, the positive slope of the contour levels in the case of the dividend yield is the counterpart of the negative slope coefficients for the logit model in Figure 1.

While one could in principle retrieve the conditional moments from the estimate of the conditional distribution, it is more natural to look at measures of location, dispersion, symmetry, and tail weight based on the conditional quantiles. The π-th conditional quantile $\theta_\pi(x)$ is just the projection onto the y-axis of the iso-probability curve with probability level π.

A measure of location of the conditional distribution is given by the conditional median $\theta_{.50}(x)$. As a measure of dispersion, one may consider the interquartile range $\theta_{.75}(x)-\theta_{.25}(x)$. A possible measure of symmetry is the ratio

$$\frac{\theta_{.75}(x)-\theta_{.50}(x)}{\theta_{.50}(x)-\theta_{.25}(x)},$$

while a possible measure of tail weight is the ratio

$$\frac{\theta_{.90}(x)-\theta_{.10}(x)}{\theta_{.75}(x)-\theta_{.25}(x)}.$$

The observed differences in the spread, symmetry, and tail weight corresponding to different values of the conditioning variable give a clear indication of conditional heteroskedasticity.

We now compare the linear specification with the additive and the semi-additive. In the latter, only the rate of growth of money enters nonlinearly. To estimate the additive and semi-additive model we considered both smoothing splines and loess. While the overall shape of the estimated relation between Y and X is similar for the two smoothers, 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. Consequently, we only present results for the case of splines. The degree of smoothing is based on the equivalent number of degrees of freedom [Hastie and Tibshirani (1990)] and is set equal to 5 after an informal comparison of Akaike criteria.

In Table 1 we compare the fit of the three models. Overall, the semi-additive model including only one nonlinear term in the money growth rate outperforms both the linear and the additive models.

Figures 4 compares the impact of money growth in the linear (panels a and b) and the semi-additive model (panel c and d). The estimates are quite different in the two cases. While money is only borderline significant in the linear logit model, it becomes strongly significant in the semi-additive [Table 1]. Further, the 3D graphs and the iso-probability plots show that the more flexible semi-additive model reveals a systematic pattern of nonlinearities at all quantiles.

For money growth rates close to the mean, we observe a negative association between money growth and the probability of excess returns below any given value. In this case, the effect of money growth is to shift the whole distribution of excess returns to the right. For extreme values of money growth rates, however, the association is positive, that is, the effect of money growth is to shift the whole distribution of excess returns to the left. To capture this complicated form of nonlinearity, a parametric model would require at least a cubic term in money growth, a quadratic term would not be sufficient.

4.1 Time invariance

The marginal distributions of both the excess returns and the predictors may be time-varying and yet the conditional distribution of excess returns may exist and be time-invariant. Our analysis is meaningful only if the conditional distribution of excess returns exists and is time-invariant. We now show some evidence in support of this assumption.

An obvious candidate for the nonlinearities that we find in the data, is the presence of structural breaks. To verify this possibility, we distinguish between three subperiods. The first goes from the beginning of the data to the end of Bretton Woods monetary system

(March 1972). The second goes from March 1972 to October 1982 and corresponds to the inflationary part of the 1970's. Although the change in the operating procedures of the Federal Reserve from October 1979 to October 1982 is sometimes also identified as a structural break, we leave this period with the inflationary 1970's. The third period goes from October 1982 to the end of the data.

In Figure 5 we consider the semi-additive model and we plot the impact of money growth on the estimated probability of negative excess returns for the whole sample (panel a) and separately for each subperiod (panel b). For comparison, money growth rates are standardized within each subperiod. Although the position of three curves is different, what is remarkable is the similarity in their shapes, and the fact that the minima and the maxima in each subperiod occur for similar values of (standardized) money growth. Thus it seems that the nonlinear relation between money and excess returns does not arise from structural breaks and can be viewed as a robust feature of the data examined.

What kind of structural model could explain the nonlinearity? The finding is consistent with the idea that future monetary policy, by lowering interest rates, may increase stock prices and returns and viceversa. For values of money growth close to its mean, the Federal Reserve is not expected to move interest rates in any particular direction. A positive relation of excess returns and money growth in this region is consistent with money leading the stock market over the business cycle. Higher money growth, however, becomes bad news for stocks because the Federal Reserve may increase interest rates in the near future to curb inflation. A further increase of money growth increases the likelihood of higher interest rates, hence the negative relation of excesss returns and money growth. In contrast, very negative money growth is good news for stocks because the Federal Reserve may reduce interest rates to stimulate the economy. A further decrease of money growth increases the likelihood of lower interest rates, hence the positive relation of excesss returns and money growth.

References

BOUDHOUKH, K., M. RICHARDSON, AND R. WHITELAW (1992), "The equity risk premium and the term structure: Two centuries of evidence," New York University, mimeo.

CHAN, K.C., S. FORESI AND L. LANG (1992), "Does money explain asset returns? Theory and empirical evidence," New York University, mimeo.

CHEN, N., R. ROLL, AND S. A. ROSS (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.

ENGLE, R.F., C.W.J. GRANGER, J.A. RICE AND A.A. WEISS (1986), "Semiparametric estimates of the relation between wheather 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.

