COVARIANCE SELECTION AND ESTIMATION AND THE VALUE/GROWTH SPREADS AS PREDICTORS OF RETURNS

NAIPING LIU

A DISSERTATION

IN

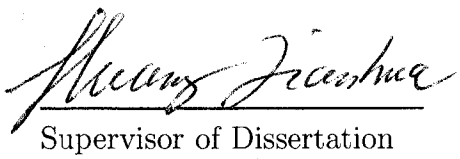
STATISTICS

for the Graduate Group in Managerial Science and Applied Economics

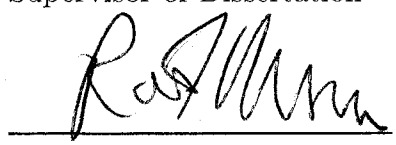
Presented to the Faculties of the University of Pennsylvania

in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2004



Supervisor of Dissertation



Graduate Group Chairperson

ACKNOWLEDGMENTS

I want to express my utmost gratitude to my advisor Jianhua Huang and Professor Lu Zhang for their invaluable guidance in the entire course of this work. It would be impossible for me to complete this dissertation without their help and support.

I would also like to thank my thesis committee members Larry Brown, Ed George, Dylan Small, Yihong Xia. I appreciate very much their insightful comments.

In the end I want to thank the faculties, staff and students in the Statistics Department of the Wharton School for providing such a superb environment for learning and doing research. Especially I want to thank Linda Zhao for her encouragement and advice.

ABSTRACT

COVARIANCE SELECTION AND ESTIMATION AND THE VALUE/GROWTH SPREADS AS PREDICTORS OF RETURNS

Naiping Liu

Jianhua Huang

This thesis is the result of efforts in three separate papers. Due to the nature of each paper, we decide to keep them separately as chapters most of the time.

In the first chapter, we propose a nonparametric and data-driven method to identify parsimony and to exploit any such parsimony to produce a statistically efficient estimator of a large covariance matrix. The approach reparameterizes the covariance matrix through the modified Cholesky decomposition of its inverse. The Cholesky factor is likely to have off-diagonal elements that are zero or close to it. Penalized normal likelihood of the new unconstrained parameters with L_1 and L_2 penalties are shown to be closely related to Tibshirani's (1996) LASSO approach and the ridge regression. Adding either penalty to the likelihood helps produce more stable estimators by introducing shrinkage to the elements in the Cholesky factor, while the L_1 penalty can also effectively identify structural zeros. The maximum penalized likelihood estimator and the sample covariance matrix are compared using simulation.

In the second chapter, we propose another approach to reparameterize the covariance matrix through the modified Cholesky decomposition of its inverse. A great deal of smoothness is observed in the Cholesky factors when the underlying data set is

longitudinal. We use quadratic splines to model the smoothness. This approach provides a simultaneous and unified method of estimation for the mean and covariance of longitudinal data. It also handles missing data naturally.

The third chapter is the result of joint work with Professor Lu Zhang. Recent rational theory predicts that the value spread is countercyclical and should be a positive predictor of future returns, and that the growth spread is procyclical and should be a negative predictor of future returns. From January 1927 to December 2001, the value spread predicts positively, and the growth spread predicts negatively future market excess returns and small firm excess returns. The value spread exhibits clearly countercyclical, and the growth spread exhibits clearly procyclical movements. However, both the cyclical properties and the predictive power of the value and growth spreads are substantially weaker in the postwar sample.

Contents

List of Tables

List of Figures

Chapter 1

Covariance Selection and Estimation via Penalized Normal Likelihood

1.1 Introduction

An estimated covariance matrix is one of the most basic ingredients needed in almost all areas of multivariate analysis and regression-based techniques in statistics such as generalized, linear and mixed models, time series and spatial data analysis. The sample covariance matrix, the most commonly used estimator is known to be positive-definite and unbiased, but highly unstable for large covariance matrices (Stein, 1975; Lin and Perlman, 1985; Wong, Carter and Kohn, 2003; Ledoit and Wolf, 2004). Lately, structured covariance matrices, with few parameters, like the compound sym-

metry, autoregressive of order one, etc. have become popular in longitudinal studies and related areas, though using a structure far from the true covariance could lead to severe bias. Between these two extremes lies an enormous wealth of covariance structures waiting to be tapped into using data-driven methods with the goal of striking a balance between the variance and bias of the covariance estimator.

Developing data-driven methods for covariance models is difficult because the number of unknown elements in the covariance matrix grows quadratically in the size of the matrix (high-dimensionality). Another challenge is how to impose the positive-definiteness constraint. Many existing methods in the literature do not adhere to an important feature of the generalized linear models (GLMs), that is, modeling unconstraint canonical parameter, but rather deal directly with the individual elements of the covariance matrix; for a review of these methods see Diggle and Verbyla (1998); Diggle, Heagerty, Liang and Zeger (2002); Boik (2002); Wong, Carter and Kohn (2003) and Pourahmadi, Daniels and Park (2004). In developing the GLM for covariance matrices, Dempster (1972) was the first to recognize the inverse covariance matrix as the canonical parameter of a multivariate normal distribution. His covariance selection method which identifies zeros in the inverse covariance matrix offers parsimony, but does not guarantee positive-definiteness of the estimator. The positive-definiteness was taken care of by Leonard and Hsu (1992) and Chiu, Leonard and Tsui (1996) who modelled linearly the matrix logarithm of a covariance matrix, and by Pourahmadi (1999, 2000) who considered generalized linear models for covariances using components of the modified Cholesky decomposition of the inverse

covariance matrix whose nonredundant entries are unconstrained and enjoy statistical interpretation as regression coefficients and variances. The latter two methods are parametric in nature.

In this chapter, we develop a nonparametric and data-driven method in the spirit of Dempster's covariance selection to identify parsimony in the covariance matrix through the unit triangular factor T of its modified Cholesky decomposition. The nonredundant entries of the rows of this matrix are the regression coefficients of one variable based on its predecessors so that the unintuitive task of modelling a covariance matrix can be reduced to that of handling several regression models simultaneously (Wu and Pourahmadi, 2003). Thus, familiar ideas for regression modeling such as ridge regression and variable selection can be applied here. Similar to ridge regression, we shrink the off-diagonal elements of T, and as in regression variable selection, since some of the off-diagonal elements of T are likely to be zero or close it, we formally identify any existing structural zeros. We propose to use penalized normal likelihood function with an L_p penalty for the nonredundant entries of T to introduce shrinkage and identify structural zeros. Since the matrix T, in essence, gauges the degrees of "dependence" in the vector of responses, imposing such penalty will curb the appetite for consuming too many parameters to capture the dependence (correlation). We show that many conceptual and computational underpinnings of our procedures are closely related to ridge regression and Tibshirani's (1996) least absolute shrinkage and selection operator (LASSO).

Our approach is flexible and closely related to the recent work of Wu and Pourah-

madi (2003) who applied local polynomial smoothing to the first few subdiagonals of the Cholesky factor and set to zero the remaining subdiagonals, so that the estimated T ends up to be a banded lower triangular matrix. In contrast, our new approach allows the zeros in the Cholesky factor to be irregularly placed, so that varying-order antedependence models (Macchiavelli and Arnold, 1994) could appear as the end result (see Section 7.1). This seems to be an advantage over Wu and Pourahmadi's (2003) approach. However, the new approach does not impose (classical nonparametric) smoothing restrictions on the Cholesky factor, so that when the classical nonparametric smoothing restriction is appropriate for the problem, the approach of Wu and Pourahmadi may work better; in this sense, the two approaches complement each other. Our approach is also related to a Bayesian approach proposed by Smith and Kohn (2002) which places a hierarchical prior to allow zero entries in T. Ledoit and Wolf (2004) considered shrinkage estimation of covariance matrices in a way rather different from us.

The outline of the chapter is as follows. In Section 2, we review the connection between the modified Cholesky decomposition of a covariance matrix and the regression coefficients when a variable is regressed on its predecessors. The equivalence of the penalized normal likelihood with L_1 and L_2 penalties and Tibshirani's (1996) LASSO procedure and ridge regression are presented in Section 3, and exploited in Sections 4 and 5 to develop algorithms to compute the estimator and select the tuning parameter. The performance of the sample covariance matrix and the maximum penalized likelihood estimator are compared in Section 6 via simulations. The new method

is illustrated using two real datasets of moderate (11×11) and large (102×102) sizes, and the latter is used to highlight the role of covariance estimation in solving an important prediction problem arising in the call-center management. Section 8 concludes the chapter.

1.2 Modified Cholesky decomposition

In this section, we review the role of the modified Cholesky decomposition in reparametrizing a covariance matrix in terms of unconstrained regression coefficients and expressing the normal likelihood as a quadratic function of these new parameters (Pourahmadi, 1999). These will serve as the basis for our proposed methodology.

For a positive-definite covariance matrix Σ, its modified Cholesky decomposition can be written as

$$T\Sigma T' = D, \tag{1.1}$$

where T is a unit lower triangular matrix having ones on its diagonal and D is a diagonal matrix. The elements of T and D are uniquely defined and have interpretations as the successive regression coefficients and prediction error variances when measurements are regressed on their predecessors. More precisely, let $y = (y_1, \ldots, y_n)$ be a time-ordered random vector with mean zero and positive-definite covariance matrix Σ. For $1 \leq t \leq n$, let $\hat{y}_t$ stand for the linear least-squares predictor of y_t based on its predecessors $y_{t-1}, \ldots, y_1$, and let $\epsilon_t = y_t - \hat{y}_t$ be its prediction error with variance $\sigma_t^2 = \mathrm{var}(\epsilon_t)$. Thus, for $t = 1$, $\hat{y}_1 = E(y_1) = 0$, and for $1 < t \leq n$, there are unique

scalars ϕ_{tj}'s so that

$$y_t = \sum_{j=1}^{t-1} \phi_{tj} y_j + \epsilon_t. \tag{1.2}$$

Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ be the vector of successive prediction errors. Then, (1.2) written in matrix form becomes

$$\boldsymbol{\epsilon} = Ty, \tag{1.3}$$

where T is a unit lower triangular matrix with $-\phi_{t,j}$ in the (t,j)th position for $2 \leq t \leq n$ and $j = 1, 2, \ldots, t-1$. Note that $\text{cov}(\boldsymbol{\epsilon}) = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) = D$. Since ϵ_t's are uncorrelated, (1) follows from (3) , i.e., the matrix T diagonalizes the covariance matrix Σ as in (1.1). The $\phi_{t,j}$'s are called the *generalized autoregressive parameters* (GARP) and σ_t^2's are the corresponding *innovation (residual) variances.*

Under the multivariate normal assumption on y, the log-likelihood function $\ell(\Sigma; y)$, ignoring an irrelevant constant, satisfies

$$-2\ell(\Sigma; y) = \log|\Sigma| + y'\Sigma^{-1}y$$

Since from (1), $|\Sigma| = |D| = \Pi_{t=1}^n \sigma_t^2$ and $\Sigma^{-1} = T'D^{-1}T$, we have

$$\begin{aligned} -2\ell(\Sigma; y) &= \log|D| + y'T'D^{-1}Ty \\ &= \sum_{t=1}^n \log\sigma_t^2 + \sum_{t=1}^n \frac{\epsilon_t^2}{\sigma_t^2}, \end{aligned} \tag{1.4}$$

which is written in terms of prediction errors and variances or the nonredundant entries of the pair (T, D) .

Thus, modified Cholesky decomposition of a covariance matrix provides a parameterization of the covariance matrix with unconstraint parameters and transfers the difficult task of modeling a covariance matrix to a sequence of simple regression problems. Now, parsimony in the Cholesky factor corresponds to zeros in the regression coefficients, and to identify such zeros is a familiar variable selection problem in regression.

1.3 Penalized likelihood

The regression representation of the modified Cholesky decomposition of a covariance matrix suggests that the familiar ideas of variable selection and regularization for least squares regression can be used for covariance matrix modeling. We explore in this section two such ideas, the ridge regression (Hoerl and Kennard, 1970ab) and LASSO (Tibshirani, 1996), using the general framework of penalized likelihood (Fan and Li, 2001).

As a motivation, consider first the linear regression problem. Suppose that we have data (x_i, y_i), $i = 1, \ldots, n$, where $x_i = (x_{i1}, \ldots, x_{ip})^T$ are the predictor variables and y_i are the responses, and that the effects of the predictor variables on the responses can be summarized by the linear regression model:

$$y_i = x_1\beta_{i1} + \cdots + x_{ip}\beta_p + \epsilon_t, \qquad i = 1, \ldots, n.$$

Ridge regression minimizes

$$\sum_{i=1}^{n}(y_i - \sum_j \beta_j x_{ij})^2 \qquad \text{subject to} \sum_j \beta_j^2 \leq u,$$

where $u \geq 0$ is a tuning parameter. By constraining the L_2 norm of the regression coefficient vector, ridge regression shrinks the regression coefficients towards 0 and yields biased estimates. The hope is that by introducing a small amount of bias, the reduction in variance can be very substantial. Ridge regression has been proved to be quite effective when there are a large number of small regression coefficients. Tibshirani's (1996) LASSO minimizes

$$\sum_{i=1}^{n}(y_i - \sum_j \beta_j x_{ij})^2 \qquad \text{subject to} \sum_j |\beta_j| \leq u.$$

By controlling the L_1 norm of the regression coefficient vector, LASSO introduces shrinkage to the estimates. Because of use of the L_1 norm, LASSO also does variable selection — it can produce coefficients that are exactly 0. LASSO is effective when there are a small to moderate number of moderate-sized regression coefficients. Ridge regression and LASSO correspond to minimization of a penalized residual sum of squares (RSS)

$$\sum_{i=1}^{n}(y_i - \sum_j \beta_j x_{ij})^2 + \lambda \sum_j |\beta_j|^p,$$

with $p = 2$ and $p = 1$, respectively. Here, λ is a tuning parameter. For normal errors, RSS can be viewed as, up to constants, the negative of the log-likelihood, and the

penalized RSS becomes a penalized log-likelihood.

We are now ready to introduce our methods for covariance matrix modeling. Suppose that we observe $y_i = (y_{i1}, \ldots, y_{in})'$, $i = 1, \ldots, m$, a random sample from $N(0, \Sigma)$. Consider the modified Cholesky decomposition of Σ as described in (1.1). According to (1.4), the log-likelihood function $\ell(\Sigma; y_1, \ldots, y_m)$ of Σ based on $y_1, \ldots, y_m$, up to an additive constant, satisfies

$$-2\ell(\Sigma; y_1, \ldots, y_m) = \sum_{t=1}^{n} \left(m \log \sigma_t^2 + \sum_{i=1}^{m} \frac{\epsilon_{it}^2}{\sigma_t^2} \right),$$

where $\epsilon_{i1} = y_{i1}$ and $\epsilon_{it} = y_{it} - \sum_{j=1}^{t-1} y_{ij}\phi_{tj}$ for $t = 2, \ldots, n$. For a given $\lambda > 0$, a form of penalized negative log-likelihood is

$$-2\ell(\Sigma; y_1, \cdots, y_m) + \lambda p(\{\phi_{tj}\}), \tag{1.5}$$

where $p(\cdot) \geq 0$ is a specified penalty function, and λ is a tuning parameter whose selection in practice will be discussed later in the chapter. For fixed λ, minimizing (1.5) with respect to $\{\phi_{tj}\}$ and σ_t^2 leads to a penalized likelihood estimate of T and D and thus estimate of Σ. When $\lambda = 0$, minimization of (1.5) simply gives the maximum likelihood estimate. This modeling framework is quite general, various choices of penalty functions are discussed in detail in Tibshirani (1996) and Fan and Li (2001). We consider in this chapter only the class of penalty functions that can be written as a L_p norm of the GARPs. For $p > 0$, the penalized likelihood objective

function with an L_p penalty has the form

$$-2\ell(\Sigma; y_1, \cdots, y_m) + \lambda \sum_{t=2}^{n} \sum_{j=1}^{t-1} |\phi_{tj}|^p. \tag{1.6}$$

The L_p penalty class has been considered for regression problems by Frank and Friedman (1993) and Fu (1998).

We focus here on two important members of the L_p penalty class: the L_2 penalty $p(\{\phi_{tj}\}) = \sum_{t=2}^{n} \sum_{j=1}^{t-1} \phi_{tj}^2$ and and the L_1 penalty $p(\{\phi_{tj}\}) = \sum_{t=2}^{n} \sum_{j=1}^{t-1} |\phi_{tj}|$. As ridge regression and LASSO, using these two penalties will introduce shrinkage estimates of the GARPs ϕ_{tj} and hence the covariance matrix. The L_1 penalty also does selection by making some GARP estimates to be exactly 0. Similar to least squares regression, the reason for shrinkage and selection is to trade off bias against variance.

Penalized likelihood estimates can be derived as Bayes estimates. Note that ϕ_{tj}^2 and $|\phi_{tj}|$ are respectively proportional to the (minus) log-density of the normal distribution and the double-exponential distribution. As a result, we can derive the penalized likelihood estimates with L_2 penalty as the Bayes posterior mode under independent diffuse prior for the innovation standard deviations σ_t and independent normal priors for the GARP ϕ_{tj}'s,

$$f(\phi_{tj}) = \frac{1}{\sqrt{\pi\tau}} \exp\left(-\frac{\phi_{tj}^2}{\tau}\right)$$

with $\tau = 1/\lambda$. Similarly, the penalized likelihood estimates with L_1 penalty is the

Bayes posterior mode under independent diffuse prior for the innovation standard deviations and independent double-exponential priors for the GARP ϕ_{tj}'s

$$f(\phi_{tj}) = \frac{1}{2\tau} \exp\left(-\frac{|\phi_{tj}|}{\tau}\right).$$

For more information on the Bayesian interpretation of the LASSO estimate, see Tibshirani (1996, Sec.5); Wong et al. (2003) and Smith and Kohn (2002) provide excellent reviews of Bayesian approach to covariance modeling.

1.4 Algorithms for calculating the penalized likelihood estimates

We describe in this section an algorithm for computing the penalized likelihood estimates of the covariance matrix for a given tuning parameter λ. It amounts to applying a similar regression algorithm repeatedly to the rows of the Cholesky factor T.

For the L_p penalty, the penalized negative loglikelihood (1.6) becomes

$$\begin{aligned}
&\sum_{t=1}^{n}\left(m\log\sigma_t^2 + \sum_{i=1}^{m}\frac{\epsilon_{it}^2}{\sigma_t^2}\right) + \lambda\sum_{t=2}^{n}\sum_{j=1}^{t-1}|\phi_{tj}|^p \\
&= \left(m\log\sigma_1^2 + \sum_{i=1}^{m}\frac{\epsilon_{i1}^2}{\sigma_1^2}\right) + \sum_{t=2}^{n}\left(m\log\sigma_t^2 + \sum_{i=1}^{m}\frac{\epsilon_{it}^2}{\sigma_t^2} + \lambda\sum_{j=1}^{t-1}|\phi_{tj}|^p\right).
\end{aligned}$$

To minimize it, we need only to minimize

$$m \log \sigma_1^2 + \sum_{i=1}^{m} \frac{\epsilon_{i1}^2}{\sigma_1^2} \tag{1.7}$$

and

$$m \log \sigma_t^2 + \sum_{i=1}^{m} \frac{\epsilon_{it}^2}{\sigma_t^2} + \lambda \sum_{j=1}^{t-1} |\phi_{tj}|^p, \qquad t = 2, \cdots, n. \tag{1.8}$$

The minimizer of (1.7) is given by $\sigma_1^2 = \sum_{i=1}^{m} y_{i1}^2/m$. For each $t = 2, \ldots, n$, the expression in (1.8) can be minimized by alternating minimization over σ_t and $\phi_{tj}, j = 1, \ldots, t-1$. To be specific, note that for fixed $\phi_{tj}, j = 1, \ldots, t-1$, (1.8) is minimized by

$$\sigma_t^2 = \frac{1}{m} \sum_{i=1}^{m} \epsilon_{it}^2 = \frac{1}{m} \sum_{i=1}^{m} \Big(y_{it} - \sum_{j=1}^{t-1} y_{ij} \phi_{tj} \Big)^2, \tag{1.9}$$

and for fixed σ_t, (1.8), as a function of ϕ_{tj}, $j = 1, \ldots, t-1$, is minimized by the minimizer of

$$\sum_{i=1}^{m} \frac{(y_{it} - \sum_{j=1}^{t-1} y_{ij} \phi_{tj})^2}{\sigma_t^2} + \lambda \sum_{j=1}^{t-1} |\phi_{tj}|^p. \tag{1.10}$$

An iterative procedure for minimizing (1.8) starts by first initializing σ_t, using for example fitted innovation standard error without the penalty. With the current value of σ_t, we minimize (1.10) to get $\phi_{tj}, j = 1, \ldots, t-1$. After getting $\phi_{tj}, j = 1, \cdots, t-1$, we set σ_t^2 as in (1.9). Iterate the process until convergence for each t, $t = 2, \cdots, n$.

We now give some implementation details of minimization of (1.10)for $p = 2$,

which is an important component of our proposed iterative procedure. Let

$$\phi_{t(t)} = \begin{pmatrix} \phi_{t1} \\ \phi_{t2} \\ \vdots \\ \phi_{t,t-1} \end{pmatrix} \quad \text{and} \quad y_{i(t)} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i,t-1} \end{pmatrix}.$$

The first term of (1.10) can be written as

$$\begin{aligned}
&\sum_{i=1}^{m} \frac{(y_{it} - \sum_{j=1}^{t-1} y_{ij}\phi_{tj})^2}{\sigma_t^2} \\
&= \frac{1}{\sigma_t^2} \sum_{i=1}^{m} (y_{it} - y'_{i(t)}\phi_{t(t)})^2 \\
&= \frac{1}{\sigma_t^2} \sum_{i=1}^{m} (y_{it}^2 - 2y_{it}y'_{i(t)}\phi_{t(t)} + \phi'_{t(t)}y_{i(t)}y'_{i(t)}\phi_{t(t)}) \\
&= \frac{1}{\sigma_t^2} \sum_{i=1}^{m} y_{it}^2 - 2\left(\frac{1}{\sigma_t^2} \sum_{i=1}^{m} y_{it}y_{i(t)}\right)' \phi_{t(t)} + \phi'_{t(t)} \left(\frac{1}{\sigma_t^2} \sum_{i=1}^{m} y_{i(t)}y'_{i(t)}\right) \phi_{t(t)} \\
&= c_t - 2g_t'\phi_{t(t)} + \phi'_{t(t)}H_t\phi_{t(t)},
\end{aligned}$$

where $c_t = (\sum_{i=1}^{m} y_{it}^2)/\sigma_t^2$ is a constant, $g_t = (\sum_{i=1}^{m} y_{it}y_{i(t)})/\sigma_t^2$ is a $(t-1) \times 1$ column vector, and $H_t = (\sum_{i=1}^{m} y_{i(t)}y'_{i(t)})/\sigma_t^2$ is a $(t-1) \times (t-1)$ positive definite symmetric matrix.

For the L_2 penalty, minimization of (1.10) has a closed-form solution. In this case,

(1.10) can be written as

$$\sum_{i=1}^{m} \frac{(y_{it} - \sum_{j=1}^{t-1} y_{ij}\phi_{tj})^2}{\sigma_t^2} + \lambda \sum_{j=1}^{t-1} \phi_{tj}^2 = c_t - 2g_t'\phi_{t(t)} + \phi_{t(t)}'H_t\phi_{t(t)} + \lambda\phi_{t(t)}'\phi_{t(t)}$$

$$= c_t - 2g_t'\phi_{t(t)} + \phi_{t(t)}'(H_t + \lambda I)\phi_{t(t)}.$$

We need the fact that $x'Ax - 2b'x$ is minimized by $x = A^{-1}b$ for a $(t-1) \times (t-1)$ positive definite matrix A and a $(t-1) \times 1$ column vector b; this fact follows from the observation that

$$x'Ax - 2b'x = (x - A^{-1}b)'A(x - A^{-1}b) - b'A^{-1}b \geq -b'A^{-1}b,$$

and the equality holds when $x = A^{-1}b$. Thus, for fixed σ_t, the minimizer of (1.10) when $p = 2$ is $\phi_{t(t)} = (H_t + \lambda I_t)^{-1}g_t$, where I_t is the $(t-1) \times (t-1)$ identity matrix.

For the L_1 penalty, minimization of (1.10) does not have a closed-form solution. Note that minimization of (1.10) is equivalent to minimization of

$$\sum_{i=1}^{m} \frac{(y_{it} - \sum_{j=1}^{t-1} y_{ij}\phi_{tj})^2}{\sigma_t^2} \qquad \text{subject to} \sum_{j=1}^{t-1} |\phi_{tj}| \leq u.$$

This is the same optimization problem for LASSO that has been considered for least squares regression. It can be thought of as a quadratic programming problem with linear inequality constraints, so standard numerical techniques could be applied; see Tibshirani (1996). However,we use an iterative algorithm that can be coded directly.

