Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Empirical Finance 14 (2007) 91 – 119
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

Conditional coskewness and asset pricing ☆

Daniel R. Smith *

Faculty of Business Administration, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A 1S6

Accepted 19 April 2006
Available online 23 August 2006

Abstract

We explore the empirical usefulness of conditional coskewness to explain the cross-section of equity returns. We find that coskewness is an important determinant of the returns to equity, and that the pricing relationship varies through time. In particular we find that when the conditional market skewness is positive investors are willing to sacrifice 7.87% annually per unit of gamma (a standardized measure of coskewness risk) while they only demand a premium of 1.80% when the market is negatively skewed. A similar picture emerges from the coskewness factor of Harvey and Siddique (Harvey, C., Siddique, A., 2000a. Conditional skewness in asset pricing models tests. Journal of Finance 65, 1263–1295.) (a portfolio that is long stocks with small coskewness with the market and short high coskewness stocks) which earns 5.00% annually when the market is positively skewed but only 2.81% when the market is negatively skewed. The conditional two-moment CAPM and a conditional Fama and French (Fama, E., French, K., 1992. The cross-section of expected returns. Journal of Finance 47,427–465.) three-factor model are rejected, but a model which includes coskewness is not rejected by the data. The model also passes a structural break test which many existing asset pricing models fail.

JEL classification: C12; C52; G12
Keywords: GMM; Asset pricing; Conditional; Nonlinear; Coskewness; Pricing kernel

☆ This paper is based on my doctoral thesis at the University of British Columbia. I thank Murray Carlson, Glen Donaldson, Wayne Ferson (the Editor), Adlai Fisher, Rob Grauer, Peter Klein, Alan Kraus, Robert Whitelaw, an anonymous referee and seminar participants at Simon Fraser University, the Queensland University of Technology, the University of Queensland, the University of British Columbia, the 2002 Northern Finance Association and the 2003 Australasian Banking and Finance Conference meetings for helpful discussion and comments. I gratefully acknowledge the financial support of a University of British Columbia Graduate Fellowship.
* Tel.: +1 604 291 4675.
E-mail address: drsmith@sfu.ca.

0927-5398/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2006.04.004

1. Introduction

The well documented failure of the Capital Asset Pricing Model (CAPM) has motivated much research into nonlinear asset pricing models. The basic linear single factor CAPM only applies when very restrictive conditions are met. These conditions imply that investors only care about the mean and variance of returns. However, if returns are non-normal and investors have non-quadratic utility then investors will generally care about all moments of returns (Rubinstein, 1973). The most popular higher moment asset pricing model is the three-moment CAPM of Kraus and Litzenberger (1976) which includes preferences over skewness. Skewness matters since investors with non-increasing absolute risk aversion have a preference for positive skewness, and an aversion to negative skewness. Harvey and Siddique (2000a) test the three-moment CAPM's implication that stocks with large negative coskewness with the market will earn higher risk premia. They construct a coskewness factor following the methodology used by Fama and French (1993) for their *smb* and *hml* factors (i.e., a long-minus-short zero-cost hedge portfolio). The mean return on this coskewness factor was 3.60% per year, suggesting that coskewness is economically significant. Harvey and Siddique (2000a) and that some of the empirical usefulness of *smb* and *hml* is because they proxy for this coskewness factor. They also demonstrate that some of the profitability of the momentum-based trading strategies (Jegadeesh and Titman, 1993), which have proved very difficult to explain using traditional factors (see, e.g., Fama and French, 1996; Grundy and Martin, 2001), can be explained by this coskewness factor.

We estimate and test conditional versions of both the two- and three-moment CAPM using GMM. Although we reject the conditional two-moment CAPM, we do not reject the conditional three-moment CAPM. We find that investors preference for coskewness varies over time with the extent of market skewness. Gamma is a standardized measure of coskewness risk analogous to beta. When the market is negatively skewed, the premium demanded for bearing gamma risk is 1.80% per year over the period 1963 to 1997. Investors are willing to sacrifice 7.87% annually when holding a positive gamma stock when the market is also positively skewed. The coskewness factor of Harvey and Siddique (1999) earns 5.00% annually when the market is positively skewed but only 2.81% when the market is negatively skewed.

We fit a conditional three-factor model in which the prices of factor risk vary through time and find that adding *smb* and *hml*, the size and book-to-market factors of Fama and French (1993), has little impact on the price of market beta risk when coskewness risk is included in the model. Fama and French (1993) found that including the loadings on *smb* and *hml* drives out the cross-sectional explanatory power of market beta when there is no coskewness term. At least part of the ability of *smb* and *hml* to explain returns is related to coskewness. The conditional three-factor model of Fama and French (1993) is rejected, but a combined model which includes the size and book-to-market factors and preference over both the market variance and skewness is not rejected.

Previous studies estimate stochastic discount factors (SDFs) or pricing kernels which are quadratic functions of market returns. Quadratic SDFs are consistent with the three-moment CAPM (see, e.g., Dittmar, 2002). It is common in the SDF literature to model the time-varying coefficients in the SDF as linear functions of the conditioning variables. In contrast we directly model the conditional moments of market returns as linear functions. We then use theory to relate these conditional moments to the SDF coefficients. Our three-moment CAPM-based SDF fits the data better than the equivalent traditional conditional SDF.

The remainder of the paper proceeds as follows. Section 2 discusses the conditional two- and three-moment CAPMs. Section 3 develops the empirical methodology. Section 4 then discusses the data used and presents some evidence on the importance of the conditioning information. Our results are presented in Section 5. Section 6 presents the results from estimating a multi-factor asset pricing model that includes the *smb* and *hml* factors. We conclude in Section 7 with a summary of the results. In Appendix A we compare the linear volatility model we employ with a GARCH(1,1) model and compare our methodology and reconcile our results with a recent paper by Dittmar (2002).

2. Model development and motivation

The conditional specification of the two-moment CAPM of Sharpe (1964) and Linter (1965) for asset $i=1,\ldots,N$ is:

$$E_{t-1}(r_{i,t}) = \beta_{i,t-1}E_{t-1}(r_{m,t}) \quad \text{for } i = 1,\ldots,N, \tag{1}$$

where $r_{i,t}$ is the excess return on asset i, $r_{m,t}$ is the excess return on the market, $E_{t-1}(\cdot)=E(\cdot|Z_{t-1})$ is the conditional expectation operator with the L-vector of time $t-1$ observable conditioning variables Z_{t-1},

$$\beta_{i,t-1} = \frac{\mathrm{Cov}_{t-1(r_{i,t},r_{m,t})}}{\sigma^2_{m,t}} \tag{2}$$

is the conditional beta of asset i, $\mathrm{Cov}_{t-1}(r_{i,t}, r_{m,t})=E_{t-1}[(r_{i,t}-E_{t-1}(r_{i,t}))(r_{m,t}-E_{t-1}(r_{m,t}))]$ is asset i's conditional covariance with the market, and $\sigma^2_{m,t}=E_{t-1}[(r_{m,t}-E_{t-1}(r_{m,t}))^2]$ is the conditional variance of market returns.

A problem with the basic two-moment CAPM is its poor empirical performance (see Campbell et al., 1997; Campbell, 2000 for recent surveys of the literature documenting the empirical shortcomings of the CAPM). Early empirical tests were unconditional in nature (e.g., Fama and MacBeth, 1973; Black et al., 1972). There is also a large literature which conducts conditional tests of the CAPM using a range of alternative methodologies.[1] If the CAPM holds conditionally, there is no reason for surprise when it fails unconditional tests. Unfortunately the conditional CAPM performs poorly despite the extensive evidence that covariances, betas and the market variance are all time-varying.

It is well known that mean–variance-based asset pricing models such as the CAPM require either normally distributed returns, or quadratic utility when returns are non-normal. There is much evidence that returns are non-normal—they are both skewed and leptokurtic. Furthermore, quadratic utility is theoretically unattractive because it requires investors to exhibit increasing absolute risk aversion. In light of these two observations, the empirical failure of the CAPM may indicate that nonlinear asset pricing models, such as the three-moment CAPM of Kraus and Litzenberger (1976), are needed.

[1] Gibbons and Ferson (1985), Campbell (1987) and Harvey (1989, 1991) use an instrumental variable approach to modelling some moments, while other moments are left unmodelled. Bollerslev, Engle, and Wooldridge (1988), Bodurtha and Mark (1991) and Ng (1991) pursue GARCH-based tests. Schwert and Seguin (1990) and Ferson and Harvey (1991, 1993a,b, 1999) allow for time-varying betas using a range of different techniques which include rolling window estimation, modelling beta as a linear function of lagged conditioning variables, and the method of Davidian and Carrol (1987). Jagannathan and Wang (1996) pursue yet another approach by demonstrating that an unconditional version of a conditional CAPM will include variables that predict variation in the price of risk (such as the default spread) as priced "factors". Still another strand of the literature models time-varying beta using varying parameter regression techniques such as in Fabozzi and Francis (1978), and Bos and Newbold (1984).