FRIEDMAN, J.H., AND W. STUETZLE (1981), "Projection pursuit regression," *Journal of the American Statistical Association*, 76, 817-823.

GERTLER, M., AND E. L. GRINOLS (1982), "Unemployment, inflation, and common stock returns," *Journal of Money, Credit, and Banking*, 14, 216-233.

HASTIE, T.J., AND R.J. TIBSHIRANI (1990), *Generalized additive models*, Chapman and Hall, New York.

KOENKER, R. AND G. BASSETT (1978), "Regression quantiles," *Econometrica*, 46, 33-50.

KOENKER, R. AND P.T. NG (1992), "Quantile smoothing splines," University of Illinois, Urbana-Champaign, mimeo.

LUCAS, R.E. (1978), "Asset prices in an exchange economy," *Econometrica*, 46, 1429-1445.

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

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

STOCK, J.H., AND M. W. WATSON (1990), "Business cycle properties of selected U.S. time series, 1959–1988," National Bureau of Economic Research, Working Paper No. 3376.

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

Table 1

The table presents informal tests based on the difference of the deviances, defined as twice the value of the maximized likelihoods for two models. By analogy with the definition of degrees of freedom in linear models, the number of degrees of freedom in the additive and semi-additive models is defined as

$$d.f. = n - \text{trace}\,(2\mathbf{R} - \mathbf{R}^T\mathbf{A}\mathbf{R}\mathbf{A}^{-1}),$$

where n is the number of observations, $\mathbf{R}$ is the weighted additive-fit operator, and $\mathbf{A}$ is the estimated information matrix [Hastie and Tibshirani (1990)]. The last three columns of the table report the p-values for a χ^2 criterion based on the difference of the deviances and $d.f.$ defined above.

		Linear (L)		Additive (A)		Semi-additive (S)		p-value		
Quantile	*Exc. Ret.*	*Deviance*	*d.f.*	*Deviance*	*d.f.*	*Deviance*	*d.f.*	*L vs A*	*L vs S*	*S vs A*
1	$\leq$ -5.53	201.97	343	167.23	323.54	196.29	339.23	0.02	0.20	0.02
2	$\leq$ -3.07	329.43	343	295.98	323.49	318.18	339.03	0.02	0.02	0.12
3	$\leq$ -1.71	402.97	343	374.39	323.19	391.12	339.29	0.09	0.01	0.41
4	$\leq$ -0.79	448.31	343	426.53	323.67	437.65	339.15	0.31	0.03	0.77
–	$\leq$ 0.00	456.99	343	432.46	323.75	445.97	339.12	0.19	0.02	0.59
5	$\leq$ 0.81	461.56	343	441.07	323.47	452.18	339.09	0.39	0.05	0.78
6	$\leq$ 1.98	448.20	343	418.83	323.76	434.10	339.15	0.06	0.10	0.46
7	$\leq$ 3.35	405.75	343	383.23	323.09	397.65	339.27	0.31	0.07	0.58
8	$\leq$ 4.80	339.29	343	312.82	323.19	330.44	339.02	0.07	0.06	0.34
9	$\leq$ 6.85	218.66	343	197.17	323.84	213.87	339.11	0.32	0.29	0.36

Figure 1

Scatterplot matrix of excess returns and predictors. Sample period: January 1960 to December 1990.