It works well in our simulation study and data analysis. The main idea of the algorithm is an iterative local quadratic approximation of $\sum_{j=1}^{t-1}|\phi_{tj}|$ (Fan and Li, 2001, Öjelund, Madson and Thyregod, 2001). The initial value of the iteration is taken to be the minimizer of (1.10) without the penalty term, that is, $\phi_{t(t)}^{(0)} = H_t^{-1}g_t$ or, when H_t is singular, the minimizer of (1.10) with the L_2 penalty. Denote the value of $\phi_{t(t)}$ at step k of the iteration as $\phi_{t(t)}^{(k)} = (\phi_{t1}^{(k)}, \phi_{t2}^{(k)}, \cdots, \phi_{t,t-1}^{(k)})'$. Since $|\phi_{tj}|$ can be approximated by

$$\frac{|\phi_{tj}^{(k)}|}{2} + \frac{\phi_{tj}^2}{2|\phi_{tj}^{(k)}|}$$

in the neighborhood of $\phi_{tj}^{(k)}$, $\sum_{j=1}^{t-1}|\phi_{tj}|$ can be approximated by

$$\sum_{j=1}^{t-1}\frac{|\phi_{tj}^{(k)}|}{2} + \sum_{j=1}^{t-1}\frac{\phi_{tj}^2}{2|\phi_{tj}^{(k)}|} = c_t^k + \phi_{t(t)}'L_t^k\phi_{t(t)}$$

in the neighborhood of $\phi_{t(t)}^{(k)} = (\phi_{t1}^{(k)}, \phi_{t2}^{(k)}, \cdots, \phi_{t,t-1}^{(k)})'$, where $c_t^k = \sum_{j=1}^{t-1}|\phi_{tj}^{(k)}|/2$ is a constant and

$$L_t^k = \begin{pmatrix} \frac{1}{2|\phi_{t1}^{(k)}|} & 0 & \cdots & 0 \\ 0 & \frac{1}{2|\phi_{t2}^{(k)}|} & \cdots & 0 \\ . & . & . & . & . & . \\ 0 & 0 & \cdots & \frac{1}{2|\phi_{t,t-1}^{(k)}|} \end{pmatrix}$$

is a $(t-1)\times(t-1)$ diagonal matrix. (Note that $|\phi_{i,j}^{(k)}|$ appears in the denominator. When any of $|\phi_{i,j}^{(k)}|$ goes below a preset threshold, typically 10^{-10}, replace it by the

threshold value.) Thus, (1.10) can be approximated by

$$\frac{1}{\sigma_t^2}\sum_{i=1}^{m}(y_{it} - y'_{i(t)}\phi_{t(t)})^2 + \lambda\sum_{j=1}^{t-1}|\phi_{tj}|$$

$$= c_t - 2g'_t\phi_{t(t)} + \phi'_{t(t)}H_t\phi_{t(t)} + \lambda c_t^k + \lambda\phi'_{t(t)}L_c^k\phi_{t(t)}$$

$$= c_t + \lambda c_t^k - 2g'_t\phi_{t(t)} + \phi'_{t(t)}(H_t + \lambda L_t^k)\phi_{t(t)}.$$

Hence, at step $(k+1)$, minimizer of (1.10) for $p = 1$ is $\phi_{t(t)}^{(k+1)} = (H_t + \lambda L_t^k)^{-1}g_t$. Repeat this process until convergence.

1.5 Selection of the tuning parameter

To implement the penalized likelihood method described above, we need to specify the tuning parameter λ. We discuss two approaches: cross-validation and generalized cross-validation (GCV).

The idea of cross-validation is readily applicable. For fast computation, we prefer the K-fold cross-validation to the delete-one-out cross-validation. Typical choices of the number of folds K are five or ten in practical implementation. Denote the full dataset by S. We randomly split the full dataset into K subsets about the same size, denoted as S^ν, $\nu = 1, \ldots, K$. For each ν, we use the data $S - S^\nu$ to estimate the parameters and S^ν to validate. The log-likelihood is used as the performance

measure. For each λ, the K-fold cross-validated log-likelihood criterion is defined as

$$\mathrm{CV}(\lambda) = \frac{1}{K}\sum_{\nu=1}^{K}\left(s_\nu \log|\widehat{\Sigma}_{-\nu}| + \sum_{i\in I_\nu} y_i'\widehat{\Sigma}_{-\nu}^{-1}y_i\right),$$

where I_ν is the index set of data in S^ν, s_ν is the size of I_ν, and $\widehat{\Sigma}_{-\nu}$ is the variance-covariance matrix estimated using the training data set $S - S^\nu$. Note that, for data in S^ν, the expected log-likelihood for variance-covariance Σ is given by

$$E\left(s_\nu \log|\Sigma| + \sum_{i\in I_\nu} y_i'\Sigma^{-1}y_i\right).$$

The function $\mathrm{CV}(\lambda)$ can thus be thought as an estimate of the expected log-likelihood when the tuning parameter is varying. We find the tuning parameter $\hat{\lambda}$ that minimizes $CV(\lambda)$. Our final estimate of Σ is based on $\hat{\lambda}$ and the full dataset.

Following Craven and Wahba (1979), we derive the GCV criterion as an approximation to the delete-one-out cross-validation criterion

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}(y_{it} - \hat{y}_{it}^{(-i)})^2 = \frac{1}{mn}\sum_{t=1}^{n}\sum_{i=1}^{m}(y_{it} - \hat{y}_{it}^{(-i)})^2,$$

where $\hat{y}_{it}^{(-i)}$ are fitted values when the ith vector of observations y_i is removed from

the sample. For $t = 1, \ldots, n$, let

$$X_t = \frac{1}{\sigma_t} \begin{pmatrix} y'_{1(t)} \\ \vdots \\ y'_{m(t)} \end{pmatrix} = \frac{1}{\sigma_t} \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1,t-1} \\ y_{21} & y_{22} & \cdots & y_{2,t-1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{m,t-1} \end{pmatrix}.$$

For the L_2 penalty, using the connection to ridge regression, it is easily seen that

$$\begin{pmatrix} \hat{y}_{1t} \\ \vdots \\ \hat{y}_{mt} \end{pmatrix} = X_t(H_t + \lambda I_t)^{-1} X'_t \begin{pmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{pmatrix} = S_t \begin{pmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{pmatrix},$$

and that

$$y_{it} - \hat{y}_{it}^{(-i)} = \left(\frac{y_{it} - \hat{y}_{it}}{1 - S_{t,ii}} \right)^2,$$

where $S_t = X_t(H_t + \lambda I_t)^{-1} X_t$ with its (i,i)-element being $S_{t,ii}$. Then we approximate $S_{t,ii}$ by $\sum_{i=1}^{m} S_{t,ii}/m = tr(S_t)/m$ in the delete-one-out cross-validation criterion to obtain the following GCV criterion

$$\mathrm{GCV}(\lambda) = \frac{1}{mn} \sum_{t=1}^{n} \sum_{i=1}^{m} \left(\frac{y_{it} - \hat{y}_{it}}{1 - tr(S_t)/m} \right)^2.$$

In the calculation of GCV criterion, σ_t's should be replaced by their estimated values. For the L_1 penalty, there is no closed-form expression that links $(y_{1t}, \ldots, y_{mt})$ to their

predicted values $(\hat{y}_{1t}, \ldots, \hat{y}_{mt})$. Using outcomes from the last iteration of minimization of (1.10), we have approximately that

$$\begin{pmatrix} \hat{y}_{1t} \\ \vdots \\ \hat{y}_{mt} \end{pmatrix} = X_t(H_t + \lambda L_t^{(k)})^{-1} X_t' \begin{pmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{pmatrix}.$$

We thus define the GCV criterion for the L_1 penalty case using the same formula as for the L_2 penalty case except that, in the definition of S_t we replace I_t by the matrix $L_t^{(k)}$ used in the last iteration of minimization of (1.10).

1.6 Simulations

In this section via simulations, we compare the performance of the sample covariance matrix (maximum likelihood estimator) to that of maximum penalized likelihood estimator in estimating the true population covariance matrix. We used our computer code, written in Compaq Visual Fortran 6 (Compaq Computer Corporation), that implements the iterative algorithm described in Section 1.4. IMSL Fortran subroutine `UVMIF` (IMSL Math/Library, Visual Numerics, Inc.) is used for optimization to select the tunining parameter.

To compare the performance of covariance matrix estimators, we consider two loss

functions:

$$Loss_1(\Sigma, G) = tr\Sigma^{-1}G - \log|\Sigma^{-1}G| - n \quad \text{and} \quad Loss_2(\Sigma, G) = tr(\Sigma^{-1}G - I)^2,$$

where Σ is the true covariance matrix and G is a positive definite matrix. The first loss is usually called entropy loss, while the second is typically called quadratic loss. Each of these losses is 0 when $G = \Sigma$ and is positive when $G \neq \Sigma$. Both loss functions are invariant with respect to transformations $G^* = CGC'$, $\Sigma^* = C\Sigma C'$ for a nonsingular matrix C (Anderson 1984, Sec. 7.8). The corresponding risk functions are defined by

$$R_i(\Sigma, G) = E_\Sigma\{Loss_i(\Sigma, G)\}, \qquad i = 1, 2.$$

An estimator $\hat{\Sigma}$ is considered better than the sample covariance matrix S if its risk function is smaller, that is, $R_i(\Sigma, \hat{\Sigma}) < R_i(\Sigma, S)$. For more information on simulation-based comparison of covariance estimators, see Lin and Perlman (1985). The risk function is approximated by Monte Carlo simulation. For the results presented below, $N = 100$ simulation runs are used.

We consider the following four covariance matrices:

- $\Sigma_1 = I$ (the identity matrix);
- $\Sigma_2 = \text{diag}(n, n-1, n-2, \ldots, 1)$ (diagonal matrix);
- $\Sigma_3^{-1} = T'D^{-1}T$, where $D = 0.01 * I$, and $T = (\phi_{t,s})$, $\phi_{t,t} = 1$, $\phi_{t+1,t} = -0.8$, the rest of entries of T are zeros (AR_1 model);

- $\Sigma_4^{-1} = T'D^{-1}T$, where $D = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ with $\sigma_1^2 = \sigma^2$, $\sigma_t^2 = \sigma^2\{1 - \frac{(t-1)\rho^2}{1+(t-1)\rho}\}$, $t \geq 2$, and $T = (\phi_{t,s})$ with $\phi_{t,t} = 1$, $\phi_{t,t-j} = -\rho\{1+(t-1)\rho\}^{-1}$, $t \geq 2$, $j = 1, \ldots, t-1$, and $\rho = 0.5$ (the compound symmetry model).

For each Σ from the above list, we simulate m i.i.d. $N(0, \Sigma)$ n-vectors for (m, n) equals $(40, 5)$, $(40, 15)$, and $(100, 30)$. We compute the sample covariance matrix S and the covariance matrix estimators using the penalized likelihood with the L_2 and L_1 penalties. Both the GCV and 5-fold CV are used in selecting the tuning parameters in the penalized likelihood. For a given loss function $L(\cdot, \cdot)$ and an estimator $\hat{\Sigma}$, the value of $L(\Sigma, \hat{\Sigma})$ is computed. For each simulation setup, we report the sample mean, standard deviation (Tables 1 and 3), median, lower and upper quartiles (Tables 2 and 4) of the calculated losses from the 100 simulation runs. The mean of the losses is a Monte Carlo estimate of the risk.

The simulation results can be summarized as follows:

- The penalized likelihood estimates outperform the sample covariance matrix in almost all cases and in most cases the improvements are substantial; the only cases that the penalized likelihood estimates do not do better than the sample covariance matrices are when $n = 5$ for Σ_3 and Σ_4.

- For the L_2 penalty, performance of penalized likelihood estimate using GCV are similar to that of using 5-fold CV.

- For the L_1 penalty, performance of the penalized likelihood estimate using 5-fold CV is better than that of using GCV for estimating Σ_1 and Σ_2, and comparable

for estimating Σ_3 and Σ_4. This suggests that the approximation involved in deriving the GCV formula for the L_1 penalty case might be too crude.

- In regard to penalized likelihood estimates, the performance of the L_1 and L_2 penalties are comparable for estimating Σ_1 and Σ_2; for Σ_3, where there are many zeroes in the T factor of the covariance matrix, L_1 penalty does better than the L_2 penalty; for Σ_4, where there are many small values in the T factor of the Cholesky decomposition of the covariance matrix, L_2 penalty does better than the L_1 penalty. These are all as expected.

1.7 Real data examples

1.7.1 Cattle data

Kenward's (1987) reports an experiment in which cattle were assigned randomly to two treatment groups A and B, and their weights were recorded to study the effect of treatments on intestinal parasites. The animals were weighed $n = 11$ times over a 133-day period and the data are balanced. No observation was missing. Of 60 cattle $m = 30$ received treatment A and the other 30 received treatment B. Zimmerman and Núñez-Antón (1997) rejected equality of the two within treatment-group covariance matrices using the classical likelihood ratio test. Thus, it is advisable to study each treatment group's covariance matrix separately; here we report our results for the group A cattle.

For estimating the covariance structure of a dataset it is generally believed (Diggle et al. 2002, p.65) that a sensible strategy is to use an over-elaborate or saturated model for the mean response profile. Thus, we apply the proposed penalized likelihood method to identify the 11×11 covariance matrix based on the residuals from an OLS fit of the saturated mean model with $n = 11$ parameters. Penalized likelihood estimates with both the L_2 and L_1 penalty are calculated. The L_1 penalty is preferred to the L_2 penalty based on the 5-fold cross-validation and has helped identify a parsimonious structure of the T matrix in the generalized Cholesky decomposition of Σ.

Table 1.5 gives the estimated factor T using penalized likelihood estimate with the L_1 penalty. Only the lower-triangular elements that are bigger in absolute value than 0.01 are reported. The results suggest that about 1/3 of the 55 lower triangular elements of T are effectively non-zero. Furthermore, most of the non-zero elements appear on the first and second subdiagonals, lending support to an ante-dependence model of order 2 (Macchiavelli and Arnold,1994). To create the results, the tuning parameter $\lambda = 11.84$ was selected using the 5-fold cross-validation. Different partition of the data in the cross-validation yields similar results.

1.7.2 Telephone call center data

Telephone call centers have become an integral part of the operations of many large organizations. With their growing presence and importance in the organization, managing call center operations more efficiently has become an issue of significant eco-

nomic interest (Brown et al. 2002). In the modeling and analysis of call centers with quantitative methods, an important issue is the modeling of external customer demand. In this section we apply the proposed method for analysis of data from one call center that belongs to a major US northeastern financial organization. The original database has detailed information about every call that got connected to this call center during 2002. We are interested in understanding the arrival pattern of calls to the service queue. In this example, we show that covariance matrix modeling can be used for forecasting the call arrival pattern to a call center. Such forecasts are useful for call center management such as staffing and scheduling.

The data that we analyze is derived from the original database, focusing on the information about the time every call arrives to the service queue. For each day in 2002 (except 6 days where the data collecting equipment went out of order), the records of phone calls start from 7:00AM until midnight. We divided the 17-hour period into 102 10-minute intervals, and counted the number of calls arrived to the service queue during each interval. Here the length of the intervals, 10 minutes, is chosen rather subjectively as a way of smoothing the data and for illustration. One nice thing about having a call-by-call database is that we can easily construct counts for intervals of arbitrary length. Since the arrival patterns of weekdays and weekends differ, we focus on weekdays here. Using the singular value decomposition to screen out outliers that include holidays and recording equipment ill-functioning days (Shen and Huang, 2004), we obtain observations for 239 regular days.

Denote the data for day i as $N_i = (N_{i1}, \ldots, N_{i,102})'$, $i = 1, \ldots, 239$, where N_{it} is

the number of calls arriving to the call center for the t-th 10-minute interval for day i. Let $y_{it} = \sqrt{N_{it} + 1/4}$, $i = 1, \ldots, 239$, $t = 1, \ldots, 102$. The square root transformation is used to make data distribution close to normal (Brown et al. 2002). We apply the proposed penalized likelihood method to estimate the 102×102 covariance matrix based on the residuals from an OLS fit of the saturated mean model. The L_1 penalty is preferred to the L_2 penalty based on the 5-fold cross-validation and has helped identify a parsimonious structure of the T matrix in the generalized Cholesky decomposition of covariance matrix. The selected tuning parameters and 5-fold CV values are given in Table 1.6. Of the 5151 elements below the main diagonal of the estimated T matrix, 4144 are essentially zero (absolute value less than 0.01). Several different random partitions of the data for the 5-fold cross-validation have been tried and yield similar results. With a Pentium III PC runing our Fortran code, the computing time for calculating the penalized likelihood estimate including tuning parameter selection using 5-fold CV, is about 20 minutes.

The estimated covariance matrix can be used for forecasting the number of arrivals during later time of a day using arrival patterns in early time of the day. Denote $y_i = (y_{i1}, \ldots, y_{i,102})'$. Form the partition $y_i = (y_i^{1'}, y_i^{2'})'$, where y_i^1 and y_i^2 measure the arrival patterns in the early and later time of day i. For example, we can take $y_i^1 = (y_{i1}, \ldots, y_{i,51})'$ and $y_i^2 = (y_{i,52}, \ldots, y_{i,102})'$, which measure respectively the arrival patterns in the early and later half of a day. The corresponding partition of the mean

and covariance matrix are denoted as

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \Sigma = \begin{pmatrix} \Sigma_{11}, \Sigma_{12} \\ \Sigma_{21}, \Sigma_{22} \end{pmatrix}.$$

Assume multivariate normality, the best mean squared error forecast of y_i^2 using y_i^1 is

$$E(y_i^2|y_i^1) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_i^1 - \mu_1). \tag{1.11}$$

Without the normality assumption, this formula gives the best mean squared error linear forecast. In order to use the formula to get a desired forecast in practice, we need plug in an estimate of μ and Σ. We can fit a saturated mean model for μ and use either the sample covariance matrix or the penalized likelihood covariance matrix estimate as an estimate of Σ.

To compare the forecast performance by using different covariance matrix estimates, we perform the following out-of-sample forecasting exercise. We split the 239 days into two groups — the training and test data sets. The data from the first 205 days (corresponding to January to October) form the training data set that is used to estimate the mean and covariance structure. The estimates are then applied for forecasting using formula (1.11) for the 34 regular days in test set (corresponding to November and December). We used the 51 (square root transformed) arrival counts in the early half of a day to forecast the (square root transformed) arrival counts in the later half of the day. For each 10-minute interval in the later half of the day,

we have the actual observed value and the forecast value for each of the 34 day in the test data set, denote them as y_{it} and $\hat{y}_{it}$, $i = 206, \ldots, 239$, $t = 52, \ldots, 102$. For each $t = 52, \ldots, 102$, define the average absolute forecast error (AAFE), forecast bias (F-Bias), and forecast standard deviation (F-SD) as

$$\text{AAFE} = \frac{1}{34} \sum_{i=206}^{239} |\hat{y}_{it} - y_{it}|,$$

$$\text{F-Bias} = \frac{1}{34} \sum_{i=206}^{239} (\hat{y}_{it} - y_{it}),$$

$$\text{F-SD} = \left\{ \frac{1}{33} \sum_{i=206}^{239} (\hat{y}_{it} - y_{it})^2 \right\}^{1/2}.$$

In Figure 1.1, we plot the AAFE, F-Bias, F-SD for the forecast using the sample covariance matrix and the penalized likelihood covariance matrix estimate. We also plot the percentage of times, among 34 days in the test data set, that the penalized likelihood based forecast has smaller absolute forecast error. The forecast based on penalized likelihood covariance matrix estimates outperforms that based on the sample covariance matrix. Measured using AAFE, F-Bias, or F-SD, the former does better in 50, 46, and 49 out of 51 time intervals (corresponding to later half of a day) where we do the forecast. The percentage of times in 34 test days the former has smaller absolute forecast error exceeding 50% is 46 out of 51 forecast points.

1.8 Discussions

In this chapter we have proposed a nonparametric method for selection and estimation of $n \times n$ covariance matrices which indirectly shrinks the sample covariance matrix. It relies on the modified Cholesky decomposition and reduces the problem to that of dealing with a sequence of regression problems. Parsimony is achieved by shrinking to zero the smaller regression coefficients or the entries of the Cholesky factor using penalized normal likelihood with L_1 and L_2 penalties. This amounts to applying the LASSO and ridge regression procedures almost verbatim n times. For moderate to large n, our simulation study supports the penalized likelihood (shrinkage) estimator relative to the sample covariance matrix. The methodology and the computational procedure are illustrated by applying them to two real datasets of sizes 11×11 and 102×102, the latter is quite large for most existing methods. Our method is ideal for longitudinal studies or situations where the variables are ordered over time or otherwise. Extensions to objective functions other than the normal likelihood and computing the standard errors of the estimates need more research.

Table 1.1 Simulation comparison of methods. Sample, L_2 penalty, and L_1 penalty in the table represent respectively the sample covariance matrix (MLE), the covariance matrix estimate using the penalized likelihood with L_2 and L_1 penalties. GCV and 5-fold CV denote the method used for selecting the tuning parameters in the penalized likelihood. Reported are the mean and standard deviation (in the parenthesis) of observed losses in 100 simulation runs, using the entropy loss.

	m	n	Sample	L_2 Penalty		L_1 Penalty	
				GCV	5-fold CV	GCV	5-fold CV
	40	5	0.417	0.162	0.213	0.212	0.215
			(0.176)	(0.111)	(0.161)	(0.129)	(0.154)
Σ_1	40	15	3.682	0.471	0.465	1.381	0.545
			(0.493)	(0.208)	(0.218)	(0.380)	(0.349)
	100	30	5.237	0.380	0.384	1.572	0.353
			(0.331)	(0.090)	(0.091)	(0.210)	(0.135)
	40	5	0.401	0.148	0.200	0.210	0.212
			(0.148)	(0.093)	(0.151)	(0.120)	(0.153)
Σ_2	40	15	3.603	0.536	0.506	1.612	0.476
			(0.512)	(0.258)	(0.189)	(0.424)	(0.234)
	100	30	5.265	0.785	0.785	2.089	0.336
			(0.388)	(0.130)	(0.130)	(0.301)	(0.113)
	40	5	0.398	0.356	0.437	0.308	0.405
			(0.142)	(0.121)	(0.196)	(0.125)	(0.248)
Σ_3	40	15	3.584	2.468	2.806	1.662	1.470
			(0.452)	(0.344)	(0.882)	(0.362)	(0.416)
	100	30	5.245	3.619	3.768	1.892	1.239
			(0.354)	(0.270)	(0.385)	(0.283)	(0.215)
	40	5	0.379	0.293	0.372	0.361	0.453
			(0.141)	(0.119)	(0.164)	(0.122)	(0.204)
Σ_4	40	15	3.505	1.378	1.357	1.977	2.067
			(0.447)	(0.282)	(0.342)	(0.314)	(0.441)
	100	30	5.234	1.571	1.444	2.475	2.426
			(0.339)	(0.161)	(0.212)	(0.190)	(0.206)

Table 1.2 Simulation comparison of methods. Sample, L_2 penalty, and L_1 penalty in the table represent respectively the sample covariance matrix (MLE), the covariance matrix estimate using the penalized likelihood with L_2 and L_1 penalties. GCV and 5-fold CV denote the method used for selecting the tuning parameters in the penalized likelihood. Reported are the median and lower and upper quartiles (in the parenthesis) of the calculated losses for 100 simulation runs, using the entropy loss.

	m	n	Sample	L_2 Penalty		L_1 Penalty	
				GCV	5-fold CV	GCV	5-fold CV
	40	5	0.369	0.131	0.167	0.172	0.192
			(0.281, 0.523)	(0.072, 0.225)	(0.087, 0.287)	(0.119, 0.281)	(0.091, 0.287)
Σ_1	40	15	3.726	0.423	0.421	1.365	0.471
			(3.318, 4.041)	(0.320, 0.582)	(0.302, 0.552)	(1.097, 1.579)	(0.345, 0.600)
	100	30	5.243	0.378	0.382	1.594	0.335
			(5.023, 5.449)	(0.320, 0.435)	(0.323, 0.460)	(1.420, 1.730)	(0.268, 0.393)
	40	5	0.382	0.112	0.152	0.178	0.170
			(0.290, 0.504)	(0.076, 0.227)	(0.084, 0.264)	(0.120, 0.287)	(0.095, 0.276)
Σ_2	40	15	3.492	0.473	0.475	1.615	0.422
			(3.257, 3.904)	(0.351, 0.662)	(0.361, 0.641)	(1.251, 1.865)	(0.311, 0.589)
	100	30	5.259	0.783	0.783	2.073	0.320
			(4.981, 5.524)	(0.694, 0.848)	(0.694, 0.848)	(1.885, 2.267)	(0.263, 0.395)
	40	5	0.379	0.360	0.405	0.297	0.360
			(0.302, 0.480)	(0.261, 0.435)	(0.300, 0.543)	(0.212, 0.401)	(0.248, 0.458)
Σ_3	40	15	3.627	2.451	2.625	1.676	1.442
			(3.254, 3.903)	(2.286, 2.657)	(2.359, 3.020)	(1.439, 1.952)	(1.184, 1.707)
	100	30	5.241	3.598	3.705	1.873	1.214
			(5.014, 5.501)	(3.463, 3.781)	(3.566, 3.901)	(1.711, 2.044)	(1.095, 1.406)
	40	5	0.352	0.263	0.342	0.331	0.412
			(0.285, 0.451)	(0.207, 0.370)	(0.253, 0.447)	(0.276, 0.434)	(0.312, 0.567)
Σ_4	40	15	3.519	1.376	1.281	1.984	1.964
			(3.250, 3.780)	(1.174, 1.528)	(1.134, 1.539)	(1.772, 2.161)	(1.789, 2.242)
	100	30	5.246	1.540	1.414	2.467	2.396
			(5.077, 5.436)	(1.475, 1.661)	(1.313, 1.523)	(2.353, 2.585)	(2.294, 2.527)

