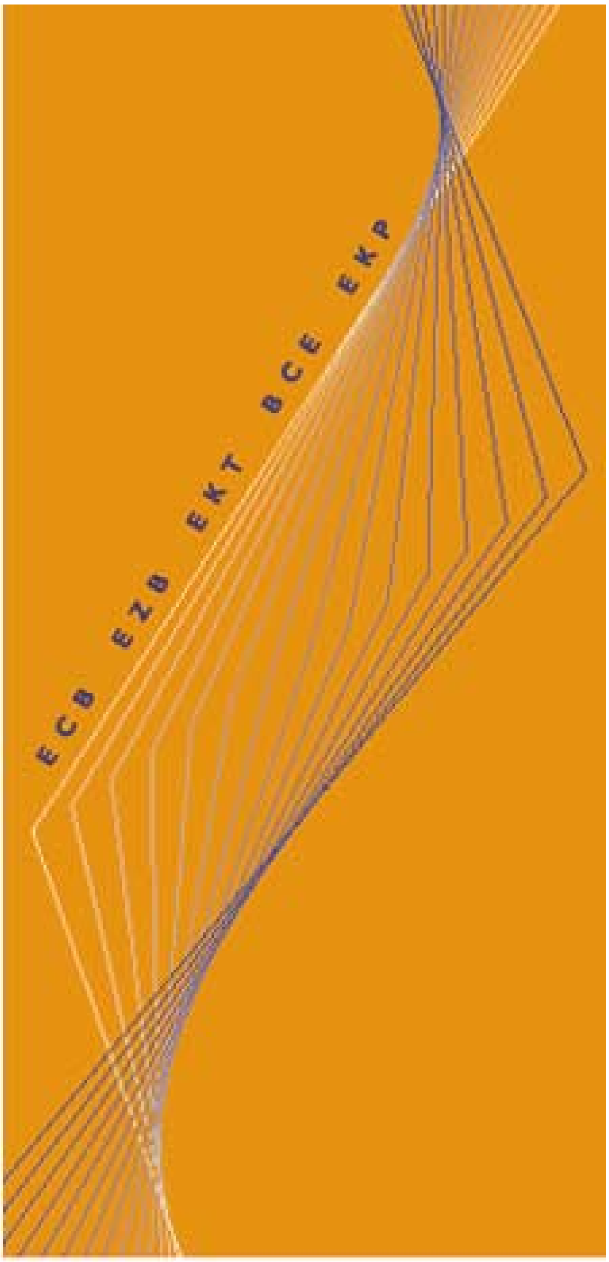
EUROPEAN CENTRAL BANK

WORKING PAPER SERIES



WORKING PAPER NO. 58

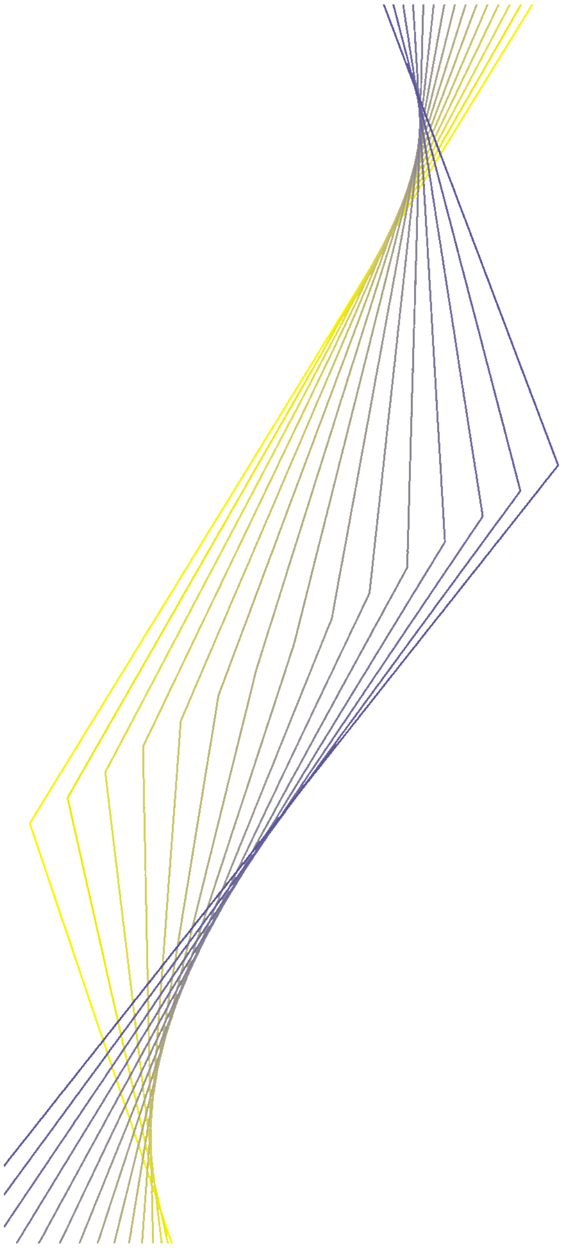
BUSINESS CYCLE ASYMMETRIES IN STOCK RETURNS: EVIDENCE FROM HIGHER ORDER MOMENTS AND CONDITIONAL DENSITIES

BY GABRIEL PEREZ-QUIROS AND ALLAN TIMMERMANN

April 2001

EUROPEAN CENTRAL BANK

WORKING PAPER SERIES



WORKING PAPER NO. 58

BUSINESS CYCLE ASYMMETRIES IN STOCK RETURNS: EVIDENCE FROM HIGHER ORDER MOMENTS AND CONDITIONAL DENSITIES*

BY GABRIEL PEREZ-QUIROS AND ALLAN TIMMERMANN

April 2001

* *We thank Justin McCrary for excellent research assistance. Two anonymous referees provided numerous helpful suggestions for improvements of the paper. Comments on an earlier version of the paper from Wouter den Haan, Rob Engle, Gloria Gonzalez-Rivera, Clive Granger, Jim Hamilton, Bruce Lehmann, Hashem Pesaran, Josh Rosenberg, Hal White and seminar participants at Camp Econometrics, University of Aarhus, UC Riverside, UC San Diego, University of Southern California and the NBER/NSF conference on forecasting and empirical methods in macro are gratefully acknowledged. The views expressed are those of the authors and do not necessarily reflect those of the European Central Bank or the European System of Central Banks.*

Address	**Kaiserstrasse 29**
	D-60311 Frankfurt am Main
	Germany
Postal address	**Postfach 16 03 19**
	D-60066 Frankfurt am Main
	Germany
Telephone	**+49 69 1344 0**
Internet	**http://www.ecb.int**
Fax	**+49 69 1344 6000**
Telex	**411 144 ecb d**

ISSN 1561-0810

Contents

Abstract

Markov switching models with time-varying means, variances and mixing weights are applied to characterize business cycle variation in the probability distribution and higher order moments of stock returns. This allows us to provide a comprehensive characterization of risk that goes well beyond the mean and variance of returns. Several mixture models with different specifications of the state transition are compared and we propose a new mixture of Gaussian and student-t distributions that captures outliers in returns. The models produce very similar expected returns and volatilities but imply very different time series for conditional skewness, kurtosis and predictive density. Consistent with economic theory, the gains in predictive accuracy from considering two-state mixture models rather than a single-state specification are higher for small firms than for large firms.

JEL classification system: C22, C52

Keywords: Markov Switching; Density Modelling; Mixtures of Distributions; Business Cycle risk

1. Introduction

Predictability of the mean and volatility of stock returns has been reported in a growing body of empirical studies.[1] The literature predominantly assumes a time-invariant relationship linking stock returns to a set of publicly known factors. However, some studies indicate that the conditional volatility of stock returns depends on the underlying state of the economy. For example, Schwert (1989) and Hamilton and Lin (1996) find that the volatility of stock returns is higher during recessions than during expansions. Additional evidence on asymmetry in the conditional distribution of stock returns is reported by Mcqueen and Roley (1993). Classifying three states according to the level of growth in industrial production, they find that announcement effects of macroeconomic news on daily stock prices critically depend on the state of the economy.

These studies indicate separate directions in which cyclical asymmetries in the variation of stock returns appear to be important. However, they fall short of investigating asymmetries in other than the first two moments of stock returns. Asymmetries need to be assessed in the context of the full conditional density of stock returns as emphasized by recent portfolio risk models. For example, Value at Risk (VaR) models characterize risk through the probability that a loss of a certain size occurs with a certain probability over a given investment horizon, c.f. Duffie and Pan (1997). In contrast, more traditional finance models such as the CAPM require information solely on the first two moments of asset returns. Neither the distributional assumptions required for the CAPM to hold (that returns are drawn from an elliptical distribution) nor the restrictions on agents' preferences (that utility is quadratic) are likely to be valid, however.[2]

We investigate in this paper the time-series properties of higher order moments and of the full conditional density of stock returns when these are modeled as a two-state Markov mixture process with time-varying transition probabilities and state-dependent coefficients in the mean and volatility equation. Stock returns are heavily influenced by outliers and statistical models have been developed to account for these through time-varying volatility, error densities more general than

[1]See, e.g., Fama and French (1989), Kandel and Stambaugh (1990), Glosten, Jagannathan and Runkle (1993) and Pesaran and Timermann (1995).

[2]Early theories of risk preceding the CAPM in fact relied on the full probability distribution of asset returns. Roy (1952) considered asset holdings subject to a safety-first constraint controlling the probability of 'disastrous' loss scenarios. Implementation of this strategy requires knowledge of the left tail of the density of asset returns. Markowitz (1959) proposed a measure of risk, the semi-variance, which treats positive and negative deviations from the mean return very differently. This measure is justified when asymmetries are present in the asset return distribution. More recently Fishburn (1977) has generalized the standard mean-variance model to measure risk as probability-weighted dispersions of payoffs below a target wealth level.

the Gaussian or some combination of these. Nonlinearities in asset returns are also widely recognized but the empirical asset pricing literature almost exclusively models these within the context of single-state models. We present new statistical evidence suggesting that a two-state specification is necessary to capture nonlinearities and outliers in the conditional distribution of stock returns.

Besides demonstrating the need for a two-state specification, the paper also fills out some important gaps in the growing literature that estimates Markov switching models using financial data. This literature has moved from first generation Markov switching models with constant transition probabilities to second generation models that allow the transition probability to be time-varying. However, little is known about how the specification of state transition probabilities and undelying densities affect the properties of the mixture model. We perform a range of specification tests that compare how different mixture specifications separate stock returns into states and how they perform along a range of predictive performance criteria. We also propose a new mixture model that combines a Gaussian and a student-t density to capture return dynamics by mixing underlying state densities that can display kurtosis. Effectively this model captures outliers by modeling these as drawn from a fat-tailed t-distribution with few degrees of freedom.

Finally we explore the economic implications of the time-series evolution in the higher order moments and predictive densities implied by the various mixture models. We use this information to perform model specification tests and also to characterize more fully the cyclical variation in risk, as measured for example by the conditional skewness or kurtosis. While the statistical performance of the models differs little in terms of their conditional mean, inspection of the higher order moments implied by different mixture specifications shows far greater ability to discriminate between models. Likewise, the predictive densities of the mixture models show substantial differences particularly when markets are volatile.

The mixture models provide a characterization of the dynamic patterns in risk that goes well beyond what can be achieved through Gaussian models. For the small firms, the mixture models identify negative expected returns and negative conditional skewness from the late expansion to the early recession stage of the economic cycle. Conditional volatility and kurtosis increase rapidly prior to and during recessions. This means that small firms' risk is particularly high around the peak of the business cycle at a time where it is very poorly approximated by a single Gaussian model. While similar patterns are found for the large firms, cyclical variation in their conditional moments is generally weaker for these firms.

The plan of the paper is as follows. Section 2 presents the econometric model and Section 3 reports estimation and forecasting results. Section 4 analyses the time-series variation in higher order conditional moments, while Section 5 looks at the evolution in the conditional density of returns. Section 6 concludes.

2. An Econometric Model of Asymmetries in Stock Returns

It is common in studies on predictability of stock returns to specify the conditional mean of returns on stocks in excess of a T-bill rate (ρ_t) as a linear function of a vector of predetermined instruments that are known at the time of the prediction ($\mathbf{X}_{t-1}$):

$$\rho_t = \beta' \mathbf{X}_{t-1} + \epsilon_t. \tag{1}$$

Here ϵ_t is a zero-mean error term. Some studies also allow for non-linear effects, typically by explicitly modeling time-dependence in the second conditional moment of stock returns (h_t):

$$\begin{aligned} \rho_t &= \beta' \mathbf{X}_{t-1} + \gamma h_t + \epsilon_t, \qquad\qquad \epsilon_t \sim (0, h_t) \\ h_t &= \vartheta(\{h_{t-i}\}_{i=1}^{p}, \{\epsilon_{t-i}\}_{i=1}^{q}, \mathbf{X}_{t-1}). \end{aligned} \tag{2}$$

Here ϑ is some time-invariant function.[3] To capture additional kurtosis in the error term, often a student-t density with few degrees of freedom (v) is adopted for ε_t :

$$\varepsilon_t \sim t(0, v, h_t) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\pi v h_t}(1 + \frac{\varepsilon_t^2}{v h_t})^{(v+1)/2}}. \tag{3}$$

Although these models allow for quite rich non-linear dynamics in stock returns, they assume that the functional relationship between excess returns and the predetermined factors $\{\mathbf{X}_{t-1}\}$ stays constant across different states of the economy.

Recent economic theories provide reasons to expect stock returns to display strong asymmetries with regard to the underlying economic state. They also suggest that these asymmetries are related to firm size. Because of information asymmetries, firms' access to capital markets tends to be based on their collateral. Small firms typically have far less collateral than large firms and will find it more difficult to raise capital. This is likely to be critical around economic recessions when small firms' capital base is particularly low. During such periods we would expect that tighter credit market conditions - as evidenced by higher interest rates and a higher credit spread - would affect small firms disproportionately more than large firms. As a result, the risk premium on these firms' equity should also rise in recessions. Theory thus predicts asymmetries in the relation between predetermined factors,

[3] Several non-linear specifications such as (2) are estimated for daily returns data in Engle and Ng (1993), while Glosten, Jagannathan and Runkle (1993) analyze monthly return series.

X_{t-1}, and excess returns, with factor sensitivities that are largest for small firms during recessions.[4]

Following these suggestions, we adopt an econometric model that allows the regression coefficients to be state-dependent. We briefly describe the Markovian latent state mixture model that forms the basis for our empirical analysis. Let s_t be a latent state variable and suppose that this can take one of k possible values, i.e. $s_t = 1,...., k$. Our specification uses the model originally proposed by Hamilton (1989) as a starting point and generalizes it by letting the intercept term, regression coefficients and variance of excess returns be state dependent:

$$\rho_t = \beta_{0s_t} + \boldsymbol{\beta}'_{s_t}\mathbf{X}_{t-1} + \epsilon_t, \qquad \epsilon_t \sim (0, h_{s_t}). \tag{4}$$

The state transition probabilities between periods $t-1$ and t are assumed to follow a first-order Markov chain. Although numerous papers have proposed different specifications for the state transition probabilities, very little work has been undertaken on comparing the resulting models. To accomplish this, we consider three alternative specifications. The simplest model assumes that state transitions are constant over time:

$$p_{ij}(\Omega_{t-1}) = P(s_t = j|s_{t-1} = i, \Omega_{t-1}) = p_{ij}, \tag{5}$$

where Ω_{t-1} is the information set available at time $t-1$. This 'first generation' mixture model is similar to that originally adopted by Hamilton (1989). It benefits from being tractable and simple to estimate.

Recent empirical experience with Markov switching models suggests that the flexibility gained by allowing the state transitions to vary over time as a function of a vector of predetermined variables, $\mathbf{y}_{t-1}$, can be very substantial and thus we also consider such models:[5]

$$p_{ij}(\Omega_{t-1}) \equiv p_{ij}(\mathbf{y}_{t-1}). \tag{6}$$

The statistical model implies a density of ρ_t conditional on Ω_{t-1} which we denote by $\phi(\rho_t|\Omega_{t-1};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters entering the likelihood function of the data. The log-likelihood function can be decomposed as follows:

[4] Of course there is no guarantee that a particular mixture model will separate states along business cycle lines. For example, many mixture models split security returns into a high and a low volatility state.

[5] Time-varying transition probabilities have been considered in an extensive literature. Filardo (1994) adopts the Composite Leading Indicator as a key explanatory variable of the transition probabilities in estimating a switching model for industrial production. Durland and McCurdy (1994) argue that the transition probabilities should depend on the duration of the state.

$$LL(\rho_T, \rho_{T-1,} ..\rho_1; \boldsymbol{\theta}) = \sum_{t=1}^{T} \ln(\phi(\rho_t|\Omega_{t-1}; \boldsymbol{\theta})), \tag{7}$$

where the information set Ω_{t-1} contains $\mathbf{X}_{t-1}$, ρ_{t-1}, $\mathbf{y}_{t-1}$, and lagged values of these variables: $\Omega_{t-1} = \{\mathbf{X}_{t-1},\ \rho_{t-1},\ \mathbf{y}_{t-1},\ \Omega_{t-2}\}$. The mixture density $\phi(\rho_t|\Omega_{t-1}; \boldsymbol{\theta})$ is obtained by summing the density functions conditional on the state $\eta(\rho_t|\Omega_{t-1}, s_t = j; \boldsymbol{\theta})$, using the respective state probabilities as weights:

$$\phi(\rho_t|\Omega_{t-1}; \boldsymbol{\theta}) = \sum_{j=1}^{k} \eta(\rho_t|\Omega_{t-1}, s_t = j; \boldsymbol{\theta}) P(s_t = j|\Omega_{t-1}; \boldsymbol{\theta}), \tag{8}$$

where $P(s_t = j|\Omega_{t-1}; \boldsymbol{\theta})$ is the conditional probability of being in state j at time t given information at time $t-1$. Under assumptions about the state densities of the innovations, ϵ_t, and a law specifying how the state evolves over time, the parameters of this model can be obtained by maximum likelihood estimation. We investigate two alternative sets of densities. The first assumes that the underlying state densities $\eta(\rho_t|s_t,\mathbf{X}_{t-1};\boldsymbol{\theta})$ are Gaussian. The conditioning factors, $\mathbf{X}_{t-1}$, enter linearly in the excess return equation within each state, but we allow their coefficients to vary between states:

$$\eta(\rho_t|\Omega_{t-1}, s_t = j; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi h_{jt}}} \exp(\frac{-(\rho_t - \beta_{0j} - \boldsymbol{\beta}_j' \mathbf{X}_{t-1})^2}{2h_{jt}}), \tag{9}$$

for $j = 1, ..., k$. Since mixtures of normals can approximate a very broad set of density families, this assumption is not very restrictive.[6]

Nevertheless, samples of asset returns contain outliers and empirical applications almost invariably use two-state models, so we propose a new mixture of student-t state densities

[6] Another closely related approach, advocated by Gallant and Tauchen (1989) and recently implemented in a study of stock returns by Harrison and Zhang (1997) uses Hermite polynomials to model the conditional density semi-nonparametrically. The two approaches are close substitutes. How 'parametric' our mixture of normals approach is depends of course on the rule used to decide how many normal densities to use. Our particular application uses a large conditioning information set in conjunction with a relatively small sample size and we partly choose the finite-mixture approach on grounds of parsimony. Furthermore, Sections 4 and 5 show that mixtures of two Gaussian densities or a mixture of a student-t and a Gaussian density can generate substantial variation over time in the shape of the conditional density.

$$\eta(\rho_t|\Omega_{t-1}, s_t = j; \boldsymbol{\theta}) = \frac{\Gamma(\frac{v_j+1}{2})}{\Gamma(\frac{v_j}{2})\sqrt{\pi v_j h_{jt}}(1 + \frac{(\rho_t - \beta_{0j} - \boldsymbol{\beta}_j' \mathbf{X}_{t-1})^2}{v_j h_{jt}})^{(v_j+1)/2}}, \tag{10}$$

where v_j is the degree of freedom parameter for the jth state and $\Gamma(.)$ is the gamma function. From the total probability theorem it follows that, for both densities, the conditional state probabilities can be obtained recursively:

$$P(s_t = i|\Omega_{t-1}; \boldsymbol{\theta}) = \sum_{j=1}^{k} P(s_t = i|s_{t-1} = j, \Omega_{t-1}; \boldsymbol{\theta}) P(s_{t-1} = j|\Omega_{t-1}; \boldsymbol{\theta}). \tag{11}$$

Finally, by Bayes' rule the conditional state probabilities can be written as

$$\begin{aligned} P(s_{t-1} &= j|\Omega_{t-1}; \boldsymbol{\theta}) = P(s_{t-1} = j|\rho_{t-1}, \mathbf{X}_{t-1}, \mathbf{y}_{t-1}, \Omega_{t-2}; \boldsymbol{\theta}) = \\ &\frac{\eta(\rho_{t-1}|s_{t-1} = j, \mathbf{X}_{t-1}, \mathbf{y}_{t-1}, \Omega_{t-2}; \boldsymbol{\theta}) P(s_{t-1} = j|\mathbf{X}_{t-1}, \mathbf{y}_{t-1}, \Omega_{t-2}; \boldsymbol{\theta})}{\sum_{j=1}^{k} \eta(\rho_{t-1}|s_{t-1} = j, \mathbf{X}_{t-1}, \mathbf{y}_{t-1}, \Omega_{t-2}; \boldsymbol{\theta}) P(s_{t-1} = j|\mathbf{X}_{t-1}, \mathbf{y}_{t-1}, \Omega_{t-2}; \boldsymbol{\theta})}. \end{aligned} \tag{12}$$

Equations (11) and (12) can be iterated on to recursively derive the state probabilities $\mathrm{P}(S_t|\Omega_{t-1}; \boldsymbol{\theta})$ and obtain the parameters of the likelihood function. Since the purpose of our analysis is to model the state dependence in the conditional distribution of excess returns given a set of publicly known predictors, we do not extract any information about the state of the economy from these variables.[7] The resulting state probabilities will entirely reflect the variation in the conditional distribution of stock returns.