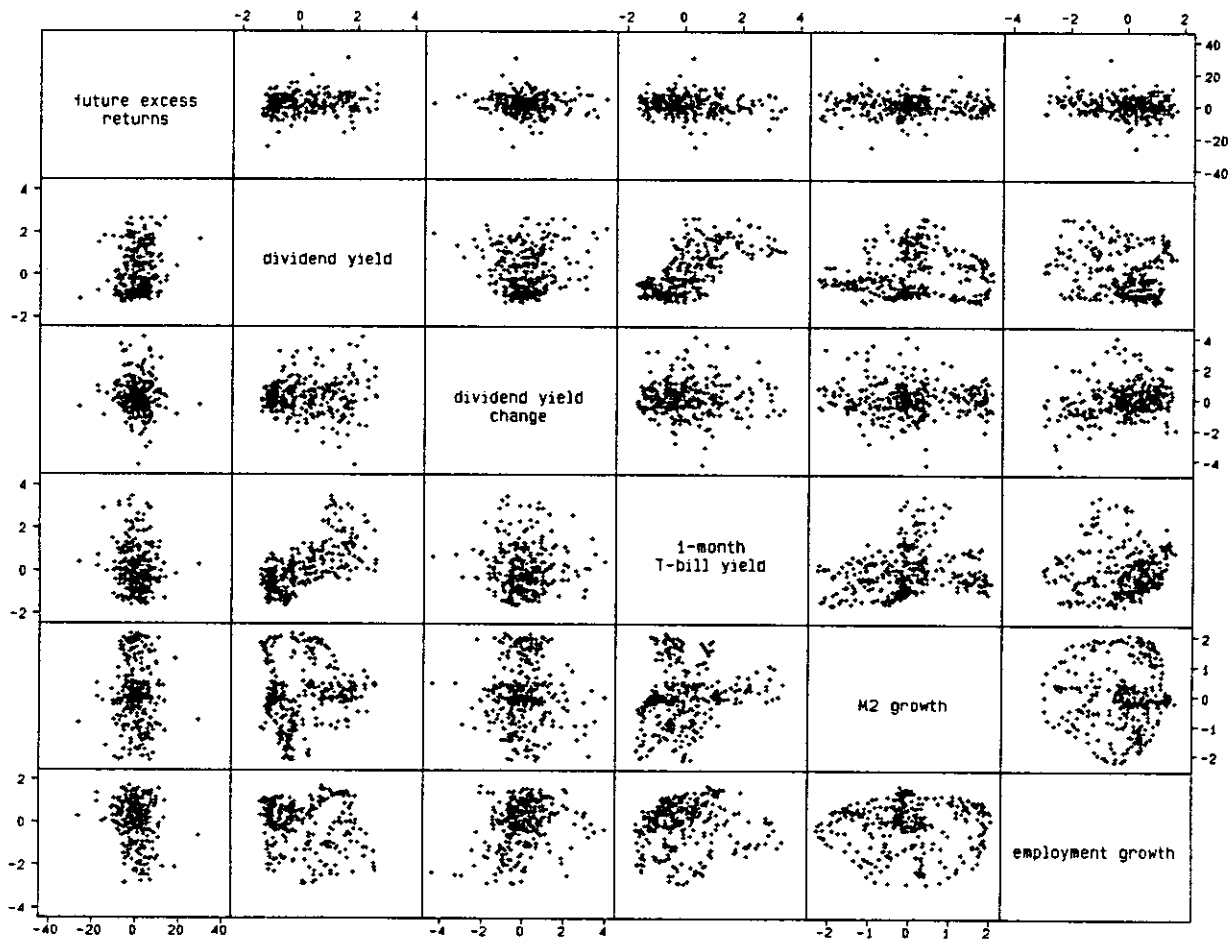

Figure 2

Benchmark case: linear logit model. We estimate ten logit models with dependent variable $Z_{j,t+1} = 1(Y_{t+1} \leq y_j)$, where $y_1, \ldots, y_{10}$ are equal to

$$\{-5.53,\ -3.07,\ -1.71,\ -0.79,\ 0.00,\ 0.81,\ 1.98,\ 3.35,\ 4.80,\ 6.85\}$$

The figure plots the estimates of the ten logit coefficients for each predictor and the constant term, with the associated ± 2 standard-error intervals. The vertical line corresponds to zero excess returns.

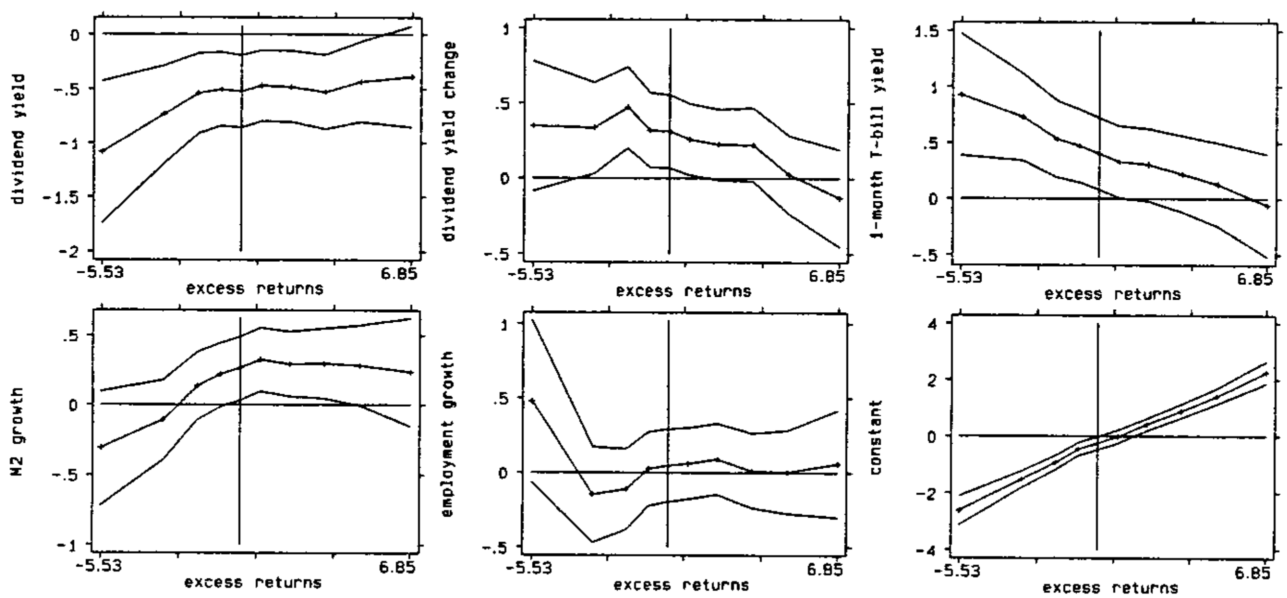

Figure 3.a

Linear logit model: 3D surface of the estimated conditional distribution of excess returns as a function of the dividend yield (**fsdxp**). The dividend yield is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