Table 1.3 Simulation comparison of methods. Sample, L_2 penalty, and L_1 penalty in the table represent respectively the sample covariance matrix (MLE), the covariance matrix estimate using the penalized likelihood with L_2 and L_1 penalties. GCV and 5-fold CV denote the method used for selecting the tuning parameters in the penalized likelihood. Reported are the mean and standard deviation (in the parenthesis) of observed losses in 100 simulation runs, using the quadratic loss.

	m	n	Sample	L_2 Penalty		L_1 Penalty	
				GCV	5-fold CV	GCV	5-fold CV
	40	5	0.841	0.334	0.437	0.429	0.442
			(0.438)	(0.258)	(0.358)	(0.302)	(0.350)
Σ_1	40	15	6.422	0.892	0.887	2.330	1.020
			(1.190)	(0.368)	(0.380)	(0.619)	(0.546)
	100	30	9.311	0.720	0.727	2.750	0.688
			(0.762)	(0.163)	(0.166)	(0.360)	(0.241)
	40	5	0.758	0.288	0.391	0.393	0.409
			(0.326)	(0.192)	(0.321)	(0.236)	(0.319)
Σ_2	40	15	6.299	0.982	0.941	2.653	0.918
			(1.228)	(0.439)	(0.357)	(0.707)	(0.434)
	100	30	9.452	1.382	1.382	3.563	0.658
			(0.797)	(0.215)	(0.215)	(0.491)	(0.212)
	40	5	0.745	0.666	0.886	0.568	0.844
			(0.293)	(0.242)	(0.531)	(0.230)	(0.756)
Σ_3	40	15	6.127	4.304	6.064	2.792	2.695
			(1.171)	(0.745)	(3.711)	(0.644)	(0.867)
	100	30	9.365	6.570	7.285	3.311	2.224
			(0.773)	(0.594)	(1.212)	(0.483)	(0.388)
	40	5	0.751	0.573	0.744	0.728	0.937
			(0.328)	(0.275)	(0.417)	(0.303)	(0.547)
Σ_4	40	15	6.052	2.365	2.391	3.587	4.029
			(1.130)	(0.527)	(0.835)	(0.685)	(1.459)
	100	30	9.280	2.729	2.452	4.592	4.522
			(0.833)	(0.327)	(0.424)	(0.437)	(0.471)

Table 1.4 Simulation comparison of methods. Sample, L_2 penalty, and L_1 penalty in the table represent respectively the sample covariance matrix (MLE), the covariance matrix estimate using the penalized likelihood with L_2 and L_1 penalties. GCV and 5-fold CV denote the method used for selecting the tuning parameters in the penalized likelihood. Reported are the median and lower and upper quartiles (in the parenthesis) of the calculated losses for 100 simulation runs, using the quadratic loss.

	m	n	Sample	L_2 Penalty		L_1 Penalty	
				GCV	5-fold CV	GCV	5-fold CV
	40	5	0.752	0.291	0.353	0.360	0.356
			(0.518, 1.000)	(0.139, 0.451)	(0.172, 0.531)	(0.219, 0.548)	(0.197, 0.547)
Σ_1	40	15	6.365	0.818	0.822	2.245	0.909
			(5.459, 7.212)	(0.609, 1.133)	(0.582, 1.096)	(1.935, 2.693)	(0.639, 1.217)
	100	30	9.255	0.731	0.734	2.763	0.655
			(8.792, 9.790)	(0.610, 0.839)	(0.615, 0.838)	(2.514, 2.944)	(0.536, 0.751)
	40	5	0.709	0.226	0.281	0.339	0.331
			(0.511, 0.918)	(0.142, 0.412)	(0.162, 0.480)	(0.209, 0.509)	(0.181, 0.565)
Σ_2	40	15	6.130	0.873	0.892	2.598	0.832
			(5.464, 7.108)	(0.640, 1.264)	(0.671, 1.138)	(2.210, 2.977)	(0.587, 1.131)
	100	30	9.386	1.374	1.374	3.555	0.641
			(8.856, 9.994)	(1.256, 1.508)	(1.256, 1.508)	(3.251, 3.866)	(0.520, 0.766)
	40	5	0.710	0.616	0.737	0.540	0.640
			(0.589, 0.849)	(0.509, 0.793)	(0.573, 1.021)	(0.395, 0.705)	(0.493, 0.879)
Σ_3	40	15	5.963	4.153	5.001	2.737	2.481
			(5.196, 6.763)	(3.811, 4.813)	(4.194, 6.891)	(2.360, 3.168)	(2.045, 3.198)
	100	30	9.316	6.509	7.066	3.242	2.163
			(8.770, 9.882)	(6.192, 6.913)	(6.546, 7.726)	(2.969, 3.598)	(1.980, 2.499)
	40	5	0.680	0.509	0.638	0.656	0.782
			(0.525, 0.927)	(0.369, 0.694)	(0.465, 0.916)	(0.519, 0.865)	(0.578, 1.080)
Σ_4	40	15	5.936	2.283	2.223	3.481	3.751
			(5.127, 6.691)	(2.027, 2.645)	(1.877, 2.686)	(3.019, 4.051)	(3.163, 4.258)
	100	30	9.157	2.718	2.408	4.578	4.511
			(8.734, 9.827)	(2.492, 2.900)	(2.131, 2.701)	(4.295, 4.849)	(4.146, 4.838)

Table 1.5 Estimated Cholesky factor T for the cattle data.

Column number										
1	2	3	4	5	6	7	8	9	10	11
1										
-0.90	1									
-0.01	-0.89	1								
	-0.01	-0.94	1							
		-0.04	-1.01	1						
			-0.23	-0.81	1					
					-0.94	1				
				-0.05	-0.55	-0.43	1			
-0.02			0.15			-0.30	-0.85			
							-0.15	-0.90	1	
-0.02					0.12			-0.23	-0.83	1

Table 1.6 Call center data: selection of tuning parameters using 5-fold CV.

	λ	5-fold CV
L_1 penalty	72.97	579.45
L_2 penalty	367.49	1370.40

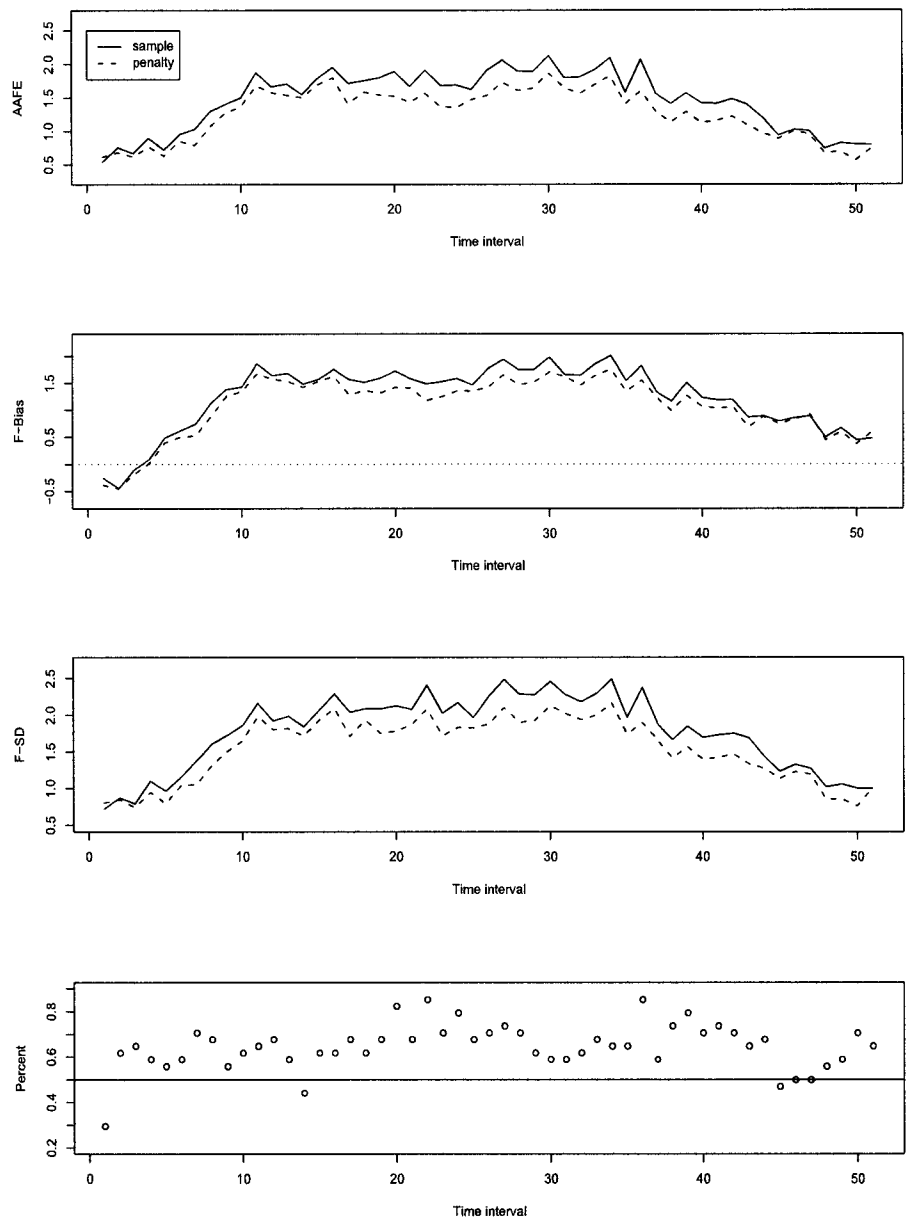


Figure 1.1 The first three graphs show the AAFE, F-Bias, F-SD for the forecast using the sample covariance matrix and the penalized likelihood covariance matrix estimate. The fourth graph shows the percentage of times, among 34 days in the test data set, that the penalized likelihood based forecast has smaller absolute forecast error.

Chapter 2

Basis Function Approximations and Nonparametric Estimation of Large Covariance Matrices

2.1 Introduction

While considerable effort is expended on the nonparametric estimation of the mean function of a longitudinal data set (Hart and Wehrly, 1986; Diggle et al. 1994; Huang et al. 2002), such effort for estimation of the covariance matrix of the data is virtually nonexistent except for the work of Chen (1994) and Diggle and Verbyla (1998), where the estimated covariance matrix may not be positive-definite. Recently, using an unconstrained reparametrization of a covariance matrix and Fan and Zhang's (2000) two-step nonparametric estimation procedure for the mean of functional lin-

ear models, Wu and Pourahmadi (2003) have proposed nonparametric estimators of covariance matrices that are guaranteed to be positive-definite. For smoothing local polynomial estimators (Fan and Gijbles, 1996) are applied to the subdiagonals of the unit lower triangular matrix obtained from the modified Cholesky decomposition of the covariance matrix (Pourahmadi, 1999).

An improvement of the Fan and Zhang's (2000) two-step procedure proposed by Huang, Wu and Zhou (2000) relies on basis function approximations (such as regression splines) to estimate the mean of functional data, it has better capabilities for handling repeated measurements made at irregular time points for different subjects in a longitudinal study. In this chapter, we apply Huang et al.'s (2002) technique to the nonparametric estimation of the covariance function. Some of the advantages of this approach compared to that in Wu and Pourahmadi (2002) are:

- Simultaneous and unified method of (nonparametric) estimation for the mean and covariance of longitudinal data. The simultaneous maximum likelihood estimation methodology is similar to Pourahmadi (2000).

- Use of (abstract) basis functions as covariates for modeling covariance matrices. This is important because identifying relevant covariates for modeling the mean is guided by our intuition and scatterplots. Such devices for identifying covariates for covariance matrices are virtually nonexistent.

- Improved capabilities for handling missing values. Since we are in the framework of maximum likelihood, EM algorithm can be developed to handle missing data.

2.2 The method

Our method relies on a key result (Newton, 1988, p.359) that a symmetric matrix Σ is positive definite if and only if there exists a unique unit lower triangular matrix T, with 1's as diagonal entries, and a unique diagonal matrix D with positive diagonal entries such that

$$\Sigma^{-1} = T'D^{-1}T \text{ or } T\Sigma T' = D.$$

Pourahmadi (1999) called the above decomposition as the modified Cholesky decomposition and discussed the computation and statistical interpretation of T and D. There is an attractive feature of this covariance matrix parameterization, that is, unlike the entries of a covariance matrix, the subdiagonal entries of T, are not constrained in any way. If we allow these entries to depend on some known covariates, we could greatly reduce the number of parameters need to be estimated, which is very useful in estimation of big covariance matrix with little data. In this chapter we explore the connections between different entries of T when the data set is a longitudinal data set. We utilize these connections to build covariates with splines.

Suppose $Y_i \sim N(X\alpha, \Sigma)$, where $i = 1, \cdots, n$, and Y_i is of dimension p. Denote by S the "sample" covariance matrix: $S = S(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X\alpha)(Y_i - X\alpha)'$, where the data are centered by the unknown mean vector $X\alpha$. Then the log-likelihood is

$$l(\Sigma, \alpha) = -\frac{n}{2}\log|\Sigma| - \frac{n}{2}tr(\Sigma^{-1}S(\alpha)),$$

up to an additive constant that can be neglected. The maximization of log-likelihood can proceed by iterating between maximizing over α and Σ, using

$$-\frac{1}{2}\,\mathrm{tr}\left(\hat{\Sigma}^{-1}\sum_{i=1}^{n}[(Y_i - X\alpha)(Y_i - X\alpha)']\right)$$

for α (which results in the generalized least squares estimate $\hat{\alpha}$ for α) and

$$-\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}\hat{S}) \tag{2.1}$$

for the Σ, where $\hat{S} = \sum_{i=1}^{n}[(Y_i - X\hat{\alpha})(Y_i - X\hat{\alpha})']$.

To impose some structure on the covariance matrix Σ, consider the generalized Cholesky decomposition (Pourahmadi, 1999)

$$\Sigma^{-1} = T'D^{-1}T \text{ or } T\Sigma T' = D,$$

where T is a lower-triangular matrix with 1's on the diagonal and $-\phi_{t,j}$ in the (t,j)th position for $2 \leq t \leq n$ and $j = 1, 2, \ldots, t-1$, and $D = \mathrm{diag}(d_1^2, \cdots, d_p^2)$. We can rewrite (2.1) as

$$-\frac{n}{2}\sum_{t=1}^{p}\log d_t^2 - \frac{1}{2}\,\mathrm{tr}(\Sigma^{-1}\hat{S}). \tag{2.2}$$

To maximize over the parameters of Σ, we iterate between maximizing over the in-

novation variances, d_t^2 and the GARP parameters, ϕ_{tj}. For the former, note that

$$\mathrm{tr}(\Sigma^{-1}\hat{S}) = \mathrm{tr}(T'D^{-1}T\hat{S}) = \mathrm{tr}(D^{-1}T\hat{S}T').$$

Define $G = T\hat{S}T'$, with (j,k)th element g_{jk}. Then (2.2) becomes

$$-\frac{n}{2}\sum_{t=1}^{p}\log d_t^2 - \frac{1}{2}\sum_{t=1}^{p}\frac{g_{tt}}{d_t^2} = -\frac{1}{2}\sum_{t=1}^{p}\Big(n\log d_t^2 + \frac{g_{tt}}{d_t^2}\Big). \tag{2.3}$$

Suppose $\log d_t^2 = z_t'\lambda$. Thus $d_t^2 = \exp(z_t'\lambda)$. Then the first derivative of (2.3) relative to λ is

$$g(\lambda) = -\frac{1}{2}\sum_{t=1}^{p}\Big(nz_t - \frac{g_{tt}}{d_t^2}z_t\Big) = -\frac{1}{2}\sum_{t=1}^{p}\Big(n - \frac{g_{tt}}{d_t^2}\Big)z_t.$$

The second derivative of (2.3) relative to λ is

$$H(\lambda) = \frac{1}{2}\sum_{t=1}^{p}\Big(-\frac{g_{tt}}{d_t^2}z_tz_t'\Big) = -\frac{1}{2}\sum_{t=1}^{p}\frac{g_{tt}}{d_t^2}z_tz_t'.$$

Thus using the Newton's method, we get the updating equation

$$\lambda_{l+1} = \lambda_l - H(\lambda_l)^{-1}g(\lambda_l).$$

For the GARP parameters, we first denote the (j,k)th element of $\hat{S}$ as s_{jk}. The

relevant pieces of the log likelihood with respect to the GARP parameters, can be written as

$$-\sum_{t=2}^{p}\frac{1}{d_t^2}\sum_{j=1}^{t}\sum_{k=1}^{t}\phi_{tj}\phi_{tk}s_{jk}, \tag{2.4}$$

where $\phi_{tt}=-1$. Suppose $\phi_{tj}=z_{tj}'\delta$. Set the first derivative of (2.4) to 0, we have

$$-\sum_{t=2}^{p}\frac{1}{d_t^2}\left\{\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}s_{jk}(z_{tj}z_{tk}'\delta+z_{tk}z_{tj}'\delta)+(-1)\sum_{j=1}^{t-1}s_{jt}z_{tj}+(-1)\sum_{k=1}^{t-1}s_{tk}z_{tk}\right\}=0 \tag{2.5}$$

This implies that $\delta=A^{-1}b$, where

$$A=\sum_{t=2}^{p}\frac{1}{d_t^2}\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}s_{jk}(z_{tj}z_{tk}'+z_{tk}z_{tj}')$$

and

$$b=\sum_{t=2}^{p}\frac{2}{d_t^2}\sum_{j=1}^{t-1}s_{tj}z_{tj}.$$

In the above discussion, we have allow the mean, innovation variance and GARP parameters to depend on some covariates: $\mu_t=x_t'\alpha$, $\log d_t^2=z_t'\lambda$ and $\phi_{tj}=z_{tj}'\delta$. We have considered generating these covariates using B-spline function. In particular, we use a quadratic splines to model the dependence on t of the mean μ_t and the log of innovation variances $\log d_t^2$. The number of basis of the quadratic splines are set to be tuning parameters. The parameterization of ϕ_{tj} is much more complex. In longitudinal data, a great deal of smoothness is observed along the sub-diagonals; in addition, the further away from the main diagonal, the smaller the magnitude of

ϕ_{tj}. This is especially true for ante-dependence models AD(p), where $\phi_{j+s,j}$ is zero when $s > p$. We model each subdiagonal of the T matrix as a spline. We use two tuning parameters to model ϕ_{tj}. The first is the number of sub-diagonals that are not zero. The second is the number of basis of the quadratic spline used to model each sub-diagonal. For computational simplicity, different sub-diagonals are set to have the same number of basis of splines. All these tuning parameters are selected using BIC, defined as

$$BIC = -\frac{2}{n}l + q\frac{\log n}{n}, \tag{2.6}$$

where n is the sample size, l is the maximized loglikelihood for a covariance model and q is the number of free parameters. A smaller value of BIC is associated with a better fitting model.

2.3 Incomplete data and the EM algorithm

Since spline smoothing as we discussed in the previous section is in a parametric framework. It is natural to apply the Expectation-Maximization (EM) algorithm to compute the MLE when there are missing data.

Suppose $Y_i \sim N(X\alpha, \Sigma)$, where $i = 1, \cdots, n$, and Y_i is of dimension p. Denote by S the "sample" covariance matrix: $S = S(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X\alpha)(Y_i - X\alpha)'$, where the data are centered by the unknown mean vector $X\alpha$. Then the log-likelihood is

$$l(\Sigma, \alpha) = -\frac{n}{2}\log|\Sigma| - \frac{n}{2}tr(\Sigma^{-1}S(\alpha)),$$

up to an additive constant that can be neglected. Suppose some data are missing at random. We now give the EM algorithm for computing the MLE.

Expectation step:

Let $\Theta = (\Sigma, \alpha)$. We have $l(\Sigma, \alpha) = l(\Theta) = l(Y_{obs}, Y_{mis}, \Theta)$. We need to calculate

$$E[l(Y_{obs}, Y_{mis}, \Theta)|Y_{obs}, \Theta^{(\tau-1)}], \tag{2.7}$$

where $\Theta^{(\tau-1)}$ are the current parameters estimates that we used to evaluate the expectation. Notice that

$$\begin{aligned} & E[Y_{i,mis}{Y_{i,mis}}'|Y_{i,obs}, \Theta^{(\tau-1)}] \\ = \; & E[Y_{i,mis}|Y_{i,obs}, \Theta^{(\tau-1)}]E[Y_{i,mis}|Y_{i,obs}, \Theta^{(\tau-1)}]' + C_i, \end{aligned}$$

where $C_i = Var[Y_{i,mis}|Y_{i,obs}, \Theta^{(\tau-1)}]$. Therefore,

$$\begin{aligned} & E[nS(\alpha)|Y_{obs}, \Theta^{(\tau-1)}] \\ & = E\left[\sum_{i=1}^{n}(Y_i - X\alpha)(Y_i - X\alpha)'|Y_{obs}, \Theta^{(\tau-1)}\right] \\ & = \sum_{i=1}^{n}[(Y_{i,fil} - X\alpha)(Y_{i,fil} - X\alpha)' + C_i], \end{aligned}$$

where $Y_{i,fil}$ is obtained by filling the missing values of Y_i with $E[Y_{i,mis}|Y_{i,obs}, \Theta^{(\tau-1)}]$.

Hence, we have that

$$
\begin{aligned}
&E[l(Y_{obs}, Y_{mis}, \Theta)|Y_{obs}, \Theta^{(\tau-1)}] \\
&= E[-\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}nS(\alpha))|Y_{obs}, \Theta^{(\tau-1)}] \\
&= -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}E[nS(\alpha)|Y_{obs}, \Theta^{(\tau-1)}]) \\
&= -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}\sum_{i=1}^{n}[(Y_{i,fil} - X\alpha)(Y_{i,fil} - X\alpha)' + C_i]).
\end{aligned}
$$

Maximization Step:

The maximization step can proceed as in the case with no missing data, that is, by iterating between maximizing over α and Σ, using

$$
-\frac{1}{2}\mathrm{tr}(\hat{\Sigma}^{-1}\sum_{i=1}^{n}[(Y_{i,fil} - X\alpha)(Y_{i,fil} - X\alpha)'])
$$

for α ($\hat{\Sigma}$ is the current estimate of Σ) and

$$
-\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}C^*) \tag{2.8}
$$

for the Σ, where $C^* = \sum_{i=1}^{n}[(Y_{i,fil} - X\hat{\alpha})(Y_{i,fil} - X\hat{\alpha})' + C_i]$ and $\hat{\alpha}$ is the current estimate of α.

The maximization over the parameters of Σ is the almost same as before with no missing data. The difference is that we need to substitute C^* for $\hat{S}$. The calculation of $C_i = Var[Y_{i,mis}|Y_{i,obs}, \Theta^{(\tau-1)}]$ in the definition of C^* is based on the conditional

distribution of one vector given the other when they are jointly multivariate normally distributed. Standard formula can be used. Specifically, suppose

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right). \tag{2.9}$$

Then

$$\mathbf{x}_1 | \mathbf{x}_2 \sim N(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}),$$

where

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2),$$

and

$$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

2.4 Simulations

In this section via simulations, we compare the performance of the sample covariance matrix S to that of nonparametric estimator $\hat{\Sigma}$, obtained from (1.1) by smoothing the first few subdiagonals of T and replacing the rest by zero. The results confirm our intuition that for large T and smooth covariance matrices, $\hat{\Sigma}$ should perform better than S for a variety of reasonable loss functions and sample sizes.

To compare the performance of covariance matrix estimators, we consider two loss

functions:

$$L_1(\Sigma, G) = \operatorname{tr}\Sigma^{-1}G - \log|\Sigma^{-1}G| - T \text{ and } L_2(\Sigma, G) = \operatorname{tr}(\Sigma^{-1}G - I)^2, \qquad (2.10)$$

where Σ is the true covariance matrix and G is a positive definite matrix (see Section 7.8 of Anderson 1984). Each of these is 0 when $G = \Sigma$ and is positive when $G \neq \Sigma$. Both loss functions are invariant with respect to transformations $G^* = CGC'$, $\Sigma^* = C\Sigma C'$ for nonsingular matrix C. The corresponding risk function are defined by $E_\Sigma\{L_i(\Sigma, G)\}$, $i = 1, 2$. A (smooth) estimator $\hat{\Sigma}$ is considered better than the (sample) covariance matrix S if its risk function is smaller, that is, $E_\Sigma\{L_i(\Sigma, \hat{\Sigma})\} < E_\Sigma\{L_i(\Sigma, S)\}$. For more information on simulation-based comparison of covariance estimators, see Lin and Perlman (1985). The risk function is approximated by Monte Carlo simulation. To produce the results presented below, N=100 simulation runs are used.

We consider the following three covariance matrices with varying level of smoothness as indicated by the smoothness of the functions associated with their (T, D) components:

- $\phi_{t,t-j} \equiv 0$, $\sigma_t^2 \equiv 1$. (The identity covariance matrix.)
- $\phi_{t,t-1} = 2(t/p)^2 - .5$, $\phi_{t,t-j} \equiv 0, j \geq 2$, $\sigma_t = \log(t/10 + 2)$. (Varying coefficient AR(1).)
- $\phi_{t,t-j} = p^{-2}\min\{t + j, t^{1.5}\}\exp\{-j/4\}$, $\sigma_t = \log(t/10 + 2)$.