3. Empirical Results

3.1. Model Specification

Our analysis uses monthly excess returns on the decile portfolios comprising the smallest and largest firms sorted by market capitalization and provided by the Center for Research in Security Prices (CRSP). We consider small (decile 1) and large (decile 10) firms separately because of the theoretical prediction that cyclical asymmetries should be much stronger for the less collateralized small firms. Following standard practice, monthly excess returns rather than nominal returns are studied to get a measure of returns relative to the benchmark risk-free T-bill rate.

[7] Hence we assume that the state transition probabilities are not influenced by the information in X_{t-1} and Ω_{t-2}, c.f. equations (5) and (6), and that $P(s_t = i|\Omega_{t-1}; \boldsymbol{\theta}) = P(s_t = i|\mathbf{X}_t, y_t, \Omega_{t-1}; \boldsymbol{\theta})$.

Returns at higher frequencies are more noisy and hence it would be more difficult to identify cyclical variation in returns at such horizons. The sample period begins in January 1954 and ends in December 1997, giving a total of 528 monthly observations.

Using variables linked to the business cycle, a large empirical literature has presented evidence that stock returns are predictable by means of single-state specifications and our choice of regressors is guided by this literature. We model excess returns as a function of an intercept term and lagged values of a 1-month T-bill rate ($I1$), obtained from the Fama-Bliss risk-free rates file on the CRSP tapes, a default premium (Def), measured as the difference between yields on Baa and Aaa-rated corporate bond portfolios, both obtained from the DRI Basic Economics data base, and the dividend yield ($Yield$). The one-month T-bill rate has been used by Fama (1981) and many others as a proxy for shocks to expected growth in real economic activity. The default premium tracks cyclical variation in the risk premium on stocks. The dividend yield, measured as dividends over the previous twelve months divided by the stock price at the end of the month and again obtained from CRSP, is commonly used as a proxy for time-varying expected returns. Fama and French (1989) find that the dividend yield tracks a mean-reverting component in expected stock returns whose variation extends beyond the business cycle. For this reason we impose a priori that the coefficients on these regressors remain the same across states.

In addition to these regressors that are common to small and large firms, we also include two additional factors in the model for small firms. Many studies have found that small firms' returns tend to be higher in January and are serially correlated so we include a January dummy and a single lag of returns for these firms. Serial correlation is believed to reflect non-trading effects and thus does not track the cycle. The source of the January effect is less well understood so we allow its coefficient to vary across states. For the small firm portfolio the mean equation thus becomes

$$\rho_t = \beta_{0s_t} + \beta_{1s_t} I_{t-1} + \beta_{2s_t} Def_{t-1} + \beta_3 Yield_{t-1} + \beta_4 \rho_{t-1} + \beta_{5s_t} Jan_t + \varepsilon_t, \quad (13)$$

while for the large firms we adopt the specification

$$\rho_t = \beta_{0s_t} + \beta_{1s_t} I_{t-1} + \beta_{2s_t} Def_{t-1} + \beta_3 Yield_{t-1} + \epsilon_t. \quad (14)$$

For the conditional variance of excess returns, h_{s_t}, we follow Glosten Jagannathan and Runkle (1993) and consider an exponential ARCH specification that varies with the state, the level of the 1-month T-bill rate and the absolute value of past shocks, $|\varepsilon_{t-1}|$, divided by the average of the lagged state volatilities weighted by the respective state probabilities at time $t-1$, σ_{t-1}, to take care of scaling effects[8]

$$\ln(h_{s_t,t}) = \lambda_{0s_t} + \lambda_{1s_t} I_{t-1} + \lambda_{2s_t}(|\varepsilon_{t-1}|/\sigma_{t-1} - \frac{2}{\pi}). \tag{15}$$

The mixture model assumes that there are two states, i.e. $k = 2$. Given the sample size relative to the number of parameters and their highly nonlinear effect, it is natural to choose a model restricted to two states.[9] Three state transition probability models are considered and we refer to these as MSI, MSII and MSIII, respectively. The first model simply assumes the state transition probabilities are constant.[10]

$$MSI: \quad p_{ii}(\Omega_{t-1}) = \Phi(\pi_i), \ i = 1, 2 \tag{16}$$

The second model assumes that transition probabilities vary with a single forward-looking summary measure of the state of the economy but omits a constant:

$$MSII: \quad p_{ii}(\Omega_{t-1}) = \Phi(\pi_i \Delta CLI_{t-2}), \ i = 1, 2 \tag{17}$$

where ΔCLI_{t-2} is the two-month lagged value of the year-on-year log-difference in the Composite Leading Indicator and $\Phi(.)$ is the cumulative density function of a standard normal variable.[11] The change in CLI seems a natural choice of state variable in this context.

[8]Notice that we do not scale by the volatility in the state at time $t-1$, $h_{s_{t-1}}$, since doing so complicates the model by introducing path dependence. The lagged variance turned out not to be significant when added as a separate regressor in the volatility equation, so we go with the simpler ARCH specification for the mixture models.

[9]Under the null of a single state, state transition probabilities are unidentified nuissance parameters and standard results on the distribution of likelihood ratio tests no longer apply. We tested for state dependence in the constants entering into the mean and volatility specifications, using the approach of Hansen (1992) and Hansen (1996). Besides varying the intercept coefficients in the mean and volatility equations across a grid, the transition probability parameters must also be varied in our setup with time-varying transition probabilities and several conditioning variables in both states. We found that the null of a single state was rejected at the 5 percent critical level or lower even when the state dependence only shows up in the intercept terms.

[10]Since there are only two states, $p_{12} = 1 - p_{11}$ and $p_{21} = 1 - p_{22}$, so two transition parameters have to be estimated for this model.

[11]Two lags of ΔCLI are used to account for the publication delay in this variable.

Notice that we have omitted a constant in the transition equation of this second model. Applications of Markov switching models to financial data suggest that states are effectively separated by levels of volatility. Typically a high volatility state with little persistence and a low volatility state with high persistence are identified. Naturally this reflects the many outliers in financial returns. If not controlled for, mixture models mechanically classify data into the high volatility state after an outlier is observed. However, this classification has little predictive power over returns. To constrain the transition probabilities to evolve more smoothly over time, we restrict these by setting the intercept equal to zero so that the transition probabilities are forced to smoothly track cyclical variations in the leading indicator.

To evaluate the effect of imposing this 'smoothness constraint' we also consider a general model that includes both a constant and ΔCLI_{t-2} in the state transitions: [12]

$$MSIII: \quad p_{ii}(\Omega_{t-1}) = \Phi(\pi_0 + \pi_i \Delta CLI_{t-2}), \; i = 1, 2 \tag{18}$$

Finally we consider both mixtures of two Gaussian densities and a mixture of a student-t and a Gaussian. Our estimations suggest that the density in one state is well characterized as a student-t with few degrees of freedom, while the degrees of freedom parameter for the second state continued to rise without reaching an upper bound. Since a student-t with a high degree of freedom parameter is very similar to a Gaussian density, we used the Gaussian distribution in the second state to ensure convergence of the estimations. We will refer to the Gaussian and student-t mixture model as MSt.

These Markov mixture models are compared to the corresponding single-state specification for the small firms

$$\rho_t = \beta_0 + \beta_1 I_{t-1} + \beta_2 Def_{t-1} + \beta_3 Yield_{t-1} + \beta_4 \rho_{t-1} + \beta_5 Jan_t + \varepsilon_t, \tag{19}$$

while the single-state model assumed for the large firms is

$$\rho_t = \beta_0 + \beta_1 I_{t-1} + \beta_2 Def_{t-1} + \beta_3 Yield_{t-1} + \epsilon_t, \tag{20}$$

[12] We impose the constraint that the intercept term is the same in both states, whereas the effect of the leading indicator can be state-dependent. This is done in order to achieve stability and convergence of the algorithm used to estimate the model. We found severe problems with convergence to multiple local optima when both the constant and the slope coefficient in the state transitions were allowed to be state-dependent.

and $\varepsilon_t \sim t(0, h_t, \upsilon)$ in both cases. Conditional volatility is modeled as

$$\ln(h_t) = \lambda_0 + \lambda_1 I_{t-1} + \lambda_2(|\varepsilon_{t-1}|/\sqrt{h_{t-1}} - \frac{2}{\pi})) + \lambda_3 \ln(h_{t-1}). \tag{21}$$

Conditional on Ω_{t-1}, the density of the single-state model is student-t and only its dispersion and center vary over time.

Other papers have utilized model specifications that are similar to ours in some respects. In the context of a Markov mixture model Hamilton and Susmel (1994) and Hamilton and Lin (1996) allow for ARCH effects although their models assume constant transition probabilities between two states. Filardo (1994) lets the transition probability of a Markov mixture specification vary over time but excludes ARCH effects. Chauvet and Potter (1997) construct leading and coincident indicators of stock returns through a factor-plus-noise approach that assumes the factor follows a latent Markov chain with constant state transition probabilities. Perez-Quiros and Timermann (2000) analyze the mean and variance of stock returns on ten size-sorted portfolios using a mixture of Gaussian densities.

3.2. Model Estimates and Data

Table 1 presents maximum likelihood estimates of the Markov mixture and single-state models fitted to the small firms' excess returns. In the single state model the coefficient on the interest rate is negative and highly statistically significant. The default premium and dividend yield coefficients are positive, although they fail to be significant at conventional critical levels. Finally the estimated coefficients of the lagged return and the January dummy are both positive and highly significant. In the volatility equation, lagged shocks ($|\varepsilon_{t-1}|$) and past volatility (h_{t-1}) are both significant while the interest rate coefficient fails to be. At 4.1, the degree of freedom estimate of the student-t distribution indicates that the first four moments are finite.

Turning to the two-state models, Table 1 shows that the coefficient estimates in the two states are very different from the single state benchmark. Across mixture specifications, the coefficient of the nominal interest rate is strongly negative and statistically significant in state 1 and closer to zero in state 2. Likewise, for three of four mixture models, the coefficient of the default premium is positive, highly significant and larger in state 1 than in state 2. Interestingly, the January dummy seems most important in state 1. This may provide a clue to the source of this poorly understood effect.

The overall level of volatility is generally higher in state 1 as indicated by the larger (less negative) intercept term in the volatility equation for this state. Interest rate effects tend to be of similar magnitude in the two states. The coefficients in the state transitions also provide important information. For the constant transition model, state 1 is less persistent ($p_{11} = 0.95$) than state 2 ($p_{22} = 0.97$). In the transition model with a leading indicator but no constant the coefficient on ΔCLI is negative in state 1 and positive in state 2, although none of the coefficients is individually significantly different from zero. It is clear from the log-likelihood values that exclusion of an intercept term leads to a substantial decline in the in-sample fit of the mixture model. In the most general transition specification (MSIII), the intercept term is highly significant while the negative coefficient on ΔCLI in state 1 is borderline significant.

Finally consider the mixture Gaussian and student-t model. To keep the number of models as low as possible, we only consider the state transition model with an intercept and ΔCLI. The conditional mean and volatility coefficients follow the same patterns as for the pure Gaussian mixtures. The coefficient on ΔCLI is again negative and significant in state 1 and close to zero in state 2. The degree of freedom parameter estimate is 7.0, suggesting fat tails in state 1. This estimate is somewhat larger than that from the single-state model (4.1), indicating that regime switching effects account for some of the leptokurtosis in returns.

A comparison of the single-state and Markov switching models in Table 1 shows that the estimates of the single-state model tend to lie between the corresponding Markov switching estimates for the two states. Since the coefficient estimates of the Markov switching model are very different across states, assuming a constant-coefficient model results in misleading conclusions. For example, a researcher might conclude from the single-state specification that the default premium is not significantly correlated with stock returns. The more correct conclusion would be that default risk is significantly positively correlated with stock returns, but only so in state 1.

Table 2 presents the estimates for the large firms across the same set of models. The signs of the coefficients in the single state specification are similar to those for the small firms: the coefficient on the interest rate is negative while the default premium and dividend yield obtain positive coefficients. For the mixture models, the large firms again display asymmetries across states in the conditional mean equation. The interest rate coefficient and the default premium coefficient are larger in absolute value in state 1 although the degree of asymmetry between the two states is less pronounced than for the small firms.

To assist in the economic interpretation of the two states identified by the mixture models, Figure 1 shows the time series of the predicted state-1 probability ($p_{1t} = \Pr(s_t = 1|\Omega_{t-1}; \boldsymbol{\theta})$). Also shown in shaded areas are the official recession periods tracked by the NBER. When a constant is included in the state transitions, the mixture model separates the data into a state that, while clearly related to recession periods, also picks up more isolated episodes of high volatility such as October 1987. For most periods, it is relatively clear which state generated the data and p_{1t} is far from 0.5. In contrast, the model without a constant in the state transitions generates smoother state probabilities and tracks the economic cycle more closely.

These plots effectively demonstrate how sensitive the state classification can be depending on which state transition and underlying density is used. While the time series of the state-1 probabilities are quite similar for the three mixture models with a constant in the transition probabilities, there is a distinct difference between how first and second generation mixture models classify states. For example, the squared correlation between the time series of p_{1t} generated by the model with a constant transition probability (MSI) and the mixture Gaussian-t model with a constant and the leading indicator (MSt) is less than 0.5 for the small firms. When a constant is included in the state transitions, the data gets separated less into business cycle states and more into high and low volatility states.

3.3. Forecasting Performance

To assess the statistical performance of the single- and two-state models we compare them along a variety of criteria. Initially we focus on traditional forecasting measures such as mean squared forecast error (MSFE) while the next sections consider precision in forecasting higher order moments and conditional densities. We first provide full-sample results in order to summarize the performance of the models when parameter estimation uncertainty is not too important. However, there is always a danger of in-sample overfitting with nonlinear models as complicated as ours, so we also present out-of-sample results that do not condition on the information embodied in the full-sample parameter estimates.

First consider the mean squared forecast errors reported in Table 3 for all models under consideration. Diebold and Mariano (1995) have suggested a test of the significance of the difference in the forecasting performance of two models and, for each sample, the table reports the values of this statistic in the second column. We set up the statistic so that a positive value means that the Markov switching model does better than the single-state model and apply a one-sided test. All mixture models fitted to the small firms generate lower MSFE values than the single-state model, and two of the test statistics are significant at the 5 percent level while a third test is borderline significant. The MSFE performance is relatively poorer for the large firms where only one of the mixture models improves on the single-state specification although none of the test statistics is significant at conventional levels. Interestingly, all four mixture models generate a positive test statistic for the large firms in the official recession periods, but all mixture models underperform during expansions.

In the out-of-sample forecasting experiment we are careful to avoid conditioning on information that was not known historically. Diebold and Rudebusch (1991) observe that the composite leading indicator has been revised numerous times, so it is important to avoid using the information implicit in later revisions. For this reason we use the originally released historical values to compute ΔCLI_{t-2}.[13] More specifically, let ΔCLI_t^{τ} be the $\tau'th$ vintage of the change in CLI applied to time t, so that $t \leq \tau$. Then we base our prediction for period t on $\{X_{t-1}, \Delta CLI_{t-2}^{\tau-2}, \widehat{\theta}_{t-1}\}$ where $\widehat{\theta}_{t-1}$ is the vector of parameters based on Ω_{t-1}. We begin the sample in 1976:1 to avoid the disruptive effects of a set of major revisions of the CLI in 1975. The parameters of the forecasting models are re-estimated at the beginning of each year using an expanding window of data. Convergence of the parameter estimates at each point in time is difficult to achieve, particularly at the beginning of the sample, but we found that this updating scheme made it feasible to explore the effect of using different starting values at each estimation point. Out-of-sample forecasting was not feasible to implement for the most complicated Gaussian, student-t mixture model whose convergence is more difficult to achieve. For this reason we only consider the three Gaussian mixture models in the out-of-sample experiment.

Unsurprisingly the mixture models do not outperform relative to the single state model in the out-of-sample experiments. Although the mixture models correct for biases in the single state model, they also add to parameter estimation variability, particularly at the beginning of the experiment where the sample is very short and it is difficult to precisely estimate the mixture models. In this situation it is commonplace to find that the MSFE performance of nonlinear models is worse than that of simpler linear alternatives, even in circumstances where the nonlinear specification is correct, see e.g. Pesaran and Potter (1997) and Clements and Smith (2000).

Although MSFE is by far the most common statistical measure of forecasting performance, it may not reflect the economic value of the predictions. A statistic designed to measure market timing information in a sequence of predictions is the nonparametric sign test proposed by Pesaran and Timermann (1992). This statistic is asymptotically normally distributed under the null of independence between the sign of the predicted and realized values of excess returns. Table 4 reports the value of this test statistic. In-sample all models produce a significant value of this test statistic. While there is little evidence to separate the performance of the single-state model from the Gaussian mixture models, the mixture Gaussian-t model generates a somewhat higher test statistic than the other models.

Out-of-sample, the single state and mixture models generate very similar values of the market timing test. However, while the predictability of the direction of the market continues to hold out-of-sample for the small firms, there is no evidence of predictability of the sign of large firms' returns out-of-sample.