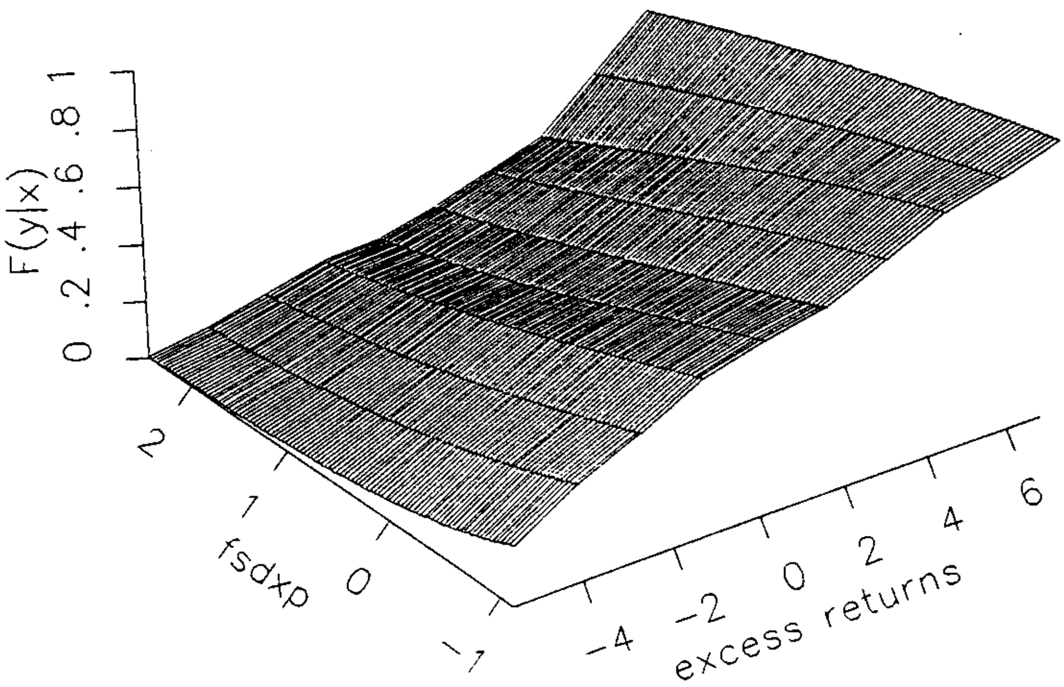

Figure 3.b

Linear logit model: Iso-probability contours of the estimated conditional distribution of excess returns as a function of the dividend yield (**fsdxp**). The dividend yield is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