Although there are some asset pricing models that account for moments higher than skewness,[2] we will focus on the three-moment CAPM. Investors who exhibit non-increasing absolute risk aversion, which is reasonable from a behavioral perspective, will exhibit a preference for positive skewness. Furthermore, preference for positive skewness is known to aggregate. Specifically, Kraus and Litzenberger (1983, Theorem 1) show that when all investors exhibit non-increasing absolute risk aversion then, in a Pareto efficient allocation, the representative agent will also exhibit non-increasing absolute risk aversion. However, preferences over higher moments like kurtosis may not aggregate. In particular, even if individual investors are averse to kurtosis, it does not necessarily follow that a representative investor will be averse to kurtosis.[3]

The conditional version of the three-moment CAPM of Kraus and Litzenberger (1976) is

$$E_{t-1}[r_{i,t}] = \beta_{i,t-1}\mu_{1,t} + \gamma_{i,t-1}\mu_{2,t}, \tag{3}$$

where

$$\gamma_{i,t-1} = \frac{\mathrm{Coskew}_{t-1}(r_{i,t}, r_{m,t})}{\kappa^3_{m,t}} \tag{4}$$

is the conditional gamma of asset i, $\mathrm{Coskew}_{t-1}(r_{i,t}, r_{m,t}) = E_{t-1}[(r_{i,t} - E_{t-1}(r_{i,t}))(r_{m,t} - E_{t-1}(r_{m,t}))^2]$ is asset i's conditional coskewness with the market, and $\kappa^3_{m,t} = E_{t-1}[(r_{m,t} - E_{t-1}(r_{m,t}))^3]$ is the conditional skewness of market returns, $\mu_{1,t}$ is the price of beta risk, and $\mu_{2,t}$ is the price of gamma risk. This type of model is consistent with a pricing kernel that is a quadratic function of the market return (see also Dittmar, 2002). Because investors will prefer positive skewness, the sign of $\mu_{2,t}$ should be opposite to the sign of the conditional skewness of market returns, $\kappa^3_{m,t}$, since in equilibrium investors are willing to sacrifice mean return for positive skewness, but demand a premium when returns are negatively skewed. Since $\beta_{m,t-1} = \gamma_{m,t-1} = 1$, the three-moment CAPM implies that the expected excess return on the market portfolio is

$$E_{t-1}(r_{m,t}) = \mu_{1,t} + \mu_{2,t}\,.^4 \tag{5}$$

[2] See, for example, Fang and Lai (1997), who develop a four-moment CAPM and present some empirical tests of their model; and Dittmar (2002) who tests an SDF model which is a cubic function of the return to a value weighted stock index and growth in labor income that is equivalent to a four-moment CAPM. Rubinstein (1973) develops a model which accounts for all comments (see also Leland, 1999 who discusses nonnormality on performance evaluation). Post, Levy, and van Vliet (2003) point out that much of the empirical work on the three-moment CAPM fails to impose risk aversion when estimating the model's parameters. The prices of risk can be linked with the primitive parameters of the representative investors cubic utility function. When the constraint that investors be risk averse over some suitable range of wealth levels is imposed (the cubic utility function is never globally risk averse) the significance of coskewness is diminished. Post, Levy, and van Vliet (2003) observe that the empirically estimated utility function implied by the unconstrained price of beta and gamma risk is an inverted "S" shape. It is interesting that this utility function implies that investors are risk averse with respect to losses but risk seeking with respect to gains. Note also that Poti (2005) demonstrates that even after imposing these constraints, coskewness is still useful for explaining the returns on hedge funds and momentum strategies.

[3] Scott and Horvath (1980) derive the direction of preferences for moments higher than the variance that is quite general, but it does not directly apply in a representative agent framework with Pareto efficient allocations.

[4] This has some very useful empirical implications. A puzzling result in empirical conditional asset pricing literature is the tendency of conditional expected return on the market to be negative over extended periods (see, e.g., He et al., 1996 and Harvey and Siddique, 2000b). This is inconsistent with investors being risk averse and caring only about mean and variance. However, when investors have a preference for positive skewness, if the market is sufficiently positively skewed, the negative price of skewness risk may dominate the positive price of variance risk, resulting in a negative market risk premium. This idea is explored by Harvey and Siddique (2000b).

There are a number of unconditional tests of the three-moment CAPM, including Kraus and Litzenberger (1976), Friend and Westerfield (1980) and Lim (1989). The results from unconditional tests are mixed. Kraus and Litzenberger (1976) and Lim (1989) find evidence supporting the three-moment CAPM, while Friend and Westerfield (1980) find evidence inconsistent with the three-moment CAPM.

The objection to performing unconditional tests of the two-moment CAPM, when the model holds conditionally, applies to unconditional tests of the three-moment CAPM.[5] This motivates a conditional analysis. Harvey and Siddique (1999) present an extensive analysis of the conditional effect of coskewness on asset prices. Also, Harvey and Siddique (2000b) test for time-variation in skewness; and Harvey and Siddique (2000a) test whether, in the context of a three-moment conditional CAPM, the market risk premium changes over time. Dittmar (2002) presents a test of the SDF model, where the discount factor is a cubic function of the return on the NYSE value weighted stock index, and labor growth following Jagannathan and Wang (1996). He finds that one needs to include cokurtosis with labor growth rates to produce an admissible pricing kernel.

3. An empirical specification

There are many econometric techniques available to test the conditional CAPM. Some tests are restrictive in the assumptions they place on the data generating process. For example, Bodurtha and Mark (1991) and Ng (1991) explicitly model covariances and prices of risk. Other techniques place fewer restrictions on the data. One approach is to directly model beta as a linear function of conditioning variables (see, e.g., Ferson and Harvey, 1999). Another approach, which we pursue in this paper, is to explicitly model the prices of risk. These two approaches are neither more general nor more restrictive than each other. We favor modelling prices of risk, rather than betas, for a number of reasons. Firstly, it allows the dynamics of covariances to follow arbitrary data generating processes. The underlying distributional theory only requires the moments to be stationary and says nothing about the shape of the conditional density beyond the modelled moments. Secondly, it is more parsimonious, dramatically reducing the dimensionality of the estimation problem.

The moment conditions implied by the three-moment CAPM can be written as follows

$$E\left[\left(r_{i,t}-r_{i,t}e_{m,t}\frac{\mu_{1,t}}{\sigma^2_{m,t}}-r_{i,t}(e^2_{m,t}-\sigma^2_{m,t})\frac{\mu_{2,t}}{\kappa^3_{m,t}}\right)\otimes Z_{t-1}\right]=0_{L\times 1},\quad \text{for } i=1,\dots,N, \tag{6}$$

by using a simplified expression for the conditional coskewness

$$\begin{aligned}\text{Coskew}_{t-1}(r_{i,t},r_{m,t}) &= E_{t-1}[(r_{i,t}-E_{t-1}r_{i,t})(r_{m,t}-E_{t-1}r_{m,t})^2]\\ &= E_{t-1}[r_{i,t}(r_{m,t}-E_{t-1}r_{m,t})^2]-E_{t-1}[r_{i,t}]\sigma^2_{m,t}\\ &= E_{t-1}[r_{i,t}\{(r_{m,t}-\mu_{m,t})^2-\sigma^2_{m,t}\}],\end{aligned}$$

and covariance

$$\text{Cov}_{t-1}(r_{i,t},r_{m,t}) = E_{t-1}[r_{i,t}(r_{m,t}-\mu_{m,t})],$$

[5] Ferson, Siegel, and Xu (in press) extend to multifactor models what Cochrane (2001) calls the Hansen–Richard critique. Even if the CAPM holds conditional on information set Ω_{t-1} (and the market portfolio is therefore efficient) it is entirely possible that tests conditioning on Z_{t-1} (a strict subset of Ω_{t-1}) may reject the conditional CAPM even when it holds exactly.

where $\mu_{m,t}=\mu_{1,t}+\mu_{2,t}$ is the expected return on the market. Following Whitelaw (1994) we model the conditional variance of the market return as

$$\sigma_{m,t}^2=(Z_{t-1}^{\top}\sigma_m)^2, \tag{7}$$

and we model the conditional skewness of the market return as[6]

$$\kappa_{m,t}^3=Z_{t-1}^{\top}\kappa_m, \tag{8}$$

and the price of beta risk as

$$\mu_{1,t}=Z_{t-1}^{\top}\mu_1, \tag{9}$$

where σ_m, κ_m, and μ_1 are L-dimensional parameter vectors. Investors who exhibit non-increasing absolute risk aversion will have a preference for positive skewness and an aversion to negative skewness. The price of gamma risk, $\mu_{2,t}$, will have the opposite sign of the conditional skewness of the market's return: when the market is negatively skewed, $\mu_{2,t}$ will be positive, and when the market is positively skewed, $\mu_{2,t}$ will be negative (see Kraus and Litzenberger, 1976). For this reason we model $\mu_{2,t}$ as

$$\mu_{2,t}=Z_{t-1}^{\top}\mu_2+\delta 1_{\{Z_{t-1}^{\top}\kappa_m>0\}}, \tag{10}$$

where $1_{\{Z_{t-1}^{\top}\kappa_m>0\}}$ is an indicator variable taking the value one when the conditional skewness of market returns is positive, and zero otherwise. Because we include a constant, the effect of the indicator variable δ is to allow the intercept to change sign depending on the sign of the conditional skewness of market returns. Imposing Eq. (5) allows us to write the conditional mean return on the market portfolio as $\mu_{m,t}=\mu_{1,t}+\mu_{2,t}$, or

$$\mu_{m,t}=Z_{t-1}^{\top}\mu_1+Z_{t-1}^{\top}\mu_2+\delta 1_{\{Z_{t-1}^{\top}\kappa_m>0\}}.$$

We can combine the moment restrictions on the individual assets with moments to identify the first three-moments of the market return to give the error vector[7]

$$u_t=\begin{pmatrix} r_{m,t}-Z_{t-1}^{\top}\mu_1-Z_{t-1}^{\top}\mu_2-\delta 1_{\{Z_{t-1}^{\top}\kappa_m>0\}} \\ e_{m,t}^2-(Z_{t-1}^{\top}\sigma_m)^2 \\ e_{m,t}^3-Z_{t-1}^{\top}\kappa_m \\ r_t-r_t e_{m,t}\dfrac{Z_{t-1}^{\top}\mu_1}{(Z_{t-1}^{\top}\sigma_m)^2}-r_t(e_{m,t}^2-(Z_{t-1}^{\top}\sigma_m)^2)\dfrac{Z_{t-1}^{\top}\mu_2+\delta 1_{\{Z_{t-1}^{\top}\kappa_m>0\}}}{Z_{t-1}^{\top}\kappa_m} \end{pmatrix}. \tag{11}$$

[6] An alternative to our choice of the linear specification is the autoregressive conditional skewness model of Harvey and Siddique (1999).