A contentious issue in the literature on predictability of stock returns is how

[13] The real-time CLI series used by Diebold and Rudebusch (1991) terminated in 1988:12, so we updated their series for the period 1989-1997.

to interpret negative values of predicted excess return. It is difficult to imagine a meaningful equilibrium model in which risk averse investors are willing to hold stocks during periods where they expect negative risk premia. Previous discussions have been based on misspecified single state models so it is clearly important to investigate whether negative expected excess returns also show up in the mixture models. A natural measure of the occurrence of negative risk premia is the proportion of months where the expected excess return is negative and more than two standard errors below zero. Computing this statistic requires knowing the standard errors of the predicted values which is surprisingly complicated due to the recursive structure of the mixture models. In Appendix A we show how to derive the standard errors of the predictions. To save space we only report full sample results for the most general mixture specification (MSIII). For the small firms, this mixture model generated negative predictions for 254 months, or 48 percent of all cases, with 71 or 13 percent being more than two standard errors below zero. In contrast there was very weak evidence of negative expected returns for the large firms. Only 103 cases, or 19 percent, were negative and these were statistically significant only in three percent of the periods. These findings suggest that the negative risk premium puzzle is related to firm size. Negative expected returns are more or less absent for large firms, but occur surprisingly often for small firms.

Economic theory suggests a trade-off between expected returns and conditional volatility or other proxies for risk. Standard finance models specify expected returns as proportional to the conditional variance of the residual component. This suggests a constant squared coefficient of correlation between expected and realized returns, independently of the underlying state or the level of volatility. We provide new evidence on this issue by analyzing whether the dgree of predictability of stock returns - commonly interpreted as reflecting time-variations in risk premia - is related to the underlying state or the conditional volatility. For this purpose Figure 2 provides smoothed plots of the squared correlation between predicted and realized excess returns computed in different neighbourhoods of the level of state-1 probability and the level of conditional volatility. The plots are based on the Gaussian mixture model with a constant and the leading indicator in the state transitions (MSIII).

The figure shows that the proportion of the small firms' returns that is predictable increases systematically with the probability of being in the recession state and as a function of the level of volatility. Variations in small firms' expected returns hence matter disproportionately more when the volatility of returns is high. These findings are more consistent with a risk premium interpretation of the time-varying expected returns than with a model that assumes a time- and state-invariant risk premium. However, they also suggest that the standard risk premium model cannot fully explain the time-varying risks. While the degree of predictability of large firms' returns also varies with the state of the economy and the level of volatility, a less clearcut pattern is found for these firms.

4. Cyclical Variation in Higher Order Moments

Stock returns are clearly not normally distributed so volatility is an insufficient measure of risk. Testing the implications of imperfect capital market theories that small firms' equity is riskier around recessions thus requires inspecting higher order moments as well as the predictive density of stock returns. For completeness, this section initially studies time-variation in the conditional mean and variance and then proceeds to analyze the evolution in the conditional skewness and kurtosis. These moments are increasingly used to characterize risk. For example, Kraus and Litzenberger (1976) extend the standard two-moment CAPM to a setting where investors' preferences are also defined over the skew of the distribution of asset returns and Harvey and Sidique (1999) investigate the empirical importance of accounting for conditional skewness in the context of a cross-sectional model of stock returns. We know of no prior study of the time-series properties of the third and fourth conditional moments of returns and their link to the economic cycle. In a recent discussion of density forecasting Tay and Wallis (2000) emphasize the importance of considering this question although very little evidence exists.

The conditional moments of the single-state model are, of course, simple to derive. Conditional on Ω_{t-1}, excess returns are generated by a single student-t distribution with a mean following from (20) or (21), conditional variance $\frac{v}{v-2}h_t$, zero conditional skewness and kurtosis of $h_t^2\Gamma(\frac{5}{2})\Gamma(\frac{v-4}{2})v^2/(\Gamma(\frac{v}{2})\sqrt{\pi})$.

Deriving the first four centered, conditional moments of the Markov mixture distributions is less straightforward. First consider the mixture of Gaussian densities. Recall that $p_{1t} = \Pr(s_t = 1|\Omega_{t-1};\boldsymbol{\theta})$ is the probability of being in state 1 at time t given information at time $t-1$. Let $\mu_{jt} = \beta_{0j} + \beta_j' X_{t-1}$, and $\sigma_{jt}^2 = h_{j,t}$ be the first two conditional moments of the $j'th$ state. Corollary 1 in Timmermann (1999) which characterizes the moments of the ergodic distribution of markov switching models can easily be extended to cover the first four conditional moments:

Conditional mean:

$$E[\rho_t|\Omega_{t-1};\boldsymbol{\theta}] \equiv \mu_t = p_{1t}\mu_{1t} + (1-p_{1t})\mu_{2t}. \tag{22}$$

Conditional variance:

$$E[(\rho_t-\mu_t)^2|\Omega_{t-1};\boldsymbol{\theta}] = (1-p_{1t})\sigma_{2t}^2 + p_{1t}\sigma_{1t}^2 + p_{1t}(1-p_{1t})(\mu_{1t}-\mu_{2t})^2. \tag{23}$$

Conditional skewness:

$$E[(\rho_t-\mu_t)^3|\Omega_{t-1};\boldsymbol{\theta}] = p_{1t}(1-p_{1t})(\mu_{1t}-\mu_{2t})\left(3(\sigma_{1t}^2-\sigma_{2t}^2)+(1-2p_{1t})(\mu_{1t}-\mu_{2t})^2\right) \tag{24}$$

Conditional kurtosis:

$$
\begin{aligned}
E[(\rho_t - \mu_t)^4|\Omega_{t-1};\boldsymbol{\theta}] &= p_{1t}(1-p_{1t})(\mu_{1t}-\mu_{2t})^2 \cdot \\
&\left((\mu_{1t}-\mu_{2t})^2(1-3p_{1t}(1-p_{1t})) + 6(1-p_{1t})\sigma_{1t}^2 + 6p_{1t}\sigma_{2t}^2\right) \\
&+3\left(p_{1t}\sigma_{1t}^4 + (1-p_{1t})\sigma_{2t}^4\right) \qquad (25)
\end{aligned}
$$

These expressions show that different means of the underlying state densities are important to the volatility, skewness and kurtosis. For example, if the means in state 1 and 2 are identical, then the mixture model cannot generate skewness. Hence the time-series of the conditional skew generated by the mixture models reflect the discrepancy between the conditional mean in the two states.

The mixture of the student-t and Gaussian density leads to somewhat different conditional moments. Suppose the t-distribution occurs in state 1 and let v_1 be the degree of freedom parameter for this state. Appendix B proves the following result

Proposition

The first four conditional moments of the two-state mixture of a student-t distribution with v_1 degrees of freedom occurring in state 1 and a Gaussian distribution for state 2 are given by

$$
\begin{aligned}
E[y_t|\Omega_{t-1};\boldsymbol{\theta}] &= \mu_t = p_{1t}\mu_{1t} + (1-p_{1t})\mu_{2t} \\
E[(y_t-\mu_t)^2|\Omega_{t-1};\boldsymbol{\theta}] &= p_{1t}\left(\frac{v_1}{v_1-2}\right)\sigma_{1t}^2 + (1-p_{1t})\sigma_{2t}^2 \\
&+p_{1t}(1-p_{1t})(\mu_{1t}-\mu_{2t})^2.
\end{aligned}
$$

$$
\begin{aligned}
E[(y_t-\mu_t)^3|\Omega_{t-1};\boldsymbol{\theta}] &= p_{1t}(1-p_{1t})(\mu_{1t}-\mu_{2t})\{3\left(\sigma_{1t}^2(\frac{v_1}{v_1-2}) - \sigma_{2t}^2\right) \\
&+(1-2p_{1t})(\mu_{1t}-\mu_{2t})^2\}
\end{aligned}
$$

$$
\begin{aligned}
E[(\rho_t - \mu_t)^4|\Omega_{t-1};\boldsymbol{\theta}] &= p_{1t}(1-p_{1t})(\mu_{1t}-\mu_{2t})^2 \cdot \\
&\Big((\mu_{1t}-\mu_{2t})^2(1-3p_{1t}(1-p_{1t})) + 6(1-p_{1t})(\frac{v_1}{v_1-2})\sigma_{1t}^2 + 6 \\
&+\frac{\Gamma(\frac{5}{2})\Gamma(\frac{v_1-4}{2})v_1^2 p_{1t}\sigma_{1t}^4}{\sqrt{\pi}\Gamma(\frac{v_1}{2})} + 3(1-p_{1t})\sigma_{2t}^4
\end{aligned}
$$

Figure 3 presents time-series plots of the conditional mean for the single-state and two of the mixture models under consideration[14]. First consider the plots for

[14] The time-series plots for the remaining mixture models were very similar and are therefore omitted.

the small firms. The five series are clearly strongly correlated, with correlations that vary between 0.80 and 0.98. Although there is a clear cyclical pattern in expected returns which decline towards the end of expansions and rise during the course of the recession periods, there is also substantial short-run variation around the cyclical component. This is a result of including the lagged return as a regressor in the small firm model. Such short-run variation is largely absent in the plots for the large firms' expected returns which do not include a lag. For the large firms the first-moment time-series correlations are even higher across models and range from 0.91 to 0.99.

Turning to the second moment plots, Figure 4 shows that these display relatively more variation across econometric specifications. Correlations between second moments now range from 0.38 to 0.88 for the small firms and from 0.57 to 0.99 for the large firms. Interestingly, the student-t mixture fitted to the small firms generates quite different conditional volatility compared to the Gaussian mixture while the two series are almost identical for the large firms. This of course has to do with the low degree of freedom estimate for the t-distribution fitted to the small firms. Clear counter-cyclical patterns in volatility that lead recessions emerge for all models.

Similarities between models are further weakened once higher order moments are considered. The conditional skewness is of course zero for the single-state model. However, the skew generated by the Markov mixture models fitted to the small firms follows a pronounced cyclical path with negative conditional skewness in the late expansion and early recession stage, c.f. Figure 5. This finding matches well with the frequent occurence of negative returns at this stage of the cycle. It is interesting that such negative returns show up as negative conditional skewness. Symmetric density models would not identify such effects, so the advantage of using a mixture model is clear in this case. The difference between the third moment of first generation (constant transitions) and second generation (time-varying transitions) mixture models is also clear. The constant transition model generates negative conditional skewness for almost all time periods, while the other models produce both positive and negative skews. Across models, correlations between the time-series of skewness range from -0.12 to 0.06 for the small firms. For the large firms there is less of a difference between the time series plots of the different mixture specifications, which generate weak cyclical patterns centered around a small negative skew.

A strong cyclical pattern also emerges from the conditional kurtosis. Figure 6 shows that this has a tendency to rapidly increase prior to and during recessions. It also decreases in the early stages of the ensuing expansions. Again first and second generation mixture specifications produce very different conditional kurtosis with correlations across models as low as 0.4.

These results allow us to characterize the cyclical variations in the risk of stock returns as follows: right before a turning point in the business cycle, i.e. during the late expansion and early recession stage, stocks are particularly risky to hold since their returns have lower conditional mean, higher conditional volatility and lower (sometimes even negative) conditional skewness. With a short delay, i.e. typically in the early expansion state, kurtosis of returns rises rapidly.

A formal statistical assessment of the models' characterization of the higher order moments of returns can be obtained from regressing various powers of the residuals from the excess return equation on the corresponding conditional moments, c.f. Pagan and Schwert (1990):

$$\varepsilon_t^n = a + bE[\varepsilon_t^n | \Omega_{t-1}; \boldsymbol{\theta}] \tag{26}$$

Forecasts of the nth moment are unbiased if $a = 0$ and $b = 1$. Results from in-sample and out-of-sample estimations are provided in Table 5. First consider the second conditional moment. For the small firms, all mixture models generate values of a closer to 0 and of b closer to 1 than the single-state model. For the large firms, the mixture model with a constant transition probability and the mixture Gaussian-t dominate the single state model in this regard.

Turning next to the conditional skewness, by construction the single state model cannot capture time-variation in this moment. In contrast, six of eight mixture models generate a positive in-sample estimate of b. There is an interesting difference between the estimates of third-moment slopes for the small and large firms. While the slope estimates exceed one for the small firms, they are well below one for the large firms. Finally the table shows that the mixture models provide a much better fit for the time series of the small firms' fourth moment than the single-state model. Tracking of the fourth conditional moment is particularly impressive for the three mixture models that include a constant in the state transitions. In contrast, when it comes to the large firms, the single-state model generates the estimate of b closest to one.

Out-of-sample the single-state model is completely unable to track time-series variation in the small firms' second and fourth conditional moments. While a similar picture emerges for the mixture model with only the leading indicator in the state transitions, the two mixture specifications with a constant in their transition probabilities generate larger and positive slope estimates for the second and third moments. The simplest model with just a constant does particularly well with estimates of a and b that are within two standard errors of zero and one, respectively, for all three moments. None of the mixture models shows any evidence of out-of-sample predictability of time-series variations in the large firms' higher order moments.

5. Time Variations in Conditional Densities

As mentioned in the introduction, general decision theories characterize the risk of a financial asset by means of the predictive density of the asset's returns. For the Gaussian mixture model this means computing

$$\phi(\rho_t|\Omega_{t-1};\boldsymbol{\theta}) = \sum_{j=1}^{2} \frac{p_{jt}}{\sqrt{2\pi h_{j,t}}} \exp(\frac{-(\rho_t - \mu_j - \boldsymbol{\beta}_j'\mathbf{X}_{t-1})^2}{2h_{j,t}}), \tag{27}$$

while the mixture of a student-t with a Gaussian distribution requires calculating

$$\begin{aligned}\phi(\rho_t|\Omega_{t-1};\boldsymbol{\theta}) \quad = \quad & \frac{p_{1t}\Gamma(\frac{v_1+1}{2})}{\Gamma(\frac{v_1}{2})\sqrt{\pi v_1 h_{1,t}}(1+\frac{(\rho_t-\mu_1-\boldsymbol{\beta}_1'\mathbf{X}_{t-1})^2}{v_1 h_{1,t}})^{(v_1+1)/2}} \\ & + \frac{(1-p_{1t})}{\sqrt{2\pi h_{2,t}}} \exp(\frac{-(\rho_t - \mu_2 - \boldsymbol{\beta}_2'\mathbf{X}_{t-1})^2}{2h_{2,t}}). \end{aligned} \tag{28}$$

The single-state model assumes a student-t density with time-varying scale and location parameters. Insights into how much the conditional density of returns varies from month to month and how the single-state and two-state specifications differ, can be gained from Figure 7 which plots the sequence of monthly densities during the volatile period 1982:09 - 1982:12 around the change in the Federal Reserve's operating procedures. First consider the conditional densities of the small firms. While the location and dispersion of the single-state density varies considerably from month to month, the shapes of densities obtained through the mixture models cover a much wider range. Throughout the full sample the mixture model generates a variety of single-peaked, hump-shaped, and bi-modal densities that vary significantly from one month to the next.

Plots for the large firms revealed far less variation in the density shapes than was found for small firms. The reason for this is clear from the earlier expression for the conditional coefficient of skewness: mixtures of normal densities can only generate skewness provided that there is a sizeable difference in the means of the marginal densities in the two states. Small firms' mean equations display the highest degree of asymmetry across the two states (c.f. Tables 1 and 2) and will thus generate the highest time-series variation in skewness. Again this is consistent with the theories on the relationship between firm size and cyclical risk exposures.

These differences in density plots for the models under consideration underline the necessity of using more general tests of model fit than, say, mean square forecast error. Diebold, Gunther and Tay (1998) propose to apply the probability integral transform to the realizations of a time series as a way of evaluating the density implied by the forecasting model. If the conditional density, $\phi_{t-1}(.)$ is correctly specified, then $\int_{-\infty}^{\rho_t} \phi_{t-1}(x)dx$ should be drawn from a uniform distribution. Failure to correctly model, say, tail probabilities would result in a disproportional number of observations of these probability transforms of excess returns near zero or one.

Formal tests of the predictive densities are provided in Table 6. We compare the inverted probability transforms to the uniform distribution using a Kolmogorov-Smirnov test. In-sample, none of the single-state or mixture models leads to a rejection of the null at the 5% critical level. Out-of-sample the picture is very different. For the small firms the only model not to be rejected is the mixture model with a constant and the leading indicator in the transition probability. For the large firms only the model with a constant transition probability passes the test at the 5% critical level. Recalling that the mixture models did not improve on the out-of-sample MSFE performance of the single-state model, these findings also show that nonlinear mixture models may be of particular use to decision makers with non-quadratic loss functions.

Berkowitz (1999) has proposed a test to evaluate the accuracy of density forecast that is suitable for small samples. Under the null of correct density forecasts, the test is believed to have good power properties for testing the null of a correctly specified density function. We implement this test by forming the predictive density function in each period, t, (a normal distribution, a mixture of normals or a mixture of a normal and a t, depending on the model) calculating the p-values of each realization and testing if the probability integral transform of these p-values follow an identical and independently distributed standard normal through a likelihood ratio test. We divide this test into three separate components, testing for serial independence (autoregressive parameter equal to zero), zero mean, and unit variance. We also report the outcome of a joint test. In-sample again none of the individual or joint tests rejects the null of a correctly specified density. While most models are rejected out-of-sample, the general Gaussian mixture (MSIII) fitted to the small firms and the mixture models with a constant in the state transitions fitted to the large firms are not rejected at the 1% level (small firms) and 5% level (large firms).

6. Conclusion

A variety of new conclusions about how to understand time-variations in risk and its relation to firm size has emerged from this paper. Most obviously, it seems that commonly used single-state specifications for stock returns that adopt the same model in recessions and expansions are misspecified and can be strongly rejected against our two-state model. During recessions single-state specifications underestimate the size of the correlation between stock returns and variables such as short interest rates and default premia. Likewise, they overestimate the correlation between these variables and stock returns during expansion periods. We also find that the shape of the conditional density of stock returns and the higher order conditional moments vary considerably over time and that this variation is closely linked to the state of the business cycle. These findings suggest that a comprehensive characterization of risk must go well beyond the standard analysis of the mean and variance of returns.

There is strong evidence that cyclical asymmetries in the predictive density of stock returns is closely related to firm size. Gains in predictive accuracy resulting from introducing two-state mixture models are far larger for small than for large firms, both in-sample and out-of-sample. The more complicated two-state models do not improve on the out-of-sample mean squared forecast error statistic of the single-state model. However, tests of predictive accuracy based on third and fourth moments or the full predictive density indicate that the mixture models lead to better forecasts. Likewise, gains from considering different transition probability models is related to firm size. For the small firms, on balance a model with a time-varying state transition linked to the leading indicator performs best, while for the large firms a constant transition probability model is preferred. This is again consistent with the predictions of imperfect capital market theories that small firms' risk exposure is most sensitive to cyclical variations.