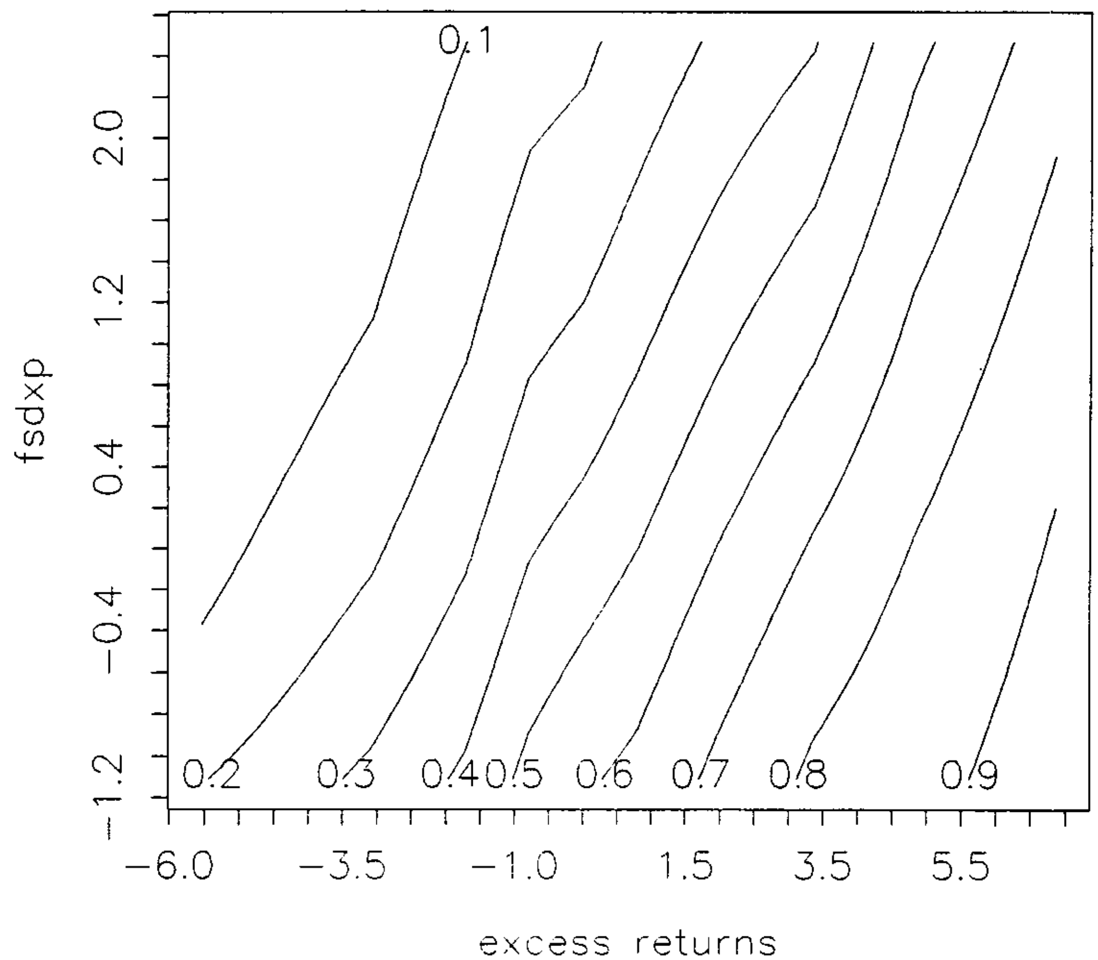

Figure 3.c

Linear logit model: 3D surface of the estimated conditional distribution of excess returns as a function of the one-month Treasury bill rate (**fygm1**). The one-month bill rate is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

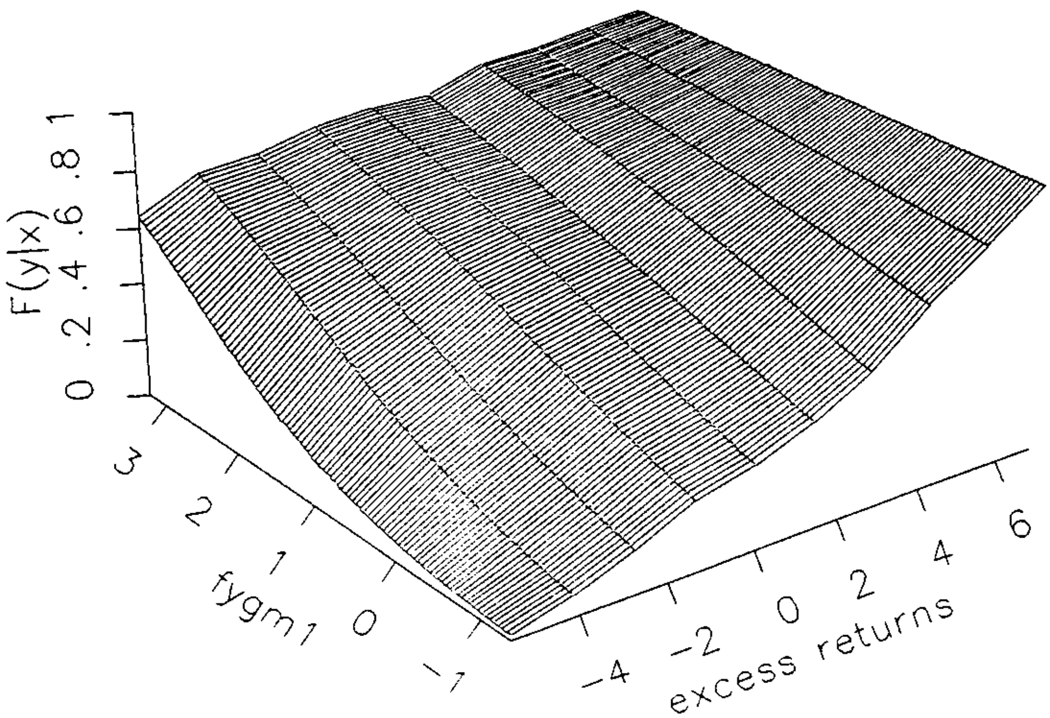

Figure 4.a

Linear logit model: 3D surface of the estimated conditional distribution of excess returns as a function of the money growth rate (**ddfm2**). The money growth rate is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