These three matrices have been used in Wu and Pourmahmadi (2003) to illustrate their method. For each Σ from the above list, we simulate n i.i.d. $N(0, \Sigma)$ random T-vectors and compute their sample covariance matrix S. The proposed method is used to yield a nonparametric estimate $\hat{\Sigma}$ of Σ. The above scheme is repeated a total of $N = 100$ times, and for a given loss function $L(\cdot, \cdot)$, the values of $L(\Sigma, S)$ and $L(\Sigma, \hat{\Sigma})$ are computed. The corresponding risks are obtained by averaging the values of the losses across simulation runs.

The results of the simulation study are presented in Tables 2.1 to 2.3 for L_1 and L_2 respectively. Based on the estimated risks in these tables, it is evident that the spline smoothed covariance estimators outperform the sample covariance matrix for every combination of (Σ, n, p). The spline-smoothed covariance estimators also outperform the two-step kernel smoothed covariance estimators of Wu and Pouramadi (2003). Consider estimation of Σ_1, for exmaple. For sample sizes $n = 50$ and $n = 100$, and matrices of dimensions $p = 10, 20, 30, 40$, the estimated risks for the kernel smoothed covariance estimators are respectively, 0.246, 0.384, 0.399, 0.448, 0.118, 0.178, 0.194, 0.216. The results for local polynomial smoothed covariance matrix are based on a technical reprot by Wei Biao Wu and Mohsen Pourahmadi.

2.5 Cattle data

Here we revisit our cattle data with the new method developed in this chapter. As usual, BIC is used to select the tuning parameters. However, instead of using sat-

Table 2.1 Simulations for Σ_1. Risks, i.e., average losses, of three estimators (sample, spline smoothed and local polynomial smoothed covariance matrices), at three test matrices for two loss functions. The results are based on 100 simulation runs.

			Entropy loss			Quadratic loss		
	n	p	Sample	Spline	Local	Sample	Spline	Local
		10	1.215	0.145	0.246	2.281	0.308	0.648
			(0.241)	(0.088)		(0.559)	(0.207)	
		20	5.023	0.164	0.384	8.679	0.341	1.036
			(0.493)	(0.092)		(1.142)	(0.201)	
	50	30	12.410	0.159	0.399	19.050	0.328	1.101
			(0.950)	(0.091)		(2.066)	(0.200)	
		40	25.326	0.139	0.448	33.606	0.284	1.151
			(1.442)	(0.083)		(2.701)	(0.171)	
Σ_1		10	0.575	0.078	0.118	1.127	0.159	0.291
			(0.115)	(0.043)		(0.262)	(0.097)	
		20	2.290	0.076	0.178	4.263	0.154	0.456
			(0.237)	(0.044)		(0.579)	(0.094)	
	100	30	5.237	0.085	0.194	9.311	0.171	0.517
			(0.331)	(0.050)		(0.762)	(0.102)	
		40	9.647	0.091	0.216	16.500	0.184	0.661
			(0.516)	(0.043)		(1.133)	(0.089)	

urated means as was done in chapter 1, we model the mean and covariance of the cattle data simultaneously with quadratic splines. BIC is optimized when there are 2 subdiagonals, with each subdiagonal having a smoothing spline of 3 basis. The diagonal elements of D are smoothed with a spline of 4 basis and the means are fitted with a spline of 9 basis. In the first chapter, we noticed that in the T estimated using penalized likelihood most of the non-zero elements appear on the first and second subdiagonals. The optimal result above also happens to have 2 subdiagonals, lending more support to an ante-dependence model of order 2 for our cattle data(Macchiavelli and Arnold,1994).

We then examine the effect of missing data. First we randomly drop 10% of the

Table 2.2 Simulations for Σ_2. Risks, i.e., average losses, of three estimators (sample, spline smoothed and local polynomial smoothed covariance matrices), at three test matrices for two loss functions. The results are based on 100 simulation runs.

			Entropy loss			Quadratic loss		
	n	p	Sample	Spline	Local	Sample	Spline	Local
		10	1.198	0.152	0.274	2.252	0.316	0.809
			(0.255)	(0.087)		(0.516)	(0.198)	
		20	5.039	0.140	0.439	8.629	0.287	1.150
			(0.531)	(0.076)		(1.165)	(0.171)	
	50	30	12.472	0.148	0.462	19.045	0.303	1.244
			(0.841)	(0.077)		(1.971)	(0.162)	
		40	25.672	0.169	0.506	33.616	0.346	1.316
			(1.407)	(0.088)		(2.708)	(0.190)	
Σ_2		10	0.562	0.079	0.159	1.099	0.162	0.501
			(0.110)	(0.047)		(0.253)	(0.102)	
		20	2.269	0.071	0.213	4.242	0.143	0.609
			(0.225)	(0.040)		(0.521)	(0.082)	
	100	30	5.265	0.088	0.242	9.452	0.177	0.703
			(0.388)	(0.041)		(0.797)	(0.088)	
		40	9.685	0.098	0.271	16.569	0.197	0.764
			(0.484)	(0.042)		(1.122)	(0.088)	

cattle data. Then we apply the EM algorithm to the rest with the same parameters as above. T, D and column means are estimated for this data set with missing values. We compare them with their counterparts when there is no missing data. The corresponding pairs show remarkable similarities as are shown in the graphs.

Table 2.3 Simulations for Σ_3. Risks, i.e., average losses, of three estimators (sample, spline smoothed and local polynomial smoothed covariance matrices), at three test matrices for two loss functions. The results are based on 100 simulation runs.

			Entropy loss			Quadratic loss		
	n	p	Sample	Spline	Local	Sample	Spline	Local
		10	1.220	0.195	0.274	2.264	0.388	0.626
			(0.235)	(0.085)		(0.550)	(0.199)	
		20	5.059	0.176	0.411	8.700	0.353	1.086
			(0.500)	(0.086)		(1.204)	(0.183)	
	50	30	12.493	0.159	0.389	19.164	0.320	1.136
			(0.790)	(0.087)		(1.965)	(0.181)	
		40	25.740	0.187	0.472	33.861	0.378	1.491
			(1.328)	(0.096)		(2.854)	(0.197)	
Σ_3		10	0.563	0.138	0.154	1.101	0.270	0.448
			(0.122)	(0.043)		(0.286)	(0.096)	
		20	2.271	0.097	0.206	4.229	0.191	0.690
			(0.234)	(0.045)		(0.554)	(0.090)	
	100	30	5.245	0.106	0.196	9.365	0.211	0.663
			(0.354)	(0.047)		(0.773)	(0.097)	
		40	9.726	0.101	0.202	16.595	0.202	0.744
			(0.523)	(0.044)		(1.169)	(0.090)	

2.6 Telephone call center data

The call center data is another good candidate to test our new method. In chapter 1, there is only 1 tuning parameter needed to be estimated, i.e, the amount of penalty. In this chapter, we have three tuning parameters, i.e, the number of sub-diagonals that are not zero, the number of basis of the quadratic spline used to model each sub-diagonal, the number of basis of the quadratic spline used to model innovation variances. Sometimes we have a fourth parameter depending on whether you use a saturated mean model or not. This increase of number of parameters sometimes presents some challenge in estimating them precisely, especially if you are faced with a

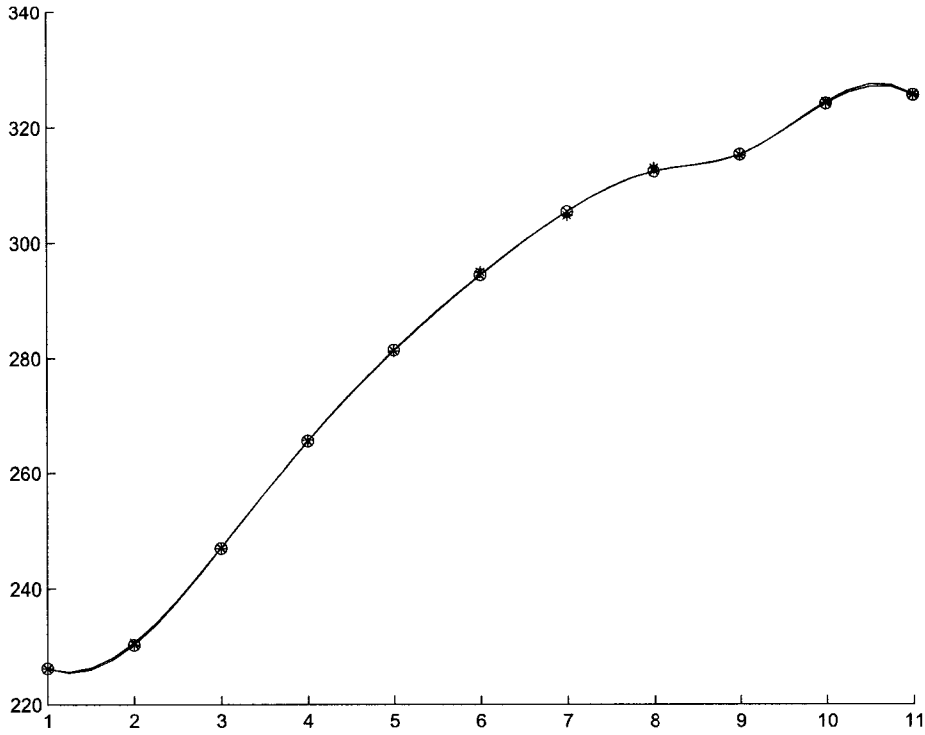


Figure 2.1 Raw means and the estimated means with spline smoothing. '*'s are for raw means. 'o's are for means estimated without missing data. 'x's are for means estimated with 10% missing data.

real data set such as call center data whose covariance matrix is as large as 102×102. Therefore, we abandon the previous covariance selection criterion of BIC and employ a more practical one. Since in the end we want to evaluate the performance of our new method in out-of-sample forecasting, we decide to choose the set of tuning parameters that are precisely good in this regard. Here is how we do it.

We split the 239 days into three groups — the training(the first 205 days, corresponding to January to October) , test/training(the middle 17 days, corresponding to November), and test(the last 17 days, corresponding to December) data sets. First we use the training data set to estimate the mean and covariance structure. For each set

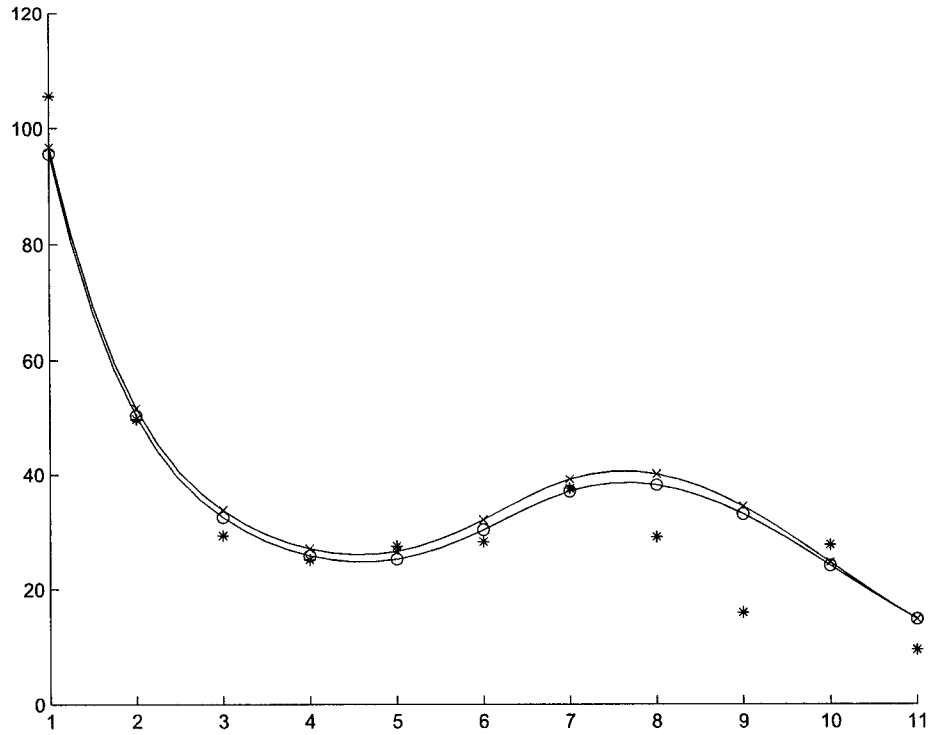


Figure 2.2 Diagonal elements of the three D matrices. '*'s are for D obtained from sample covariance matrix. 'o's are for D estimated without missing data. 'x's are for D estimated with 10% missing data.

of tuning parameters, there is a corresponding covariance matrix estimate. We then evaluate the forecasting performance of these estimates. The details of forecasting is the same as those in chapter 1. We apply the estimates for forecasting using formula (1.11) for the middle 17 days in test/training set (corresponding to November). We use the 51 (square root transformed) arrival counts in the early half of a day to forecast the (square root transformed) arrival counts in the later half of the day. For each 10-minute interval in the latter half of the day, we have the actual observed value and the forecast value for each of the 17 day in the test/training data set, denote them as y_{it} and $\hat{y}_{it}$, $i = 206, \ldots, 222$, $t = 52, \ldots, 102$. Define the total average absolute

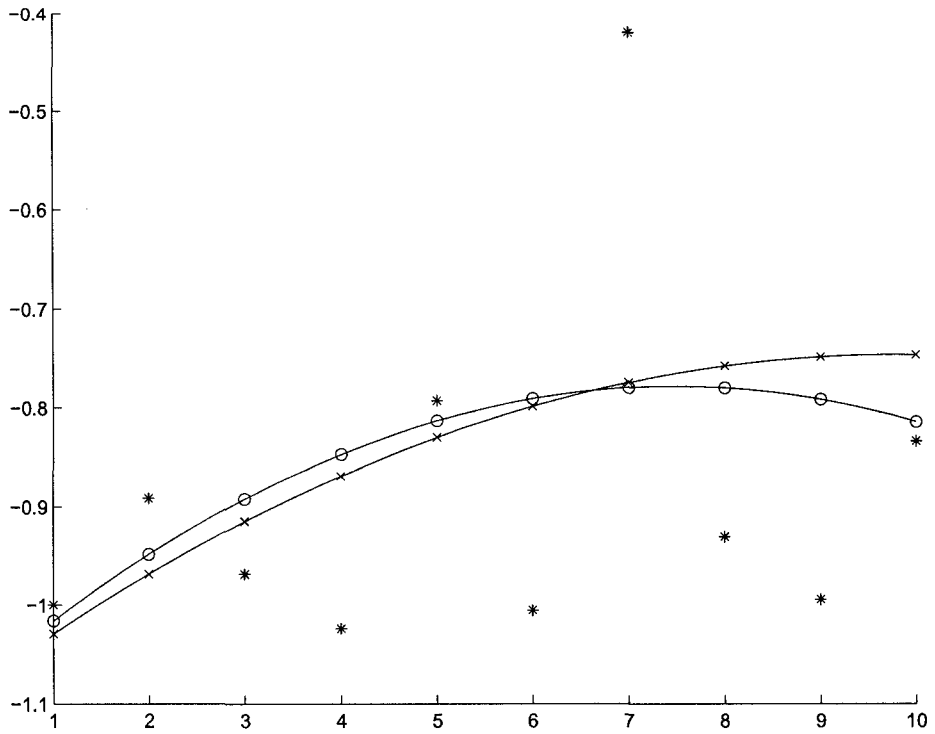


Figure 2.3 The first subdiagonal of the three T matrices. '*'s are for T obtained from sample covariance matrix. 'o's are for T estimated without missing data. 'x's are for T estimated with 10% missing data.

forecast error TAAFE_{206}^{222} as

$$\text{TAAFE}_{206}^{222} = \frac{1}{51 \times 17} \sum_{t=52}^{102} \sum_{i=206}^{222} |\hat{y}_{it} - y_{it}|.$$

We choose the covariance matrix estimate that has the smallest TAAFE_{206}^{222}. Its corresponding tuning parameters are saved for next phase.

Now we combine the training and test/training data sets into one big training set(corresponding to January to November). Mean and covariance structure are reestimated for this big training set with previously obtained tuning parameters. Finally

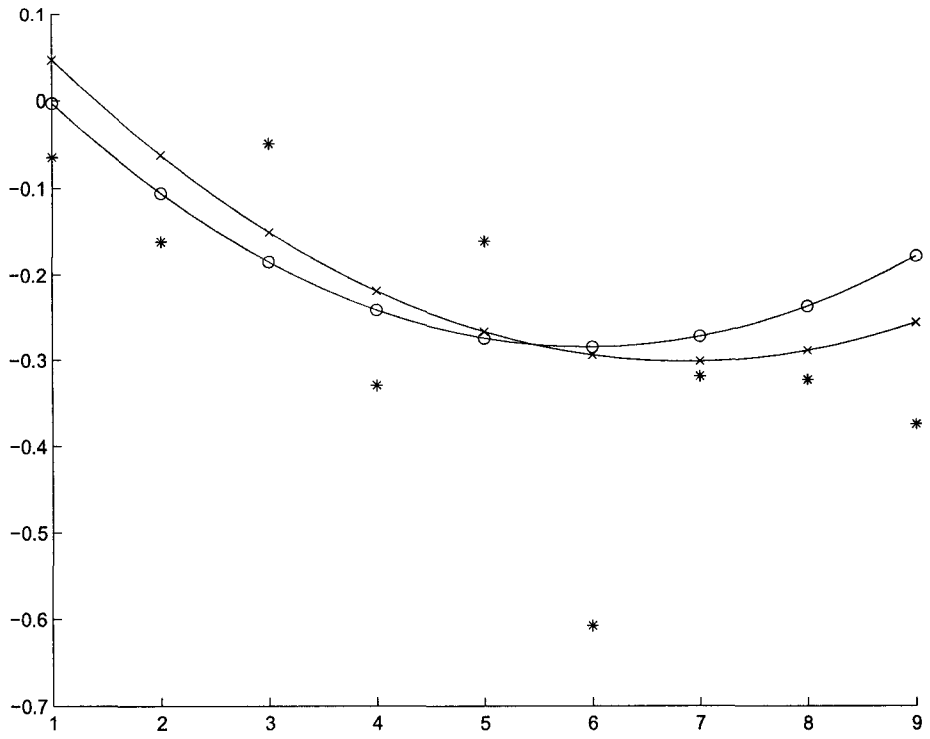


Figure 2.4 The second subdiagonal of the three T matrices. '*'s are for T obtained from sample covariance matrix. 'o's are for T estimated without missing data. 'x's are for T estimated with 10% missing data.

we use test data set(the last 17 days, corresponding to December) to assess the forecasting performance of the estimates. The measure is TAAFE_{223}^{239}, which is defined as

$$\text{TAAFE}_{223}^{239} = \frac{1}{51 \times 17} \sum_{t=52}^{102} \sum_{i=223}^{239} |\hat{y}_{it} - y_{it}|.$$

The result is compared with what we get if only the sample covariance matrix for the big training set is used. Also we are very interested in finding out how well our new method performs relative to the methods of previous chapter, especially the one with L_1 penalty. To be fair, we use TAAFE_{206}^{222} as well to determine optimal L_1

Table 2.4 Call center data

	Sample	Spline Smoothing	L_1 penalty
		2 subdiagonals	
Parameters estimates		15 basis for innovation variances	$\lambda = 279.36$
		8 basis for each subdiagonal	
TAAFE_{223}^{239}	1.3667	0.8437	0.8495

penalty amount instead of the usual 5-fold CV or GCV. The final results are shown in Table 2.4.

It's amazing to see how well spline smoothing and L_1 penalty perform relative to the sample covariance matrix. The improvement is very significant. Furthermore, it's interesting to note that these two methods achieve almost the same good forecasting results despite their apparently different approaches and different covariance matrix estimates.

2.7 Conclusion

In this chapter, we propose an alternative approach to reparameterize the covariance matrix through the modified Cholesky decomposition of its inverse. A great deal of smoothness is observed in the Cholesky factors when the underlying data set is longitudinal. We use basis functions, or quadratic splines to model this smoothness. As such, the daunting task of modelling covariance matrices becomes much easier and intuitive. What's more, this new approach provides a simultaneous and unified method of (nonparametric) estimation for the mean and covariance of longitudinal data. It also handles missing data in a natural way. The methodology and the

computational procedure are illustrated by applying them to two real datasets of sizes 11×11 and 102×102, the latter is quite large for most existing methods. Our method is ideal for longitudinal studies or situations where the variables are ordered over time or otherwise.

Chapter 3

The Value/Growth Spreads As Predictors of Returns

3.1 Introduction

Recent rational theory (e.g., Zhang (2003)) predicts that the value spread (defined as the average book-to-market of value stocks minus the average book-to-market of growth stocks) is countercyclical, and should be a positive predictor of future returns. But the growth spread (defined as the average market-to-book of growth stocks minus the average market-to-book of value stocks) is procyclical, and should be a negative predictor of future returns. Finally, the log spread (defined as the log book-to-market of value stocks minus the log book-to-market of growth stocks) is weakly countercyclical, and should be a weak, positive predictor of future returns. We test these hypotheses using predictive regressions.

The evidence is largely supportive. From January 1927 to December 2001, the slopes from regressing future market excess returns and small firm excess returns onto the value spread over different horizons are positive and mostly significant. In contrast, the slopes of the growth spread are negative and mostly significant. The evidence is somewhat weaker in the post-1945 subsample. Finally, the log spread yields mixed predictability results. Their slopes are mostly positive in the long sample, but are mostly negative in the postwar sample.

To shed light on potential sources of the predictability, we examine the comovement of the spreads with conditioning variables that are often used to predict future market returns and macroeconomic activities. We find that, in the long sample, the value spread covaries positively with countercyclical variables such as dividend yield, term premium, default premium, and aggregate book-to-market, but covaries negatively with the procyclical short-term interest rate. And the growth spread covaries negatively with the countercyclical variables and positively with the procyclical short rate. The evidence is again weaker in the postwar sample. The cross-correlation structure of the log spread with the conditioning variables is similar to that of the value spread, but the correlations are generally smaller. In all, the results suggest that the value spread is countercyclical, the growth spread is procyclical, and the log spread is weakly countercyclical.

The issue we study is important. In an intriguing article, Campbell and Vuolteenaho (2004) break market beta into two components, cash flow beta (reflecting news about

the market's future cash flows) and discount rate beta (reflecting news about the market's discount rates). They find that value stocks have higher cash flow betas than growth firms. And this helps explain the value anomaly because cash flow beta has a higher price of risk (see also Bansal, Dittmar, and Lundblad (2004) and Hansen, Heaton, and Li (2004)).

The empirical success of Campbell and Vuolteenaho's (2004) two-beta model depends critically on their use of the small-stock value spread to predict aggregate stock returns. (Campbell and Vuolteenaho define the value spread to be the log book-to-market of value stocks minus the log book-to-market of growth stocks, which we call the log spread, to be distinguished from the value spread in book-to-market levels.) It is natural to ask why this variable can be used to predict future returns. In this regard, investment-based models can complement the ICAPM-framework underlying the two-beta model because these models directly tie stock returns with firm characteristics such as book-to-market.

Our results lend some support to the choice of predictive variables in Campbell and Vuolteenaho (2004). We show that the log spread is related to the value and the growth spreads. Both are strong predictors of future returns. However, we also point out that because the log spread is equivalent to the log market-to-book of growth stocks minus the log market-to-book of value stocks, the log spread contains information on both the value and the growth spreads. But since the value and growth spreads have opposite cyclical properties, the log spread appears to contain

less information on future market returns.

Our work is also related to the time series predictability literature. Kothari and Shanken (1997) and Pontiff and Schall (1998) find that the aggregate book-to-market ratio is a positive predictor of future market excess returns and small firm excess returns. And the predictive relations are much weaker in the post-1960 period. We differ in that we use the value and growth spreads to predict stock returns. Although these spreads and the aggregate book-to-market are correlated, we find that they contain different information on future returns. The aggregate book-to-market is a stronger predictor of future market excess returns, but the value spread is a stronger predictor of small firm excess returns. And while the aggregate book-to-market dominates the growth spread in predicting returns in the long sample, the growth spread dominates the aggregate book-to-market in the postwar sample. Another related paper is Eleswarapu and Reinganum (2003) who find that annual market excess returns correlate negatively with the returns of growth stocks in the previous three years.

Finally, we contribute to the literature on cross-sectional predictability. Overreaction and time-varying expected returns have been long proposed as two competing explanations of the value anomaly. While rational theory predicts explicitly that the value spread is countercyclical and should predict positively future stock returns and that the opposite is true for the growth spread, these patterns related to business cycle fluctuations are not predicted by overreaction-based explanations (e.g., DeBondt and Thaler (1985)).

The rest of the chapter is organized as follows. Section 3.2 develops testable hypotheses based on recent rational theory. Sections 3.3 and 3.4 describe our data and estimation methods, respectively. Section 3.5 presents our empirical results. Finally, Section 3.6 concludes.