Appendix A
Calculation of Standard Errors for the Predictions from the Markov Switching Model

Equations (4) - (12) form the Markov switching model according to which the predicted value of excess returns is given by

$$\begin{aligned}\widehat{\rho}_t &= P(S_t=1|\Omega_{t-1})E[\rho_t|S_t=1,\Omega_{t-1}]+(1-P(S_t=1|\Omega_{t-1}))E[\rho_t|S_t=2,\Omega_{t-1}]\\ &= P(S_t=1|\Omega_{t-1})\left(E[\rho_t|S_t=1,\Omega_{t-1}]-E[\rho_t|S_t=2,\Omega_{t-1}]\right)+E[\rho_t|S_t=2,\Omega_{t-1}]\end{aligned} \tag{A1}$$

or, in a more compact form,

$$\widehat{\rho}_t = f_t(\boldsymbol{\theta},\Omega_{t-1}), \tag{A2}$$

where $\boldsymbol{\theta}=(\beta_{01},\beta_{02},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\lambda_{01},\lambda_{02},\lambda_{11},\lambda_{12},\lambda_{21},\lambda_{22})'$ is the parameter vector. This equation can be approximated by a first-order Taylor expansion evaluated at $\boldsymbol{\theta}_0$:

$$\widehat{\rho}_t = f_t(\boldsymbol{\theta}_0,\Omega_{t-1})+(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\frac{\partial f_t}{\partial\theta}. \tag{A3}$$

so that the variance of the forecast of excess returns can be approximated by

$$Var(\widehat{\rho}_t) = \left(\frac{\partial f_t}{\partial \boldsymbol{\theta}}\right)' Var(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \left(\frac{\partial f_t}{\partial \boldsymbol{\theta}}\right). \tag{A4}$$

In the present case we have (from equation (A1))

$$\begin{aligned} f_t(\boldsymbol{\theta}, \Omega_{t-1}) &= P(S_t = 1|\Omega_{t-1})\left(E[\rho_t|S_t = 1, \Omega_{t-1}] - E[\rho_t|S_t = 2, \Omega_{t-1}]\right) \\ &\quad + \left(E[\rho_t|S_t = 2, \Omega_{t-1}]\right). \end{aligned} \tag{A5}$$

Therefore,

$$\begin{aligned} \frac{\partial f_t}{\partial \boldsymbol{\theta}} &= \frac{\partial P(S_t = 1|\Omega_{t-1})}{\partial \boldsymbol{\theta}}\left(E[\rho_t|S_t = 1, \Omega_{t-1}] - E[\rho_t|S_t = 2, \Omega_{t-1}]\right) \\ &\quad + P\left(S_t = 1|\Omega_{t-1}\right)\frac{\partial\left(E[\rho_t|S_t = 1, \Omega_{t-1}] - E[\rho_t|S_t = 2, \Omega_{t-1}]\right)}{\partial \boldsymbol{\theta}} \\ &\quad + \frac{\partial E[\rho_t|S_t = 2, \Omega_{t-1}]}{\partial \boldsymbol{\theta}}. \end{aligned} \tag{A6}$$

Applying equation (11) from the main text and using (5) or (6) gives

$$\begin{aligned} P(S_t &= 1|\Omega_{t-1}) = P_{11}(\Omega_{t-1})P(S_{t-1} = 1|\Omega_{t-1}) + (1 - P_{22}(\Omega_{t-1}))P(S_{t-1} = 2|\Omega_{t-1}) \\ &= (P_{11}(\Omega_{t-1}) + P_{22}(\Omega_{t-1}) - 1)P(S_{t-1} = 1|\Omega_{t-1}) + (1 - P_{22}(\Omega_{t-1})), \end{aligned} \tag{A7}$$

where the state transitions depend on the particular transition model.

Differentiating (A7),

$$\begin{aligned} \frac{\partial P(S_t = 1|\Omega_{t-1})}{\partial \boldsymbol{\theta}} &= \frac{\partial(P_{11}(\Omega_{t-1}) + P_{22}(\Omega_{t-1}) - 1)}{\partial \boldsymbol{\theta}} P(S_{t-1} = 1|\Omega_{t-1}) \\ &\quad + (P_{11}(\Omega_{t-1}) + P_{22}(\Omega_{t-1}) - 1)\frac{\partial P(S_{t-1} = 1|\Omega_{t-1})}{\partial \boldsymbol{\theta}} \\ &\quad + \frac{\partial(1 - P_{22}(\Omega_{t-1}))}{\partial \boldsymbol{\theta}}. \end{aligned} \tag{A8}$$

In equation (A8) we need to calculate $\frac{\partial P_{11}(\Omega_{t-1})}{\partial \boldsymbol{\theta}}$, $\frac{\partial P_{22}(\Omega_{t-1})}{\partial \boldsymbol{\theta}}$, and $\frac{\partial P(S_{t-1}=1|\Omega_{t-1})}{\partial \boldsymbol{\theta}}$. For the most general transition probability specification, the first two derivatives are

$$\frac{\partial P_{11}(\Omega_{t-1})}{\partial \theta_i} = \begin{cases} \phi(\pi_0 + \pi_1 y_{t-1})y_{t-1} & \text{when } \theta_i = \pi_1 \\ \phi(\pi_0 + \pi_1 y_{t-1}) & \text{when } \theta_i = \pi_0 \\ 0 & \text{otherwise} \end{cases} \tag{A9}$$

and

$$\frac{\partial P_{22}(\Omega_{t-1})}{\partial \theta_i} = \begin{cases} \phi(\pi_0 + \pi_2 y_{t-1}) y_{t-1} & \text{when } \theta_i = \pi_2 \\ \phi(\pi_0 + \pi_2 y_{t-1}) & \text{when } \theta_i = \pi_0 \\ 0 & \text{otherwise} \end{cases} \tag{A10}$$

where θ_i is the i'th element of $\boldsymbol{\theta}$. In order to calculate $\frac{\partial P(S_{t-1}=1|\Omega_{t-1})}{\partial \boldsymbol{\theta}}$, we use equation (12) which can be rewritten as

$$P(S_{t-1} = 1|\Omega_{t-1}) = \frac{B_{t-1}}{A_{t-1}}, \tag{A11}$$

where

$$B_{t-1} = \eta(\rho_{t-1}|S_{t-1} = 1, \mathbf{X}_{t-1}, y_{t-1}, \Omega_{t-2}) P(S_{t-1} = 1|\mathbf{X}_{t-1}, y_{t-1}, \Omega_{t-2}), \tag{A12}$$

$$A_{t-1} = B_{t-1} + \eta(\rho_{t-1}|S_{t-1} = 2, \mathbf{X}_{t-1}, y_{t-1}, \Omega_{t-2}) P(S_{t-1} = 2|\mathbf{X}_{t-1}, y_{t-1}, \Omega_{t-2}). \tag{A13}$$

Differentiating (A11) it follows that

$$\frac{\partial P(S_{t-1} = 1|\Omega_{t-1})}{\partial \boldsymbol{\theta}} = \frac{\partial B_{t-1}/\partial \boldsymbol{\theta}}{A_{t-1}} - \frac{B_{t-1} \partial A_{t-1}/\partial \boldsymbol{\theta}}{A_{t-1}^2}, \tag{A14}$$

where

$$\begin{aligned} \frac{\partial B_{t-1}}{\partial \boldsymbol{\theta}} &= \frac{\partial \eta(\rho_{t-1}|S_{t-1} = 1, X_{t-1}, y_{t-1}, \Omega_{t-2})}{\partial \boldsymbol{\theta}} P(S_{t-1} = 1|\Omega_{t-2}) \\ &\quad + \eta\left(\rho_{t-1}|S_{t-1} = 1, X_{t-1}, y_{t-1}, \Omega_{t-2}\right) \frac{\partial P(S_{t-1} = 1|\Omega_{t-2})}{\partial \boldsymbol{\theta}}, \end{aligned} \tag{A15}$$

$$\begin{aligned} \frac{\partial A_{t-1}}{\partial \boldsymbol{\theta}} &= \frac{\partial B_{t-1}}{\partial \boldsymbol{\theta}} + \frac{\partial \eta(\rho_{t-1}|S_{t-1} = 2, X_{t-1}, y_{t-1}, \Omega_{t-2})}{\partial \boldsymbol{\theta}} (1 - P(S_{t-1} = 1|\Omega_{t-2})) \\ &\quad + \eta\left(\rho_{t-1}|S_{t-1} = 2, X_{t-1}, y_{t-1}, \Omega_{t-2}\right) \frac{\partial (1 - P(S_{t-1} = 1|\Omega_{t-2}))}{\partial \boldsymbol{\theta}}, \end{aligned} \tag{A16}$$

where we used the assumption that $P(S_{t-1} = 1|X_{t-1}, y_{t-1}, \Omega_{t-2}) = P(S_{t-1} = 1|\Omega_{t-2})$.

Because of the ARCH effect in the variance, the term $\frac{\partial\eta(\rho_{t-1}|S_{t-1}=1,X_{t-1},y_{t-1},\Omega_{t-2})}{\partial\boldsymbol{\theta}}$ deserves special interest. To calculate this derivative, in addition to the "standard" first order effect of each of the parameters, we have to consider the term $\frac{\partial\frac{|\varepsilon_{t-1}|}{h_{t-1}}}{\partial\boldsymbol{\theta}}$ and, obviously, this term is a function of $\frac{\partial h_{t-1}}{\partial\boldsymbol{\theta}}$. To calculate this derivative, we repeat the same derivations that we use in this appendix, now applied to the function $h_t = w_t(\boldsymbol{\theta}, \Omega_{t-1})$.

Finally, plugging (A14) into (A8) we obtain:

$$\begin{aligned}\frac{\partial P(S_t=1|\Omega_{t-1})}{\partial\boldsymbol{\theta}} &= \frac{\partial(P_{11}(\Omega_{t-1})+P_{22}(\Omega_{t-1})-1)}{\partial\boldsymbol{\theta}}P(S_{t-1}=1|\Omega_{t-1}) \\ &\quad +(P_{11}(\Omega_{t-1})+P_{22}(\Omega_{t-1})-1)\frac{\partial B_{t-1}/\partial\boldsymbol{\theta}}{A_{t-1}}-\frac{B_{t-1}\partial A_{t-1}/\partial\boldsymbol{\theta}}{A_{t-1}^2} \\ &\quad +\frac{\partial(1-P_{22}(\Omega_{t-1}))}{\partial\boldsymbol{\theta}}. \qquad \text{(A17)}\end{aligned}$$

Equations (A15) - (A17) can be iterated on to calculate $\partial P(S_t = 1|\Omega_{t-1})/\partial\boldsymbol{\theta}$ provided that an initial value is assigned to $\partial P(S_t = 1|\Omega_0)/\partial\boldsymbol{\theta}$. Starting the original algorithm by using a fixed value of the first state of the economy, as we do, implies that $\partial P(S_t = i|\Omega_0)/\partial\boldsymbol{\theta} = \mathbf{0}$, ($i$ = 1,2). Finally, substituting $\partial P(S_t = i|\Omega_{t-1})/\partial\boldsymbol{\theta}$ (i = 1,2) into equation (A6), and noting that $\partial E[\rho_t|S_t = 1, \Omega_{t-1}]/\partial\boldsymbol{\theta}$ is easy to calculate from (4), we obtain $\partial f_t/\partial\boldsymbol{\theta}$ and, by (A4), the variance of ρ_t, the square root of which gives the standard errors of the predictions from the Markov switching model.

Appendix B
Conditional moments of the student-t and Gaussian mixture model

The first conditional moment of the mixture model is given by

$$\begin{aligned}E[y_t|\Omega_{t-1}] &= \mu_t = \mu_{1t}\Pr(S_t=1|\Omega_{t-1})+\mu_{2t}(1-p_{1t})\Pr(S_t=2|\Omega_{t-1}) \\ &= \mu_{1t}p_{1t}+\mu_{2t}(1-p_{1t}).\end{aligned}$$

The second conditional moment can be derived as follows:

$$\begin{aligned}\sigma_t^2 &= E[(y_t-\mu_t)^2|\Omega_{t-1}] = p_{1t}E[(\mu_{1t}+\sigma_{1t}\varepsilon_t-\mu_t)^2]+(1-p_{1t})E[(\mu_{2t}+\sigma_{2t}\varepsilon_t-\mu_t)^2 \\ &= p_{1t}\sigma_{1t}^2+(1-p_{1t})\sigma_{2t}^2+p_{1t}(\mu_{1t}-\mu_t)^2+(1-p_{1t})(\mu_{2t}-\mu_t)^2.\end{aligned}$$

Because of the ARCH effect in the variance, the term $\frac{\partial\eta(\rho_{t-1}|S_{t-1}=1,X_{t-1},y_{t-1},\Omega_{t-2})}{\partial\boldsymbol{\theta}}$ deserves special interest. To calculate this derivative, in addition to the "standard" first order effect of each of the parameters, we have to consider the term $\frac{\partial\frac{|\varepsilon_{t-1}|}{h_{t-1}}}{\partial\boldsymbol{\theta}}$ and, obviously, this term is a function of $\frac{\partial h_{t-1}}{\partial\boldsymbol{\theta}}$. To calculate this derivative, we repeat the same derivations that we use in this appendix, now applied to the function $h_t = w_t(\boldsymbol{\theta}, \Omega_{t-1})$.

Finally, plugging (A14) into (A8) we obtain:

$$\begin{aligned}\frac{\partial P(S_t = 1|\Omega_{t-1})}{\partial\boldsymbol{\theta}} &= \frac{\partial(P_{11}(\Omega_{t-1}) + P_{22}(\Omega_{t-1}) - 1)}{\partial\boldsymbol{\theta}}P(S_{t-1} = 1|\Omega_{t-1}) \\ &\quad +(P_{11}(\Omega_{t-1}) + P_{22}(\Omega_{t-1}) - 1)\frac{\partial B_{t-1}/\partial\boldsymbol{\theta}}{A_{t-1}} - \frac{B_{t-1}\partial A_{t-1}/\partial\boldsymbol{\theta}}{A_{t-1}^2} \\ &\quad +\frac{\partial(1 - P_{22}(\Omega_{t-1}))}{\partial\boldsymbol{\theta}}. \qquad \text{(A17)}\end{aligned}$$

Equations (A15) - (A17) can be iterated on to calculate $\partial P(S_t = 1|\Omega_{t-1})/\partial\boldsymbol{\theta}$ provided that an initial value is assigned to $\partial P(S_t = 1|\Omega_0)/\partial\boldsymbol{\theta}$. Starting the original algorithm by using a fixed value of the first state of the economy, as we do, implies that $\partial P(S_t = i|\Omega_0)/\partial\boldsymbol{\theta} = \mathbf{0}$, ($i$ = 1,2). Finally, substituting $\partial P(S_t = i|\Omega_{t-1})/\partial\boldsymbol{\theta}$ (i = 1,2) into equation (A6), and noting that $\partial E[\rho_t|S_t = 1, \Omega_{t-1}]/\partial\boldsymbol{\theta}$ is easy to calculate from (4), we obtain $\partial f_t/\partial\boldsymbol{\theta}$ and, by (A4), the variance of ρ_t, the square root of which gives the standard errors of the predictions from the Markov switching model.

Appendix B
Conditional moments of the student-t and Gaussian mixture model

The first conditional moment of the mixture model is given by

$$\begin{aligned}E[y_t|\Omega_{t-1}] &= \mu_t = \mu_{1t}\Pr(S_t = 1|\Omega_{t-1}) + \mu_{2t}(1 - p_{1t})\Pr(S_t = 2|\Omega_{t-1}) \\ &= \mu_{1t}p_{1t} + \mu_{2t}(1 - p_{1t}).\end{aligned}$$

The second conditional moment can be derived as follows:

$$\begin{aligned}\sigma_t^2 &= E[(y_t - \mu_t)^2|\Omega_{t-1}] = p_{1t}E[(\mu_{1t} + \sigma_{1t}\varepsilon_t - \mu_t)^2] + (1 - p_{1t})E[(\mu_{2t} + \sigma_{2t}\varepsilon_t - \mu_t)^2 \\ &= p_{1t}\sigma_{1t}^2 + (1 - p_{1t})\sigma_{2t}^2 + p_{1t}(\mu_{1t} - \mu_t)^2 + (1 - p_{1t})(\mu_{2t} - \mu_t)^2.\end{aligned}$$

Bibliography

Berkowitz, J., 1999, "Evaluating the forecast of risk models," Federal Reserve Board. Finance and Economics Discussion Series. 1999-11.

Chauvet, M. and S. Potter, 2000, "Coincident and leading indicators of the stock market," Journal of Empirical Finance, Vol. 7, 87-111.

Clements, M.P. and J. Smith, 2000, "Evaluating the forecast densities of linear and nonlinear models: Applications to output growth and unemployment," Journal of Forecasting, 19, 225-276.

Diebold, F.X. and G.D. Rudebusch, 1991, "Forecasting output with the Composite Leading Index: An ex ante analysis," Journal of the American Statistical Association, 86, 603-610.

______ and R.S. Mariano, 1995, "Comparing predictive accuracy," Journal of Business and Economic Statistics, 13, 253-263.

______ , T.A. Gunther and A.S. Tay. 1998, "Evaluating density forecasts with applications to financial risk management," International Economic Review, 39, 863-884.

Duffie, D. and J. Pan, 1997, "An overview of value at risk," The Journal of Derivatives, Spring, 7-49.

Durland, J.M. and T.H. McCurdy, 1994, "Duration dependent transitions in a market model of US GNP growth," Journal of Business and Economic Statistics, 12, 279-288.

Fama, E.F., 1981, "Stock returns, real activity, inflation, and money," American Economic Review, 71, 545-565.

Fama, E. F. and K. R. French, 1989, "Size and book-to-market factors in earnings and returns," Journal of Finance, 50, 131-155.

Filardo, A.J., 1994, "Business cycle phases and their transitional dynamics," Journal of Business and Economic Statistics, 12, 299-308.

Fishburn, P., 1977, Mean-Risk analysis with risk associated with below market returns," American Economic Review, 67, 116-126.

Gallant, R., and G. Tauchen, 1989, Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications,", Econometrica, 57, 1091 - 1120.

Glosten, L.R. and R. Jagannathan and D. Runkle, 1993, "On the relation between the expected value and the volatility of the nominal excess return on stocks," Journal of Finance, 48, 1779-1801.

Hamilton, J.D., 1989, "A new approach to the economic analysis of non-stationary time series and the business cycle," Econometrica, 57, 357-384.

______ and G. Lin, 1996 "Stock market volatility and the business cycle," Journal of Applied Econometrics, 11, 573-593.

______ and R. Susmel, 1994, "Autoregressive conditional heteroskedasticity and changes in regime," Journal of Econometrics, 64, 307-333.

Hansen, B.E., 1992, "The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP," Journal of Applied Econometrics, 7, S61-S82.

Hansen, B.E., 1996, "Erratum: The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP," Journal of Applied Econometrics, 11, 195-98.

Harrison, P. and H.H. Zhang, 1997, "Cyclical variation in the risk and return relation," Mimeo, Brandeis University.

Harvey, C.R. and A. Siddique, 2000, "Conditional skewness in asset pricing tests," Journal of Finance, 55, 1263-1295.

Kandel, S. and R.F. Stambaugh, 1990, "Expectations and volatility of consumption and asset returns," Review of Financial Studies, 3, 207-232

Kraus, A. and R. Litzenberger, 1976, "Skewness preference and the valuation of risky assets," Journal of Finance, 31, 1085-1100.

Markowitz, H.M, 1959, "Portfolio selection", John Wiley and Sons.

Pagan, A. and G. W. Schwert, 1990, "Alternative models for conditional stock volatility," Journal of Econometrics, 45, 267-290.

Perez-Quiros, G. and A. Timmermann, 2000, "Firm size and cyclical variations in stock returns," Journal of Finance, 55, 1229-1263.

Pesaran, M.H. and A. Timmermann, 1992, "A simple non-parametric test of predictive performance," Journal of Business and Economic Statistics, 10, 461-465.

______ and ______, 1995, "Predictability of stock returns: Robustness and economic significance," Journal of Finance, 50, 1201-1228.

______ and S.M. Potter, 1997, "A Floor and ceiling model of US output," Journal of Economic Dynamics and Control, 21, 661-695.

Roy, A.D.,"Safety-first and the holding of assets," Econometrica, 20, 431-449.

Schwert, G.W., 1989, "Why does stock market volatility change over time?," Journal of Finance, 44, 1115-1153.

Tay, A.S. and K.F. Wallis, 2000, "Density forecasting: A survey," 19, 235-254.

Timmermann, A. 2000, "Moments of Markov switching models, " Journal of Econometrics, 96, 75-111.

.