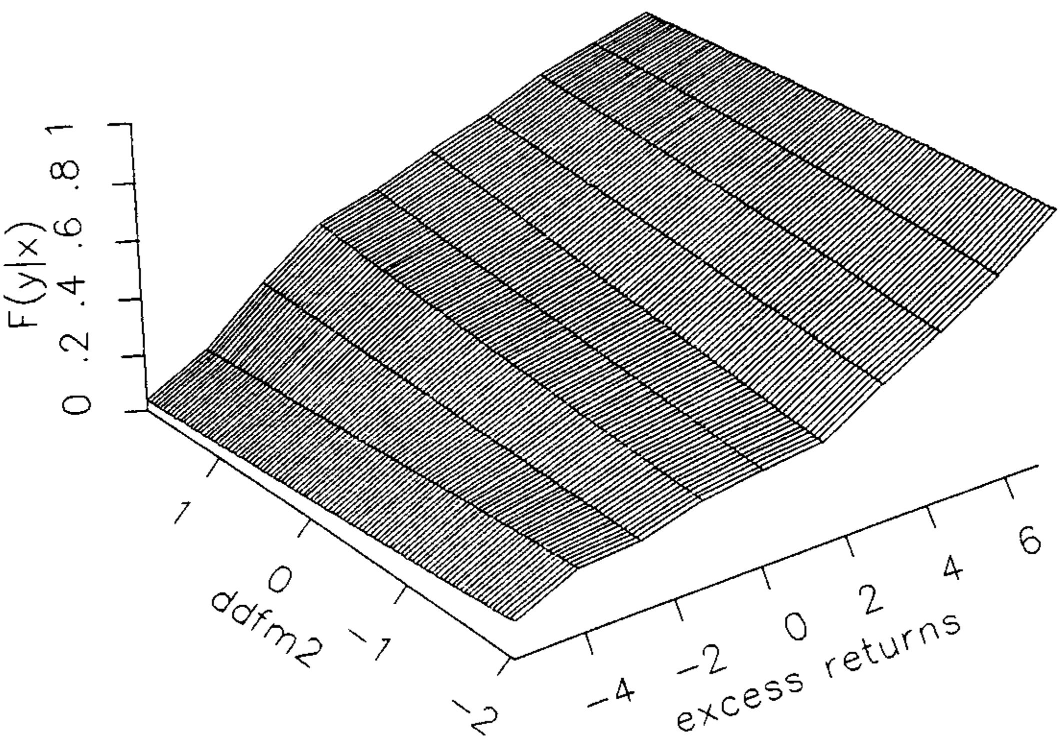

Figure 4.b

Linear logit model: Iso-probability contours of the estimated conditional distribution of excess returns as a function of money growth rate (**ddfm2**). The money growth rate is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

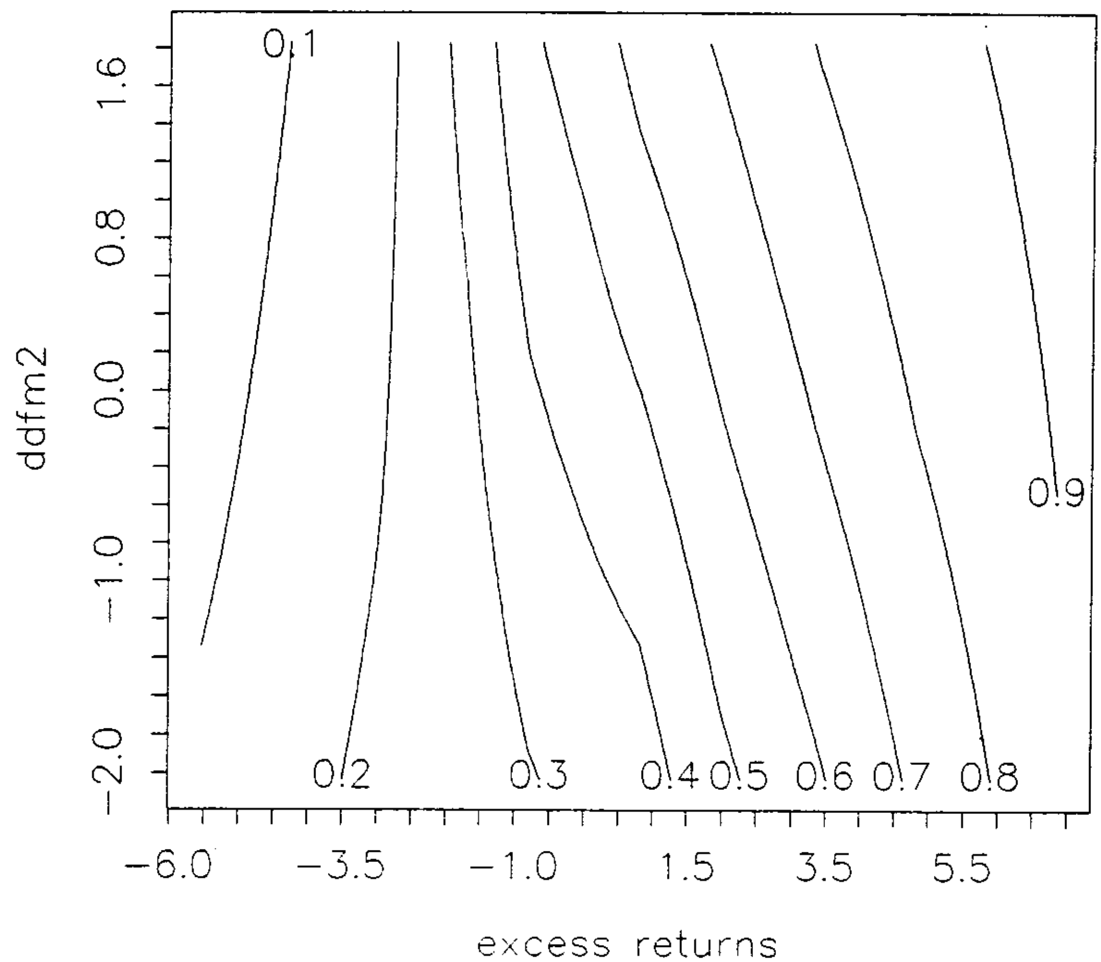

Figure 4.c

Semi-additive logit model: 3D surface of the estimated conditional distribution of excess returns as a function of money growth rate (**ddfm2**). The money growth rate is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