[7] Note that the moment conditions require dividing the appropriate price of risk by the conditional variance and skewness. Numerical problems will arise when either the conditional variance or skewness is very close to zero. In the three-moment CAPM when the market returns are symmetric (i.e., $\kappa_{m,t}^3=0$) the price of coskewness risk $\mu_{2,t}$ will equal zero (Kraus and Litzenberger, 1976, footnote 7, p.1088). To ensure numerical stability we set $\mu_{2,t}=0$ when $|\kappa_{m,t}|<0.1$, and set $\sigma_{m,t}^2=\max(0.5, (Z_t^{\top}\sigma_m)^2)$. To gain some perspective on our lower bound on the variance of returns note that the unconditional variance of returns is 17.5 and the minimum monthly realized volatility calculated following French, Schwert, and Stambaugh (1987) using daily returns is 0.6703. Further, the parameter estimates for the volatility model reported in Table 1 are calculated without imposing the lower bound and the minimum fitted conditional variance is 2.97. Neither constraint is binding in any of the models reported below and is employed simply to avoid numerical problems during the parameter minimization stage. Finally, we experimented with multiplying the relevant moment conditions by the product of the variance and skewness but found quite disappointing results. In particular, the standard errors on the prices of risk parameters are on average twice as large as the standard errors reported in Table 3.

This is used to form the $(N+3)L$ moment restrictions as $E(u_t \otimes Z_{t-1}) = 0_{(N+3)L\times 1}$, where $\otimes$ denotes the Kronecker product. The model has $4L+1$ parameters which we estimate using Hansen's (1982) Generalized Method of Moments (GMM) which leaves NL - 1 overidentifying moment conditions which we use to construct Hansen's (1982) J_T specification test.

The two-moment CAPM is a special case which we can test using the error vector

$$u_t = \begin{pmatrix} r_{m,t} - Z_{t-1}^\top \mu_m \\ e_{m,t}^2 - (Z_{t-1}^\top \sigma_m)^2 \\ r_t - r_t(r_{m,t} - Z_{t-1}^\top \mu_m) \dfrac{Z_{t-1}^\top \mu_m}{(Z_{t-1}^\top \sigma_m)^2} \end{pmatrix} \tag{12}$$

where the expected return on the market is modelled as

$$\mu_{m,t} = Z_{t-1}^\top \mu_m. \tag{13}$$

There are $2L$ parameters which are estimated by GMM which leaves NL overidentifying moment conditions which are used to construct Hansen's (1982) J_T model specification test.

4. Data

Our empirical analysis is conducted using two alternative portfolio strategies, but to conserve space we only report one set of results. We analyze portfolios rather than individual assets because we need to invert the moment covariance matrix and the number of observations places a limit on the number of assets that can be used. We report our results for 17 value—weighted industry portfolios whose returns were obtained from Ken French's Website[8]. We also analyze, but do not report, the results using the 25 portfolios (FF25) formed by sorting stocks on their market capitalization and book-equity to market-equity ratio. It has become common to use portfolios formed by sorting stocks on characteristics (see, e.g., Fama and French, 1992, 1993, 1996; Lakonishok et al., 1994), and the FF25 are by far the most widely analyzed. We choose to focus our discussion on the industry data because of the debate surrounding characteristic-based portfolios. In particular, Lo and MacKinlay (1990) and Conrad, Cooper, and Kaul (2003) argue that many of the interesting empirical regularities observed in the FF25 portfolios may be overstated due to data snooping. It is reassuring that our results are quite similar in both datasets, and our conclusions are robust to the choice of portfolio formation strategy.

We use a set of six conditioning information variables that are fairly standard in the literature, in fact they are directly taken from He et al. (1996) and are similar to those used in many other studies. We include a constant; the *S&P*500 index return (*S&P*); the dividend price ratio on the *S&P*500 index, defined as the cumulative 12-month dividends divided by the current price level (Div); the term spread, measured by the difference in yields on a three-month Treasury bill and a one-month Treasury bill rate (Term); the default spread, defined as the difference between the yield on Baa rated bonds and Aaa rated bonds (Default); and finally the one-month Treasury bill rate (Tbill). To ensure that inference is conducted using only publicly available information, we lag the information variables one-month. The time subscript t is used for factor and portfolio returns, e.g. r_t, and the time subscript $t-1$ is used for the information variables Z_{t-1}. Our data covers the period

[8] The data and a discussion the SIC codes used to form them is available online at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

July, 1963 to December, 1997. To ease interpretation of the parameters, we transform them to correspond to the loadings on instruments with zero mean and unit variance. Note, however, that the moment conditions are defined on the raw data to ensure that the standard errors are appropriate.

We avoid modelling covariances following Harvey (1991) by including the term $r_{i,t}e_{m,t}$ in place of the conditional covariances when defining the moment conditions. This is contrasted with the approach of Harvey (1989) who uses the more traditional definition of covariance which suggests using $e_{i,t}e_{m,t}$ when defining the moment conditions. The advantage of our approach is that it requires 102 fewer parameters to be estimated and thus 102 fewer moment conditions to impose! This dramatically increases the feasible number of portfolios and conditioning variables that can tested. To determine if our results depend on this alternative definition we calculate the correlation between $r_{i,t}e_{m,t}$ and $e_{i,t}e_{m,t}$ (where $e_{i,t}$ and $e_{m,t}$ are the residuals from an OLS regression of respectively $r_{i,t}$ and $r_{m,t}$ on Z_{t-1}). These correlations range from 98.0% to 99.3% across all portfolios. The correlation is similar between the two versions of coskewness.

5. Empirical results

We report the parameter estimates for the instrumental variables models for the market mean, variance and skewness in Table 1. In the first row of panel B we test whether each moment is constant and are able to reject this null hypothesis for all three-moments. The plots of the conditional moments in Fig. 1 exhibit substantial time-variation. The time-variation in each moment is both statistically significant and economically meaningful as will be demonstrated below.

In Appendix A we compare the linear volatility model with the widely applied GARCH(1,1) conditional volatility model and find that the models are remarkably similar in their overall performance. To further assess the linear specification we have chosen we test this against two more general models.[9] The first generalization is to a quadratic function of the instruments. We find no evidence of nonlinearity in the conditional mean or conditional skewness but there is some evidence of nonlinearity in the conditional variance. The second model tests a specification similar to Bodurtha and Mark (1991) including the lagged squared market innovation to the conditional variance model, and the lagged cubed market innovation to the conditional skewness model. The results from these models are virtually identical to those reported below.[10]

Table 2 reports the parameter estimates for the two-moment CAPM when the price of covariance risk and the variance of market returns are explicitly modelled. The parameters are estimated using Hansen's (1982) GMM. Unless stated to the contrary, we use the efficient weighting matrix and follow Ferson and Foerster (1994) and update the estimate of the efficient weighting matrix iteratively using consistent estimates of the parameter vector.

The positive constant in the conditional mean μ indicates that the unconditional price of covariance risk is positive, as expected, and is almost two standard errors from zero (the p-value of the Wald test is

[9] We thank Wayne Ferson and an anonymous referee for suggesting these tests.

[10] We undertake a small Monte Carlo experiment to assess the consequences of ignoring this nonlinearity. In particular we compare the frequency of rejections using the J_T test on the two-moment CAPM with linear mean and standard deviation when the moments are linear and when the standard deviation is a quadratic function of the information variables. Consistent with previous research (Ferson and Foerster, 1994) we find that the conditional models tend to overreject in moderate sample sizes. The correctly specified linear moment model is rejected in 11% of samples. When we approximate the quadratic standard deviation using a linear specification we reject only slightly more frequently in 13% of samples.

Table 1
Parameter estimates of the linear moment model

Par.	Constant	S&P	Div	Term	Default	Tbill
μ_m	0.5034	−0.0157	0.5637	−0.7965	0.8024	−1.3738
	(0.2100)	(0.2985)	(0.3969)	(0.2884)	(0.3569)	(0.3498)
σ_m	4.0734	−0.7976	−0.2548	0.0794	0.3742	0.4390
	(0.1879)	(0.2711)	(0.5294)	(0.2930)	(0.3963)	(0.2751)
κ_m	−31.2706	23.7165	90.3285	−67.6148	−22.3929	−58.2384
	(25.8381)	(43.7364)	(74.7492)	(39.8327)	(53.1115)	(25.6704)

Panel B: specification tests

	$\mu_{m,t}$	$\sigma_{m,t}$	$\kappa_{m,t}$
Constant	21.2118	33.9373	12.6751
	(0.0007)	(0.0000)	(0.0266)
Nonlinearity	6.4834	16.1172	2.6319
	(0.2620)	(0.0065)	(0.7565)
ARCH(1)		0.0165	0.0206
		(0.0559)	(0.0470)

GMM parameter estimates and standard errors (in parenthesis) are reported for a model of excess returns on the value weighted market portfolio in which the conditional mean, standard deviation and skewness are linear functions of instrumental variables. The data covers the period July 1962 to December 1997. The moment conditions are:

$$E[(r_{m,t}-\mu_{m,t}) \otimes Z_{t-1}] = 0_{L\times 1}$$

$$E[\{(r_{m,t}-\mu_{m,t})^2-\sigma^2_{m,t}\} \otimes Z_{t-1}] = 0_{L\times 1}$$

$$E[\{(r_{m,t}-\mu_{m,t})^3-\kappa^3_{m,t}\} \otimes Z_{t-1}] = 0_{NL\times 1}$$

and $\mu_{m,t}=Z^{\top}_{t-1}\mu_m, \sigma_{m,t}=Z^{\top}_{t-1}\sigma_m,$ *and* $\kappa^3_{m,t}=Z^{\top}_{t-1}\kappa_m$.
In panel B we report a Wald test (with p-value below) for the null hypothesis that each moment is constant; a Wald test for linearity, i.e. testing for a nonzero quadratic term in each conditional moment (with p-value below), and report an estimated ARCH(1) and autoregressive conditional skewness term with standard error below.