3.2 Hypothesis Development

We develop our testable hypotheses from the dynamic asset pricing model of Zhang (2003). His model features both costly reversibility and time-varying price of risk. Cost reversibility means that it is more costly for firms to divest than to invest (e.g., Ramey and Shapiro (2001)), and countercyclical price of risk means that discount rates are higher in bad times than in good times (e.g., Fama and French (1988, 1989)). The model predicts that the value spread is countercyclical and that the growth spread is procyclical.[1]

Based on Zhang's (2003) model, Figure 3.1 plots the value spread (Panel A), the growth spread (Panel B), and the log spread (Panel C) between value and growth firms against aggregate economic conditions modeled as aggregate productivity, denoted x. High aggregate productivity indicates booms and low aggregate productivity indicates recessions. The solid lines represent the benchmark model with costly reversibility and time-varying price of risk, and the broken lines represent the simplified model

[1]Gomes, Kogan, and Zhang (2003) construct a related dynamic model and show that the cross-sectional dispersion of book-to-market is countercyclical (see Panel d of their Figure 5). Other related papers include Berk, Green, and Naik (1999), Cooper (2003), Carlson, Fisher, and Giammarino (2003), and Kogan (2003). However, these authors do not discuss explicitly the time series properties of the value spread or their underlying economic mechanism.

without these two features. Panel A is identical to Panel B of Figure 4 in Zhang, but the other two panels are new.

From Panel A in Figure 3.1, the value spread is clearly countercyclical in the benchmark model, but is largely constant in the simplified model. From Panel B, the growth spread is clearly procyclical, and is even more so in the simplified model. Finally, Panel C shows that the log spread is weakly countercyclical. In terms of stock return predictability, Zhang (2003) therefore predicts that the value spread is a positive predictor of future returns, the growth spread is a negative predictor of future returns, and the log spread is a weak, positive predictor of future returns.

The economic mechanism driving these predictions in Zhang (2003) is as follows. First, why is the value spread countercyclical? In recessions, all firms invest at lower rates than average. Because of their relatively low profitability (e.g., Fama and French (1995)), value firms are more likely to scrape capital than growth firms. When investment is reversible, value firms can scale down easily. With costly reversibility value firms face higher adjustment costs when divesting; as a result, they are stuck with more unproductive assets, leading to higher book-to-market ratios in bad times.

Further, time-varying price of risk propagates the effects of costly reversibility. Higher discount rates in bad times lower firms' expected net present values. As future prospects become even gloomier, value firms want to scrap even more capital, giving rise to even higher book-to-market ratios in bad times. On the other hand, growth firms are less prone to costly reversibility and time-varying price of risk. Their assets

are more productive; thus they have less incentives to scaling down in recessions. In all, the value spread is high in recessions and low in booms, as shown in Panel A of Figure 3.1.

Why is the growth spread procyclical? In good times, growth firms invest more and grow faster than value firms. Investing and growing are less urgent for value firms because their previously unproductive assets become more productive given positive aggregate shocks. As a result, the dispersion in growth options between value and growth firms is widened in booms. And time-varying price of risk again propagates the effects. A lower discount rate in good times increases firms' expected net present values, causing growth firms to invest even more and grow even faster. In all, the growth spread is high in good times and low in bad times, as shown in Panel B of Figure 3.1.[2]

Finally, what explains the behavior of the log spread? Notice that the value spread in logs is mathematically equivalent to the growth spread in logs:

$$\underbrace{\log\left(\frac{\text{BE}}{\text{ME}}\right)_{\text{value}} - \log\left(\frac{\text{BE}}{\text{ME}}\right)_{\text{growth}}}_{\textbf{The Value Spread In Logs}} = \underbrace{\log\left(\frac{\text{ME}}{\text{BE}}\right)_{\text{growth}} - \log\left(\frac{\text{ME}}{\text{BE}}\right)_{\text{value}}}_{\textbf{The Growth Spread In Logs}} \tag{3.1}$$

where BE denotes book value and ME denotes market value. The log spread therefore reflects both the countercyclical movements of the value spread and the procyclical

[2]The panel also shows that the growth spread in the simplified model without costly reversibility is higher and exhibits stronger procyclical movements than the growth spread in the benchmark model. The intuition is that, although firms do not face high cost when expanding capital, the mere possibility of high cost when scaling down in future recessions reduces firms' growth rates in good times in the benchmark model.

movements of the growth spread. This makes the economic interpretation of the log spread less clear, and explains why it is weakly countercyclical, as shown in Panel C of Figure 3.1.

3.3 Data and Descriptive Statistics

This section discusses sample construction and basic properties of the data.

3.3.1 Sample Construction

We measure the value spread, denoted S_{val}, as the average book-to-market ratio of portfolio ten minus the average book-to-market ratio of portfolio one from the ten deciles sorted in ascending order on book-to-market. We measure the growth spread, denoted S_{grw}, as the average market-to-book ratio of portfolio one minus the average market-to-book ratio of portfolio ten. And we measure the log spread, denoted S_{log}, as the log book-to-market of portfolio ten minus the log book-to-market of portfolio one.

We obtain the Fama-French portfolio data from Kenneth French's website. The data set contains the calendar year-end book-to-market ratios for all the book-to-market deciles. For months from January to December of year t, the book-to-market ratio of a given portfolio is constructed by dividing its book-to-market ratio at the end of December of year $t-1$ (where book value and market value are both measured at the end of December of year $t-1$) by its compounded gross return from the end of

December of year $t-1$.

Our definition of the value spread is different from that of previous studies (e.g., Asness, Friedman, Krail, and Liew (2000), Campbell and Vuolteenaho (2004), Cohen, Polk, and Vuolteenaho (2003), and Yogo (2003)). There the value spread is defined as the log book-to-market of value stocks minus the log book-to-market of growth stocks — the log spread in our definition. As we argued in Section 3.2, the interpretation of the log spread as the value spread is not economically precise because it can be equally interpreted as the growth spread in logs. We hence reserve the term, the value spread, only for the difference in book-to-market levels between value and growth portfolios.

We follow Cohen, Polk, and Vuolteenaho (2003) and use the entire CRSP universe to construct the value spread. Campbell and Vuolteenaho (2004) construct their measures using the small-high and small-low portfolios from the two-by-three sort on size and book-to-market (e.g., Fama and French (1993)). We have tried to construct our spreads using the small-cap portfolios; the resulting measures and the entire-sample measures have extremely similar properties. For example, the correlation between the small-cap value spread and the entire-sample value spread is 0.98 from January 1927 to December 2001, and is 0.97 in the sample after 1945. We therefore omit the results with only the small-cap portfolios.

3.3.2 Time Series Properties

Figure 3.2 plots the time series of the value spread (Panel A), the growth spread (Panel B), and the log spread (Panel C) along with NBER recession dates in shadowed area. From Panel A, there is a structural break around 1945 for the value spread. Before 1945, the value spread is relatively high and volatile, especially in the Great Depression periods; after 1945, the value spread is relatively low and stable. We thus conduct empirical analysis both in the full sample from January 1927 to December 2001 and in the subsample after January 1945.

From Panels B and C of Figure 3.2, no structural breaks similar to that in the value spread are evident for the growth spread and the log spread. The growth spread is relatively low in the recessionary 1930s and 1970s and is relatively high in the expansionary 1960s and 1990s, suggesting that the growth spread is procyclical. Finally, Panel C shows that the log spread is relatively high in both the Great Depression and in the expansionary 1990s, suggesting that the log spread exhibits no clear-cut cyclical patterns. This is not surprising because the log spread contains information on both the value spread and the growth spread.

3.3.3 Descriptive Statistics

Table 3.1 reports descriptive statistics for the value spread, the growth spread, the log spread, and for the portfolio returns used as the dependent variables in predictive regressions. These portfolio returns include the equal-weighted market excess re-

turn, denoted r_{ew}^{mkt}; the value-weighted market excess return, denoted r_{vw}^{mkt}; the equal-weighted small-cap (quintile) excess return, denoted r_{ew}^{sml}; and the value-weighted small-cap (quintile) excess return, denoted r_{vw}^{sml}. Our choice of portfolio returns in predictive regressions follows previous studies in the time series predictability literature (e.g., Pontiff and Schall (1998)).

From the first row of Panels A and B, the value spread is on average 4.57 with a monthly volatility of 5.45 from January 1927 to December 2001, and the average value spread reduces to 2.32 and its volatility reduces to 1.13 in the postwar sample. The average growth spread is more stable across samples: it is 4.32 in the full sample and is only slightly higher, 4.62, in the postwar sample. And the average log spread is also stable across samples: it is 2.66 in the full sample and is 2.39 in the postwar sample. Finally, the log spread is less volatile than the other two spreads.

All three spreads are fairly persistent. From the first row of Panel A in Table 3.1, the first-order autocorrelation of the value spread in the long sample is 0.95; it reduces to 0.67 at the 12-th order, but is still 0.50 even at the 60-th order. The persistence structure of the value spread remains basically unchanged in the subsample. The log spread and the growth spread have slightly higher autocorrelations than the value spread in the full sample, but have comparable autocorrelations in the postwar sample.

Importantly, although these monthly autocorrelations are high, they are much lower than the autocorrelations of a set of commonly used predictive variables such dividend yield, aggregate book-to-market, and earnings price ratio. Lewellen (2004,

Table 1) reports that the first-order autocorrelations of these variables range from 0.988 to 0.999. The less persistence in the value and growth spreads relative to traditional predictive variables has important implications for our choice of estimation methods, as we discuss later in Section 3.4.

We have also conducted the Augmented Dickey-Fuller (ADF) unit root tests with an intercept and 12 lags for the three spreads. The p-value of the ADF test for the value spread is 0.08 in the long sample; thus the null hypothesis of a unit root cannot be rejected at the five percent level, but can be rejected at the ten percent level. In the postwar sample, the p-value for the value spread is basically zero. The unit root hypothesis cannot be rejected for the growth spread in both samples; for the log spread, the null cannot be rejected in the long sample, but can be rejected in the postwar sample.

The descriptive statistics of market excess returns and small firm excess returns are well known, but they are reported in Table 3.1 for completeness. In particular, all the portfolio returns are positively autocorrelated over the one-month horizon, but generally uncorrelated thereafter, consistent with Lo and MacKinlay (1988).

3.4 Estimation

To investigate the stock return predictability associated with the value, growth, and log spreads, we follow Fama and French (1988, 1989) and adopt the simple regression

framework:

$$r_{t+\tau} = \alpha_\tau + \beta_\tau S_t + \epsilon_{t+\tau} \tag{3.2}$$

where S_t is one of the spreads (S_{val}, $S_{\log}$, or S_{grw}) measured at the beginning of time t. $r_{t+\tau}$ is the simple excess return of either the market portfolio or the small-stock portfolio return from time t to time $t+\tau$, where τ denotes different horizons including one-month, one-quarter, one-year, two-year, and five-year holding period.

Since all three spreads are fairly persistent, the standard inferences are biased because returns depend on changes in stock prices; thus changes in the value spread are likely to be negatively related to its contemporaneous returns. This correlation induces a spurious bias in the estimates from regressing future returns on a persistent regressor (e.g., Stambaugh (1999)). To obtain the correct p-values for the slope coefficients, we have to account for the small sample bias.

We use two methods in this regard. The first is the randomization method of Nelson and Kim (1993). We estimate the following first-order autoregressive process for the regressor:

$$S_{t+\tau} = \theta + \rho S_t + \eta_{t+\tau}. \tag{3.3}$$

We then retain both the estimated $\eta_{t+\tau}$ and the contemporaneous excess returns $r_{t+\tau}$ to control for their contemporaneous correlations. The pairs $(\eta_{t+\tau}, r_{t+\tau})$ are then randomized by resampling without replacement. From the randomized series, we create pseudo series of the regressor by substituting the randomized $\eta_{t+\tau}$ in Eq. (3.3)

along with estimated $\hat{\theta}$ and $\hat{\rho}$. The initial value, S_0, is picked randomly from the original series of S_t in the data.

This procedure creates pseudo series of the regressor and the excess returns that have similar time-series properties as the actual data do. However, these pseudo data are generated under the null hypothesis that there is no return predictability associated with S_t. We then estimate Eq. (3.2) using these pseudo data and store the coefficients. This process is repeated for 1000 times. Bias is defined as the sample mean of these 1000 coefficient estimates. The one-sided p-value is the estimated probability of obtaining a coefficient that is at least as large as the coefficient estimated from the actual data. A p-value less than 0.05 implies that the coefficient is significantly positive at the five percent significance level, and a p-value greater than 0.95 implies the coefficient is significantly negative at the five percent significance level.

The second method we use to obtain correct slopes and their p-values is due to Stambaugh (1999). Assume the vector of residuals from Eqs. (3.2) and (3.3), $[\epsilon_t \ \eta_t]'$, follows an i.i.d. multivariate normal distribution, $N(0, \Sigma)$. Stambaugh shows that the finite-sample distribution of the bias in slope, $\hat{\beta} - \beta$, depends on ρ and Σ but not on α, β, or θ. After setting ρ and Σ to be their respective sample estimates, we use simulations to obtain the finite-sample distribution of $\hat{\beta} - \beta$. We then use it to calculate the one-sided p-value and bias in slope. Despite the distributional assumption of $[\epsilon_t \ \eta_t]'$, the p-values from this method are quite similar to those obtained from the randomization method of Nelson and Kim (1993).

Lewellen (2004) argues that Stambaugh's (1999) method may greatly understate the significance of the slopes. This effect is likely to be quantitatively substantial if ρ is very close to one. From Table 3.1, the monthly autocorrelations of the spreads are mostly below 0.98. Because Lewellen argues that a persistent regressor must have a monthly first-order autocorrelation above 0.98 to have a sizable impact on the significance of the slopes,[3] we therefore do not use his method to adjust for the small-sample bias.

3.5 Empirical Results

Section 3.5.1 reports results of using the value spread, the log spread, and the growth spread to predict future returns. Section 3.5.2 studies potential sources of the predictability. And Section 3.5.3 examines the incremental predictive ability of the value and the growth spreads relative to traditional predictive variables.

3.5.1 Univariate Regressions

This subsection reports univariate, predictive regressions. In general, we find that the value spread is a positive predictor of returns, the growth spread is a negative predictor of returns, and the log spread predicts returns in either direction depending on specific sample used. The evidence is broadly consistent with the testable hypotheses

[3]More specifically, Lewellen (2004, p. 7) states that "with 25 years of data, this [method] requires a monthly autocorrelation around 0.98 and an annual autocorrelation around 0.85; with 50 years of data, the values are 0.99 and 0.90, respectively.

developed in Section 3.2.

Predicting Returns with the Value Spread

Table 3.2 presents the predictive regressions of future stock returns onto the value spread, S_{val}, both in the full sample from January 1927 to December 2001 (Panel A) and in the postwar sample from January 1945 to December 2001 (Panel B). We perform predictive regressions over different horizons, denoted τ, including monthly (M), quarterly (Q), annual (Y), two-year (2Y), and five-year (5Y) horizons. For regressions with two-year and five-year horizons, we use overlapping quarterly observations.

To test zero slopes, we report three p-values: p_{NW}, the p-value associated with the Newey-West t-statistic adjusted for heteroscedasticity and autocorrelations up to 12 lags; p_{NK}, the p-value constructed from the finite-sample distribution of the slopes using Nelson and Kim's (1993) method; and finally, p_{S}, the p-value obtained from Stambaugh's (1999) method. All the p-values are one-sided values that are the estimated probabilities of obtaining a coefficient at least as large as that estimated from the actual data. A value less than five percent implies that the coefficient is significantly positive at the five percent significance level, and a value greater than 0.95 implies that the coefficient is significantly negative at the five percent significant level.

A few interesting results emerge from Panel A of Table 3.2. First, the subpanel denoted β_τ shows that all the slopes are positive, suggesting that the value spread

predicts positively future market excess return and small firm excess return. Second, the magnitudes of both the slopes and the goodness-of-fit coefficients (from the subpanel denoted R^2) increase with the horizon τ, suggesting that the amount of the predictability rises with the regression horizon. Third, the slopes are mostly significant: the subpanel denoted p_{NW} shows that the one-sided p-values associated with Newey-West t-statistics are mostly significant at five percent level. And using Nelson and Kim's (1993) and Stambaugh's (1999) methods to adjust for the small-sample bias yields quantitatively similar results, as shown in the subpanels denoted p_{NK} and p_{S}. Finally, the subpanels denoted b_{NK} and b_{S} report the estimated biases in the slopes using Nelson and Kim's and Stambaugh's methods, respectively. The biases are generally small and are less than ten percent of the slopes.

Panel B of Table 3.2 replicates the predictive regressions using the value spread as in Panel A, but in the postwar sample. All the slopes β_τ are again positive and increasing in the regression horizon τ. A comparison with Panel A reveals that the slopes in the subsample are much higher than those in the full sample; in some cases, the slopes are more than doubled, indicating a lower average value spread in the postwar sample than that in the long sample. However, these slopes are also estimated with less precision; as a result, their significance is substantially lower in the subsample than that in the full sample. Except for a few significant p_{NW}'s in short-horizon regressions, significant p-values appear only in the two-year and five-year horizons. Finally, for the most part, the regression R^2's are substantially lower

in the postwar sample.

Predicting Returns with the Growth Spread

Table 3.3 regresses future returns onto the growth spread, S_{grw}. From the first five rows of the table, all the slopes are negative in both the full sample and the postwar sample, suggesting that the growth spread is a negative predictor of future returns. From the p-values in Panel A, the slopes estimated from the full sample are mostly significant with only a few exceptions. And from Panel B, the growth spread is largely significant when predicting market excess returns in the postwar sample, but it is only significant in the long horizons when predicting small firm excess returns.

Predicting Returns with the Log Spread

From Panel A of Table 3.4, the slopes of the log spread in predictive regressions are mostly positive in the full sample. The slopes are significant in predicting equal-weighted small firm excess returns, and are also significant in predicting the equal-weighted market excess returns and the value-weighted small firm excess returns in long horizons, but not in predicting the value-weighted market excess returns. Strikingly, Panel B shows that the predictive power of the log spread is not stable across samples. The slopes of the log spread become mostly negative in the postwar sample, although mostly insignificant.

In sum, Tables 3.2–3.4 show that, first, the value spread is a positive predictor of future returns from 1927 to 2001, but its predictive power is substantially weaker in

the postwar sample. Second, the growth spread is a significantly negative predictor of future market excess returns in both samples, and is significant when predicting small firm excess returns except in the long horizon regressions in the postwar sample. And finally, the log spread is a weak positive predictor of returns in the full sample and a weak negative predictor in the postwar sample. Overall, except for the negative correlations between the log spread and future returns in the postwar sample, our evidence is broadly consistent with the testable hypotheses developed in Section 3.2.

Our results contrast with those of Campbell and Vuolteenaho (2004) who use the log spread to predict market returns, but finds the slope to be negative. Two reasons may explain the difference. First, the log spread contains information on both the value and the growth spreads, which have opposite cyclical properties and predict returns with opposite signs. As another manifestation of the mixed properties of the log spread, Yogo (2003) regresses returns onto the log spread and three other variables, and finds the slope on the log spread to be significantly positive. Second, Campbell and Vuolteenaho use multiple regressions of future returns onto the log spread and three other variables that are correlated with the log spread. This makes direct interpretation and comparison to our results somewhat difficult.

3.5.2 Potential Sources of the Predictability

One potential source of stock return predictability associated with the value and growth spreads is the cyclical variations in these variables (see Section 3.2). To

investigate this possibility, we examine the cross correlations between the various spreads and a set of conditioning variables commonly used to predict market excess returns and aggregate economic conditions. Overall, our evidence suggests that the value spread is countercyclical, the growth spread is procyclical, and the log spread is weakly countercyclical in the long sample but is weakly procyclical in the postwar sample.

Conditional Variables

We use five variables: dividend yield (div); default premium (def); term premium (term); short-term interest rate (rf); and the aggregate book-to-market (b/m). Our choice of these conditional variables is motivated from the time series predictability literature.[4]

The data sources are described as follows. The dividend yield is the sum of dividend payments accruing to the CRSP value-weighted portfolio over the previous 12 months, divided by the contemporaneous level of the index. The default premium is the yield spread between Moody's Baa and Aaa corporate bonds. Data on the default yield are obtained from the monthly database of the Federal Reserve Bank of Saint Louis. The term premium is defined as the yield spread between a long-term and a one-year Treasury bond. Data on bond yields are obtained from the Ibbotson database. The one-month Treasury bill rate is obtained from CRSP. And finally, to

[4]An incomplete list of papers includes: Fama and French (1988), dividend yield; Keim and Stambaugh (1986), default premium; Campbell (1987) and Fama and French (1989), term premium; Fama and Schwert (1977) and Fama (1981), short-term T-bill rate; Kothari and Shanken (1997) and Pontiff and Schall (1998), aggregate book-to-market.

construct the aggregate book-to-market ratio, we obtain the data on book value by combining the Compustat annual research file and Moody's book equity data collected by Davis, Fama, and French (2000) available from Kenneth French's website. The monthly data on market value are from the CRSP monthly stock file.

Cross Correlations in the Full Sample

Panel A of Table 3.5 shows the cross correlations for the sample from January 1927 to December 2001. From the fifth row of the panel, the value spread displays positive correlations with all the well-known countercyclical variables including the dividend yield (0.45), the default premium (0.61), the term premium (0.39), and the aggregate book-to-market (0.85), where the pairwise cross correlations are reported in parentheses. And the value spread also displays a negative correlation of -0.52 with the procyclical short-term interest rate.

From the sixth row of Panel A, the growth spread correlates negatively with all the countercyclical variables including the dividend yield (-0.69), the default premium (-0.30), the term premium (-0.11), and the aggregate book-to-market (-0.75); it also correlates positively with the procyclical short-term interest rate (0.20). The seventh row of Panel A shows that the cross correlation structure of the log spread with the conditioning variables is similar to that of the value spread, although the correlations are mostly smaller in magnitude. And finally, the log spread correlates highly with the value spread (0.82) and correlates weakly with the growth spread (-0.09).

In sum, the full-sample evidence shows that the value spread is clearly countercyclical, and the growth spread is clearly procyclical; the log spread is also countercyclical, but the degree of its cyclical variation seems weaker than that of the value spread.

Cross Correlations in the Postwar Sample

From the fifth row of Panel B in Table 3.5, there are important changes in the cross correlations between the value spread and the conditioning variables in the postwar sample. Notably, the correlation between the value spread and the term premium is now close to zero, and that between the value spread and the default premium switches sign, from 0.61 in the long sample to -0.23 in the subsample. But the value spread continues to correlate positively with the countercyclical dividend yield (0.51) and aggregate book-to-market (0.73), and negatively with the procyclical short-term interest rate (-0.50).

The sixth row of Panel B shows that the growth spread continues to exhibit procyclical movements: its correlation with the dividend yield is -0.80, that with the default premium is -0.14, and that with the aggregate book-to-market is -0.88. However, its correlations with the term premium and the short term interest rate are close to zero.

Finally, the seventh row of Panel B shows that the cyclical variation of the log spread is less clear in the postwar sample. Although the log spread correlates neg-

atively with the countercyclical aggregate book-to-market (-0.16), it also correlates negatively with the procyclical short term interest rate (-0.51).

Correlations between the Value/Growth/Log Spreads and Future Returns

Table 3.5 also reports the cross correlations between the various spreads and future stock returns. We measure all the returns at the end of period t or at the beginning of period $t+1$ and all the conditioning variables and the spreads at the beginning of period t. From the first four rows of Table 3.5, the value spread is positively correlated with the next period market excess returns and small-stock excess returns in the long sample (Panel A). The same holds in the postwar sample (Panel B), although the correlations are somewhat lower than those in the long sample. The growth spread is negatively correlated with next period returns in both samples and the correlations are again higher in the long sample. The log spread correlates weakly and positively with future returns in the long sample, but correlates weakly and negatively with future returns in the postwar sample.

3.5.3 Relative Predictive Power

Given the cross correlations reported in Table 3.5, a natural question is whether the value spread and the growth spread contain incremental information about future returns that is not captured by other predictive variables already in the literature. We do not consider the log spread because its predictive ability is weaker and its economic interpretation is less clear than either the value or the growth spread. This

subsection uses bivariate and multiple regressions to evaluate the relative predictive power of the value and growth spreads.

Bivariate Regressions with Aggregate Book-to-Market

We first study predictive power relative to aggregate book-to-market ratio; this is important given its high correlations with both the value spread and the growth spread (Table 3.5).

Table 3.6 reports the bivariate regressions with the value spread and the aggregate book-to-market (Panels A and C) and those with the growth spread and the aggregate book-to-market (Panels B and D), both in the long sample and in the postwar sample. We only report the p-values from the Nelson and Kim (1993) method; those from Stambaugh's (1999) method are quantitatively similar and are omitted to save space.

From Panel A of Table 3.6, in the presence of the aggregate book-to-market, the value spread loses its ability to predict future market excess returns, but retains most of its ability to predict small firm excess returns. More important, the ability of the aggregate book-to-market to predict small firm excess returns documented in Pontiff and Schall (1998) diminishes substantially in the presence of the value spread. The two variables therefore contain different information on future returns in the long sample. In the postwar sample, Panel C shows that both variables have mostly positive but insignificant slopes, and that they have comparable predictive power.

From Panel B of Table 3.6, the growth spread loses almost all of its predictive abil-

ity in the presence of the aggregate book-to-market in the long sample. In particular, the slopes of the growth spreads when predicting small firm excess returns are mostly positive and in a few cases even significant. But Panel D shows that in the postwar sample the predictive power of the growth spread is much stronger than that of the aggregate book-to-market. All the slopes with the growth spread are negative and mostly significant; the slopes of the aggregate book-to-market often become negative and sometimes even significantly negative.

In sum, the value spread contains incremental information on the small firm excess returns but not on the market excess returns relative to aggregate book-to-market; the aggregate book-to-market dominates the growth spread in predicting returns in the long sample, but is dominated by the growth spread in the postwar sample.