Table 1
Small Firms

The following models were estimated for excess returns on the small firm portfolio

Single State Model:

$\rho_t = \beta_0 + \beta_1 \, I1_{\,t-1} + \beta_2 \, Def_{\,t-1} + \beta_3 \, Yield_{\,t-1} + \beta_4 \, \rho_{t-1} + \beta_5 \, \mathrm{Jan}_t + \varepsilon_t$

$\varepsilon_t \sim t(0,v,h_t)$, $\ln(h_t)=\lambda_0 + \lambda_1 \, I1_{\,t-1} + \lambda_2[(\,|\varepsilon_{t-1}|/(h_{t-1})^{1/2})-(2/\pi)] + \lambda_3 \ln(h_{t-1})$

Gaussian Mixture:

$\rho_t = \beta_{0,St} + \beta_{1,St} \, I1_{\,t-1} + \beta_{2,St} \, Def_{\,t-1} + \beta_3 \, Yield_{\,t-1} + \beta_4 \rho_{t-1} + \beta_{5,St} \, Jan + \varepsilon_t$

$\varepsilon_t \sim N(0,h_{St})$, $\ln(h_{St})=\lambda_0 + \lambda_{1,St} \, I1_{\,t-1} + \lambda_{2,St}[(\,|\varepsilon_{t-1}|/(h_{t-1})^{1/2})-(2/\pi)]$

Model I: $p_t=\mathrm{Prob}(s_t=1|s_{t-1}=1)=\Phi(\pi_0)$, $q_t=\mathrm{Prob}(s_t=2|s_{t-1}=2)=\Phi(\pi_1)$

Model II: $p_t=\mathrm{Prob}(s_t=1|s_{t-1}=1)=\Phi(\pi_1 \Delta CLI_{\,t-2})$, $q_t=\mathrm{Prob}(s_t=2|s_{t-1}=2)=\Phi(\pi_2 \Delta CLI_{\,t-2})$

Model III: $p_t=\mathrm{Prob}(s_t=1|s_{t-1}=1)=\Phi(\pi_0+\pi_1 \Delta CLI_{\,t-2})$, $q_t=\mathrm{Prob}(s_t=2|s_{t-1}=2)=\Phi(\pi_0+\pi_2 \Delta CLI_{\,t-2})$

Gaussian and student-t Mixture:

Same transition probability specification as Model III with densities:

$\varepsilon_t \sim N(0,h_{St})$ if $S_t=1$

$\varepsilon_t \sim t(0,v,h_{St})$ if $S_t=2$.

ρ_t is the monthly excess returns on the small firm portfolio, *I1* is the one-month T- bill rate, *Def* is the default premium, *Yield* is the dividend yield, *Jan* is a January dummy and ΔCLI is the annual rate of growth of the Composite Index of Leading Indicators. The sample period is 1954-1997.

Mean Parameters	Single-State Model		Gaussian Mixture Model I		Gaussian Mixture Model II		Gaussian Mixture Model III		Gaussian, Student-t Mixture Model	
Constant, State 1	-0.013	(0.011)	-0.016	(0.024)	-0.034	(0.019)	-0.030	(0.020)	-0.028	(0.016)
Constant, State 2			-0.017	(0.011)	-0.005	(0.012)	-0.010	(0.010)	0.003	(0.010)
Interest Rate, State 1	-4.006	(1.196)	-6.025	(3.022)	-7.588	(3.122)	-7.722	(3.054)	-6.721	(2.238)
Interest Rate, State 2			-4.656	(1.394)	0.057	(1.741)	-2.116	(1.291)	1.417	(1.587)
Default Premium, State 1	15.648	(10.836)	19.234	(21.684)	50.023	(20.022)	47.578	(18.462)	49.309	(15.304)
Default Premium, State 2			26.996	(11.839)	-31.553	(13.378)	-9.211	(10.777)	-42.039	(12.296)
Dividend Yield	0.592	(0.361)	0.600	(0.342)	0.763	(0.337)	0.754	(0.312)	0.551	(0.318)
Lag Term	0.233	(0.040)	0.196	(0.036)	0.204	(0.036)	0.177	(0.036)	0.188	(0.036)
January Premium State 1	0.079	(0.009)	0.198	(0.025)	0.160	(0.018)	0.191	(0.021)	0.146	(0.016)
January Premium, State 2			0.060	(0.008)	0.055	(0.008)	0.066	(0.007)	0.052	(0.007)
Variance Parameters										
Constant, State 1	-0.842	(0.463)	-5.361	(0.321)	-5.694	(0.257)	-5.105	(0.356)	-6.216	(0.350)
Constant, State 2			-7.003	(0.202)	-7.177	(0.311)	-6.778	(0.206)	-7.228	(0.301)
Interest Rate, State 1	15.368	(10.070)	81.685	(54.354)	103.639	(41.905)	33.101	(56.448)	125.306	(66.572)
Interest Rate, State 2			88.572	(32.389)	85.874	(56.561)	48.985	(37.698)	120.508	(62.371)
ARCH, State 1	0.185	(0.058)	-0.218	(0.110)	-0.208	(0.117)	-0.197	(0.104)	-0.021	(0.115)
ARCH, State 2			0.196	(0.192)	0.510	(0.136)	0.177	(0.167)	0.429	(0.188)
Lagged Variance	0.884	(0.065)								
Transition Probability Parameters										
Constant, State 1			1.638	(0.213)			1.816	(0.166)	1.453	(0.184)
Constant, State 2			1.816	(0.183)						
Leading Indicator, State 1					-0.144	(0.123)	-0.251	(0.129)	-0.255	(0.121)
Leading Indicator, State 2					0.174	(0.125)	-0.069	(0.088)	-0.006	(0.090)
Degrees of Freedom	4.113	(0.628)							7.010	(3.560)
Log Likelihood	773.235		787.408		768.676		786.973		792.596	

Note: Standard errors appear in parentheses to the right of the parameter estimates.

Table 2
Large Firms

The following Markov switching model was estimated for excess returns on the large firm portfolio

Single State Model:

$\rho_t = \beta_0 + \beta_1 I1_{t-1} + \beta_2 Def_{t-1} + \beta_3 Yield_{t-1} + \varepsilon_t$

$\varepsilon_t \sim t(0,v,h_t)$, $\ln(h_t)=\lambda_0 + \lambda_1 I1_{t-1} + \lambda_2[(|\varepsilon_{t-1}|/(h_{t-1})^{1/2})-(2/\pi)] + \lambda_3 \ln(h_{t-1})$

Gaussian Mixture:

$\rho_t = \beta_{0,St} + \beta_{1,St} I1_{t-1} + \beta_{2,St} Def_{t-1} + \beta_3 Yield_{t-1} + \beta_4 \rho_{t-1} + \beta_{5,St} Jan_t + \varepsilon_t$

$\varepsilon_t \sim N(0,h_{St})$, $\ln(h_{St})=\lambda_0 + \lambda_{1,St} I1_{t-1} + \lambda_{2,St}[(|\varepsilon_{t-1}|/(h_{t-1})^{1/2})-(2/\pi)]$

Model I: $p_t=Prob(s_t=1|s_{t-1}=1)=\Phi(\pi_0)$, $q_t=Prob(s_t=2|s_{t-1}=2)=\Phi(\pi_1)$

Model II: $p_t=Prob(s_t=1|s_{t-1}=1)=\Phi(\pi_1 \Delta CLI_{t-2})$, $q_t=Prob(s_t=2|s_{t-1}=2)=\Phi(\pi_2 \Delta CLI_{t-2})$

Model III: $p_t=Prob(s_t=1|s_{t-1}=1)=\Phi(\pi_0+\pi_1 \Delta CLI_{t-2})$, $q_t=Prob(s_t=2|s_{t-1}=2)=\Phi(\pi_0+\pi_2 \Delta CLI_{t-2})$

Gaussian and student-t Mixture:

Same transition probability specification as Model III with densities:

$\varepsilon_t \sim N(0,h_{St})$ if $S_t=1$

$\varepsilon_t \sim t(0,v,h_{St})$ if $S_t=2$.

ρ_t is the monthly excess returns on the large firm portfolio, *I1* is the one-month T- bill rate, *Def* is the default premium, *Yield* is the dividend yield and ΔCLI is the annual rate of growth of the Composite Index of Leading Indicators. The sample period is 1954-1997.

	Single-State Model		Gaussian Mixture Model I		Gaussian Mixture Model II		Gaussian Mixture Model III		Gaussian, Student-t Mixture Model	
Mean Parameters										
Constant, State 1	-0.008	(0.008)	-0.022	(0.009)	-0.031	(0.012)	-0.021	(0.009)	-0.021	(0.008)
Constant, State 2			-0.009	(0.006)	0.007	(0.010)	-0.008	(0.006)	-0.007	(0.006)
Interest Rate, State 1	-3.688	(0.771)	-5.408	(1.414)	-4.382	(1.668)	-5.491	(1.476)	-5.559	(1.223)
Interest Rate, State 2			-2.765	(0.791)	-2.933	(1.349)	-2.898	(0.739)	-2.868	(0.722)
Default Premium, State 1	13.269	(6.494)	25.853	(10.013)	36.274	(12.384)	25.840	(10.287)	27.010	(8.795)
Default Premium, State 2			6.916	(7.742)	-4.579	(10.908)	6.900	(7.343)	6.137	(7.617)
Dividend Yield	0.659	(0.258)	0.841	(0.191)	0.662	(0.261)	0.822	(0.179)	0.813	(0.200)
Variance Parameters										
Constant, State 1	-1.456	(0.760)	-6.561	(0.263)	-6.839	(0.277)	-6.545	(0.278)	-6.619	(0.250)
Constant, State 2			-8.629	(0.206)	-7.644	(0.373)	-8.598	(0.190)	-8.588	(0.163)
Interest Rate, State 1	22.574	(13.121)	65.105	(43.003)	124.239	(44.108)	65.105	(45.251)	49.417	(41.077)
Interest Rate, State 2			188.622	(33.929)	70.823	(56.468)	188.620	(31.359)	190.627	(25.983)
ARCH, State 1	0.186	(0.064)	0.146	(0.122)	0.245	(0.122)	0.149	(0.130)	0.142	(0.098)
ARCH, State 2			-1.667	(0.276)	-0.169	(0.178)	-1.620	(0.247)	-1.620	(0.300)
Lagged Variance	0.804	(0.104)								
Transition Probability Parameters										
Constant, State 1			1.472	(0.203)			1.391	(0.180)	1.420	(0.183)
Constant, State 2			1.348	(0.201)						
Leading Indicator, State 1					-0.126	(0.119)	-0.026	(0.084)	-0.026	(0.061)
Leading Indicator, State 2					0.121	(0.208)	-0.046	(0.066)	-0.049	(0.053)
Degrees of Freedom	9.772	(3.689)							19.279	(23.023)
Log Likelihood	985.452		1001.032		988.670		1001.117		1002.840	

Note: Standard errors appear in parentheses to the right of the parameter estimates.

Table 3

Mean Squared Forecast Errors Performance

In-Sample Forecasting Results (1954:1 - 1997:12)

Small Firms

	Full Sample		Recession Periods		Expansion Periods	
	MSFE	DM Test	MSFE	DM Test	MSFE	DM Test
Single-State	4.105		6.946		3.598	
Gaussian Mixture						
Model I	3.819	1.634	5.474	1.542	3.523	0.720
Model II	3.943	1.329	5.951	1.441	3.585	0.206
Model III	3.779	1.694	5.263	1.517	3.514	0.898
Gaussian, student-t Mixture	3.833	2.020	5.681	1.667	3.503	1.300

Large Firms

	Full Sample		Recession Periods		Expansion Periods	
	MSFE	DM Test	MSFE	DM Test	MSFE	DM Test
Single-State	1.546		2.601		1.358	
Gaussian Mixture						
Model I	1.547	-0.083	2.522	1.237	1.373	-1.140
Model II	1.537	0.892	2.514	1.624	1.362	-0.727
Model III	1.549	-0.185	2.537	1.013	1.372	-1.089
Gaussian, student-t Mixture	1.547	-0.100	2.529	1.095	1.372	-1.113

Out-of-Sample Forecasting Results (1976:1 - 1997:12)

Small Firms

	Full Sample		Recession Periods		Expansion Periods	
	MSFE	DM Test	MSFE	DM Test	MSFE	DM Test
Single-State	3.982		5.171		3.829	
Gaussian Mixture						
Model I	4.240	-1.239	5.680	-0.853	4.055	-1.003
Model II	3.962	0.214	5.918	-0.836	3.710	1.142
Model III	3.769	1.135	4.941	0.748	3.618	1.012

Large Firms

	Full Sample		Recession Periods		Expansion Periods	
	MSFE	DM Test	MSFE	DM Test	MSFE	DM Test
Single-State	1.703		2.497		1.601	
Gaussian Mixture						
Model I	1.712	-0.194	2.424	0.353	1.621	-0.430
Model II	1.727	-0.689	2.428	0.298	1.637	-1.386
Model III	1.736	-0.921	2.639	-0.841	1.620	-0.556

Note: For the Gaussian mixtures, Model I assumes constant transition probabilities, Model II includes only the leading indicator in the state transition equation and Model III includes both a constant and the leading indicator. The Gaussian and student-t mixture adopts the same transition equation as model III. The DM statistic tests the null that the MSFE of the single state model is no higher than that of the mixture models.

Table 4
Market Timing Test

In-Sample Forecasting Results (1954:1 - 1997:12)

Small Firms

	Full Sample		Recession Periods		Expansion Periods	
	%Correct Signs	PT Test	%Correct Signs	PT Test	%Correct Signs	PT Test
Single-State	0.646	6.386	0.713	3.826	0.634	5.251
Gaussian Mixture						
Model I	0.636	6.132	0.688	3.402	0.627	5.261
Model II	0.644	6.860	0.700	3.605	0.634	5.947
Model III	0.644	6.569	0.688	3.428	0.636	5.622
Gaussian and student-t Mixture	0.655	7.028	0.688	3.428	0.650	6.114

Large Firms

	Full Sample		Recession Periods		Expansion Periods	
	%Correct Signs	PT Test	%Correct Signs	PT Test	%Correct Signs	PT Test
Single-State	0.598	2.306	0.663	2.905	0.587	1.029
Gaussian Mixture						
Model I	0.610	2.904	0.700	3.549	0.594	1.145
Model II	0.595	2.303	0.650	2.631	0.585	1.162
Model III	0.623	3.601	0.700	3.549	0.609	1.904
Gaussian and student-t Mixture	0.631	4.019	0.713	3.787	0.616	2.249

Out-of-Sample Forecasting Results (1976:1 - 1997:12)

Small Firms

	Full Sample		Recession Periods		Expansion Periods	
	%Correct Signs	PT Test	%Correct Signs	PT Test	%Correct Signs	PT Test
Single-State	0.580	3.143	0.700	2.233	0.564	2.656
Gaussian Mixture						
Model I	0.583	3.373	0.733	2.583	0.564	2.703
Model II	0.583	3.110	0.633	1.787	0.577	3.000
Model III	0.576	3.033	0.733	2.636	0.555	2.424

Large Firms

	Full Sample		Recession Periods		Expansion Periods	
	%Correct Signs	PT Test	%Correct Signs	PT Test	%Correct Signs	PT Test
Single-State	0.519	0.154	0.600	0.959	0.509	-0.154
Gaussian Mixture						
Model I	0.489	-1.067	0.667	1.717	0.466	-1.677
Model II	0.515	0.622	0.567	0.530	0.509	0.532
Model III	0.496	0.113	0.633	1.325	0.479	-0.276

Note: For the Gaussian mixtures, Model I assumes constant transition probabilities, Model II includes only the leading indicator in the transition equation and Model III includes both a constant and the leading indicator. The Gaussian and student-t mixture adopts the same transition equation as model III. The PT statistic tests the null of independence between the sign of the realized and predicted excess returns and is asymptotically normally distributed.

Table 5
Pagan-Schwert Tests of Predictability of Higher Order Moments

Powers of the residuals from the return equation are projected on their conditional expectation
$(\varepsilon_t)^n = a + b(E(\varepsilon_t)^n) + \upsilon_t$

In-Sample Forecasting Results (1954:1 - 1997:12)

Small Firms

	Single-State Model		Gaussian Mixture Model I		Gaussian Mixture Model II		Gaussian Mixture Model III		Gaussian and student-t Mixture	
n=2	1.292	(0.879)	0.569	(0.894)	0.972	(0.892)	0.761	(0.792)	0.182	(0.826)
	0.664	(0.230)	0.861	(0.254)	0.720	(0.241)	0.770	(0.216)	1.117	(0.297)
n=3	0.018	(0.162)	-0.048	(0.129)	-0.046	(0.137)	-0.113	(0.125)	-0.076	(0.132)
	0.000	(0.000)	1.665	(1.225)	1.827	(1.138)	2.413	(1.584)	4.247	(2.868)
n=4	0.107	(0.051)	0.040	(0.063)	0.104	(0.043)	0.033	(0.061)	0.054	(0.043)
	0.033	(0.033)	0.993	(0.806)	0.212	(0.178)	0.970	(0.759)	0.915	(0.470)

Large Firms

	Single-State Model		Gaussian Mixture Model I		Gaussian Mixture Model II		Gaussian Mixture Model III		Gaussian and student-t Mixture	
n=2	0.145	(0.393)	0.033	(0.374)	0.316	(0.388)	0.204	(0.377)	0.034	(0.394)
	0.905	(0.277)	0.969	(0.268)	0.783	(0.265)	0.867	(0.269)	1.005	(0.290)
n=3	-0.036	(0.026)	-0.031	(0.023)	-0.023	(0.026)	-0.029	(0.022)	-0.031	(0.023)
	0.000	(0.000)	-0.230	(0.978)	0.639	(1.269)	0.131	(0.914)	-0.124	(0.927)
n=4	0.005	(0.007)	0.008	(0.005)	0.008	(0.006)	0.010	(0.005)	0.009	(0.005)
	0.679	(0.434)	0.418	(0.277)	0.403	(0.291)	0.291	(0.213)	0.316	(0.232)

Out-of-Sample Forecasting Results (1976:1 - 1997:12)

Small Firms

	Single-State Model		Gaussian Mixture Model I		Gaussian Mixture Model II		Gaussian Mixture Model III	
n=2	3.256	(0.930)	2.011	(1.542)	3.153	(0.929)	2.618	(1.164)
	0.100	(0.059)	0.467	(0.332)	0.100	(0.068)	0.282	(0.227)
n=3	-0.070	(0.232)	-0.149	(0.239)	-0.193	(0.252)	-0.165	(0.223)
	0.000	(0.000)	0.841	(0.802)	0.018	(0.068)	1.557	(1.351)
n=4	0.174	(0.084)	0.109	(0.083)	0.169	(0.084)	0.150	(0.095)
	-0.005	(0.004)	0.474	(0.550)	-0.009	(0.016)	-0.021	(0.435)

Large Firms

	Single-State Model		Gaussian Mixture Model I		Gaussian Mixture Model II		Gaussian Mixture Model III	
n=2	1.297	(0.424)	1.262	(0.427)	1.566	(0.394)	1.287	(0.389)
	0.208	(0.148)	0.227	(0.155)	0.084	(0.128)	0.209	(0.130)
n=3	-0.013	(0.045)	-0.028	(0.042)	-0.013	(0.045)	-0.020	(0.046)
	0.000	(0.000)	-0.077	(0.529)	0.099	(0.310)	-0.365	(0.202)
n=4	0.020	(0.014)	0.017	(0.011)	0.018	(0.011)	0.017	(0.011)
	-0.184	(0.255)	0.009	(0.094)	-0.102	(0.130)	-0.015	(0.068)

Note: For the Gaussian Mixtures, Model I assumes constant transition probabilities, Model II includes only the Leading Indicator in the transition equation and Model III includes both a constant and the Leading Indicator. The Gaussian and student-t mixture adopts the same transition equation as model III.