9%). It is interesting that the sign of most coefficients in the mean equation are opposite to the sign of the equivalent parameter in the standard deviation equation (the exception is the coefficient on *S&P* which is positive for the mean, but insignificant in the variance equation). This suggests a negative relationship between conditional mean and variance in market returns, since a change in any conditioning variable moves the mean and standard deviation in opposite directions. This result is consistent with some of the existing empirical literature, e.g. Whitelaw (1994).[11] The positive coefficients on Div and Default and the negative coefficients on Term and Tbill suggest that expected returns are pro-cyclical; and the point estimates on the interest rate variables (Term, Default and Tbill) suggest that conditional volatility is counter-cyclical (again consistent with Whitelaw, 1994).

We are interested in testing several hypotheses about the model specification. The J_T test is a general test of goodness of fit or specification. It is calculated as $J_T = Tg^{\top}_T S^{-1}_T g_T$ where S_T is the

[11] The evidence on the risk-return relation in equity returns is mixed. Many papers find a negative relation whereas some other recent papers find a positive relation. Some recent papers finding a positive relation include Ghysels, Santa-Clara, and Valkanov (2005), Guo and Whitelaw (2006), Pastor, Sinha, and Swaminathan (2005), Scruggs (1998), and Smith (2005).

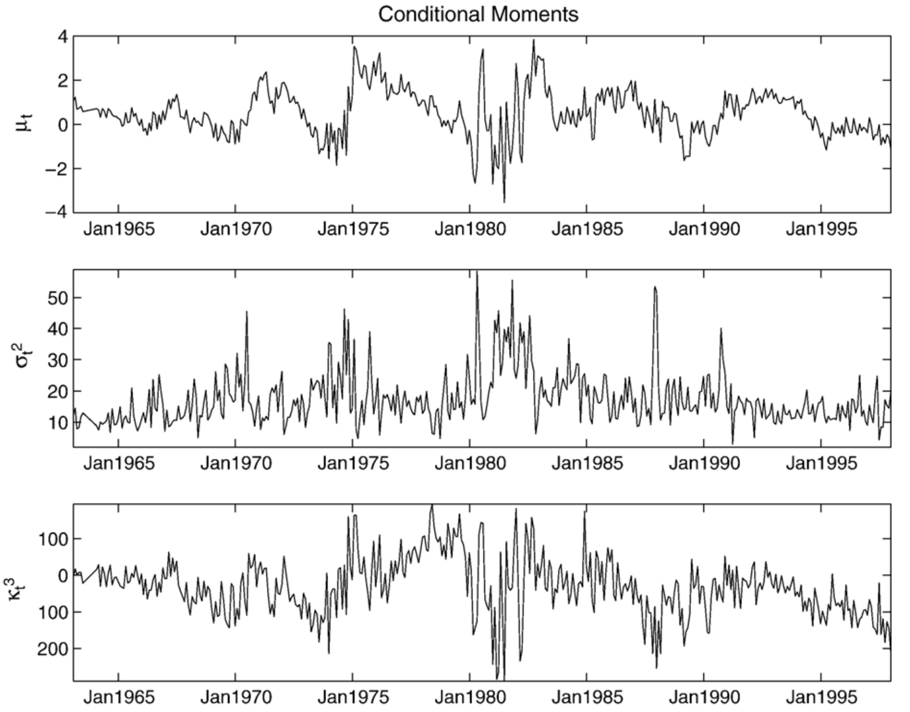

Fig. 1. Conditional moments. The fitted conditional mean, variance and skewness are plotted. The parameter estimates are from Table 1.

sample variance–covariance matrix of the moments and g_T are the pricing errors. A correctly specified model delivers pricing errors close to zero, yielding a small J_T statistic. We can conclusively reject the conditional two-moment CAPM as the J_T test is significant at better than the 0.1% level.

Secondly, is covariance risk priced (H_0: $\mu = 0_{L\times 1}$)? We reject this hypothesis with a Wald test, and conclude that covariance risk is indeed priced. It is also clear that the expected market returns, the conditional variance, and the price of covariance risk are all time-varying, i.e., these hypotheses are all rejected at better than 0.1%. In particular, when the conditional covariance is modelled as a linear function of the conditioning variables, and a Wald test for the hypothesis of zero coefficients on the five conditioning variables is examined, we can reject the joint null hypothesis that all industries have constant covariances at a confidence level of 0.1% (using the Hochberg (1988) lower bound on the p-value).[12] We also test if the price of covariance risk is constant using a Wald test, and as reported in Table 2, we clearly reject this hypothesis. Both components of the price of risk, the conditional mean and the conditional variance, appear to be time-varying.

[12] Let P_i be the p-values testing m hypothesis. The composite hypothesis that all m hypothesis hold jointly is rejected at significance level α when at least one of the tests satisfies the well known Bonferonni bound $P_i \leq \alpha = m$. The individual significance levels are adjusted recognizing that we are conducting multiple tests. The Hochberg (1988) bound is a slightly tighter bound $P_{(i)} \leq \alpha = (m - i + 1)$ where $P_{(i)}$ is the i-th largest individual p-value.

Table 2
Parameter estimates of the two-moment model GMM parameter estimates and standard errors (in parenthesis) are reported for the test of the conditional two-moment CAPM with linear means and variances

Parameter	Constant	S&P	Div	Term	Default	Tbill
μ_m	0.2629	0.2702	0.8208	−0.7544	0.4597	−1.0689
	(0.1563)	(0.0862)	(0.1405)	(0.1350)	(0.1523)	(0.1935)
σ_m	3.6677	0.4375	−0.6155	0.9054	−0.0472	1.4947
	(0.1245)	(0.1285)	(0.1974)	(0.1336)	(0.1876)	(0.1669)
J_T-statistic	154.0911	0.0007	102 *d.f.*			
Constant $\mu_{m,t}$	66.8048	0.0000	5 *d.f.*			
Constant $\sigma^2_{m,t}$	142.6327	0.0000	5 *d.f.*			
Constant $\lambda_{m,t}$	379.5360	0.0000	10 *d.f.*			
Zero $\lambda_{m,t}$	117.6614	0.0000	6 *d.f.*			

The 17 industry portfolio returns cover the period July 1962 to December 1997. The moment restrictions implied by the model are:

$$E[(r_{m,t}-\mu_{m,t})\otimes Z_{t-1}]=0_{L\times 1}$$

$$E[\{(r_{m,t}-\mu_{m,t})^2-\sigma^2_{m,t}\}\otimes Z_{t-1}]=0_{L\times 1}$$

$$E\left[\left(r_t-r_t(r_{m,t}-\mu_{m,t})\frac{\mu_{m,t}}{\sigma^2_{m,t}}\right)\otimes Z_{t-1}\right]=0_{NL\times 1}$$

and $\mu_{m,t}=Z^{\top}_{t-1}\mu_m$ and $\sigma_{m,t}=Z^{\top}_{t-1}\sigma_m$. The parameters are estimated by GMM. Hansen's (1982) J_T-statistic and a number of Wald tests are also reported along with their p-values and degrees of freedom.

The conditional mean and variance parameters are exactly identified using only the moments given by the first two elements in u_t in Eq. (12). The standard errors reported in Table 1 are estimated using only market returns and are much higher than the corresponding standard errors reported in Table 2 which use both the time-series returns on the market and individual portfolio returns. The differences between these two sets of standard errors indicate significant efficiency gains from using both time-series and cross-sectional data to estimate time-series parameters.[13]

We next test the three-moment CAPM. The parameter estimates are reported in Table 3. The three-moment CAPM provides a much better fit to the data than the two-moment CAPM. The p-value for the J_T test of the three-moment CAPM is around 5%. We are unable to reject the three-moment CAPM whereas Dittmar (2002) rejects a similar model. In Appendix B we explore the reasons for the different conclusions. We also find strong evidence that both covariance and coskewness risk are priced, and the null hypotheses that the prices of both covariance and coskewness risk are constant is easily rejected. We also find strong evidence that both the markets' variance and skewness are time-varying. We also test whether the conditioning variables Z_{t-1} can explain variation in coskewness. There is some evidence that coskewness is time-varying. We can reject the null hypothesis of constant coskewness in fourteen of the industry portfolios at the 10% significance level, and in seven portfolios at the 5% significance level. The joint hypothesis of constant coskewness has a p-value of around 7%.

[13] The differences in the standard errors may also be due to different finite sample biases in the standard errors (Ferson and Foerster, 1994).

Table 3
Parameter estimates of the full three-moment model

Parameter	Constant	*S&P*	Div	Term	Default	Tbill
μ_1	0.5867	−0.2041	−0.2316	−0.6704	1.2110	−1.5817
	(0.1865)	(0.1957)	(0.2395)	(0.2248)	(0.2449)	(0.2652)
μ_2	0.1500	0.3504	0.5720	−0.4522	−0.3451	0.1424
	(0.0920)	(0.1129)	(0.2022)	(0.1459)	(0.0910)	(0.1483)
σ_m	4.0817	−0.5857	0.2067	0.3596	0.0635	0.5871
	(0.1043)	(0.1017)	(0.1732)	(0.1133)	(0.1554)	(0.1506)
κ_m	−29.3938	22.5929	91.2938	−66.0881	−21.3923	−58.9168
	(0.6757)	(0.4841)	(1.3495)	(1.7220)	(0.6760)	(1.5098)
δ	−0.8059					
	(0.1598)					
J-statistic	118.0552	0.0547	95 *d.f.*			
Zero $\lambda_{1,t}$	67.7703	0.0000	6 *d.f.*			
Zero $\lambda_{2,t}$	34.8643	0.0000	7 *d.f.*			
Constant $\mu_{1,t}$	49.4404	0.0000	5 *d.f.*			
Constant $\lambda_{2,t}$	34.7874	0.0000	6 *d.f.*			
Constant $\sigma^2_{m,t}$	1661.4379	0.0000	5 *d.f.*			
Constant $\kappa^3_{m,t}$	27691.1070	0.0000	5 *d.f.*			

Parameter estimates and standard errors (in parenthesis) are reported for the test of the conditional three-moment CAPM with linear prices of risk, standard deviation and skewness. The 17 industry portfolio returns cover the period July 1962 to December 1997. The moment restrictions implied by the model are:

$$E[(r_{m,t}-\mu_{1,t}-\mu_{2,t})\otimes Z_{t-1}]=0_{L\times 1}$$

$$E[(e^2-\sigma^2_{m,t})\otimes Z_{t-1}]=0_{L\times 1}$$

$$E[(e^3_t-\kappa^3_{m,t})\otimes Z_{t-1}]=0_{L\times 1}$$

$$E\left[\left(r_t-r_t e_{m,t}\frac{\mu_{1,t}}{\sigma^2_{m,t}}-r_t(e^2_{m,t}-\sigma^2_{m,t})\frac{\mu_{2,t}}{\kappa^3_{m,t}}\right)\otimes Z_{t-1}\right]=0_{NL\times 1}$$

where $e_t=r_{m,t}-\mu_{1,t}-\mu_{2,t}$, $\mu_{1,t}=Z^\top_{t-1}\mu_1$, $\mu_{2,t}=Z^\top_{t-1}\mu_2+\delta 1_{\{Z^\top_{t-1}\kappa_m>0\}}$, $\sigma^2_{m,t}=(Z^\top_{t-1}\sigma_m)^2$, and $\kappa^3_{m,t}=Z^\top_{t-1}\kappa_m$.