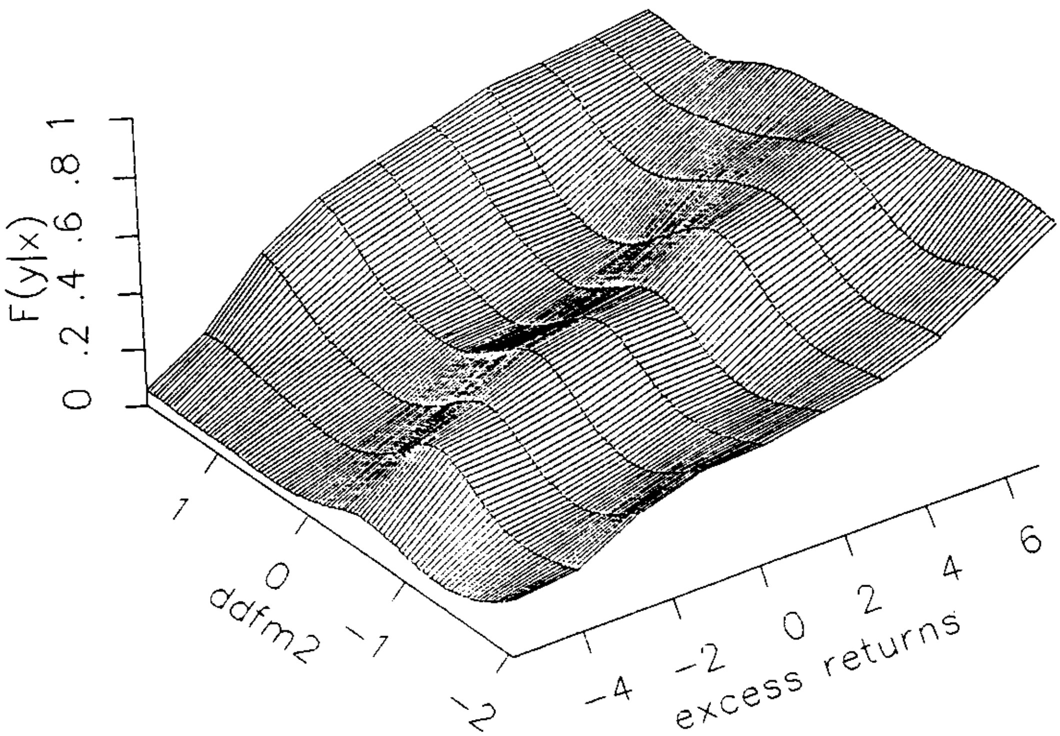

Figure 4.d

Semi-additive logit model: Iso-probability contours of the estimated conditional distribution of excess returns as a function of money growth rate (**ddfm2**). The money growth rate is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