Bivariate Regressions with Other Predictive Variables

Table 3.7 replicates the bivariate regressions in Table 3.6 but with the aggregate book-to-market replaced by the term premium. From Panel A, the value spread dominates the term premium in predicting future returns in the long sample. From Panel C, the term premium dominates the value spread in short horizons in the postwar sample, but the value spread retains some of its predictive power in long-horizon regressions. From Panels B and D, the slopes for the growth spread are all negative in both samples; some of them are significant in the presence of the term premium.

Table 3.8 reports the bivariate regression of returns onto the value spread and

the default premium (Panels A and C) and onto the growth spread and the default premium (Panels B and D). From Panel A, the value spread retains most of its predictive ability in the long sample. From Panels B and D, the growth spread is only significant in long-horizon regressions in both samples; the default premium remains strong in the long sample, but it loses almost all its predictive power in the postwar sample. Some of its slopes in the postwar sample even become negative and significant.

Table 3.9 uses the dividend yield along with the value spread or the growth spread in bivariate regressions. From Panel A, the value spread largely dominates the dividend yield in the long sample. In the postwar sample, the predictive ability of the dividend yield seems mostly stronger than that of the value spread in long-horizon regressions, but the slopes for both predictors in short horizons are mostly insignificant. From Panels B and D, the dividend yield largely dominates the growth spread in both samples.

Table 3.10 uses the short-term interest rate along with the value spread and the growth spread in the bivariate regressions. From Panel A, the value spread mostly dominates the short rate in the long sample; the short rate loses much of its predictive power except only in the five-year horizon. In the postwar sample, Panel C shows that the short rate dominates the value spread in the short horizons, but the value spread retains its predictive ability in long-horizon regressions. From Panels B and D, the growth spread has predictive ability largely comparable with that of the short

rate in both samples.

In sum, the value spread often dominates common predictive variables in the long sample. Predictability is weaker in the postwar sample for the variables. Although often dominated by other variables in the long sample, the growth spread retains its long horizon predictive ability in the postwar sample.

Multiple Regressions

Table 3.11 reports the results of regressing future returns onto the value spread and four conditioning variables including the term premium, the default premium, the dividend yield, and the short-term interest rate. From Panel A, the slopes of the value spread are mostly positive although insignificant in the long sample; no variable clearly dominates the others. From Panel B, the value spread has mostly negative although insignificant slopes in the postwar sample. This result is extremely similar to that of aggregate book-to-market documented in Pontiff and Schall (1998, Panel B of Table 3).

We interpret the negative slopes of the value spread and the aggregate book-to-market in multiple regressions as a result of multicollinearity, as opposed to the two variables being negative predictors of returns. Alternative interpretations are clearly possible. But Table 3.11 also shows that the default premium is a significantly negative predictor, and the short rate is a positive although insignificant predictor of long horizon returns in the postwar sample. These counterintuitive results seem

difficult to reconcile without multicollinearity.

Finally, Table 3.12 reports the multiple regressions of future returns onto the growth spread and the four conditioning variables. The growth spread loses most of its predictive power in the presence of the four other variables. But no variable clearly dominates others in both samples. Although the slopes of the default premium are mostly significant, but its slopes in long-horizon regressions in the postwar sample are significantly negative, again suggesting multicollinearity at work.

3.6 Conclusion

Consistent with recent theory, we find that, from January 1927 to December 2001, the value spread (defined as the book-to-market of value stocks minus the book-to-market of growth stocks) exhibits countercyclical movements and predict positively future stock returns. The growth spread (defined as the market-to-book of growth stocks minus the market-to-book of value stocks) exhibits procyclical movements and predict negatively future stock returns. But the evidence from the post-1945 sample is substantially weaker. Finally, the log spread (defined as the log book-to-market of value stocks minus the log book-to-market of growth stocks) exhibits weakly countercyclical movements and predict weakly and positively future returns in the long sample, but weakly and negatively in the post-1945 sample.

Table 3.1 Descriptive Statistics of Returns, The Value Spread, The Growth Spread, and The Log Spread

This table reports descriptive statistics including mean m (in percent), volatility σ (in percent), autocorrelations (of order 1–12, 24, 36, 48, and 60), and p-value associated with the Augmented Dickey-Fuller (ADF) unit root test with an intercept and 12 lags. We report these statistics for the value spread, S_{val}, defined as the average book-to-market of value stocks minus the average book-to-market of growth stocks; the growth spread, S_{grw}, defined as the average market-to-book of growth stocks minus the average market-to-book of value stocks; the log spread, $S_{\log}$, defined as the log book-to-market of value stocks minus that of growth stocks (equivalently, the log market-to-book of growth stocks minus the log market-to-market of value stocks; the equal-weighted and value-weighted market excess returns $r_{\text{ew}}^{\text{mkt}}$ and $r_{\text{vw}}^{\text{mkt}}$, respectively; and the equal-weighted and value-weighted small firm (quintile) excess returns $r_{\text{ew}}^{\text{sml}}$ and $r_{\text{vw}}^{\text{sml}}$, respectively. Panel A reports the results from January 1927 to December 2001, and Panel B reports the results from January 1945 to December 2001.

Panel A: January 1927–December 2001

	m	σ	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}	ρ_{11}	ρ_{12}	ρ_{24}	ρ_{36}	ρ_{48}	ρ_{60}	p_{ADF}
S_{val}	4.57	5.45	0.95	0.91	0.87	0.83	0.81	0.78	0.76	0.74	0.73	0.72	0.70	0.67	0.60	0.55	0.42	0.50	0.081
S_{grw}	4.32	1.94	0.97	0.94	0.92	0.89	0.88	0.86	0.84	0.83	0.81	0.80	0.79	0.78	0.77	0.63	0.57	0.40	0.766
$S_{\log}$	2.66	0.57	0.98	0.97	0.96	0.95	0.94	0.92	0.92	0.91	0.90	0.90	0.89	0.88	0.85	0.81	0.76	0.72	0.554
$r_{\text{ew}}^{\text{mkt}}$	0.95	7.32	0.19	0.01	-0.11	-0.06	0.00	-0.04	0.01	0.03	0.15	0.07	-0.02	0.01	0.01	0.03	-0.01	0.02	
$r_{\text{vw}}^{\text{mkt}}$	0.66	5.46	0.10	-0.02	-0.12	0.01	0.08	-0.03	0.01	0.04	0.09	0.01	-0.03	0.00	0.03	0.02	-0.02	0.02	
$r_{\text{ew}}^{\text{sml}}$	1.49	10.44	0.22	0.03	-0.07	-0.08	-0.05	-0.03	0.03	0.02	0.17	0.09	0.03	0.07	0.03	0.05	0.02	0.05	
$r_{\text{vw}}^{\text{sml}}$	1.07	9.50	0.20	0.01	-0.08	-0.08	-0.05	-0.03	0.03	0.01	0.17	0.08	0.02	0.04	0.01	0.04	-0.01	0.02	

Panel B: January 1945–December 2001

	m	σ	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}	ρ_{11}	ρ_{12}	ρ_{24}	ρ_{36}	ρ_{48}	ρ_{60}	p_{ADF}
S_{val}	2.32	1.13	0.96	0.92	0.88	0.85	0.82	0.80	0.77	0.75	0.72	0.70	0.69	0.68	0.60	0.53	0.66	0.56	0.000
S_{grw}	4.62	2.05	0.97	0.94	0.93	0.91	0.89	0.87	0.86	0.84	0.83	0.82	0.80	0.79	0.81	0.67	0.63	0.42	0.821
$S_{\log}$	2.39	0.30	0.96	0.92	0.89	0.86	0.83	0.80	0.77	0.75	0.73	0.70	0.69	0.67	0.59	0.41	0.36	0.24	0.022
$r_{\text{ew}}^{\text{mkt}}$	0.78	4.84	0.14	-0.01	-0.03	-0.01	0.03	-0.02	-0.03	-0.10	0.01	-0.01	0.01	0.05	0.02	0.01	0.06	0.02	
$r_{\text{vw}}^{\text{mkt}}$	0.66	4.11	0.04	-0.03	-0.01	0.02	0.09	-0.04	0.00	-0.04	0.01	-0.01	-0.01	0.03	0.02	-0.02	0.01	-0.03	
$r_{\text{ew}}^{\text{sml}}$	1.01	6.17	0.22	0.02	-0.03	0.01	0.00	0.00	0.00	-0.10	-0.02	-0.00	0.05	0.12	0.08	0.09	0.14	0.10	
$r_{\text{vw}}^{\text{sml}}$	0.84	5.95	0.20	0.01	-0.05	-0.00	-0.00	0.02	0.03	-0.09	-0.01	-0.01	0.03	0.05	0.03	0.04	0.07	0.04	

Table 3.2 Predictive Regressions Using the Value Spread

This table reports univariate, predictive regressions of returns onto the value spread. The value spread is measured as the book-to-market ratio of value stocks (portfolio ten) minus that of growth stocks (portfolio one) in the ten deciles sorted on book-to-market. Table 3.1 contains detailed definitions for stock returns used as dependent variables. We report the slope β_τ, p-values associated with Newey-West t-statistics p_{NW}, biases in the slope b_{NK} and p-values of the slopes p_{NK} using Nelson and Kim's (1993) method, biases in the slope b_{S} and p-values of the slopes p_{S} using Stambaugh's (1999) method, R^2, and the number of observations, T.

Panel A: January 1927 to December 2001

τ	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$
	β_τ				p_{NW}			
M	0.170	0.066	0.303	0.222	**0.043**	0.130	**0.023**	**0.038**
Q	0.617	0.257	1.102	0.818	0.079	0.131	0.059	0.074
Y	1.567	0.688	2.700	1.935	**0.000**	**0.007**	**0.001**	**0.001**
2Y	3.302	1.411	6.032	4.477	**0.000**	**0.002**	**0.000**	**0.000**
5Y	4.910	1.919	9.494	6.864	**0.001**	**0.004**	**0.001**	**0.001**
	b_{NK}				p_{NK}			
M	0.010	0.006	0.011	0.012	**0.003**	**0.047**	**0.000**	**0.003**
Q	0.033	0.019	0.054	0.052	**0.003**	**0.023**	**0.000**	**0.001**
Y	0.029	0.026	0.060	0.038	**0.007**	**0.044**	**0.001**	**0.010**
2Y	-0.023	-0.002	-0.045	-0.000	**0.000**	**0.000**	**0.000**	**0.000**
5Y	0.007	-0.002	-0.061	-0.064	**0.000**	**0.000**	**0.000**	**0.000**
	b_{S}				p_{S}			
M	0.007	0.007	0.006	0.012	**0.000**	0.054	**0.000**	**0.000**
Q	0.027	0.024	0.038	0.056	**0.002**	**0.033**	**0.000**	**0.002**
Y	0.037	0.063	0.045	0.043	**0.004**	0.051	**0.000**	**0.004**
2Y	-0.008	0.018	-0.021	-0.027	**0.000**	**0.000**	**0.000**	**0.000**
5Y	-0.026	0.031	-0.066	-0.025	**0.000**	**0.000**	**0.000**	**0.000**
	R^2				T			
M	0.016	0.004	0.025	0.016	899	899	899	899
Q	0.049	0.018	0.073	0.051	299	299	299	299
Y	0.121	0.047	0.159	0.109	74	74	74	74
2Y	0.217	0.078	0.293	0.221	292	292	292	292
5Y	0.290	0.087	0.325	0.272	280	280	280	280

Panel B: January 1945 to December 2001

τ	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$
	β_τ				p_{NW}			
M	0.330	0.225	0.267	0.316	**0.039**	0.058	0.152	0.102
Q	1.104	0.779	0.999	1.121	**0.023**	**0.036**	0.097	0.061
Y	1.787	1.032	1.968	2.007	0.213	0.277	0.244	0.228
2Y	3.216	1.701	3.353	3.713	0.107	0.209	0.159	0.123
5Y	10.704	8.484	11.100	13.247	**0.001**	**0.005**	**0.010**	**0.004**
	b_{NK}				p_{NK}			
M	0.025	0.017	0.026	0.024	0.080	0.117	0.199	0.131
Q	0.095	0.066	0.125	0.098	0.079	0.121	0.192	0.133
Y	0.387	0.213	0.572	0.361	0.298	0.336	0.355	0.306
2Y	0.064	0.049	0.061	-0.016	**0.038**	0.123	0.096	0.053
5Y	0.040	0.056	-0.181	-0.024	**0.000**	**0.000**	**0.001**	**0.000**
	b_{S}				p_{S}			
M	0.023	0.025	0.022	0.014	0.075	0.131	0.175	0.126
Q	0.083	0.080	0.126	0.113	0.079	0.093	0.175	0.135
Y	0.241	0.251	0.397	0.427	0.268	0.337	0.327	0.325
2Y	0.059	0.088	0.022	0.080	**0.029**	0.134	0.079	0.050
5Y	-0.161	0.065	-0.173	0.043	**0.000**	**0.000**	**0.000**	**0.000**
	R^2				T			
M	0.006	0.004	0.002	0.004	683	683	683	683
Q	0.017	0.013	0.008	0.011	227	227	227	227
Y	0.011	0.006	0.006	0.008	56	56	56	56
2Y	0.026	0.010	0.014	0.020	220	220	220	220
5Y	0.171	0.106	0.072	0.116	208	208	208	208

Table 3.3 Predictive Regressions Using the Growth Spread

This table reports univariate, predictive regressions of returns on the growth spread. The growth spread is measured as the market-to-book ratio of growth stocks (portfolio one) minus that of value stocks (portfolio ten) in the ten deciles sorted on book-to-market. Table 3.1 contains detailed definitions for stock returns used as dependent variables. We report the slope β_τ, p-values associated with Newey-West t-statistics p_{NW}, biases in the slope b_{NK} and p-values of the slopes p_{NK} using Nelson and Kim's (1993) method, biases in the slope b_{S} and p-values of the slopes p_{S} using Stambaugh's (1999) method, R^2, and the number of observations, T.

	Panel A: January 1927 to December 2001								Panel B: January 1945 to December 2001							
τ	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$
	β_τ				p_{NW}				β_τ				p_{NW}			
M	-0.374	-0.228	-0.461	-0.376	**0.998**	**0.995**	**0.985**	**0.979**	-0.205	-0.170	-0.105	-0.129	**0.995**	**0.992**	0.781	0.844
Q	-1.206	-0.679	-1.572	-1.291	**0.994**	**0.988**	**0.976**	**0.974**	-0.612	-0.457	-0.377	-0.441	**0.994**	**0.976**	0.826	0.873
Y	-2.893	-1.293	-4.175	-3.195	**0.990**	0.949	**0.963**	**0.957**	-1.510	-0.950	-1.299	-1.296	**0.996**	0.918	0.888	0.888
2Y	-8.406	-4.575	-11.308	-9.506	**1.000**	**0.998**	**0.998**	**0.997**	-4.264	-2.658	-3.489	-3.562	**0.999**	**0.976**	**0.955**	**0.955**
5Y	-14.901	-9.882	-17.311	-16.843	**1.000**	**1.000**	**0.994**	**0.999**	-10.099	-8.054	-6.824	-9.141	**1.000**	**0.995**	0.936	**0.983**
	b_{NK}				p_{NK}				b_{NK}				p_{NK}			
M	-0.021	-0.025	-0.029	-0.014	**0.993**	**0.967**	**0.990**	**0.979**	-0.020	-0.017	-0.023	-0.015	**0.970**	**0.955**	0.754	0.833
Q	-0.081	-0.077	-0.148	-0.098	**0.987**	**0.953**	**0.970**	**0.963**	-0.069	-0.066	-0.086	-0.081	0.944	0.921	0.750	0.817
Y	-0.264	-0.156	-0.200	-0.065	**0.950**	0.843	0.939	0.933	-0.190	-0.162	-0.310	-0.194	0.843	0.806	0.732	0.756
2Y	0.005	-0.019	-0.060	0.046	**1.000**	**1.000**	**1.000**	**1.000**	-0.044	-0.019	-0.036	-0.026	**1.000**	**1.000**	**0.999**	**0.999**
5Y	-0.048	-0.026	0.060	0.047	**1.000**	**1.000**	**1.000**	**1.000**	0.095	0.018	-0.063	-0.018	**1.000**	**1.000**	**0.998**	**1.000**
	b_{S}				p_{S}				b_{S}				p_{S}			
M	-0.018	-0.019	-0.019	-0.025	**0.994**	**0.972**	**0.992**	**0.978**	-0.021	-0.021	-0.018	-0.018	**0.968**	**0.952**	0.791	0.837
Q	-0.061	-0.077	-0.072	-0.114	**0.987**	**0.954**	**0.982**	**0.970**	-0.080	-0.060	-0.069	-0.082	0.945	0.935	0.765	0.807
Y	-0.093	-0.157	-0.129	-0.198	**0.959**	0.845	**0.953**	0.927	-0.047	-0.152	-0.262	-0.249	0.876	0.819	0.739	0.754
2Y	-0.017	-0.003	-0.041	0.063	**1.000**	**1.000**	**1.000**	**1.000**	-0.043	-0.064	-0.092	-0.083	**1.000**	**1.000**	**0.996**	**1.000**
5Y	0.117	-0.053	0.039	0.002	**1.000**	**1.000**	**1.000**	**1.000**	0.075	-0.035	0.018	0.060	**1.000**	**1.000**	**1.000**	**1.000**
	R^2				T				R^2				T			
M	0.010	0.007	0.007	0.006	899	899	899	899	0.008	0.007	0.001	0.002	683	683	683	683
Q	0.023	0.015	0.018	0.016	299	299	299	299	0.017	0.015	0.004	0.006	227	227	227	227
Y	0.048	0.019	0.044	0.035	74	74	74	74	0.031	0.019	0.011	0.014	56	56	56	56
2Y	0.144	0.083	0.105	0.102	292	292	292	292	0.128	0.065	0.041	0.050	220	220	220	220
5Y	0.193	0.166	0.078	0.118	280	280	280	280	0.297	0.187	0.053	0.107	208	208	208	208

Table 3.4 Predictive Regressions with the Log Spread

This table reports univariate, predictive regressions of returns on the growth spread. The log spread is measured as the log book-to-market of value (portfolio ten) minus that of growth (portfolio one) in the ten deciles sorted on book-to-market. Table 3.1 contains detailed definitions for stock returns used as dependent variables. We report the slope β_τ, p-values associated with Newey-West t-statistics p_{NW}, biases in the slope b_{NK} and p-values of the slopes p_{NK} using Nelson and Kim's (1993) method, biases in the slope b_{S} and p-values of the slopes p_{S} using Stambaugh's (1999) method, R^2, and the number of observations, T.

	Panel A: January 1927 to December 2001								Panel B: January 1945 to December 2001							
τ	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$
	β_τ				p_{NW}				β_τ				p_{NW}			
M	0.675	-0.021	1.755	1.028	0.159	0.519	**0.045**	0.128	-0.309	-0.518	-0.121	-0.057	0.678	0.820	0.552	0.526
Q	2.331	0.212	5.558	3.337	0.143	0.439	0.055	0.128	0.006	-0.517	0.405	0.547	0.499	0.622	0.442	0.415
Y	9.237	1.676	20.568	12.348	**0.040**	0.319	**0.012**	**0.046**	-4.083	-4.788	-0.783	-1.901	0.786	0.914	0.530	0.582
2Y	16.543	1.415	41.024	24.413	**0.042**	0.413	**0.004**	**0.031**	-13.818	-12.915	-7.436	-7.824	0.935	**0.954**	0.700	0.723
5Y	41.568	6.710	98.084	62.719	**0.000**	0.163	**0.000**	**0.000**	2.818	-7.478	31.698	27.427	0.418	0.721	0.088	0.108
	b_{NK}				p_{NK}				b_{NK}				p_{NK}			
M	0.093	0.030	0.154	0.123	0.116	0.544	**0.016**	0.079	0.036	-0.002	0.024	-0.058	0.698	0.815	0.588	0.495
Q	0.379	0.251	0.581	0.469	0.137	0.515	**0.035**	0.098	0.146	-0.062	-0.010	0.147	0.503	0.591	0.462	0.447
Y	0.777	0.827	1.370	1.152	0.086	0.398	**0.017**	0.074	0.561	0.498	1.198	-0.164	0.701	0.772	0.537	0.566
2Y	0.085	0.145	0.530	0.157	**0.000**	0.348	**0.000**	**0.001**	0.213	-0.167	0.216	0.227	**0.989**	**0.993**	0.795	0.828
5Y	0.125	0.300	-0.347	-0.225	**0.000**	0.067	**0.000**	**0.000**	-0.029	-0.049	-0.127	0.613	0.371	0.829	**0.008**	**0.017**
	b_{S}				p_{S}				b_{S}				p_{S}			
M	0.080	0.040	0.087	0.113	0.115	0.562	**0.020**	0.079	-0.007	-0.015	-0.009	0.001	0.675	0.820	0.546	0.518
Q	0.369	0.211	0.524	0.678	0.118	0.484	**0.039**	0.129	0.152	0.005	-0.036	-0.092	0.525	0.603	0.433	0.422
Y	0.950	0.974	0.845	1.113	0.078	0.407	**0.017**	0.066	0.790	0.810	0.714	0.467	0.705	0.791	0.541	0.576
2Y	0.360	-0.119	0.320	0.307	**0.000**	0.314	**0.000**	**0.000**	0.263	0.017	-0.064	0.568	**0.991**	**0.991**	0.807	0.847
5Y	0.081	0.392	0.043	0.116	**0.000**	0.074	**0.000**	**0.000**	0.193	0.129	0.840	-0.124	0.377	0.819	**0.009**	**0.014**
	R^2				T				R^2				T			
M	0.003	0.000	0.009	0.004	899	899	899	899	0.000	0.001	0.000	0.000	683	683	683	683
Q	0.008	0.000	0.020	0.009	299	299	299	299	0.000	0.000	0.000	0.000	227	227	227	227
Y	0.043	0.003	0.094	0.045	74	74	74	74	0.004	0.009	0.000	0.001	56	56	56	56
2Y	0.059	0.001	0.146	0.071	292	292	292	292	0.029	0.033	0.004	0.005	220	220	220	220
5Y	0.227	0.012	0.379	0.248	280	280	280	280	0.001	0.005	0.035	0.030	208	208	208	208

Table 3.5 Cross Correlations

This table reports the cross-correlations for the equal-weighted market excess return r_{ew}^{mkt}; the value-weighted market excess return r_{vw}^{mkt}; the equal-weighted small firm (quintile) excess return r_{ew}^{sml}; the value-weighted small firm (quintile) excess return r_{vw}^{sml}; the value spread S_{val} (the book-to-market of value stocks minus that of growth stocks); the log spread, S_{log} (the log book-to-market of value stocks minus that of growth stocks; equivalently, the log market-to-book of growth stocks minus that of value stocks); the growth spread, S_{grw} (the market-to-book of growth stocks minus that of value stocks); the dividend yield, div; the default premium, def; the term premium, term; the short-term interest rate, rf; and the aggregate book-to-market, b/m. Stock returns are measured at the end of period t or the beginning of period $t+1$, and all the conditioning variables are measured at the beginning of period t. Panel A reports the results for the sample from January 1927 to December 2001, and Panel B reports the results for the sample from January 1945 to December 2001.

Panel A: January 1927–December 2001

	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	S_{val}	S_{grw}	S_{log}	div	def	term	rf	b/m
r_{ew}^{mkt}	1.00	0.92	0.93	0.95	0.13	-0.10	0.05	0.11	0.16	0.10	-0.08	0.14
r_{vw}^{mkt}		1.00	0.76	0.81	0.07	-0.08	-0.00	0.07	0.06	0.08	-0.07	0.10
r_{ew}^{sml}			1.00	0.98	0.16	-0.09	0.10	0.11	0.19	0.11	-0.11	0.15
r_{vw}^{sml}				1.00	0.13	-0.08	0.06	0.08	0.14	0.10	-0.09	0.13
S_{val}					1.00	-0.42	0.82	0.45	0.61	0.39	-0.52	0.85
S_{grw}						1.00	-0.09	-0.69	-0.30	-0.11	0.20	-0.75
S_{log}							1.00	0.30	0.48	0.37	-0.66	0.55
div								1.00	0.51	0.12	-0.31	0.59
def									1.00	0.34	-0.09	0.52
term										1.00	-0.54	0.30
rf											1.00	-0.42
b/m												1.00

Panel B: January 1945–December 2001

	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	S_{val}	S_{grw}	S_{log}	div	def	term	rf	b/m
r_{ew}^{mkt}	1.00	0.91	0.88	0.91	0.08	-0.09	-0.02	0.10	0.10	0.14	-0.10	0.06
r_{vw}^{mkt}		1.00	0.71	0.76	0.06	-0.08	-0.04	0.10	0.06	0.14	-0.12	0.07
r_{ew}^{sml}			1.00	0.97	0.05	-0.03	-0.01	0.05	0.08	0.14	-0.11	-0.00
r_{vw}^{sml}				1.00	0.06	-0.04	-0.00	0.06	0.07	0.14	-0.12	0.02
S_{val}					1.00	-0.53	0.46	0.51	-0.23	0.01	-0.50	0.73
S_{grw}						1.00	0.44	-0.80	-0.14	0.03	0.05	-0.88
S_{log}							1.00	-0.28	-0.41	0.02	-0.51	-0.16
div								1.00	0.15	-0.11	-0.11	0.76
def									1.00	0.06	0.63	0.09
term										1.00	-0.41	-0.02
rf											1.00	-0.15
b/m												1.00

Table 3.6 Predictive Regressions: The Value (Growth) Spread and the Aggregate Book-to-Market

This table reports bivariate, predictive regressions of returns onto the value (growth) spread S_v (S_g) and the aggregate book-to-market (b/m). We report the slopes $\beta_{S,\tau}$, $\beta_{\mathrm{b/m},\tau}$, and p-values of slopes $p_{S,\mathrm{NK}}$ and $p_{\mathrm{b/m,NK}}$ using Nelson and Kim's (1993) method.