Theory predicts that the price of beta risk ($\mu_{1,t}$) is positive, and that the sign of the price of gamma risk ($\mu_{2,t}$) is opposite to the sign of the conditional skewness of market returns. The constant in $\mu_{2,t}$ is the unconditional price of gamma risk when the market is negatively skewed. It is positive but not statistically significant. The parameter δ captures the difference in the unconditional price of gamma risk when the market is positively skewed and, as theory predicts, it is negative and statistically significant. The point estimates tell an interesting economic story. Investors appear to be relatively ambivalent towards coskewness when the market is negatively skewed, only demanding on average a premium for bearing gamma risk of 1.80% annually. However, they exhibit strong preferences for positive skewness as they are willing, on average, to sacrifice 7.87% annually for bearing gamma risk when the market is positively skewed.

An alternative measure of coskewness risk is the coskewness factor constructed by Harvey and Siddique (1999). This coskewness factor, *SKS*, is a hedge portfolio which is long stocks that have a negative coskewness with the market and short stocks with a positive coskewness with the market.[14] The expected return to this hedge portfolio should be positive regardless of the sign of market skewness. To see this, note that when the market is positively skewed assets with negative coskewness serve to reduce the skewness of the market portfolio. Since investors like skewness, in equilibrium these assets must earn a higher return than skewness increasing assets, and *SKS* will have a positive expected return. However, when the market portfolio is negatively skewed, assets with negative coskewness serve to increase this negative skewness, and to entice investors to hold these assets in equilibrium they earn higher returns than positive coskewness assets. If attitudes towards coskewness vary with the magnitude of market skewness then the expected return on *SKS*, although positive, will vary too.

Over the period July 1963 to December 1993 the average return on this coskewness factor was 3.60%.[15] However, in months when the conditional skewness of the market is positive, the average annualized return on *SKS* is 5.00%, while in months where the market is negatively skewed the average premium is only 2.81%. This shows that investors are more interested in positive skewness than they are averse to negative skewness.

Our specification for the price of gamma risk only allows the intercept to change signs with the sign of conditional skewness. We also consider a more general specification in which all the parameters of μ_2 change with the sign of conditional skewness:

$$\mu_{2,t} = Z_{t-1}^{\top}(\mu_2 + 1_{\{Z_{t-1}^{\top}\kappa_m > 0\}}\delta) \tag{14}$$

where δ is now an L-column vector. This nests our chosen specification as a special case. We test the hypothesis that the coefficients on the five instrumental variables equal zero using a Lagrange Multiplier test. The test statistic is 3.4176 which has five degrees of freedom and hence a p-value of only 0.63. Estimating the more general model, the negative intercept is the only statistically significant parameter. The five slope coefficients are typically less than half a standard error from zero.

We find that allowing gamma risk to be priced increases the unconditional price of beta risk. In the two-moment model the evidence for a positive unconditional price of beta risk was marginal (a t ratio of 1.68), yet in the context of the three-moment CAPM the point estimate is much more significant both economically (the point estimate more than doubles from 3% annually to 7% annually) and statistically (being more than three standard errors from zero). The smaller estimate in the two-moment model is an evidence of an omitted variable bias that arises because it ignores coskewness. The price of covariance risk is determined by two sets of moments: the expected return on the market portfolio and the cross-sectional moment restrictions on the individual portfolios. In the two-moment CAPM the price of covariance risk equals the expected return on the market, i.e. $E(r_{m,t})=\mu_{m,t}$, while in the three-moment CAPM the expected market return depends on the price of both covariance and coskewness risk, i.e. $E(r_{m,t})=\mu_{1,t}+\mu_{2,t}$. This suggests that two two-moment estimate is biased when the three-moment CAPM holds: $\mu_{m,t}-\mu_{1,t}=\mu_{2,t}$. Because the price of coskewness risk $\mu_{2,t}$ is on average negative, $\mu_{m,t}$ is too low. Including the cross-sectional moment restrictions implied by the two-moment CAPM in a three-moment world also produces an estimator which is asymptotically biased downward. The omitted variables bias in this context is derived in Appendix C.

[14] The standardized coskewness is computed using 60 months of market model residuals and the value-weighted portfolios of the top and bottom 30% of stocks.

[15] We thank Akhtar Siddique for kindly supplying this data.

5.1. Structural break tests

Although there is ample evidence that asset pricing models need to be conditional since expected returns vary through time, Ghysels (1998) has demonstrated that even conditional models fail structural break tests. This suggests that many conditional asset pricing models are misspecified and incorrectly use conditioning information. The J_T and H_T tests are useful for detecting general model misspecification, but they have very low power to detect some specific forms of model misspecification, and structural breaks in particular (Ghysels, 1998). We therefore test for structural breaks using the Lagrange Multiplier (*LM*)-based tests of Andrews (1993) and Andrews and Ploberger (1994) which do not presuppose knowledge of the break date.[16] We report these three structural break tests for a range of conditional asset pricing models in Table 4.[17] We find evidence of a structural break in the two-moment CAPM as did Ghysels (1998). Interestingly the evidence of a structural break in the three-moment CAPM is much weaker, with only the exp*LM* test being significant.[18]

5.2. Robustness checks

The J_T statistic favors models that produce volatile pricing errors, because small values of $J_T = T g_T^\top S_T^{-1} g_T$ can arise either by small pricing errors (g_T), or by volatile pricing errors ("large" S_T). When the variance–covariance matrix of moment conditions is large, then there will also be large standard errors for the parameters. A small J_T statistic is likely to be spurious when accompanied by insignificant parameter estimates. We find that both the J_T statistic and the parameter estimates are significant, which somewhat alleviates these concerns. Nevertheless, we present a few experiments with different weighting matrices.

We re-estimate the models using two types of prespecified model independent weighting matrices—the identity matrix, which weights all moments as equally important; and a matrix W which places more weight on those moments which are measured more precisely, and takes account of the information contained in covariance between moments. It is important that W be independent of the models being compared to ensure that differences between the values of the objective function $g_T^\top W g_T$ are due to model fit and not differences in the weighting matrix. We form W by taking the dependent variable in each of the moment conditions, for example $r_{m,t}$ in the moments identifying the mean and $e_{m,t}^2$ for the moments identifying the conditional standard deviation, and stacking them in a vector y_t. With the exception of the market's variance, the fitted values of each element of y_t are estimated using a linear regression. Let $v_t = y_t - \hat{y}_t(Z_{t-1})$ be the residual from this regression, then the weight matrix W is defined as

$$W = \left(1/T \sum_{t=1}^{T} v_t v_t^\top\right)^{-1}.$$

[16] The traditional *LM* test is inappropriate when the break date is unknown because the break date is an unidentified nuisance parameter, complicating inference. Andrews (1993) develops a sup-LM test and derives its limiting asymptotic distribution and tabulates critical values of the test statistic for a range of significance levels. Andrews and Ploberger (1994) present the ave*LM* and the exp*LM* statistics which are more powerful than the sup*LM* test and tabulate critical values.

[17] The test requires that we omit a fraction of the initial and final observations for the test to be well behaved, and we arbitrarily set this to be ten percent.

[18] Note that Smith (2004) finds in a Monte Carlo study that the exp*LM* test tends to overreject the null of no structural break in GARCH models in finite samples.

Table 4
Structural break tests

Model	supLM	aveLM	expLM
Two-moment	39.9316	19.4696	21.9283
	(0.0031)	(0.0246)	(0.0000)
Three-moment	39.7623	24.6256	27.2966
	(0.3556)	(0.5085)	(0.0014)
Linear price of risk: two-moment	24.3232	13.5382	9.3774
	(0.2428)	(0.2856)	(0.1921)
Three-moment	46.5777	32.4677	28.4847
	(0.7038)	(0.7242)	(0.0416)
Three-factor	60.6774	37.8547	39.1963
	(0.0017)	(0.0048)	(0.0000)
Three-moment and three-factor	68.9867	58.9514	57.7208
	(0.3146)	(0.0575)	(0.0000)

Andrews (1993) and Andrews and Ploberger (1994) supLM, expLM and aveLM tests for structural breaks are reported for each model estimated (setting $\pi=0.15$). The naming convention refers to two- and three-moment CAPM when modelling market moments, and con. (for constrained) three-moment CAPM when using the sign corrected price of covariance and three-moment-three-factor includes the FF factors and coskewness with explicitly modelled prices of risk. Asymptotic p-values are reported in parenthesis below each test statistic and are calculated numerically following Andrews (1993, p.841) and Andrews and Ploberger (1994, p.1398).

If the conditional mean on each asset is linear in Z_{t-1} (see Harvey, 1989 for a discussion of one situation in which this will be true) and standard regularity conditions are met then W will be a consistent estimator of the efficient weighting matrix.