Table 6

Predictive Density Tests

In-Sample Forecasting Results (1954:1 - 1997:12)

		Small Firms					Large Firms				
		Single State Model	Gaussian Mixture Model I	Gaussian Mixture Model II	Gaussian Mixture Model III	Mixture Gaussian/ t-distrib.	Single State Model	Gaussian Mixture Model I	Gaussian Mixture Model II	Gaussian Mixture Model III	Mixture Gaussian/ t-distrib.
Parameter	**Constant**	-0.006	0.003	-0.038	0.005	-0.018	-0.037	0.034	0.014	0.034	0.034
Estimates	**Autoreg. term**	-0.023	-0.035	-0.011	-0.040	-0.014	0.021	0.033	0.029	0.036	0.038
	Variance	0.999	0.999	0.976	0.983	1.037	0.990	0.995	0.997	1.001	1.009
P-values	**Constant = 0**	0.603	0.418	0.809	0.359	0.744	0.625	0.444	0.505	0.412	0.379
of the test	**Autoregressive coefficient = 0**	0.894	0.938	0.378	0.913	0.691	0.397	0.436	0.753	0.440	0.436
	Variance = 1	0.971	0.981	0.439	0.589	0.229	0.747	0.876	0.923	0.963	0.777
	Joint test	0.962	0.882	0.702	0.773	0.630	0.798	0.760	0.909	0.745	0.696
Kolmogorov	**Uniform Density**	0.029	0.028	0.047	0.030	0.028	0.028	0.026	0.021	0.028	0.031
Smirnov	**Normal Density**	0.028	0.028	0.047	0.030	0.028	0.030	0.027	0.021	0.028	0.032

Kolmogorov-Smirnov Critical Values: 0.059 (5%) 0.071 (1%)

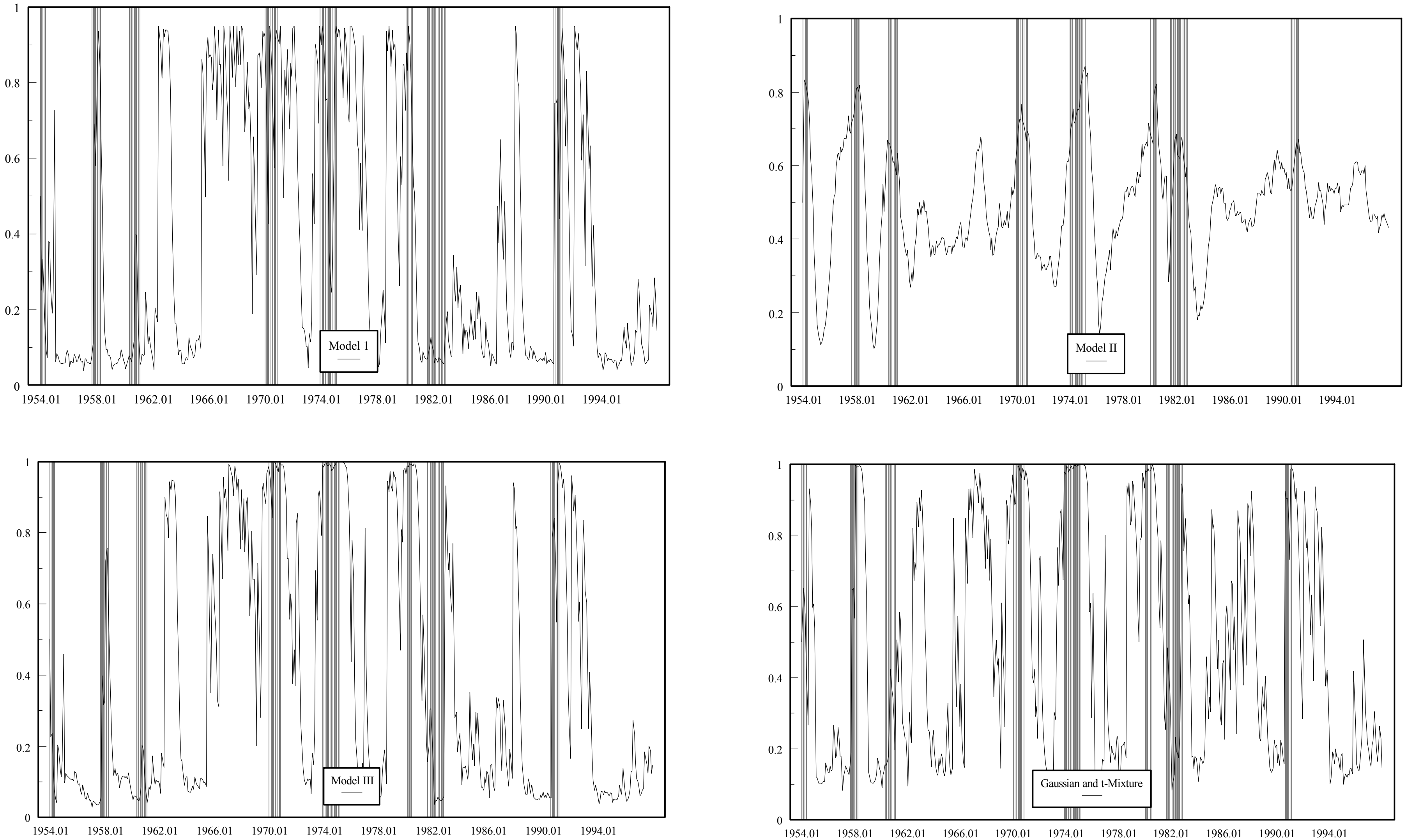
Out -of-Sample Forecasting Results (1976:1 - 1997:12)

		Small Firms				Large Firms			
		Single State Model	Gaussian Mixture Model I	Gaussian Mixture Model II	Gaussian Mixture Model III	Single State Model	Gaussian Mixture Model I	Gaussian Mixture Model II	Gaussian Mixture Model III
Parameter	**Constant**	0.112	-0.214	-0.114	-0.111	0.146	-0.117	-0.142	-0.173
Estimates	**Autoregressive**	-0.081	-0.057	-0.044	-0.134	0.034	0.005	-0.020	0.021
	Variance	0.862	1.083	0.811	0.990	0.964	1.008	1.019	0.979
P-values	**Constant = 0**	0.181	0.351	0.469	0.029	0.580	0.940	0.745	0.729
of the test	**Autoregressive coefficient = 0**	0.039	0.002	0.025	0.071	0.015	0.063	0.026	0.005
	Variance = 1	0.001	0.061	0.000	0.810	0.401	0.855	0.668	0.636
	All of them	0.001	0.000	0.000	0.028	0.100	0.309	0.120	0.048
Kolmogorov	**Uniform Density**	0.103	0.104	0.141	0.080	0.102	0.060	0.098	0.088
Smirnov	**Normal Density**	0.106	0.106	0.140	0.083	0.106	0.057	0.102	0.091

Kolmogorov-Smirnov Critical Values: 0.084 (5%) 0.101 (1%)

Note: For the Gaussian mixtures, Model I assumes constant transition probabilities, Model II includes only the Leading Indicator in the transition equation and model III includes both a constant and the Leading Indicator. The Gaussian and student-t mixture adopts the same transition equation as model III. The Kolmogorov-Smirnoff statistic tests the null that the predictive density implied by a given model is correctly specified. Under the null of no misspecification, the inverted probability transforms should be either iid normally distributed with zero mean, unit variance and zero autocorrelation or follow a uniform distribution, depending on which probability transform is used. The p-values measure the probability that the null is satisfied.

Figure 1
Probability of State 1, Small Firms



Note: Shaded areas indicate NBER recession periods. Model I assumes a constant transition probability, Model II includes only the Leading Indicator in the state transition equation, Model III includes both a constant and the Leading Indicator in the state The transitions. Gaussian and student-t mixture model uses the same state transition as model III.

Figure 1 (cont)
Probability of State 1 Large Firms

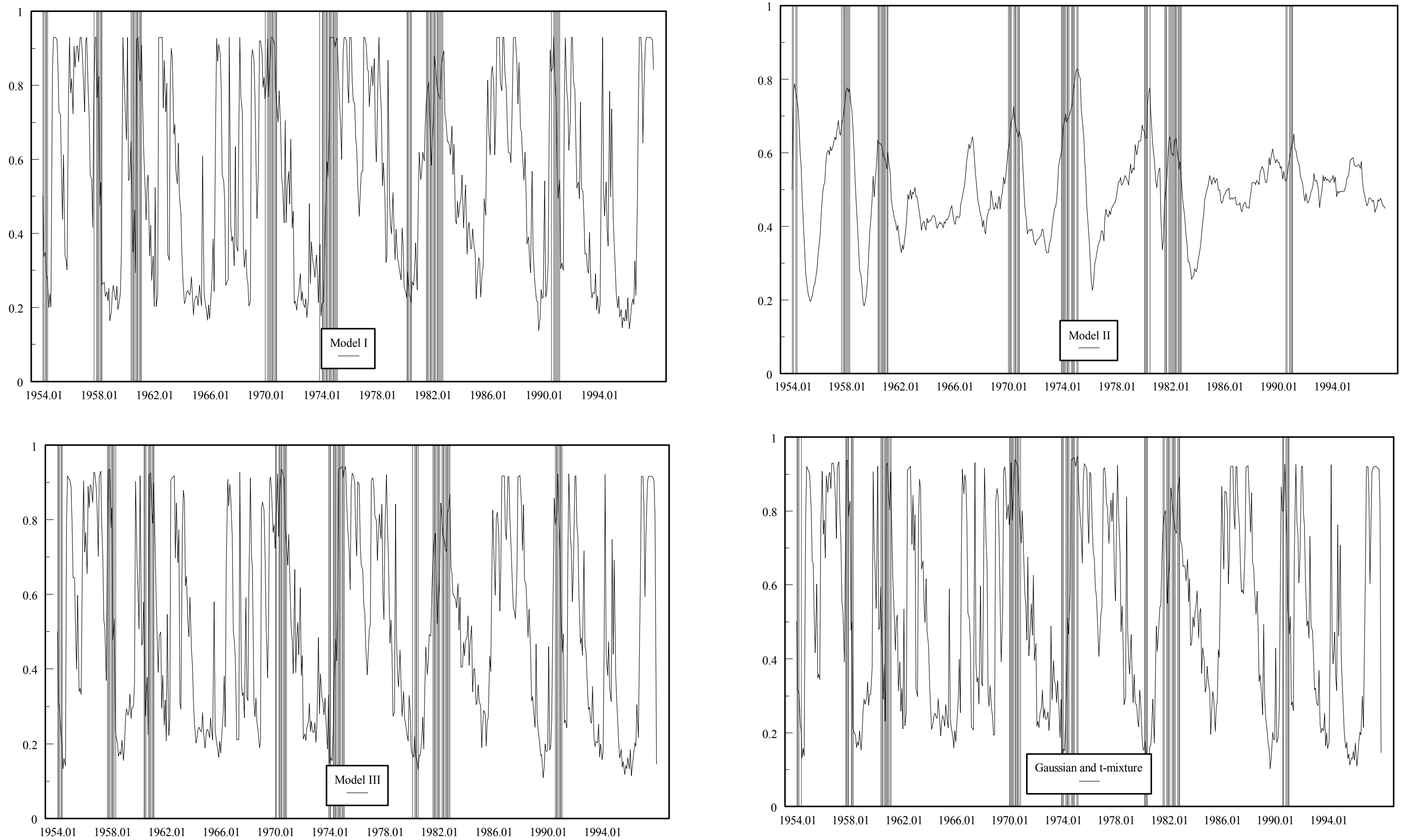
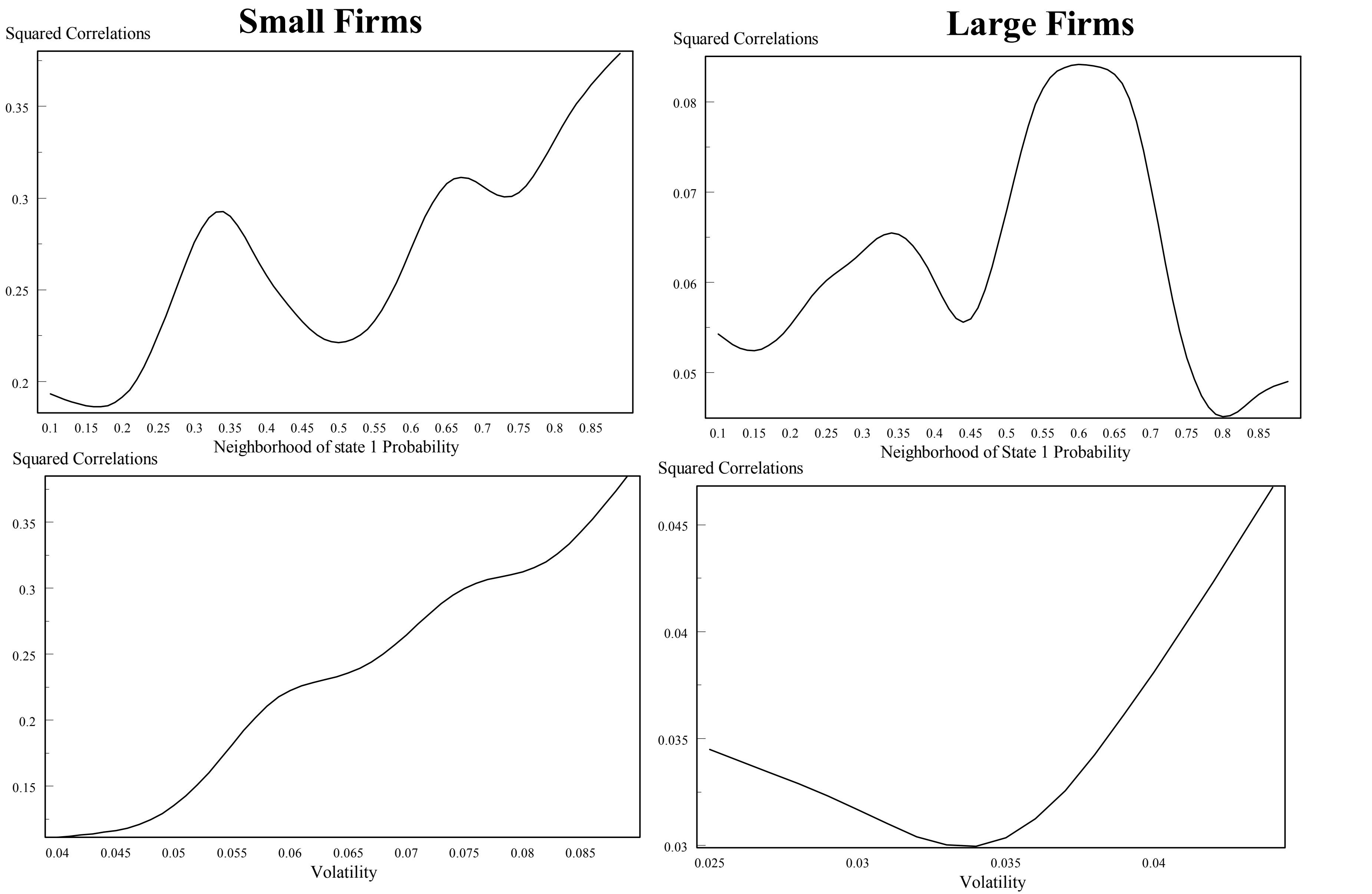
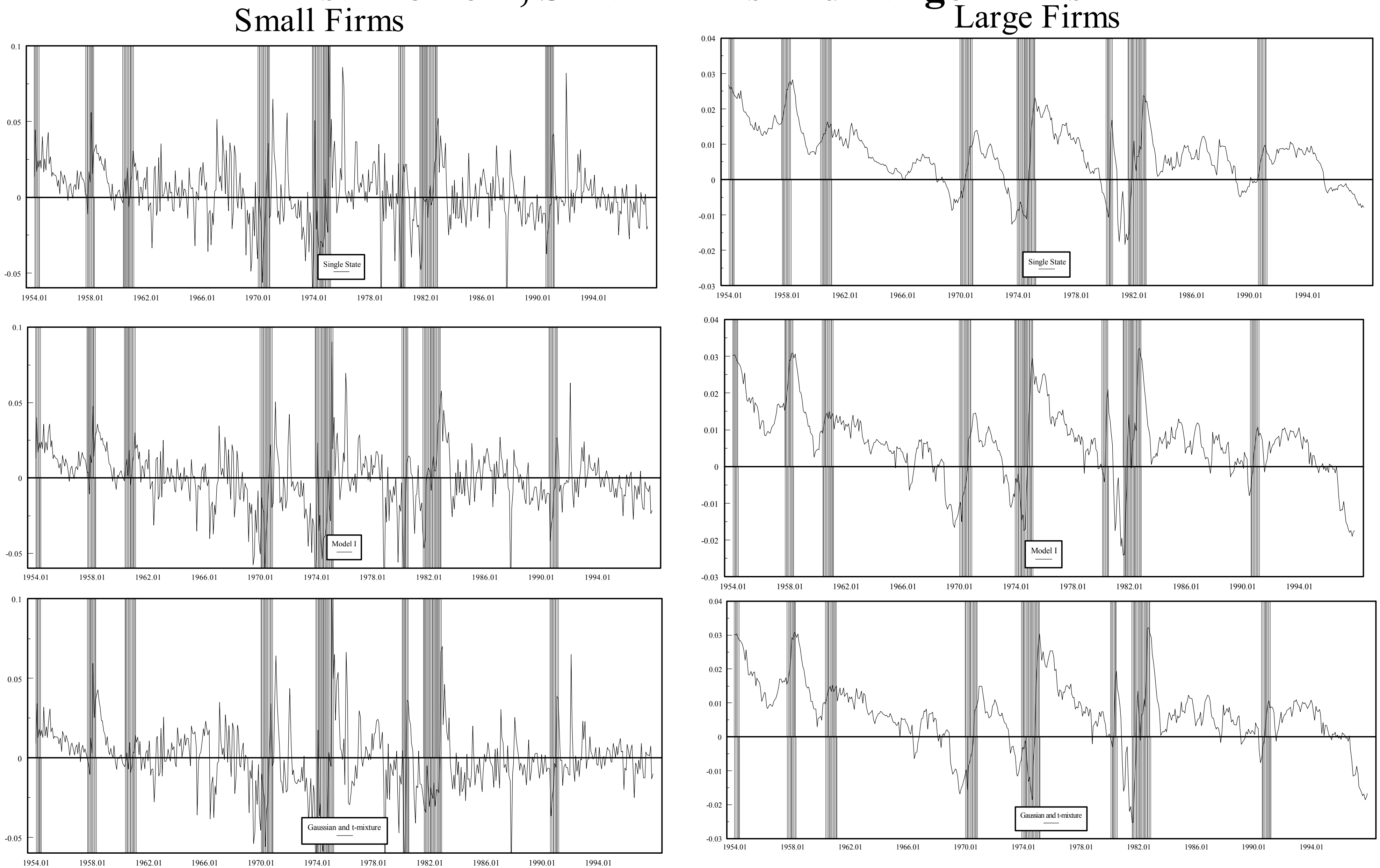