	Panel A: January 1927–December 2001, S_{val} and b/m								Panel B: January 1927–December 2001, S_{grw} and b/m							
τ	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$
	$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$				$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$			
M	0.099	-0.430	1.237	0.627	0.418	0.906	**0.023**	0.113	0.163	-0.035	0.618	0.439	0.286	0.511	0.077	0.137
Q	0.021	-1.488	3.299	1.394	0.516	0.906	0.100	0.262	1.470	0.646	3.447	2.582	0.084	0.195	**0.033**	**0.050**
Y	6.682	1.803	15.788	10.074	0.135	0.334	**0.033**	0.097	1.326	0.995	3.934	2.487	0.374	0.360	0.252	0.315
2Y	6.630	-4.418	30.153	15.646	0.059	0.939	**0.000**	**0.003**	-1.486	-1.000	4.529	2.233	0.689	0.694	0.131	0.273
5Y	22.625	-5.622	78.627	44.076	**0.000**	0.916	**0.000**	**0.000**	-8.455	-10.155	5.828	-4.286	**0.977**	**0.999**	0.221	0.788
	$\beta_{\mathrm{b/m},\tau}$				$p_{\mathrm{b/m,NK}}$				$\beta_{\mathrm{b/m},\tau}$				$p_{\mathrm{b/m,NK}}$			
M	0.972	0.932	0.490	0.687	**0.015**	**0.005**	0.175	0.116	1.179	0.540	2.006	1.551	**0.000**	**0.020**	**0.000**	**0.000**
Q	4.012	3.430	3.344	3.721	**0.015**	**0.003**	0.072	**0.042**	5.128	2.642	8.732	6.838	**0.001**	**0.007**	**0.000**	**0.001**
Y	3.207	2.713	0.591	1.858	0.315	0.267	0.486	0.405	9.962	5.003	17.139	12.400	**0.014**	0.058	**0.004**	**0.015**
2Y	13.912	14.366	4.467	11.135	**0.000**	**0.000**	0.227	**0.021**	18.524	9.825	33.769	26.255	**0.000**	**0.000**	**0.000**	**0.000**
5Y	6.237	19.125	-28.175	-5.638	0.085	**0.000**	**0.997**	0.760	19.668	6.825	44.250	29.438	**0.000**	**0.019**	**0.000**	**0.000**

	Panel C: January 1945–December 2001, S_{val} and b/m								Panel D: January 1945–December 2001, S_{grw} and b/m							
τ	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$	$r^{\mathrm{mkt}}_{\mathrm{ew}}$	$r^{\mathrm{mkt}}_{\mathrm{vw}}$	$r^{\mathrm{sml}}_{\mathrm{ew}}$	$r^{\mathrm{sml}}_{\mathrm{vw}}$
	$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$				$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$			
M	0.338	0.099	0.637	0.607	0.165	0.350	0.072	0.069	-0.718	-0.427	-0.954	-0.800	**0.990**	0.944	**0.994**	**0.988**
Q	1.169	0.512	2.048	1.950	0.169	0.268	0.106	0.094	-1.755	-0.736	-2.351	-1.987	**0.956**	0.768	**0.953**	0.928
Y	-0.152	-1.043	2.475	1.945	0.535	0.618	0.353	0.364	-3.655	-0.993	-5.964	-4.803	0.839	0.611	0.849	0.822
2Y	-3.263	-3.992	1.587	1.353	0.849	0.933	0.323	0.388	-8.904	-3.370	-11.726	-9.985	**1.000**	**0.959**	**1.000**	**0.998**
5Y	2.024	-0.871	13.355	11.772	0.305	0.576	**0.010**	**0.018**	-11.355	-4.755	-11.458	-11.762	**0.998**	0.926	**0.986**	**0.992**
	$\beta_{\mathrm{b/m},\tau}$				$p_{\mathrm{b/m,NK}}$				$\beta_{\mathrm{b/m},\tau}$				$p_{\mathrm{b/m,NK}}$			
M	0.046	0.213	-0.464	-0.346	0.507	0.241	0.921	0.881	-0.339	-0.091	-0.840	-0.609	0.897	0.666	**0.996**	**0.961**
Q	0.114	0.508	-1.250	-0.927	0.483	0.335	0.831	0.792	-0.560	0.241	-1.801	-1.232	0.731	0.374	0.926	0.873
Y	2.799	2.894	-0.455	0.311	0.250	0.226	0.548	0.512	-0.423	1.271	-3.680	-2.322	0.594	0.387	0.805	0.705
2Y	9.228	7.885	3.007	3.867	**0.000**	**0.001**	0.215	0.121	-0.901	1.972	-5.892	-3.711	0.672	0.159	**0.970**	0.913
5Y	13.868	14.326	-0.780	4.681	**0.000**	**0.000**	0.571	0.194	5.739	9.647	-0.585	3.443	**0.029**	**0.002**	0.566	0.266

Table 3.7 Predictive Regressions: The Value (Growth) Spread and the Term Premium

This table reports bivariate, predictive regressions of returns onto the value (growth) spread S_v (S_g) and the term premium (term). We report the slopes $\beta_{S,\tau}$, $\beta_{\text{term},\tau}$, and p-values of slopes $p_{S,\text{NK}}$ and $p_{\text{term},\text{NK}}$ using Nelson and Kim's (1993) method.

	Panel A: January 1927–December 2001, S_{val} and term								Panel B: January 1927–December 2001, S_{grw} and term							
τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$				$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	0.748	0.212	1.413	0.993	**0.006**	0.183	**0.000**	**0.006**	-0.650	-0.395	-0.772	-0.631	**0.985**	**0.959**	**0.976**	0.946
Q	3.252	1.197	5.896	4.306	**0.004**	0.074	**0.000**	**0.004**	-2.187	-1.219	-2.773	-2.286	**0.978**	0.937	**0.959**	0.940
Y	7.864	3.027	14.015	9.539	**0.014**	0.129	**0.007**	**0.019**	-5.251	-2.252	-7.469	-5.656	0.921	0.761	0.921	0.876
2Y	17.477	6.421	32.737	23.654	**0.000**	**0.000**	**0.000**	**0.000**	-14.227	-7.612	-18.799	-15.824	**1.000**	**1.000**	**1.000**	**1.000**
5Y	25.867	5.886	53.083	36.005	**0.000**	**0.004**	**0.000**	**0.000**	-21.286	-13.526	-24.081	-23.516	**1.000**	**1.000**	**1.000**	**1.000**
	$\beta_{\text{term},\tau}$				$p_{\text{term},\text{NK}}$				$\beta_{\text{term},\tau}$				$p_{\text{term},\text{NK}}$			
M	0.452	0.381	0.615	0.559	**0.026**	**0.020**	**0.032**	0.052	0.672	0.420	1.080	0.876	**0.003**	**0.023**	**0.002**	**0.005**
Q	0.496	0.622	0.660	0.677	0.283	0.167	0.282	0.254	1.507	0.946	2.628	2.085	**0.047**	0.063	**0.024**	**0.029**
Y	4.068	2.871	5.830	5.469	0.119	0.121	0.097	0.107	6.579	3.812	10.509	8.592	**0.027**	0.050	**0.032**	**0.020**
2Y	2.901	3.924	3.233	4.050	0.111	**0.009**	0.190	0.088	8.094	5.559	13.859	11.471	**0.000**	**0.001**	**0.000**	**0.000**
5Y	5.549	13.000	2.868	8.161	**0.034**	**0.000**	0.293	**0.035**	13.122	13.697	20.727	19.425	**0.000**	**0.000**	**0.000**	**0.000**

	Panel C: January 1945–December 2001, S_{val} and term								Panel D: January 1945–December 2001, S_{grw} and term							
τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$				$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	0.363	0.246	0.289	0.345	0.104	0.154	0.173	0.157	-0.441	-0.366	-0.243	-0.291	**0.966**	**0.960**	0.792	0.838
Q	1.255	0.886	1.137	1.275	0.089	0.113	0.150	0.146	-1.310	-0.985	-0.844	-0.970	0.945	0.929	0.769	0.815
Y	2.064	1.254	2.274	2.295	0.282	0.356	0.312	0.307	-3.336	-2.118	-2.880	-2.867	0.856	0.773	0.731	0.741
2Y	3.700	1.972	3.853	4.269	**0.027**	0.116	0.101	**0.049**	-8.151	-5.176	-6.638	-6.795	**1.000**	**0.999**	**0.999**	**1.000**
5Y	12.330	9.796	12.786	15.259	**0.000**	**0.001**	**0.002**	**0.000**	-16.353	-13.294	-11.078	-14.825	**1.000**	**1.000**	**1.000**	**1.000**
	$\beta_{\text{term},\tau}$				$p_{\text{term},\text{NK}}$				$\beta_{\text{term},\tau}$				$p_{\text{term},\text{NK}}$			
M	0.652	0.571	0.872	0.800	**0.001**	**0.000**	**0.000**	**0.000**	0.671	0.586	0.883	0.814	**0.000**	**0.001**	**0.000**	**0.000**
Q	1.525	1.351	2.250	2.003	**0.011**	**0.003**	**0.004**	**0.005**	1.559	1.377	2.272	2.028	**0.012**	**0.006**	**0.009**	**0.007**
Y	2.576	2.767	2.858	2.422	0.156	0.089	0.220	0.252	2.529	2.740	2.794	2.357	0.148	0.095	0.215	0.237
2Y	0.473	3.321	-0.643	-0.030	0.354	**0.003**	0.596	0.482	0.710	3.474	-0.454	0.162	0.284	**0.002**	0.549	0.443
5Y	3.753	11.173	3.449	4.174	**0.028**	**0.000**	0.140	0.078	4.222	11.556	3.755	4.588	**0.022**	**0.000**	0.115	0.071

Table 3.8 Predictive Regressions: The Value (Growth) Spread and the Default Premium

This table reports bivariate, predictive regressions of returns onto the value (growth) spread S_v (S_g) and the default premium (def). We report the slopes $\beta_{S,\tau}$, $\beta_{\text{def},\tau}$, and p-values of slopes $p_{S,\text{NK}}$ and $p_{\text{def,NK}}$ using Nelson and Kim's (1993) method.

Panel A: January 1927–December 2001, S_{val} and def

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	0.363	0.232	0.681	0.597	0.105	0.189	0.067	0.071
Q	1.995	1.106	3.714	3.101	**0.046**	0.089	**0.019**	**0.018**
Y	5.406	4.137	7.990	7.354	0.059	**0.048**	**0.045**	**0.049**
2Y	11.105	7.278	20.790	18.019	**0.000**	**0.000**	**0.000**	**0.000**
5Y	14.153	6.759	30.711	24.047	**0.000**	**0.002**	**0.000**	**0.000**
	$\beta_{\text{def},\tau}$				$p_{\text{def,NK}}$			
M	0.916	0.210	1.585	1.000	**0.007**	0.190	**0.001**	**0.008**
Q	2.306	0.527	3.878	2.335	**0.033**	0.291	**0.022**	0.093
Y	6.732	0.025	13.806	7.191	0.082	0.605	**0.015**	0.115
2Y	11.976	1.075	21.081	11.516	**0.000**	0.316	**0.000**	**0.001**
5Y	22.348	6.761	37.825	24.380	**0.000**	**0.003**	**0.000**	**0.001**

Panel B: January 1927–December 2001, S_{grw} and def

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	-0.426	-0.370	-0.330	-0.356	0.917	0.945	0.781	0.825
Q	-1.425	-1.057	-1.341	-1.369	0.900	0.916	0.818	0.822
Y	-3.221	-2.095	-3.237	-3.360	0.806	0.784	0.709	0.754
2Y	-10.476	-7.212	-11.299	-11.396	**1.000**	**1.000**	**0.998**	**1.000**
5Y	-15.330	-13.106	-11.523	-16.078	**1.000**	**1.000**	**0.963**	**1.000**
	$\beta_{\text{def},\tau}$				$p_{\text{def,NK}}$			
M	1.012	0.243	1.905	1.261	**0.004**	0.150	**0.000**	**0.001**
Q	3.135	0.907	5.813	3.876	**0.009**	0.143	**0.004**	**0.015**
Y	9.003	1.876	17.628	10.592	**0.024**	0.341	**0.012**	**0.030**
2Y	15.842	3.507	30.772	19.442	**0.000**	**0.040**	**0.000**	**0.000**
5Y	26.851	7.294	53.676	34.819	**0.000**	**0.004**	**0.000**	**0.000**

Panel C: January 1945–December 2001, S_{val} and def

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	0.509	0.325	0.434	0.478	**0.030**	0.084	0.116	0.084
Q	1.655	1.079	1.552	1.628	**0.032**	0.070	0.111	0.078
Y	3.271	1.650	3.502	3.293	0.145	0.258	0.229	0.226
2Y	4.302	1.746	3.693	4.056	**0.018**	0.142	0.112	0.061
5Y	13.389	10.071	11.018	13.726	**0.000**	**0.000**	**0.004**	**0.000**
	$\beta_{\text{def},\tau}$				$p_{\text{def,NK}}$			
M	0.595	0.311	0.580	0.527	**0.001**	**0.044**	**0.027**	**0.033**
Q	1.660	0.805	1.726	1.468	**0.010**	0.076	0.051	0.050
Y	5.138	2.046	5.272	4.304	0.071	0.242	0.148	0.181
2Y	2.453	-0.852	-0.662	-0.867	**0.048**	0.676	0.599	0.614
5Y	3.791	1.084	-6.231	-5.391	**0.047**	0.323	**0.960**	0.936

Panel D: January 1945–December 2001, S_{grw} and def

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	-0.358	-0.320	-0.149	-0.209	0.936	0.948	0.664	0.756
Q	-1.108	-0.886	-0.598	-0.774	0.904	0.879	0.680	0.736
Y	-2.748	-1.884	-2.268	-2.392	0.808	0.728	0.672	0.709
2Y	-8.091	-5.347	-7.003	-7.185	**1.000**	**0.999**	**0.996**	**0.999**
5Y	-16.384	-13.244	-12.011	-15.761	**1.000**	**1.000**	**0.999**	**1.000**
	$\beta_{\text{def},\tau}$				$p_{\text{def,NK}}$			
M	0.426	0.191	0.459	0.387	**0.018**	0.135	0.052	0.070
Q	1.099	0.415	1.263	0.962	0.064	0.227	0.110	0.157
Y	3.905	1.354	4.045	3.115	0.125	0.359	0.251	0.279
2Y	0.276	-2.022	-2.539	-2.859	0.398	0.885	0.828	0.877
5Y	-1.617	-3.076	-10.534	-10.832	0.752	0.907	**0.995**	**1.000**

Table 3.9 Predictive Regressions: The Value (Growth) Spread and the Dividend Yield

This table reports bivariate, predictive regressions of returns onto the value (growth) spread S_v (S_g) and the dividend yield (div). We report the slopes $\beta_{S,\tau}$, $\beta_{\text{div},\tau}$, and p-values of slopes $p_{S,\text{NK}}$ and $p_{\text{div,NK}}$ using Nelson and Kim's (1993) method.

Panel A: January 1927–December 2001, S_{val} and div

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	0.709	0.238	1.442	1.090	**0.010**	0.126	**0.000**	**0.006**
Q	2.388	0.847	4.810	3.615	**0.023**	0.133	**0.005**	**0.014**
Y	6.417	2.857	11.838	9.115	**0.038**	0.140	**0.014**	**0.025**
2Y	10.004	3.545	21.962	16.420	**0.000**	**0.038**	**0.000**	**0.000**
5Y	13.576	1.762	35.196	22.586	**0.001**	0.244	**0.000**	**0.000**
	$\beta_{\text{div},\tau}$				$p_{\text{div,NK}}$			
M	0.478	0.273	0.467	0.267	0.136	0.243	0.244	0.396
Q	2.250	1.259	2.859	2.031	0.069	0.144	0.104	0.167
Y	6.458	2.751	9.474	5.452	0.113	0.241	0.115	0.225
2Y	18.363	9.403	25.685	18.808	**0.000**	**0.000**	**0.000**	**0.000**
5Y	31.603	20.104	41.550	36.295	**0.000**	**0.000**	**0.000**	**0.000**

Panel B: January 1927–December 2001, S_{grw} and div

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	-0.335	-0.341	-0.236	-0.392	0.878	0.925	0.718	0.865
Q	-0.114	-0.370	0.774	0.032	0.539	0.664	0.280	0.460
Y	1.294	0.384	3.736	0.436	0.389	0.485	0.284	0.468
2Y	-0.316	-1.837	5.154	0.132	0.573	0.860	0.087	0.503
5Y	-1.355	-4.544	10.425	1.867	0.678	**0.976**	0.056	0.389
	$\beta_{\text{div},\tau}$				$p_{\text{div,NK}}$			
M	0.568	0.147	0.957	0.491	0.094	0.425	0.058	0.210
Q	3.293	1.404	5.644	3.750	**0.012**	0.096	**0.004**	**0.019**
Y	10.385	4.365	17.661	10.049	**0.015**	0.118	**0.008**	0.073
2Y	22.846	9.871	39.324	26.588	**0.000**	**0.000**	**0.000**	**0.000**
5Y	37.027	18.273	63.697	47.710	**0.000**	**0.000**	**0.000**	**0.000**

Panel C: January 1945–December 2001, S_{val} and div

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	0.162	0.060	0.206	0.248	0.251	0.398	0.253	0.194
Q	0.540	0.275	0.648	0.800	0.251	0.350	0.305	0.210
Y	-1.809	-1.967	-0.834	-0.764	0.711	0.787	0.589	0.574
2Y	-2.686	-3.236	-0.620	-0.180	0.895	**0.957**	0.583	0.529
5Y	4.259	2.198	9.794	10.708	0.054	0.189	**0.020**	**0.000**
	$\beta_{\text{div},\tau}$				$p_{\text{div,NK}}$			
M	0.409	0.377	0.185	0.211	0.201	0.205	0.520	0.430
Q	1.370	1.171	0.934	0.908	0.206	0.194	0.393	0.397
Y	7.284	5.997	5.769	5.715	0.102	0.093	0.236	0.206
2Y	11.951	9.720	8.378	8.329	**0.000**	**0.000**	**0.007**	**0.004**
5Y	16.250	15.256	6.011	9.152	**0.000**	**0.000**	0.091	**0.008**

Panel D: January 1945–December 2001, S_{grw} and div

τ	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$	$r^{\text{mkt}}_{\text{ew}}$	$r^{\text{mkt}}_{\text{vw}}$	$r^{\text{sml}}_{\text{ew}}$	$r^{\text{sml}}_{\text{vw}}$
	$\beta_{S,\tau}$				$p_{S,\text{NK}}$			
M	-0.074	-0.060	0.046	0.014	0.659	0.609	0.453	0.470
Q	0.116	0.258	0.610	0.368	0.480	0.353	0.318	0.361
Y	4.082	4.485	3.268	3.249	0.108	**0.032**	0.236	0.211
2Y	-0.240	2.477	-1.219	-1.208	0.579	**0.048**	0.703	0.709
5Y	-6.743	-3.114	-6.601	-8.969	**0.999**	0.918	**0.953**	**0.994**
	$\beta_{\text{div},\tau}$				$p_{\text{div,NK}}$			
M	0.433	0.360	0.328	0.349	0.212	0.222	0.355	0.294
Q	1.743	1.518	1.754	1.616	0.161	0.135	0.213	0.219
Y	9.518	8.462	7.874	7.841	**0.035**	**0.025**	0.143	0.146
2Y	10.332	9.882	7.116	7.310	**0.000**	**0.000**	**0.014**	**0.006**
5Y	13.695	14.190	6.299	8.254	**0.000**	**0.000**	0.063	**0.025**

Table 3.10 Predictive Regressions: The Value (Growth) Spread and the Short Term Interest Rate

This table reports bivariate, predictive regressions of returns onto the value (growth) spread S_v (S_g) and the short term interest rate (rf). We report the slopes $\beta_{S,\tau}$, $\beta_{\mathrm{rf},\tau}$, and p-values of slopes $p_{S,\mathrm{NK}}$ and $p_{\mathrm{rf},\mathrm{NK}}$ using Nelson and Kim's (1993) method.

	Panel A: January 1927–December 2001, S_{val} and rf								Panel B: January 1927–December 2001, S_{grw} and rf							
τ	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$
	$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$				$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$			
M	0.833	0.230	1.469	1.048	**0.005**	0.142	**0.000**	**0.004**	-0.628	-0.383	-0.696	-0.579	**0.982**	0.945	0.934	**0.954**
Q	3.582	1.283	6.220	4.638	**0.002**	0.053	**0.000**	**0.000**	-2.131	-1.183	-2.567	-2.164	**0.960**	0.921	0.943	0.926
Y	9.236	3.411	15.102	10.688	**0.007**	0.095	**0.008**	**0.015**	-4.996	-2.031	-6.574	-5.108	0.899	0.750	0.886	0.843
2Y	19.909	8.097	34.440	25.896	**0.000**	**0.000**	**0.000**	**0.000**	-14.204	-7.793	-17.834	-15.412	**1.000**	**1.000**	**1.000**	**1.000**
5Y	25.271	7.928	46.566	32.787	**0.000**	**0.000**	**0.000**	**0.000**	-20.101	-13.754	-19.942	-21.044	**1.000**	**1.000**	**1.000**	**1.000**
	$\beta_{\mathrm{rf},\tau}$				$p_{\mathrm{rf},\mathrm{NK}}$				$\beta_{\mathrm{rf},\tau}$				$p_{\mathrm{rf},\mathrm{NK}}$			
M	-0.177	-0.253	-0.357	-0.314	0.756	0.904	0.842	0.843	-0.481	-0.295	-0.976	-0.740	**0.972**	0.928	**0.994**	**0.985**
Q	0.268	-0.300	0.131	0.135	0.399	0.679	0.437	0.461	-1.154	-0.726	-2.563	-1.825	0.897	0.855	**0.961**	0.930
Y	-0.418	-1.392	-2.250	-1.868	0.559	0.729	0.654	0.631	-4.170	-2.734	-8.757	-6.367	0.879	0.838	0.948	0.900
2Y	2.531	0.284	0.860	1.290	0.190	0.463	0.467	0.357	-5.001	-2.398	-13.447	-9.088	**0.977**	0.906	**0.999**	**0.997**
5Y	-5.403	-5.930	-14.946	-12.531	**0.961**	**0.990**	**0.991**	**0.996**	-14.526	-7.389	-34.978	-25.315	**1.000**	**0.998**	**1.000**	**1.000**
	Panel C: January 1945–December 2001, S_{val} and rf								Panel D: January 1945–December 2001, S_{grw} and rf							
τ	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$	$r_{\mathrm{ew}}^{\mathrm{mkt}}$	$r_{\mathrm{vw}}^{\mathrm{mkt}}$	$r_{\mathrm{ew}}^{\mathrm{sml}}$	$r_{\mathrm{vw}}^{\mathrm{sml}}$
	$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$				$\beta_{S,\tau}$				$p_{S,\mathrm{NK}}$			
M	0.166	0.009	-0.072	0.005	0.281	0.567	0.633	0.533	-0.397	-0.325	-0.182	-0.232	0.944	0.942	0.708	0.777
Q	0.891	0.372	0.293	0.536	0.166	0.349	0.448	0.347	-1.222	-0.899	-0.709	-0.845	0.927	0.889	0.711	0.787
Y	1.429	-0.528	1.281	1.111	0.351	0.651	0.424	0.449	-3.193	-1.875	-2.689	-2.659	0.835	0.726	0.697	0.712
2Y	4.757	0.801	4.270	4.104	**0.001**	0.325	0.087	0.057	-8.131	-4.981	-6.617	-6.718	**1.000**	**1.000**	**0.999**	**0.999**
5Y	12.737	7.016	9.830	12.113	**0.000**	**0.004**	**0.010**	**0.001**	-16.021	-12.594	-10.541	-14.206	**1.000**	**1.000**	**0.998**	**1.000**
	$\beta_{\mathrm{rf},\tau}$				$p_{\mathrm{rf},\mathrm{NK}}$				$\beta_{\mathrm{rf},\tau}$				$p_{\mathrm{rf},\mathrm{NK}}$			
M	-0.409	-0.483	-0.739	-0.696	**0.985**	**0.998**	**1.000**	**0.995**	-0.474	-0.472	-0.694	-0.688	**0.994**	**0.996**	**0.993**	**0.997**
Q	-0.731	-1.035	-1.700	-1.489	0.854	**0.960**	**0.969**	**0.962**	-1.123	-1.183	-1.817	-1.720	**0.952**	**0.985**	**0.969**	**0.971**
Y	-1.041	-3.368	-1.746	-2.180	0.625	0.915	0.664	0.713	-1.528	-2.985	-2.194	-2.546	0.715	0.893	0.680	0.742
2Y	2.145	-2.339	0.839	-0.335	0.135	0.948	0.424	0.593	0.070	-2.567	-1.047	-2.135	0.529	**0.970**	0.684	0.856
5Y	0.823	-5.420	-5.809	-6.179	0.383	**0.996**	0.946	**0.961**	-4.986	-8.467	-10.366	-11.745	**0.987**	**1.000**	**1.000**	**1.000**

Table 3.11 Multiple Regressions Using the Value Spread

This table reports multiple, predictive regressions of returns onto the value spread (S_v), the term premium (term), the default premium (def), the dividend yield (div), and the short-term Treasury bill rate (rf) across different horizons (τ), including monthly (M), quarterly (Q), annual (Y), two-year (2Y), and five-year (5Y) horizons. For regressions with two-year and five-year horizons, we use overlapping quarterly observations. We report the slopes and their corresponding p-values using Nelson and Kim's (1993) method. All p-values are one-sided p-values which are the estimated probabilities of obtaining a coefficient at least as large as the coefficient estimated from the actual data series. A p-value less than 0.05 implies the coefficient is significantly positive at 0.05 level, whereas a p-value greater than 0.95 implies the coefficient is significantly negative at 0.05 level. p-value is in bold if it is greater than 95% or less than 5%.