The parameter estimates of the two- and three-moment CAPMs are reported in Table 5 for both W (panel A) and the identity matrix (panel B). The economic story told by the point estimates is the same as for the efficient weights: the price of covariance risk is positive in both models. Investors demand a small and insignificant positive premium for coskewness when returns are negatively skewed, but a large negative premium when returns are positively skewed. These conclusions are robust to different portfolio formation strategies.

The minimized value of the quadratic form, J_T, along with Zhou's (1994) H_T test statistic and its p-value, are reported in Table 5 below each set of parameter estimates. In each case including conditional coskewness adds much to the explanatory power of the empirical model. We cannot reject the model at the 10% level when the covariance matrix-based weights are used. When the identity matrix is used the p-value is 1.9%. The standard errors for the two fixed weight matrices are much larger for both the two- and three-moment CAPM than when the efficient weighting matrix is used. This is most pronounced for the three-moment CAPM where the standard errors are several times larger. This demonstrates that there are significant efficiency gains from using the efficient weighting matrix.

5.3. Price of risk dynamics

We plot the fitted values of both $\mu_{1,t}$ and $\mu_{2,t}$ from the three-moment CAPM in Fig. 2. As a reference we also plot $\mu_{m,t}$ from the price of two-moment CAPM. The price of beta risk, $\mu_{m,t}$ in the two-moment CAPM and $\mu_{1,t}$ in the three-moment CAPM, should be positive. There is only a relatively short period in the mid 1970s and the period of the Federal Reserve monetary experiment of 1979 to 1982 when both models have a negative conditional price of beta risk. However, there are a number of periods (in the 1960s and 1990s) when the two-moment CAPM

Table 5
Two- and three-moment CAPM parameter estimates: fixed weights

Parameter	Constant	*S&P*	Div	Term	Default	Tbill
Panel A: model independent covariance-based weight						
Two-moment						
μ_m	0.1057	0.3210	−0.1208	−0.4844	0.7863	−0.6149
	(0.2697)	(0.2856)	(0.2373)	(0.3012)	(0.3590)	(0.2665)
σ_m	4.2490	−0.4832	0.5958	0.2692	−0.0363	0.3111
	(0.1957)	(0.2157)	(0.3582)	(0.2243)	(0.3067)	(0.2600)
H_T	0.7091	[147.8009]	(0.0021)	102 *d.f.*		
Three-moment						
μ_1	0.6568	−0.1432	−0.5342	−0.5129	1.2276	−1.1856
	(0.2690)	(0.4113)	(1.0830)	(0.5710)	(0.5985)	(0.5210)
μ_2	0.2201	0.3199	1.0530	−0.6511	−0.5949	0.0550
	(0.1783)	(0.2203)	(0.6830)	(0.4548)	(0.3316)	(0.3438)
σ_m	4.0998	−0.5861	0.3765	0.3886	0.0171	0.5494
	(0.2776)	(0.3670)	(0.2915)	(0.2244)	(0.3602)	(0.3784)
κ_m	−30.2808	20.9311	91.2506	−67.4353	−22.4491	−59.4996
	(17.1182)	(11.0478)	(51.3525)	(39.7508)	(12.5900)	(34.6309)
δ	−1.0943					
	(0.5659)					
H_T	0.4864	[110.8729]	(0.1270)	95 *d.f.*		
Panel B: identity matrix						
Two-moment						
μ_m	0.4454	0.1913	0.7855	−0.5757	0.0412	−0.6763
	(0.2397)	(0.2571)	(0.5250)	(0.2905)	(0.3899)	(0.9045)
σ_m	4.1150	−0.6710	−0.0530	0.0863	0.1304	0.5425
	(0.2019)	(0.2929)	(0.4360)	(0.2713)	(0.4240)	(0.3215)
H_T	11.1722	[142.9680]	(0.0047)	102 *d.f.*		
Three-moment						
μ_1	0.5004	−0.1488	0.1519	−0.5456	0.8410	−1.3233
	(0.6059)	(1.0200)	(2.4210)	(1.2432)	(1.1191)	(1.4897)
μ_2	0.3842	0.1946	0.3712	−0.1606	−0.0572	0.0905
	(0.3609)	(0.4637)	(1.6740)	(0.9630)	(0.4876)	(1.2818)
σ_m	4.0189	−0.7820	−0.1543	0.1354	0.3092	0.6077
	(0.1619)	(0.2268)	(0.5424)	(0.4337)	(0.3983)	(0.7625)
κ_m	−38.1711	21.6285	116.0047	−89.6698	−32.7180	−75.1896
	(0.1777)	(1.0332)	(1.9107)	(4.0372)	(0.1662)	(1.9069)
δ	−0.7742					
	(0.5768)					
H_T	4.5240	[125.6508]	(0.0193)	95 *d.f.*		

This table presents the GMM parameter estimates and standard errors (in parenthesis) from estimating the two- and three-moment CAPM models that rely on explicit modelling of market moments. Panel A uses the inverse of a linear projection-based covariance matrix (W^{-1}), while panel B uses an identity weighting matrix. Below each set of parameter estimates we report the minimized value of the quadratic objective, the H_T statistic [in square brackets] and its p-value (in parenthesis). The 17 industry portfolio returns cover the period July 1962 to December 1997.

has a negative price of beta risk while in the three-moment CAPM it is positive. Interestingly, these periods were during the two longest post-war expansions that the US has enjoyed. Dividend yields were low during both recessions which produced low expected returns since the dividend yield is positively related to expected returns in both the two-moment CAPM, where $E_{t-1}(r_{m,t})=\mu_{m,t}$, and the three-moment CAPM, where $E_{t-1}(r_{m,t})=\mu_{1,t}+\mu_{2,t}$. However, in the three-moment

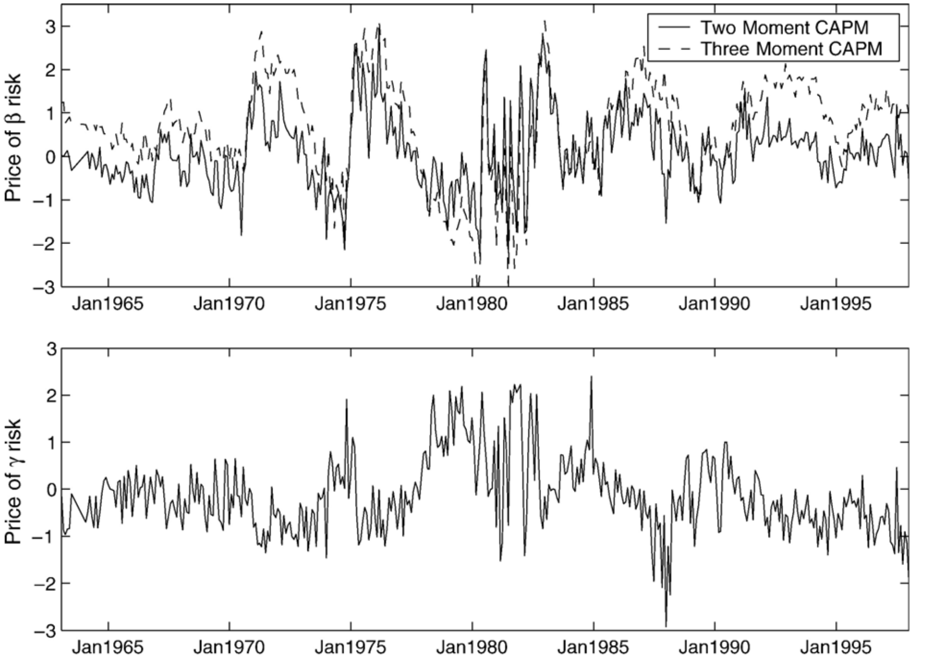

Fig. 2. Plot of price of beta and gamma risk. The upper plot is the fitted price of beta risk for both the two-($\mu_{m,t}$) and three-moment ($\mu_{1,t}$) CAPMs. The lower plot contains the fitted value for the price of gamma risk ($\mu_{2,t}$) from the three-moment CAPM. The parameter estimates use the model independent covariance matrix-based weights.

CAPM the dividend yield is negatively related to the price of covariance risk $\mu_{1,t}$ but positively related to the price of 16 coskewness risk $\mu_{2,t}$. This generates a positive price of covariance risk while market returns are low.

The lower plot demonstrates that there is significant variation in the price of gamma risk. Recall that the sign of $\mu_{2,t}$ should be opposite to the sign of $\kappa^3_{m,t}$. We calculate the ratio ($\lambda_{2,t}=\mu_{2,t}/\kappa^3_{m,t}$) and find it to be negative as predicted by theory. The smallest t ratio is -1.37 (which has an asymptotic p-value of 0.085) while using the other two weighting matrices produce t ratios around -3.

6. Multiple factor models with coskewness

In this section we consider empirical tests of the three-factor model of Fama and French (1993).[19] We also test a model which includes both the Fama–French factors and coskewness with the market returns. The motivation for this analysis is the possibility that *smb* and *hml* are related to coskewness and that the empirical success of the FF three-factor model can be

[19] This model uses three-factors: *hml* (high minus low), *smb* (small minus big), and the excess market return, to describe returns to equity. *hml* is a zero net investment hedge portfolio that is formed by taking a long position in assets with a high book equity to market equity ratio and a short position in low book-to-market stocks. *smb* is similarly formed by taking a long position in small stocks and a short position in big stocks. The returns on these factors are available from Ken French's Website.