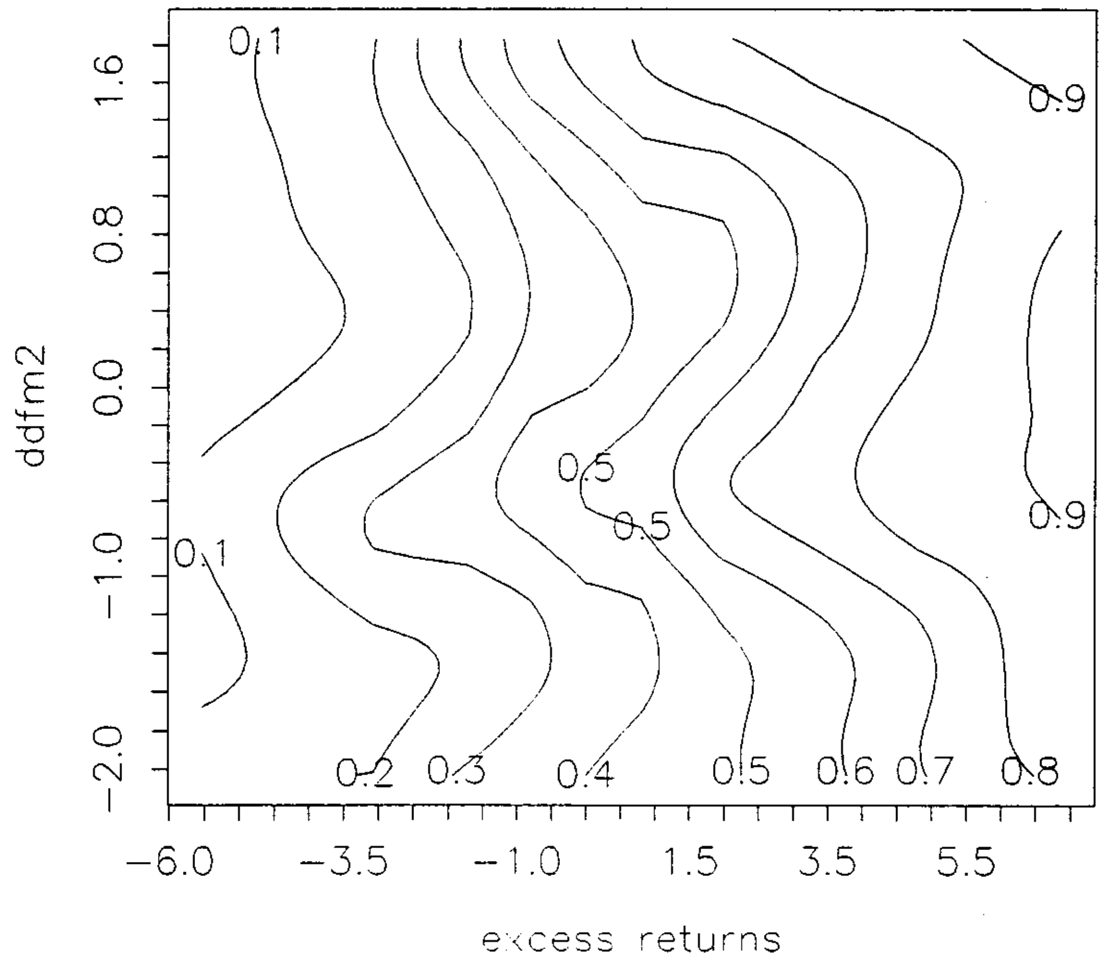

Figure 5.a

Semi-additive logit model: Estimated conditional probability of negative excess returns as a function of money growth rate (**ddfm2**) for the whole sample period. The money growth rate is standardized by subtracting the sample mean and dividing by the sample standard deviation. All other predictors are held constant to their mean value.

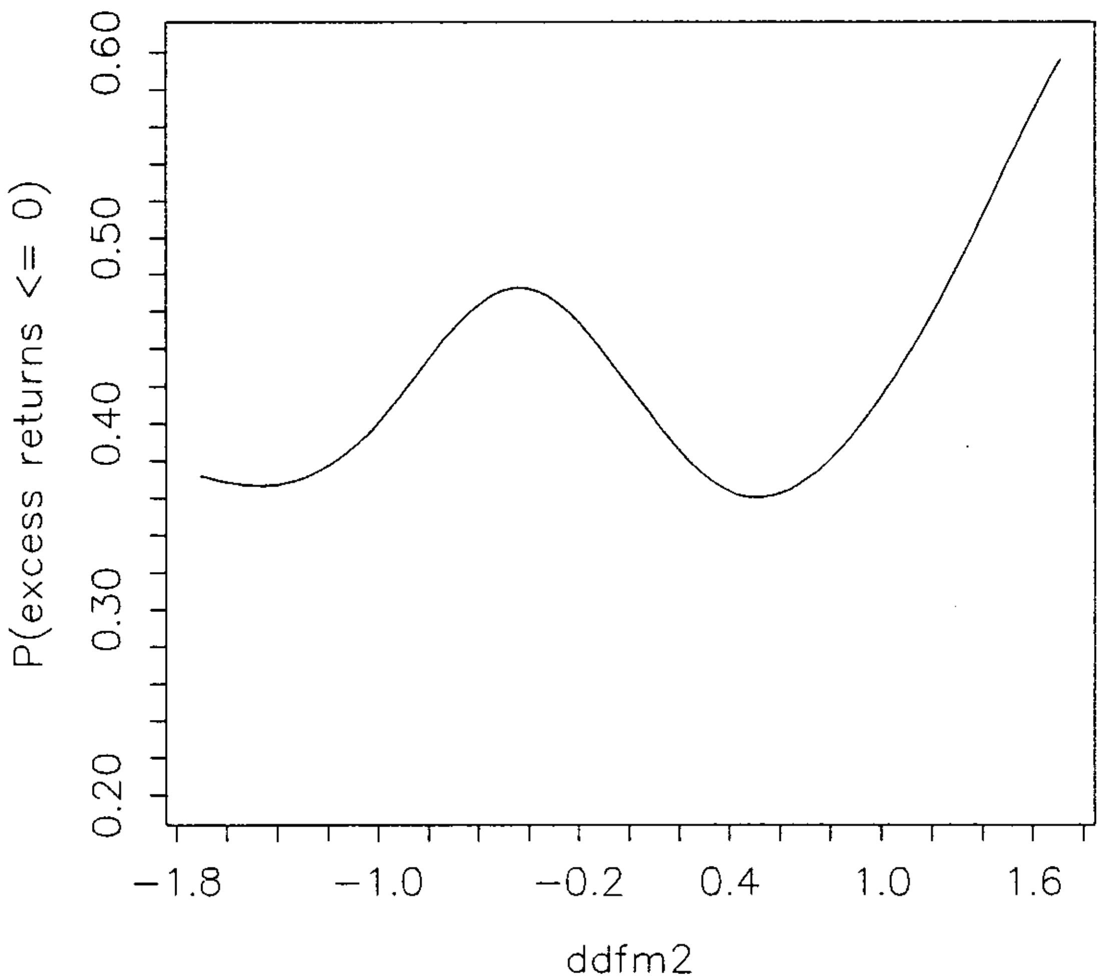

Figure 5.b

Semi-additive logit model: Estimated conditional probability of negative excess returns as a function of money growth rate (**ddfm2**) by subperiod:

1. January 1960 to March 1972
2. April 1972 to October 1982
3. November 1982 to December 1990.

All other predictors are held constant to their mean value. For comparison, the data have been standardized within each subperiod.