	Panel A: January 1927 to December 2001								Panel B: January 1945 to December 2001							
τ	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}
	$\beta_{S_{val},\tau}$				$p_{S_{val},NK}$				$\beta_{S_{val},\tau}$				$p_{S_{val},NK}$			
M	-0.01	0.00	-0.05	0.09	0.50	0.52	0.55	0.43	-0.10	-0.24	-0.24	-0.18	0.64	0.87	0.75	0.74
Q	1.55	0.74	2.63	2.55	0.10	0.18	0.07	0.06	0.14	-0.32	-0.28	-0.00	0.46	0.67	0.62	0.46
Y	3.40	3.13	2.46	4.10	0.15	0.12	0.22	0.13	-2.33	-3.94	-1.70	-2.10	0.73	0.93	0.67	0.71
2Y	9.90	7.75	14.92	15.93	**0.00**	**0.00**	**0.00**	**0.00**	-2.31	-5.05	-1.00	-1.31	0.84	**1.00**	0.63	0.67
5Y	2.69	1.14	4.35	5.46	0.21	0.33	0.25	0.14	4.49	0.27	6.44	6.96	**0.04**	0.45	0.07	**0.04**
	$\beta_{term,\tau}$				$p_{term,NK}$				$\beta_{term,\tau}$				$p_{term,NK}$			
M	0.21	0.35	-0.07	0.13	0.27	0.07	0.54	0.36	0.10	0.17	0.03	0.02	0.37	0.25	0.49	0.47
Q	0.66	0.89	0.24	0.71	0.27	0.18	0.41	0.33	0.41	0.55	0.33	0.41	0.35	0.27	0.40	0.40
Y	4.81	4.10	3.73	5.10	0.22	0.16	0.33	0.22	-0.74	-0.31	-1.74	-2.02	0.58	0.58	0.65	0.67
2Y	7.36	9.03	4.78	8.43	**0.01**	**0.00**	0.15	**0.04**	3.65	5.89	2.45	2.30	**0.05**	**0.00**	0.21	0.20
5Y	4.90	18.76	-10.78	3.75	0.11	**0.00**	0.89	0.28	7.88	15.52	6.61	7.40	**0.01**	**0.00**	0.09	**0.04**
	$\beta_{def,\tau}$				$p_{def,NK}$				$\beta_{def,\tau}$				$p_{def,NK}$			
M	0.95	0.09	2.10	1.30	**0.00**	0.33	**0.00**	**0.01**	1.19	0.74	1.58	1.44	**0.00**	**0.00**	**0.00**	**0.00**
Q	1.39	-0.25	3.49	1.61	0.16	0.59	0.07	0.21	2.96	1.69	4.20	3.50	**0.00**	**0.01**	**0.00**	**0.00**
Y	3.52	-2.56	12.82	5.38	0.28	0.81	0.06	0.26	7.50	4.35	9.79	8.64	**0.03**	0.08	**0.03**	**0.04**
2Y	0.72	-8.55	11.30	0.06	0.39	**1.00**	**0.01**	0.47	-3.84	-6.15	-6.30	-5.44	**0.96**	**1.00**	**0.98**	**0.97**
5Y	13.15	-8.15	42.44	18.20	**0.01**	**0.99**	**0.00**	**0.00**	-3.92	-7.18	-11.46	-11.07	0.93	**1.00**	**1.00**	**1.00**
	$\beta_{div,\tau}$				$p_{div,NK}$				$\beta_{div,\tau}$				$p_{div,NK}$			
M	0.18	0.26	-0.26	-0.18	0.55	0.37	0.87	0.84	0.24	0.33	-0.02	0.03	0.48	0.33	0.81	0.78
Q	1.92	1.44	1.79	1.63	0.19	0.22	0.29	0.30	0.88	1.06	0.33	0.44	0.48	0.33	0.67	0.66
Y	5.98	4.09	5.47	4.19	0.23	0.27	0.38	0.44	5.16	5.18	3.00	3.38	0.31	0.24	0.51	0.45
2Y	19.44	13.50	22.85	20.14	**0.00**	**0.00**	**0.00**	**0.00**	13.14	12.54	10.25	10.26	**0.00**	**0.00**	**0.00**	**0.00**
5Y	27.63	24.98	24.92	29.89	**0.00**	**0.00**	**0.00**	**0.00**	18.07	19.47	10.24	13.54	**0.00**	**0.00**	**0.02**	**0.00**
	$\beta_{rf,\tau}$				$p_{rf,NK}$				$\beta_{rf,\tau}$				$p_{rf,NK}$			
M	-0.36	-0.09	-1.08	-0.69	0.85	0.63	**0.98**	0.94	-1.23	-0.97	-1.81	-1.69	**1.00**	**1.00**	**1.00**	**1.00**
Q	0.27	0.31	-0.77	0.06	0.40	0.41	0.63	0.49	-2.79	-2.16	-4.57	-3.84	**1.00**	**0.99**	**1.00**	**0.99**
Y	1.44	1.83	-3.93	-0.71	0.45	0.35	0.67	0.52	-7.20	-7.16	-9.70	-9.56	0.91	0.93	0.91	0.90
2Y	7.24	8.36	1.13	6.76	**0.02**	**0.00**	0.39	0.07	3.92	2.39	4.37	2.50	0.06	0.14	0.13	0.25
5Y	-5.12	7.56	-31.91	-14.27	0.86	**0.01**	**1.00**	**0.97**	4.37	4.24	3.86	3.02	0.11	0.13	0.27	0.29

Table 3.12 Multiple Regressions Using the Growth Spread

This table reports multiple, predictive regressions of returns on the growth spread (S_g), the term premium (term), the default premium (def), the dividend yield (div), and the short-term Treasury bill rate (rf) across different horizons (τ), including monthly (M), quarterly (Q), annual (Y), two-year (2Y), and five-year (5Y) horizons. For regressions with two-year and five-year horizons, we use overlapping quarterly observations. We report the slopes and their corresponding p-values using Nelson and Kim's (1993) method. All p-values are one-sided p-values which are the estimated probabilities of obtaining a coefficient at least as large as the coefficient estimated from the actual data series. A p-value less than 0.05 implies the coefficient is significantly positive at 0.05 level, whereas a p-value greater than 0.95 implies the coefficient is significantly negative at 0.05 level. p-value is in bold if it is greater than 95% or less than 5%.

	Panel A: January 1927 to December 2001								Panel B: January 1945 to December 2001							
τ	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}	r_{ew}^{mkt}	r_{vw}^{mkt}	r_{ew}^{sml}	r_{vw}^{sml}
	$\beta_{S_{grw},\tau}$				$p_{S_{grw},NK}$				$\beta_{S_{grw},\tau}$				$p_{S_{grw},NK}$			
M	-0.44	-0.34	-0.49	-0.56	0.93	0.92	0.88	0.92	-0.11	-0.08	-0.03	-0.06	0.67	0.61	0.53	0.60
Q	-0.36	-0.34	0.16	-0.32	0.60	0.63	0.46	0.60	0.04	0.24	0.44	0.22	0.51	0.34	0.34	0.43
Y	0.94	0.69	2.32	-0.21	0.43	0.40	0.38	0.56	3.62	4.14	2.38	2.36	0.19	0.07	0.31	0.26
2Y	-0.33	-0.83	3.76	-0.13	0.56	0.66	0.18	0.52	0.64	3.68	-0.74	-0.87	0.37	**0.00**	0.58	0.63
5Y	-1.37	-2.82	8.58	1.73	0.66	0.87	0.13	0.43	-6.11	-1.45	-6.57	-8.97	**0.99**	0.75	**0.96**	**0.99**
	$\beta_{term,\tau}$				$p_{term,NK}$				$\beta_{term,\tau}$				$p_{term,NK}$			
M	0.17	0.32	-0.12	0.06	0.27	0.09	0.56	0.44	0.11	0.20	0.08	0.05	0.34	0.22	0.40	0.42
Q	0.50	0.79	0.07	0.49	0.36	0.21	0.46	0.35	0.39	0.64	0.43	0.43	0.34	0.24	0.40	0.37
Y	4.73	4.00	3.81	4.87	0.19	0.14	0.29	0.25	0.25	1.13	-1.04	-1.23	0.50	0.45	0.59	0.62
2Y	6.66	8.41	4.23	7.35	**0.02**	**0.00**	0.20	**0.05**	4.13	7.22	2.53	2.41	**0.03**	**0.00**	0.23	0.20
5Y	4.57	18.39	-10.21	3.54	0.13	**0.00**	0.88	0.27	6.32	15.29	4.65	5.06	**0.02**	**0.00**	0.15	0.13
	$\beta_{def,\tau}$				$p_{def,NK}$				$\beta_{def,\tau}$				$p_{def,NK}$			
M	1.00	0.13	2.13	1.42	**0.01**	0.29	**0.00**	**0.00**	1.19	0.72	1.55	1.43	**0.00**	**0.00**	**0.00**	**0.00**
Q	2.39	0.25	5.08	3.21	**0.05**	0.39	**0.01**	0.06	2.97	1.65	4.15	3.48	**0.00**	**0.02**	**0.00**	**0.00**
Y	5.33	-0.87	13.94	7.72	0.15	0.67	**0.04**	0.14	6.92	3.63	9.40	8.24	**0.04**	0.13	**0.04**	**0.05**
2Y	6.80	-3.72	19.97	9.80	**0.02**	0.93	**0.00**	**0.01**	-4.02	-6.76	-6.29	-5.44	**0.96**	**1.00**	**0.97**	**0.97**
5Y	14.92	-7.24	44.47	21.43	**0.00**	**0.98**	**0.00**	**0.00**	-3.13	-7.05	-10.52	-9.90	0.89	**1.00**	**1.00**	**1.00**
	$\beta_{div,\tau}$				$p_{div,NK}$				$\beta_{div,\tau}$				$p_{div,NK}$			
M	-0.16	0.00	-0.64	-0.60	0.91	0.78	**0.97**	**0.99**	0.10	0.17	-0.14	-0.10	0.73	0.59	0.92	0.91
Q	1.65	1.18	1.91	1.38	0.26	0.26	0.30	0.39	0.97	1.13	0.59	0.62	0.46	0.34	0.62	0.60
Y	6.75	4.65	7.29	4.07	0.21	0.22	0.28	0.43	7.22	7.02	4.29	4.49	0.20	0.13	0.46	0.41
2Y	19.20	12.90	25.59	20.07	**0.00**	**0.00**	**0.00**	**0.00**	12.62	13.23	9.20	8.97	**0.00**	**0.00**	**0.00**	**0.00**
5Y	26.74	23.20	30.22	30.90	**0.00**	**0.00**	**0.00**	**0.00**	15.39	18.52	8.01	9.76	**0.00**	**0.00**	**0.05**	**0.02**
	$\beta_{rf,\tau}$				$p_{rf,NK}$				$\beta_{rf,\tau}$				$p_{rf,NK}$			
M	-0.39	-0.12	-1.10	-0.78	0.87	0.66	**0.98**	**0.96**	-1.19	-0.84	-1.67	-1.59	**1.00**	**1.00**	**1.00**	**1.00**
Q	-0.55	-0.09	-2.09	-1.25	0.64	0.54	0.81	0.74	-2.86	-1.94	-4.34	-3.80	**0.99**	**0.98**	**1.00**	**1.00**
Y	-0.21	0.30	-5.00	-2.79	0.52	0.47	0.71	0.65	-5.23	-4.21	-8.32	-7.96	0.83	0.85	0.85	0.88
2Y	2.22	4.38	-6.08	-1.29	0.26	**0.04**	0.87	0.60	5.30	5.75	4.81	3.09	0.03	0.01	0.11	0.20
5Y	-6.57	6.83	-33.71	-16.98	0.92	**0.02**	**1.00**	**0.99**	0.91	3.87	-0.79	-2.28	0.40	0.14	0.57	0.69

Figure 3.1 Theoretical Properties of the Value Spread, the Growth Spread, and the Log Spread

The figure reports the cyclical properties of the value spread (book-to-market of value stocks minus book-to-market of growth stocks), the growth spread (market-to-book of growth stocks minus market-to-book of value stocks), and the log spread (log book-to-market of value minus that of growth, equivalently, log market-to-book of growth minus that of value) predicted by the model of Zhang (2003). Panel A plots the value spread; Panel B plots the growth spread; and Panel C plots the log spread. All the spreads are plotted against aggregate economic conditions modeled as aggregate productivity, denoted x. Two versions of the model are considered. The solid lines are for the benchmark model with costly reversibility and time-varying price of risk. The broken lines are for the special case with symmetric adjustment cost and constant price of risk.

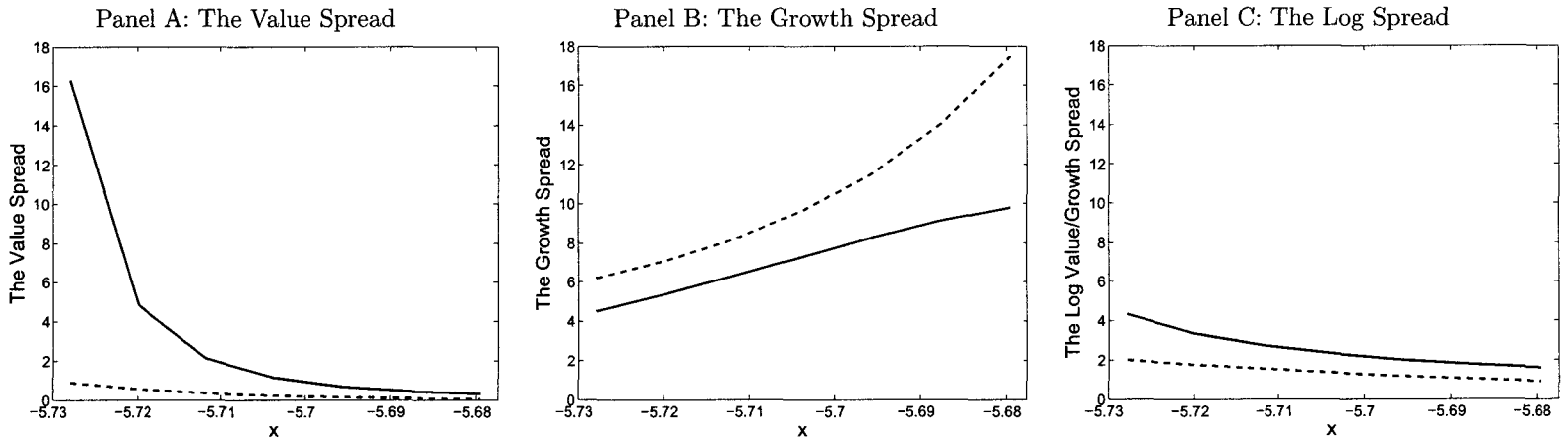
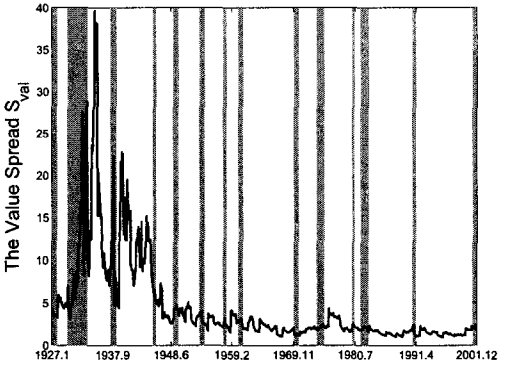


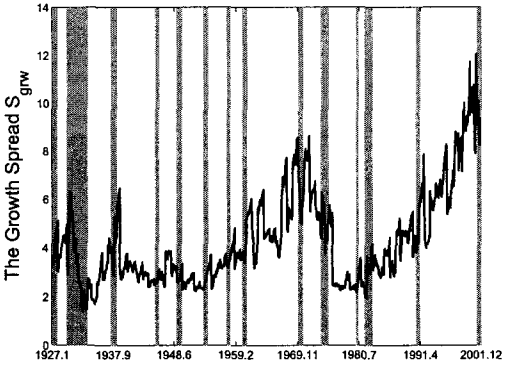
Figure 3.2 Time Series of The Value Spread, The Growth Spread, and The Log Spread

This figure plots the time series of the value spread, (S_{val}, Panel A), the growth spread (S_{grw}, Panel B), and the log spread ($S_{\log}$, Panel C) from January 1927 to December 2001. NBER recession dates are plotted in shadowed area. The value spread is measured as the average book-to-market ratio of portfolio ten (value) minus the average book-to-market ratio of portfolio one (growth) from the ten deciles sorted on book-to-market. The growth spread is measured as the average market-to-book ratio of growth portfolio minus the average market-to-book ratio of value portfolio. Finally, the log spread is measured as the log book-to-market ratio of value portfolio minus the log book-to-market of growth portfolio. The Fama-French portfolio data are obtained from Kenneth French's website. The data set contains the calendar year-end book-to-market ratios for all the portfolios. For months from January to December of year t, the book-to-market ratio of a given portfolio is constructed by dividing its book-to-market ratio at the end of December of year $t-1$ by its compounded gross return from the end of December of year $t-1$.

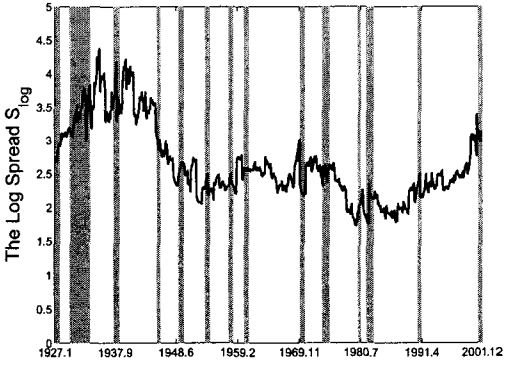
Panel A: The Value Spread, S_{val}



Panel B: The Growth Spread, S_{grw}



Panel C: The Log Spread, $S_{\log}$



Bibliography

[1] Boik, R.J. (2002). Spectral models for covariance matrices. *Biometrika*, **89**, 159–182.

[2] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. and Zhao, L. (2002). Statistical analysis of a telephone call center: a queueing-science perspective. Technical report, Department of Statistics, University of Pennsylvania

[3] Chiu, T.Y.M., Leonard, T. and Tsui, K.W. (1996). The matrix-logarithm covariance model. *J. Amer. Statist. Assoc.*, **91**, 198–210.

[4] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numer. Mathem.* **31**, 377–403.

[5] Dempster, A. (1972). Covariance selection. *Biometrics*, **28**, 157–175.

[6] Diggle, P.J. and Verbyla, A.P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, **54**, 401–415.

[7] Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data.*, 2nd ed. Oxford: Oxford University Press.

[8] Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

[9] Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chomometrics regression tools. *Technometrics* **35**, 109–148.

[10] Fu, W.J. (1998). Penalized regressions: The bridge versus the LASSO. *Journal of Computational and Graphical Statistics* **7**, 397–416.

[11] Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67.

[12] Hoerl, A.E. and Kennard, R.W. (1970b). Ridge regression: Application to nonorthogonal problems. *Technometrics* **12**, 69–82.

[13] Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. J. of Portfolio Management, 30, to appear.

[14] Leonard, T. and Hsu, J.S.J. (1992). Bayesian inference for a covariance matrix. *Ann. Statist.* **36**, 1669–1696.

[15] Lin, S.P. and Perlman, M.D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis*, **6**, Ed. P. R. Krishnaiah, 411–429. Amsterdam: North-Holland.

[16] Macchiavelli, R.E. and Arnold, S.F. (1994). Variable order antedependence models. *Commun. Statist.* **A23**, 13–22.

[17] Öjelund, H., Madsen, H. and Thyregod, P. (2001). Calibration with absolute shrinkage. *Journal of Chemometrics* **15**, 497–509.

[18] Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, **86**, 677–690.

[19] Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, **87**, 425–435.

[20] Pourahmadi, M., Daniels, M. and Park, T. (2004). Simultaneous modelling of the Cholesky decomposition of several covariance matrices with applications. Manuscript.

[21] Shen, H. and Huang, J.Z. (2004). Analysis of call center data using singular value decomposition and principal component analysis. Manuscript.

[22] Smith, M., and Kohn, R. (2002). Parsimonious Covariance Matrix Estimation for Longitudinal Data. *Journal of the American Statistical Association*, **97**, 1141–1153.

[23] Stein, C. (1975). Estimation of a covariance matrix. In *Reitz lecture*, Atlanta, Georgia, 1975. 39th annual meeting IMS.

[24] Roverato, A. (2000). Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika*, **87**, 99–112.

[25] Tibshirani, R. J. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Ser. B*, **58**, 267–288.

[26] Wermuch, N. (1980). Linear recursive equations, covariance selection and path analysis. *Journal of the American Statistical Association*, **75**, 963–972.

[27] Wong, F., Carter, C.K. and Kohn, R. (2003). Efficient estimation of covariance selection models. *Biometrika*, **90**, 809–830.

[28] Wu, W.B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, **90**, 831–844.

[29] Zimmerman, D.L. and Núñez-Antón, V. (1997). Structured antedependence models for longitudinal data. In *Modeling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions,* Springer Lecture Notes in Statistics, No. 122, Ed. T.G. Gregoire et al., pp. 63–76. New York: Springer-Verlag.

[30] Asness, Clifford S., Jacques A. Friedman, Robert J. Krail, and John M. Liew, 2000, Style timing: value versus growth, *Journal of Portfolio Management* Spring, 50–60.

[31] Bansal, Ravi, Robert F. Dittmar, and Christian T. Lundblad, 2004, Consumption, dividends, and the cross-section of equity returns, forthcoming, *Journal of Finance.*

[32] Berk, Jonathan B, Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance*, 54, 1153–1607.

[33] Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

[34] Campbell, John Y., and Tuomo Vuolteenaho, 2004, Bad beta, good beta, forthcoming, *American Economic Review.*

[35] Carlson, Murray, Adlai Fisher, and Ron Giammarino, 2003, Corporate investment and asset price dynamics: Implications for the cross-section of returns, forthcoming, *Journal of Finance.*

[36] Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2003, The value spread, *Journal of Finance* 58, 609–641.

[37] Cooper, Ilan, 2003, Asset pricing implications of non-convex adjustment costs and irreversibility of investment, working paper, Norwegian School of Management.

[38] Davis, James L., Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929–1997, *Journal of Finance* 55, 389–406.

[39] DeBondt, Werner F.M., and Richard H. Thaler, 1985, Does the stock market overreact? *Journal of Finance*, 40, 793–808.

[40] Eleswarapu, Venkat R., and Marc R. Reinganum, 2003, The predictability of aggregate stock market returns: evidence based on glamour stocks, forthcoming, *Journal of Business.*

[41] Fama, Eugene F., 1981, Stock returns, real activity, inflation, and money, *American Economic Review* 71, 545–565.

[42] Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

[43] Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

[44] Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, 33, 3–56.

[45] Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance*, 50, 131–155.

[46] Fama, Eugene F., and G. William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

[47] Gomes, Joao F., Leonid Kogan, and Lu Zhang, 2003, Equilibrium cross section of returns, *Journal of Political Economy* 111 (4), 693–732.

[48] Hansen, Lars Peter, John C. Heaton, and Nan Li, 2004, Consumption strike back? working paper, University of Chicago.

[49] Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics*17, 357–390.

[50] Kogan, Leonid, 2003, Asset prices and real investment, forthcoming, *Journal of Financial Economics.*

[51] Kothari, S.P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: a time-series analysis, *Journal of Financial Economics* 44, 169–203.

[52] Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance*, 49 (5) 1541–1578.

[53] Lewellen, Jonathan W., 2004, Predicting returns with financial ratios, forthcoming, *Journal of Financial Economics.*

[54] Lo, Andrew, and A. Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* 1, 41–66.

[55] Nelson, Charles R., and Myung J. Kim, 1993, Predictable stock returns: the role of small sample bias, *Journal of Finance* 48, 641–661.

[56] Pontiff, Jeffrey, and Lawrence D. Schall, 1998, Book-to-market ratios as predictors of market returns, *Journal of Financial Economics*, 49, 141–160.

[57] Ramey, Valerie A. and Matthew D. Shapiro, 2001, Displaced capital: A study of aerospace plant closings, *Journal of Political Economy* 109, 958–992.

[58] Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics*, 54, 375–421.

[59] Yogo, Motohiro, 2003, A consumption-based explanation of expected stock returns, working paper, Harvard University.

[60] Zhang, Lu, 2003, The value premium, forthcoming, *Journal of Finance.*