Figure 2
Variations in Predictability of Returns



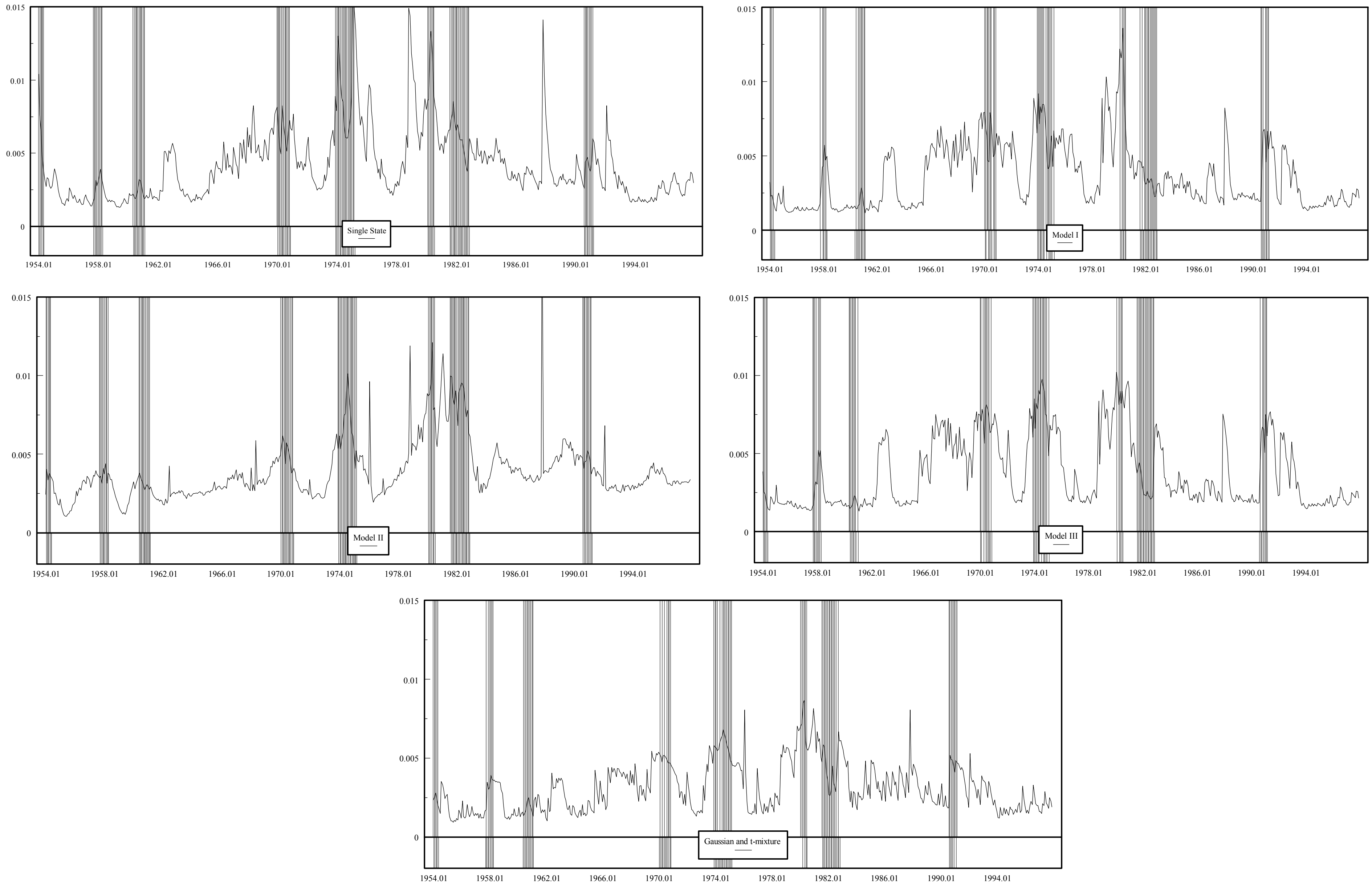
Note: These figures plot estimates of the squared correlation between predicted and realized excess returns computed in a neighborhood of state- 1 probabilities (upper windows) and conditional volatilities (lower windows). A tricubic spline was used to produce the plots.

Figure 3
First Moment, Small Firms and Large Firms



Note: Shaded areas indicate NBER recession periods. The figures plot the conditional mean of excess returns implied by the single state model and the different specifications listed in the footnote to Figure 1. Model II and III are not plotted because they are very similar to the Gaussiand and t- mixture.

Figure 4
Second Moment, Small Firms



Note: Shaded areas indicate NBER recession periods. The figures plot the conditional variance of excess returns for the single state model and the specifications listed in the footnote to Figure 1

Figure 4 (cont)
Second Moment, Large Firms

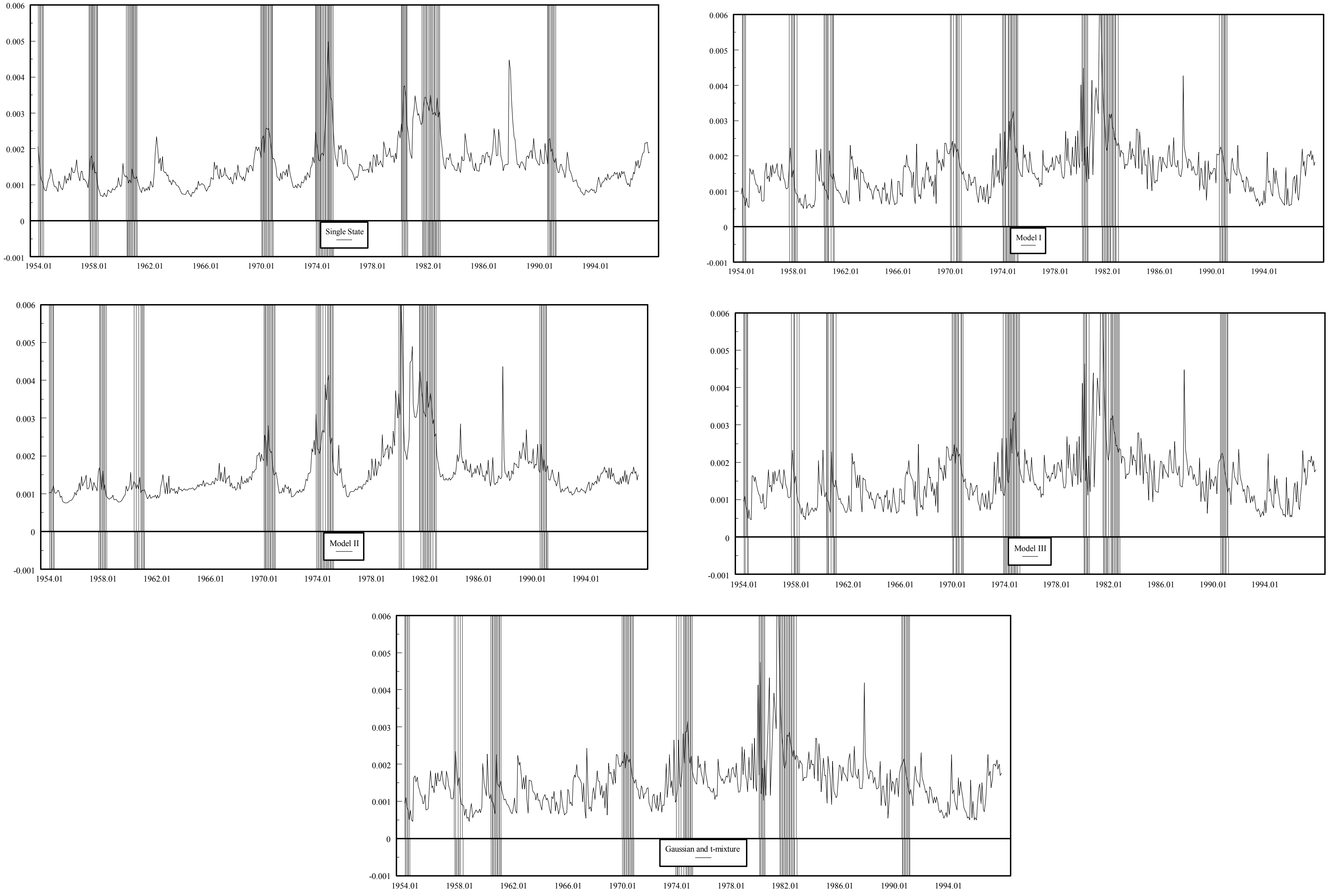
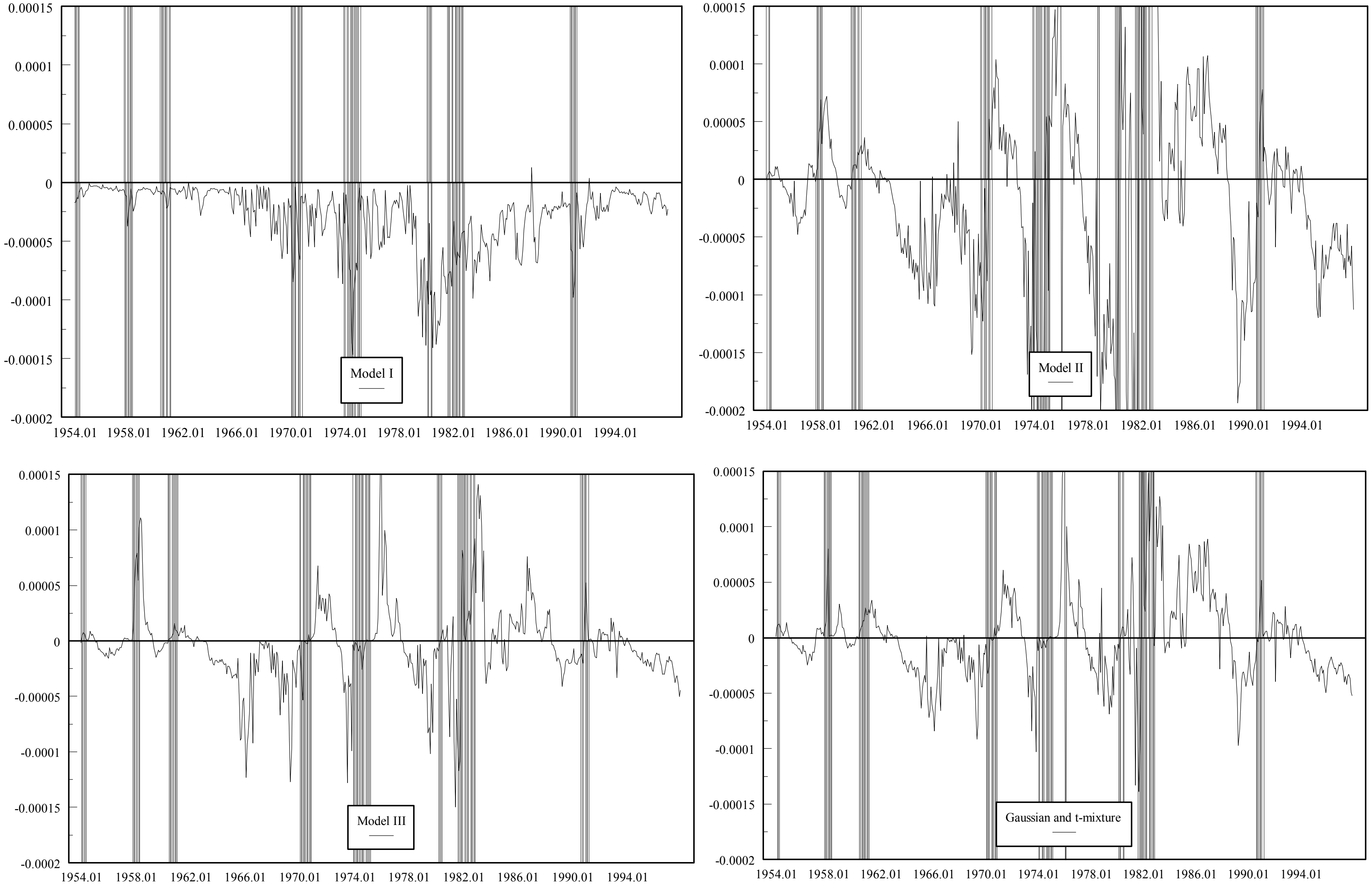


Figure 5
Third Moment, Small Firms



Note: Shaded areas indicate NBER recession periods. The figures plot the conditional skewness of excess returns for each of the specifications listed in the footnote 1

Figure 5 (cont)
Third Moment, Large Firms

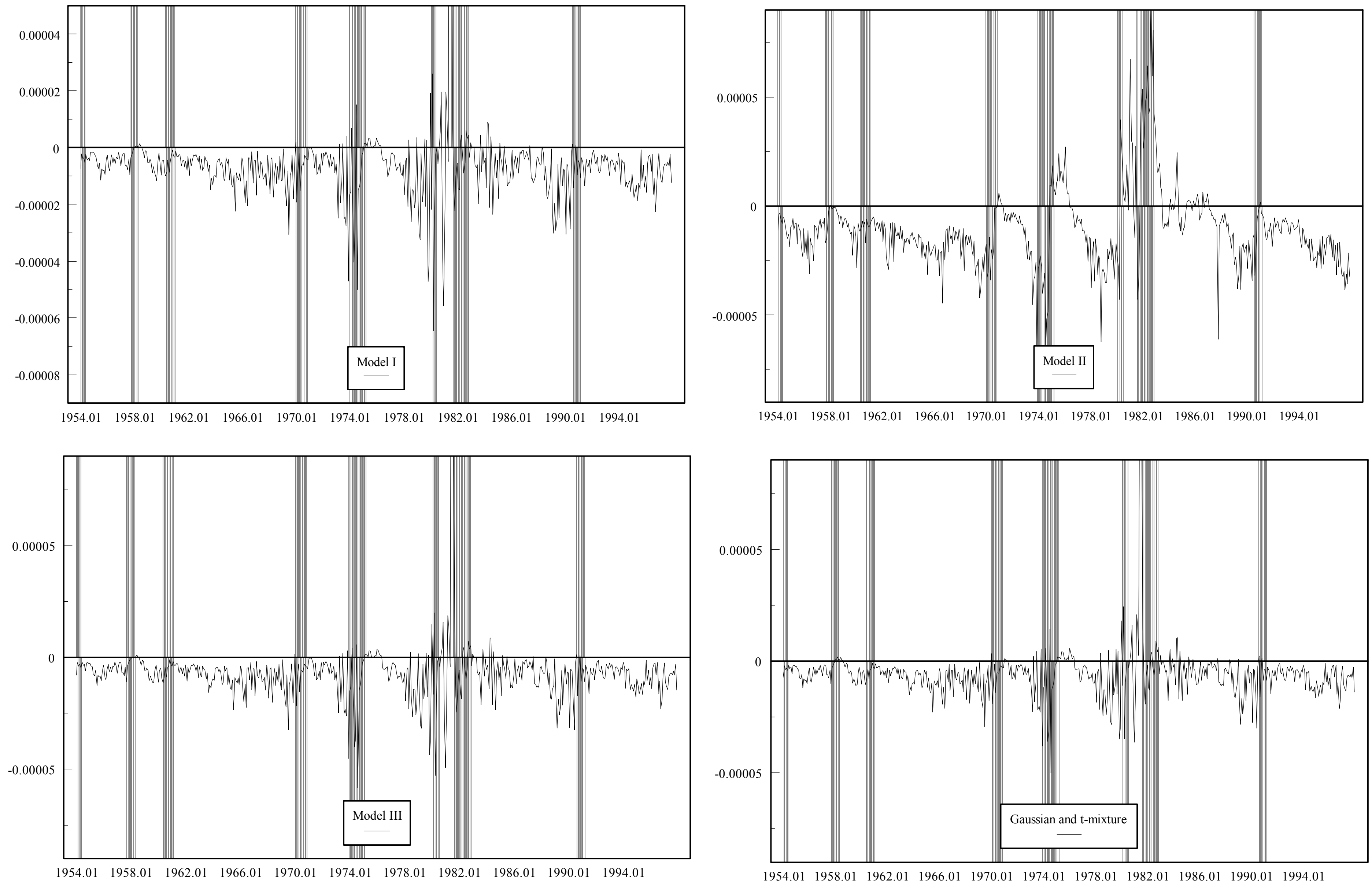
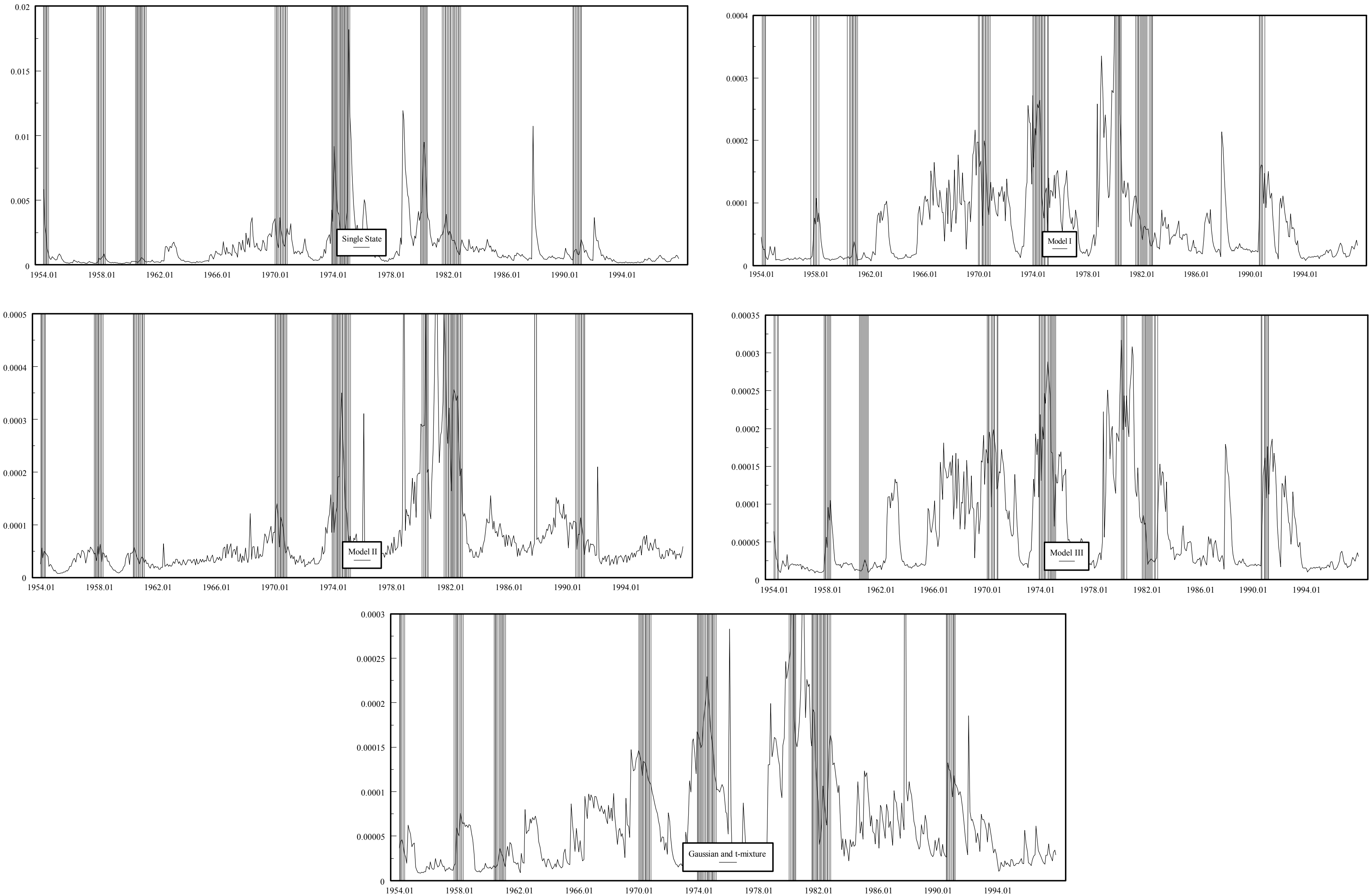
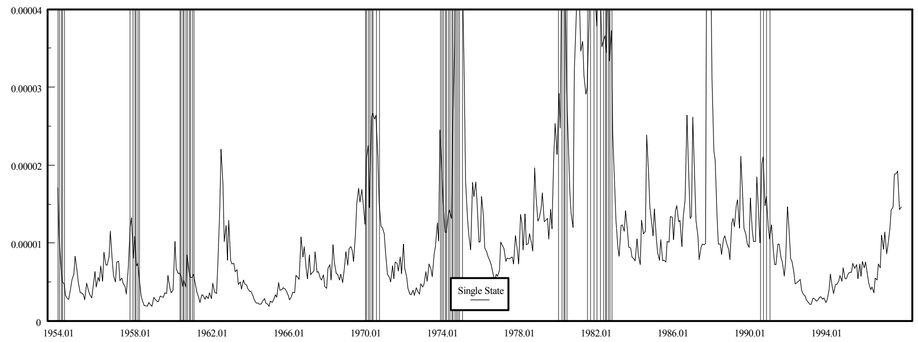