Table 6
Parameter estimates of the Fama–French three-factor asset pricing model

Parameter	Constant	S&P	Div	Term	Default	Tbill	λ=0	Constant λ	J_T test
CAPM									
λ_1	0.0494	0.0263	0.0503	−0.0429	0.0329	−0.0945	59.6454	58.4344	141.3887
	(0.0137)	(0.0094)	(0.0151)	(0.0086)	(0.0146)	(0.0160)	(0.0000)	(0.0000)	(0.0018)
FF three-factor model									
λ_1	0.0468	−0.0017	−0.0139	−0.0525	0.0943	−0.0848	37.8544	28.4213	116.7004
	(0.0170)	(0.0157)	(0.0273)	(0.0151)	(0.0252)	(0.0283)	(0.0000)	(0.0000)	(0.0106)
λ_{smb}	0.0073	0.0799	0.1699	0.0412	−0.0759	−0.0405	24.9595	24.5991	
	(0.0278)	(0.0271)	(0.0454)	(0.0328)	(0.0463)	(0.0430)	(0.0003)	(0.0002)	
λ_{hml}	0.0184	−0.0794	−0.1360	0.0295	0.2023	0.0429	19.7976	17.7466	
	(0.0282)	(0.0330)	(0.0561)	(0.0262)	(0.0523)	(0.0443)	(0.0030)	(0.0033)	
Three-moment CAPM									
λ_1	0.0487	−0.0239	−0.0657	−0.0215	0.1439	−0.1313	43.9646	43.6103	105.0632
	(0.0183)	(0.0181)	(0.0315)	(0.0200)	(0.0343)	(0.0311)	(0.0000)	(0.0000)	(0.1325)
λ_2	0.0066	−0.0089	−0.0087	−0.0215	0.0198	−0.0370	49.9908	43.6621	
	(0.0037)	(0.0028)	(0.0037)	(0.0037)	(0.0058)	(0.0068)	(0.0000)	(0.0000)	
FF three-factor model with coskewness									
λ_1	0.0362	−0.0222	−0.0409	−0.0244	0.0881	−0.1375	26.6214	26.2039	86.1516
	(0.0210)	(0.0206)	(0.0378)	(0.0218)	(0.0376)	(0.0368)	(0.0002)	(0.0001)	(0.2469)
λ_{smb}	−0.0443	0.0850	0.1806	−0.0892	−0.1142	−0.1541	23.6001	22.6911	
	(0.0335)	(0.0307)	(0.0583)	(0.0416)	(0.0605)	(0.0551)	(0.0006)	(0.0004)	
λ_{hml}	−0.0759	−0.0147	−0.0910	−0.0370	0.0508	−0.0131	10.3512	4.6564	
	(0.0342)	(0.0316)	(0.0618)	(0.0331)	(0.0573)	(0.0504)	(0.1106)	(0.4592)	
λ_2	0.0052	−0.0058	0.0003	−0.0164	0.0052	−0.0386	32.9661	32.0307	
	(0.0055)	(0.0032)	(0.0063)	(0.0039)	(0.0071)	(0.0084)	(0.0000)	(0.0000)	

Parameter estimates and standard errors (in parenthesis) are reported for the price of factor covariance risk in the conditional three-factor model of Fama and French (1993). The conditional mean return on each factor and the price of each factors covariance risk is modelled as linear functions of the information variables:

$$E[(f_{j,t}-\mu_{j,t})\otimes Z_{t-1}] = 0_{L\times 1},\ \text{for } j\in F$$

$$E[(e_{m,t}^2-\sigma_{m,t}^2)\otimes Z_{t-1}] = 0_{L\times 1}$$

$$E\left[r_t-r_t e_{m,t}\lambda_{1,t}-r_t(e_{m,t}^2-\sigma_t^2)\lambda_{2,t}-\sum\nolimits_{j\in\{F-\}} r_t e_{j,t}\lambda_{j,t}\otimes Z_{t-1}\right] = 0_{NL\times 1}$$

where $e_{j,t}=f_{j,t}-\mu_{j,t}$ and $\mu_{j,t}=Z_{t-1}^{\top}\mu_j$ for $F=\{m, smb, hml\}$, and $\lambda_{jt}=Z_{t-1}^{\top}\lambda_j$, for the factors (excluding the market) $j\in F-$ ($F-=\{smb, hml\}$), $\lambda_{1,t}=Z_{t-1}^{\top}\lambda_1$ is the price of market covariance risk, $\lambda_{2,t}=Z_{t-1}^{\top}\lambda_2$ is the price of market coskewness risk, and $\sigma_{m,t}=(Z_{t-1}^{\top}\sigma_m)^2$ is the conditional standard deviation of the market. Also presented are Wald test statistics and p-values in parenthesis for the null hypothesis that each price of risk is zero, that each price of risk is constant, and the J_T specification test.

attributed, in part, to the factors proxying for omitted coskewness. Harvey and Siddique (1999) provide some support for this idea in their unconditional analysis. Our multi-factor models are estimated following He et al. (1996), who extend Harvey's (1989) conditional test to account for multiple factors, and use the following moment conditions:

$$E[(r_t-r_t(f_t-\mu_t)^{\top}\lambda_t)\otimes Z_{t-1}] = 0_{NL\times 1}, \tag{15}$$

Table 7
Comparing various specifications of the asset pricing models

Model	Criteria	J_T test	p-value
Panel A: *model independent covariance-based weights*			
FF1	0.6572	142.1554	0.0016
FF1 + skewness	0.5399	119.2980	0.0069
FF3	0.5080	131.4396	0.0029
FF3 + skewness	0.4246	101.3670	0.0389
Panel B: *identity matrix*			
FF1	8.3816	158.6566	0.0001
FF1 + skewness	1.8125	138.5110	0.0002
FF3	2.2715	108.6459	0.0880
FF3 + skewness	0.9818	148.5649	0.0000

We compare the performance of the two-moment and three-factor asset pricing models both with and without conditional coskewness being priced. We also present results that constrain the price of market covariance risk to be positive and the market price of coskewness risk to be negative. We present: 1) the minimized value of the objective function $g_T' W_{gT}$, 2) Hansen's J_T statistic and 3. its p-value, using both the identity and model−independent covariance matrix as weighting matrices.

where $\lambda_t = Z_{t-1}^{\top}\lambda$, is the K-vector of risk prices (λ_t is an L by K matrix), and $\mu_t = E_{t-1}(f_t)$ is the K-vector of expected returns on the K factors f_t. Additional moment conditions identify the conditional means on each of the K factors:

$$E[(f_t - \mu_t) \otimes Z_{t-1}] = 0_{LK\times 1}. \tag{16}$$

This approach is parsimonious since it only explicitly models the conditional means and conditional prices of risk for each of the K factors. The only econometric restrictions placed on the dynamics of covariance are that they generate stationarity moments. This model can be extended to include the third moment of market returns

$$\begin{aligned} E_{t-1}[r_{i,t}] = {} & \mathrm{Cov}_{t-1}(r_{i,t}, r_{m,t})\lambda_{1,t} + \mathrm{Coskew}_{t-1}(r_{i,t}, r_{m,t})\lambda_{2,t} \\ & + \sum_{j\in\{smb,hml\}} \mathrm{Cov}_{t-1}(r_{i,t}, f_{j,t})\lambda_{j,t}. \end{aligned} \tag{17}$$

This model is consistent with a pricing kernel which is quadratic in the market return and linear in the other priced factors. We model $\lambda_{1,t}$ and $\lambda_{2,t}$ as linear functions of the conditioning variables Z_{t-1} and term these various models collectively as linear prices of risk. To identify the model we need moments for the conditional mean of each factor, the conditional variance of market returns, and the pricing restrictions:

$$\begin{aligned} E[r_t - r_t(r_{m,t} - Z_{t-1}^{\top}\mu_m) Z_{t-1}^{\top}\lambda_1 + r_t(e_{m,t}^2 - (Z_{t-1}^{\top}\sigma)^2 Z_{t-1}^{\top}\lambda_2 \\ - \sum_{j\in\{smb,hml\}} r_t(f_{j,t} - Z_{t-1}^{\top}\mu_j) Z_{t-1}^{\top}\lambda_j) \otimes Z_{t-1}] = 0_{NL\times 1}, \end{aligned} \tag{18}$$

where $e_{m,t}=r_{m,t}-Z_{t-1}^{\top}\mu_m$ is the market's residual. Note that we only need to explicitly model the conditional variance of the market portfolio while the empirical model in the previous section required modelling both the conditional variance and skewness.

The results of fitting a range of models with linear prices of risk are reported in Table 6. The basic two-moment CAPM is easily rejected just as in our previous setup. Including the factors

Table 8
Industry pricing errors: identity weighting matrix

	FF1	FF1 skew	FF3	FF3 skew
Panel A: average pricing errors				
Food	0.3912	0.1867	−0.0011	0.0477
Mines	0.1421	0.1950	0.1054	−0.0225
Oil	0.4681	0.1991	0.0306	0.0352
Clths	0.2001	0.0959	0.0981	−0.0717
Durbl	0.2081	0.1885	−0.0049	0.0327
Chems	0.2208	0.1106	0.0267	−0.0231
Cnsum	0.3918	0.2168	0.0127	0.0746
Cnstr	0.1417	0.1396	0.0240	−0.0138
Steel	0.0398	0.0552	−0.0199	−0.1171
FabPr	0.2511	0.1864	0.1125	0.0108
Machn	0.1772	0.2025	−0.0450	0.0906
Cars	0.2098	0.1519	0.1159	0.0046
Trans	0.2047	0.1157	0.0601	−0.0305
Utils	0.1011	−0.0568	−0.0362	−0.1213
Rtail	0.2336	0.0656	0.0366	0.0183
Finan	0.2132	0.1196	−0.0295	−0.0589
Other	0.2111	0.0919	−0.0703	0.0058
Ave abs	0.2239	0.1399	0.0488	0.0458
Panel B: average abs average managed pricing errors				
Food	0.3071	0.0763	0.0390	0.0632
Mines	0.1979	0.1354	0.0986	0.0388
Oil	0.4892	0.1036	0.0995	0.0596
Clths	0.1549	0.0817	0.0820	0.0761
Durbl	0.1634	0.0674	0.0593	0.0527
Chems	0.1744	0.0551	0.0749	0.0375
Cnsum	0.2537	0.1073	0.1142	0.0489
Cnstr	0.1214	0.0813	0.0823	0.0605
Steel	0.1986	0.1058	0.0924	0.0680
FabPr	0.1633	0.1034	0.1628	0.0693
Machn	0.1594	0.0767	0.1009	0.0666
Cars	0.1624	0.1171	0.1258	0.0659
Trans	0.1013	0.0556	0.0743	0.0810
Utils	0.1267	0.1727	0.0928	0.0837
Rtail	0.1295	0.1392	0.0876	0.0662
Finan	0.0867	0.0715	0.0693	0.0487
Other	0.1559	0.0692	0.1110	0.0206
Ave abs	0.1851	0.0952	0.0922	0.0593