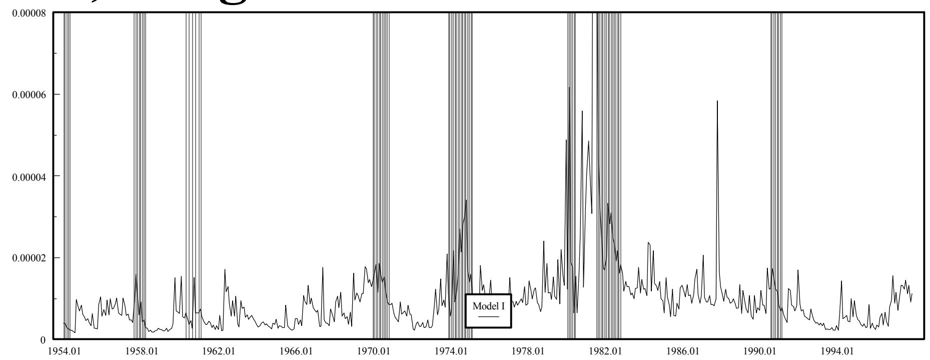
Figure 6
Fourth Moment, Small Firms

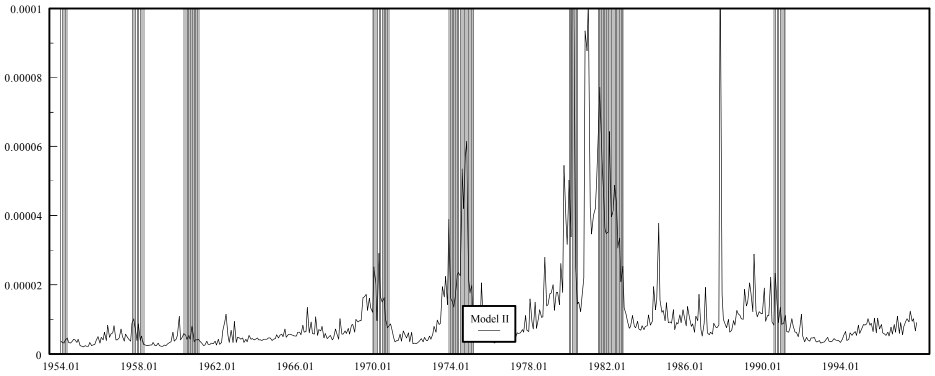


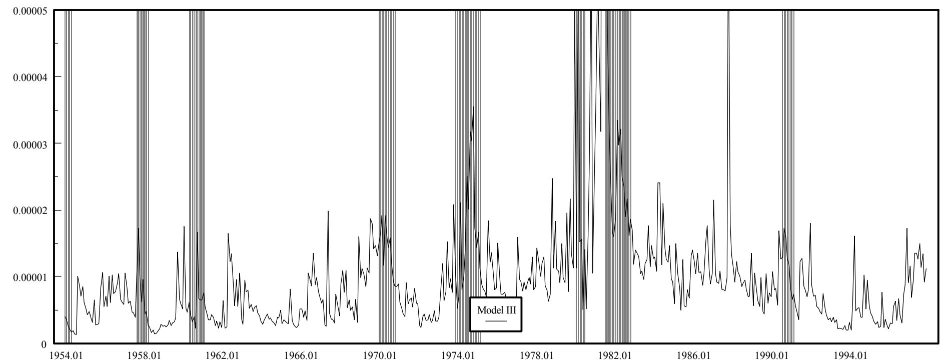
Note: Shaded areas indicate NBER recession periods. The figures plot the conditional kurtosis of excess returns for the single state model and the specifications listed in the footnote to Figure 1

Figure 6 (cont)
Fourth Moment, Large Firms









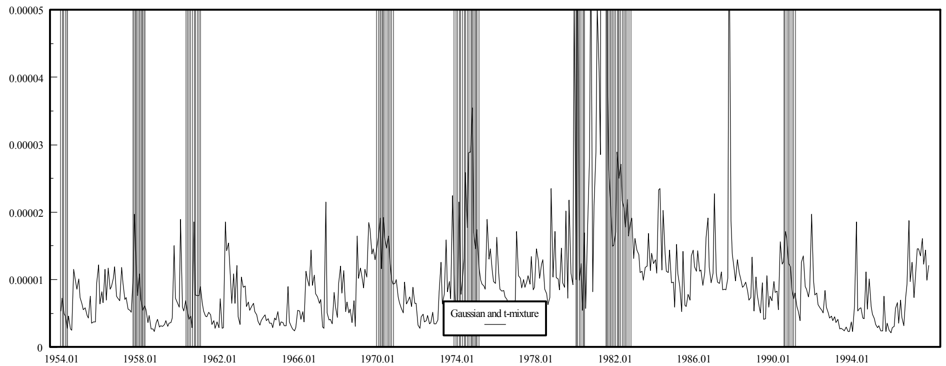
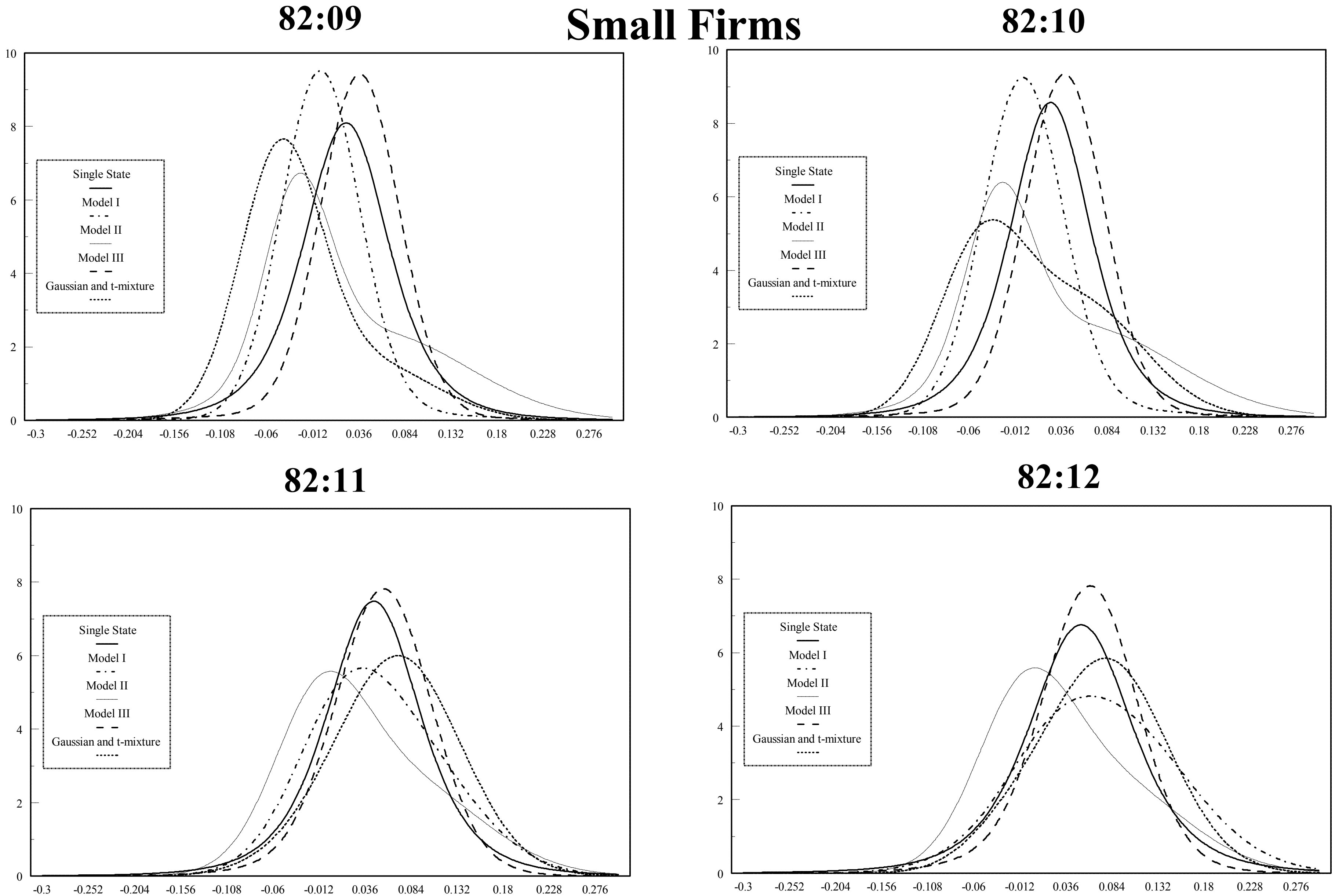


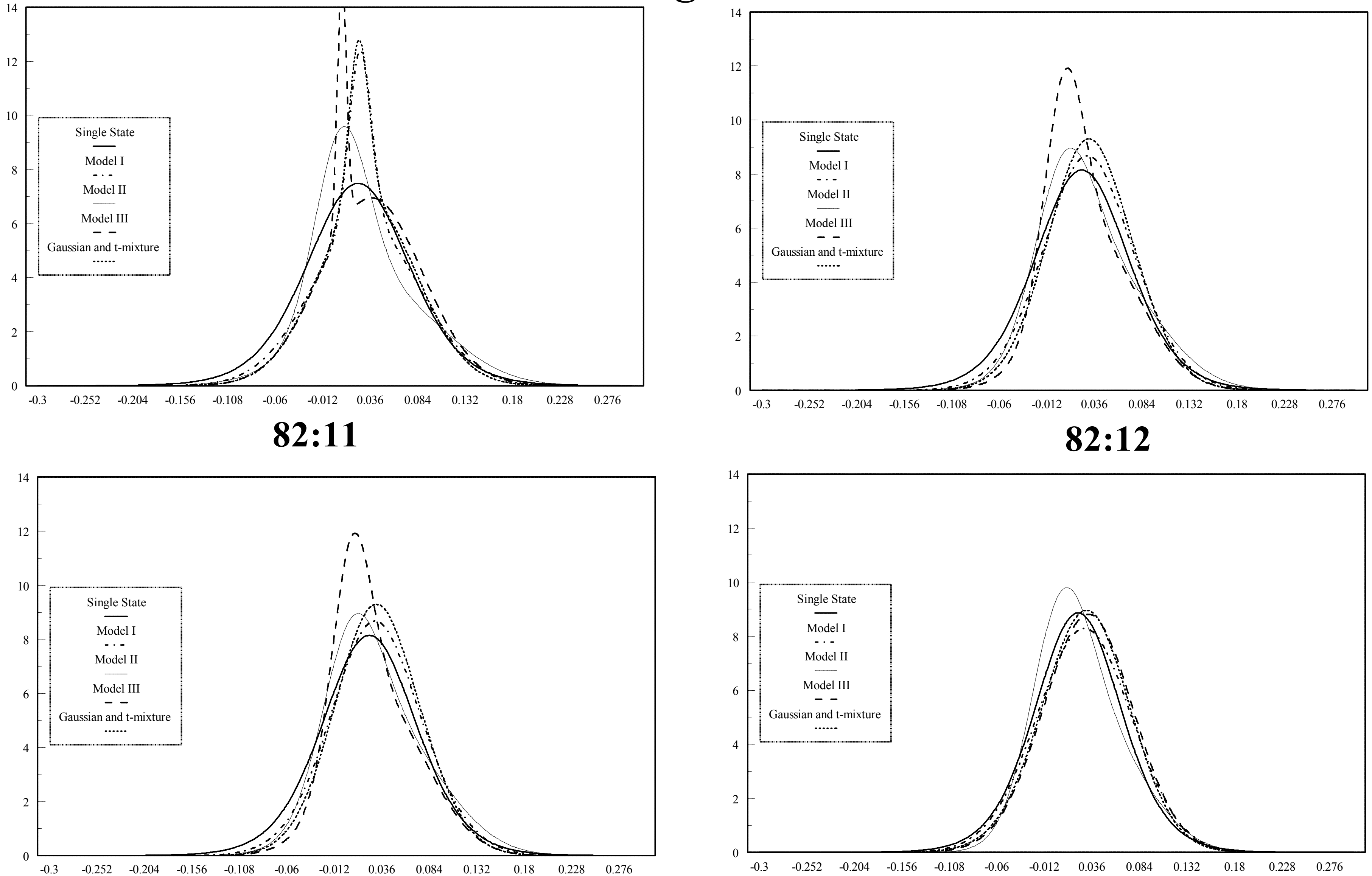
Figure 7
Density functions 82:09-82:12
Small Firms



Note: These figures plot estimates of the density function assocated with each of the different model specifications for the period 1982:09-1982:12

Figure 7 (cont)
Density functions 82:09-82:12
Large Firms

82:09 **82:10**



Note: These figures plot estimates of the density function assocated with each of the different model specifications for the period 1982:09-1982:12

European Central Bank Working Paper Series

1 "A global hazard index for the world foreign exchange markets" by V. Brousseau and F. Scacciavillani, May 1999.

2 "What does the single monetary policy do? A SVAR benchmark for the European Central Bank" by C. Monticelli and O. Tristani, May 1999.

3 "Fiscal policy effectiveness and neutrality results in a non-Ricardian world" by C. Detken, May 1999.

4 "From the ERM to the euro: new evidence on economic and policy convergence among EU countries" by I. Angeloni and L. Dedola, May 1999.

5 "Core inflation: a review of some conceptual issues" by M. Wynne, May 1999.

6 "The demand for M3 in the euro area" by G. Coenen and J.-L. Vega, September 1999.

7 "A cross-country comparison of market structures in European banking" by O. de Bandt and E. P. Davis, September 1999.

8 "Inflation zone targeting" by A. Orphanides and V. Wieland, October 1999.

9 "Asymptotic confidence bands for the estimated autocovariance and autocorrelation functions of vector autoregressive models" by G. Coenen, January 2000.

10 "On the effectiveness of sterilized foreign exchange intervention" by R. Fatum, February 2000.

11 "Is the yield curve a useful information variable for the Eurosystem?" by J. M. Berk and P. van Bergeijk, February 2000.

12 "Indicator variables for optimal policy" by L. E. O. Svensson and M. Woodford, February 2000.

13 "Monetary policy with uncertain parameters" by U. Söderström, February 2000.

14 "Assessing nominal income rules for monetary policy with model and data uncertainty" by G. D. Rudebusch, February 2000.

15 "The quest for prosperity without inflation" by A. Orphanides, March 2000.

16 "Estimating the implied distribution of the future short term interest rate using the Longstaff-Schwartz model" by P. Hördahl, March 2000.

17 "Alternative measures of the NAIRU in the euro area: estimates and assessment" by S. Fabiani and R. Mestre, March 2000.

18 "House prices and the macroeconomy in Europe: Results from a structural VAR analysis" by M. Iacoviello, April 2000.

19 "The euro and international capital markets" by C. Detken and P. Hartmann, April 2000.

20 "Convergence of fiscal policies in the euro area" by O. De Bandt and F. P. Mongelli, May 2000.

21 "Firm size and monetary policy transmission: evidence from German business survey data" by M. Ehrmann, May 2000.

22 "Regulating access to international large value payment systems" by C. Holthausen and T. Rønde, June 2000.

23 "Escaping Nash inflation" by In-Koo Cho and T. J. Sargent, June 2000.

24 "What horizon for price stability" by F. Smets, July 2000.

25 "Caution and conservatism in the making of monetary policy" by P. Schellekens, July 2000.

26 "Which kind of transparency? On the need for clarity in monetary policy-making" by B. Winkler, August 2000.

27 "This is what the US leading indicators lead" by M. Camacho and G. Perez-Quiros, August 2000.

28 "Learning, uncertainty and central bank activism in an economy with strategic interactions" by M. Ellison and N. Valla, August 2000.

29 "The sources of unemployment fluctuations: an empirical application to the Italian case" by S. Fabiani, A. Locarno, G. Oneto and P. Sestito, September 2000.

30 "A small estimated euro area model with rational expectations and nominal rigidities" by G. Coenen and V. Wieland, September 2000.

31 "The disappearing tax base: Is foreign direct investment eroding corporate income taxes?" by R. Gropp and K. Kostial, September 2000.

32 "Can indeterminacy explain the short-run non-neutrality of money?" by F. De Fiore, September 2000.

33 "The information content of M3 for future inflation" by C. Trecroci and J. L. Vega, October 2000.

34 "Capital market development, corporate governance and the credibility of exchange rate pegs" by O. Castrén and T. Takalo, October 2000.

35 "Systemic risk: A survey" by O. De Bandt and P. Hartmann, November 2000.

36 "Measuring core inflation in the euro area" by C. Morana, November 2000.

37 "Business fixed investment: Evidence of a financial accelerator in Europe" by P. Vermeulen, November 2000.

38 "The optimal inflation tax when taxes are costly to collect" by F. De Fiore, November 2000.

39 "A money demand system for euro area M3" by C. Brand and N. Cassola, November 2000.

40 "Financial structure and the interest rate channel of ECB monetary policy" by B. Mojon, November 2000.

41 "Why adopt transparency? The publication of central bank forecasts" by P. M. Geraats, January 2001.

42 "An area-wide model (AWM) for the euro area" by G. Fagan, J. Henry and R. Mestre, January 2001.

43 "Sources of economic renewal: from the traditional firm to the knowledge firm" by D. R. Palenzuela, February 2001.

44 "The supply and demand for eurosystem deposits – The first 18 months" by U. Bindseil and F. Seitz, February 2001.

45 "Testing the Rank of the Hankel matrix: a statistical approach" by G. Camba-Mendez and G. Kapetanios, February 2001.

46 "A two-factor model of the German term structure of interest rates" by N. Cassola and J. B. Luís, February 2001.

47 "Deposit insurance and moral hazard: does the counterfactual matter?" by R. Gropp and J. Vesala, February 2001.

48 "Financial market integration in Europe: on the effects of EMU on stock markets" by M. Fratzscher, March 2001.

49 "Business cycle and monetary policy analysis in a structural sticky-price model of the euro area" by M. Casares, March 2001.

50 "Employment and productivity growth in service and manufacturing sectors in France, Germany and the US" by T. von Wachter, March 2001.

51 "The functional form of the demand for euro area M1" by L. Stracca, March 2001.

52 "Are the effects of monetary policy in the euro area greater in recessions than in booms?" by G. Peersman and F. Smets.

53 "An evaluation of some measures of core inflation for the euro area" by J.-L. Vega and M. A. Wynne, April 2001.

54 "Assessment criteria for output gap estimates" by G. Camba-Méndez and D. R. Palenzuela, April 2001.

55 "Modelling the demand for loans to the private sector in the euro area" by A. Calza, G. Gartner and J. Sousa, April 2001.

56 “Stabilization policy in a two country model and the role of financial frictions” by E. Faia, April 2001.

57 “Model-based indicators of labour market rigidity” by S. Fabiani and D. Rodriguez-Palenzuela, April 2001.

58 “Business cycle asymmetries in stock returns: evidence from higher order moments and conditional densities” by G. Perez-Quiros and A. Timmermann, April 2001.