We present the average pricing errors for each portfolio for the conditional two- and three-moment CAPM and various linear price of risk models (including the three-factor model of Fama and French (1993)). Panel A reports the average pricing error, and panel B reports the average of the absolute time-series average pricing errors across all $L+1$ managed portfolios formed from each industry portfolio. The final row in each panel is the average (across the 17 industries) absolute price error. The parameters are estimated using the identity weighting matrix.

smb and *hml* does improve the model fit, but we are still able to reject the FF three-factor model with a *p*-value of around 1%. There is, however, strong evidence supporting a conditional analysis since all three prices of factor risk appear to be time-varying (consistent with the empirical evidence presented by He et al. (1996)). Adding the Fama and French (1993) factors to the market does not adversely affect the price of market covariance risk. The unconditional market covariance risk λ_m is statistically significant and positive. This is interesting because it suggests that previous conclusions that covariance with the market has no ability to explain cross-sectional variability in equity returns (e.g., Fama and French, 1993; Ferson and Harvey, 1999) may be due to restrictive assumptions on the dynamics of covariances. After we relax these assumptions, covariance with the market is important in explaining cross-sectional variation in equity returns. The factors *smb* and *hml* are also important in explaining returns, although their unconditional prices of risk are statistically indistinguishable from zero.

The final two models reported in Table 6 include conditional coskewness with the market return. Neither of these models is rejected at standard significance levels. There is ample evidence that all prices of risk are time-varying. Furthermore, we find that adding conditional coskewness in the FF three-factor model causes the unconditional price of risk on *smb* and *hml* to become negative (although *smb* is not statistically different from zero). We find no evidence that the price of *hml* risk is either time-varying or nonzero. This is particularly interesting given that previous studies generally conclude that *hml* is the more important factor (Fama and French, 1996).

The structural break tests for the linear price of risk models are also reported in Table 4. Interestingly we find that the strongest evidence of parameter instability is in those models which include the FF factors, suggesting that either the price of *smb* and *hml* factor risk or their expected returns are not linear in the instrumental variables.

It is difficult to use the parameter estimates in Table 6 to compare model fit because the weighting matrices applied to the various moments are different. We therefore estimate all the models using the fixed weighting matrices that we used in our robustness check of the previous results. In Table 7 we present the minimized value of the objective function, Zhou's H_T overidentifying moment test and its *p*-value. When the efficient weighting matrix is used (panel A) the improvement from adding coskewness is greater than the improvement from adding both *smb* and *hml*. Combining both coskewness and the FF factors improves the objective criteria further. However, when the identity weighting matrix is used (panel B) adding both *smb* and *hml* improves the pricing fit more than adding coskewness alone, whether or not we impose sign constraints. It is significant that adding coskewness to the three-factor model does improve the model fit even in this case. It is clear from these results that coskewness can explain some of the cross-section of equity returns that *smb* and *hml* cannot, and that the FF factors contribute to the explanatory power of the three-moment CAPM.

To understand which industries are priced by which moments we present the average pricing errors for each of the 17 industry portfolios and the average return on the managed portfolios in Table 8 using the identity weighting matrix for comparability. The final row reports the average absolute pricing errors. The FF three-factor model does a very good job pricing the raw portfolios. The average absolute pricing error for the FF three-factor model is less than one quarter than that of the single factor model and less than half than that of the three-moment model and there is a minor improvement when both the FF factors and market coskewness are included, suggesting an important role for *smb* in light of the weak evidence for *hml* above. As is known (Ferson and Harvey, 1999), the FF three-factor model is much less impressive in pricing dynamic managed portfolios. The three-factor and three-moment models are quite similar in pricing the managed portfolios across all industries, and the combined model reduces the average pricing errors by about another 50%. The Fabricated Products, Machinery and "Other" industries are priced better by the three-moment model, whereas the Food, Mines, Retail and Utilities are priced better by the three-factor model. The Oil, Chemicals, Consumer Products, Automotive, and Other portfolios require both the FF factors and coskewness.

7. Conclusions

In this paper we test the conditional three-moment CAPM and find that coskewness is important in explaining the cross-section of asset returns. We do not reject the conditional three-moment CAPM in our data and find strong evidence that both coskewness and the price of coskewness risk are both time-varying. Our most interesting finding is that investors care more about coskewness risk when the market is positively skewed than when the market is negatively skewed. When the market is positively skewed investors are willing to sacrifice 7.87% per unit of gamma (a standardized measure of coskewness risk) annually. However, when the market is negatively skewed they only demand a premium for bearing gamma risk of 1.80% annually. We also observe a similar pattern in the returns to the coskewness factor of Harvey and Siddique (1999). This factor earns 5.00% annually when the market is positively skewed but only earns 2.81% annually when the market is negatively skewed.

We also compare the fit from asset pricing models that include coskewness with multiple factor models. Adding conditional coskewness to the Fama–French factors does better than either the three-factor or three-moment model alone.

Finally, we reconcile our results with evidence on the importance of coskewness from SDF tests. We do not reject the three-moment CAPM while an SDF-based test does rejects a conditional quadratic pricing kernel which includes preferences over skewness. We find that including moment conditions that relate the SDF parameters directly to the moments of market returns significantly improves the performance of the model. Furthermore, the SDF implied by the three-moment CAPM fits the data better than the quadratic SDF.

Appendix A. Comparing linear and GARCH volatility models

Although the instrumental variable approach to modelling the conditional mean is common (see, e.g., Campbell, 1987; Harvey, 1989, 1991, He et al., 1996 and Whitelaw, 1994), it is less common to model volatility using instrumental variables (though see, e.g., Whitelaw, 1994), and to our knowledge the instrumental variable technique has never been used to model conditional skewness. We compare the fit of the instrumental variable volatility model with the GARCH(1,1) model of Bollerslev (1986). The GARCH(1,1) is a natural benchmark for comparison because of its wide application and empirical success in modelling volatility in numerous financial time-series (Bollerslev et al., 1992; see also Hansen and Lunde, 2001 for an interesting perspective). To ensure that the two volatility models are comparable we use the same mean parameters to construct both volatility models. In particular, we take the residuals from the linear mean model and assume these are the true residuals when estimating both the GARCH and linear instrumental variables volatility model parameters. The log-likelihood is calculated as $\mathcal{L} = \sum_{t=1}^{T} \log f(e_t|0,\sigma_t^2)$ where $f(\cdot|0,\sigma_t^2)$ is the pdf of a normal random variable with mean 0 and variance σ_t^2, where σ_t^2 is the conditional volatility from either the GARCH

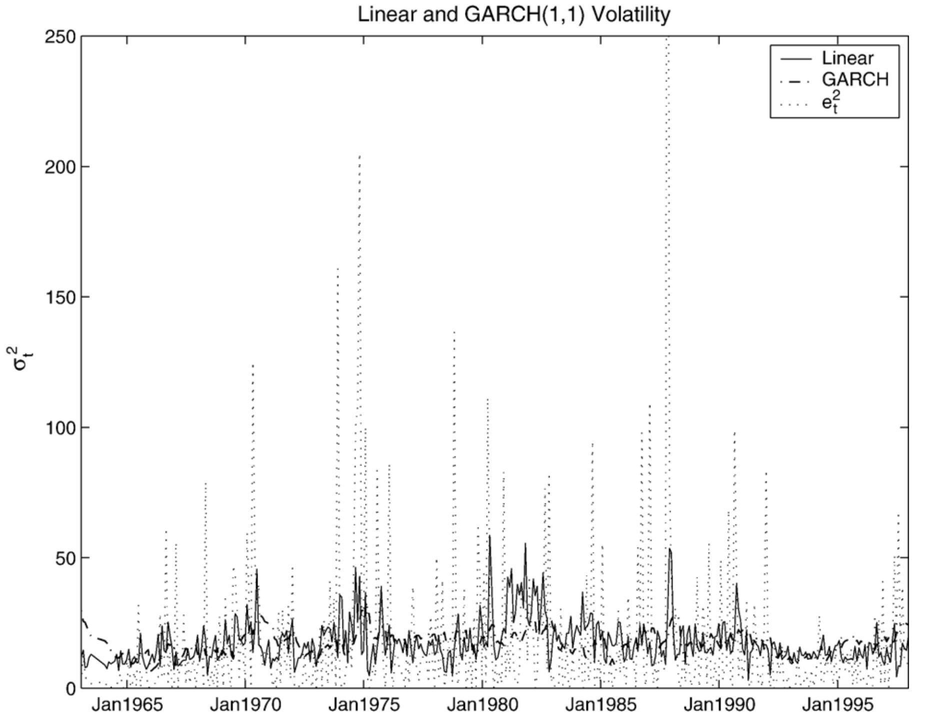

Fig. 3. GARCH and linear conditional volatility. The fitted conditional volatility from a GARCH(1,1) model (Eq. (19)) and the linear instrumental variable model (Table 1) along with the squared residual are plotted.

(1,1) model ($\mathcal{L}^{\text{GARCH}}$) or the linear instrumental variable volatility model ($\mathcal{L}^{IV}$). We initially estimate a GARCH(1,1) model

$$\sigma_t^2 = \underset{(0.0908)}{0.4678} + \underset{(0.0186)}{0.0590}\, e_{t-1}^2 + \underset{(0.0372)}{0.8032}\, \sigma_{t-1}^2 \tag{19}$$

(with robust standard errors in parenthesis) which has a log-likelihood of $\mathcal{L}^{\text{GARCH}} = -1172.74$. The log-likelihood implied by the fitted linear conditional standard deviations is $\mathcal{L}^{IV} = -1159.49$ which suggests a better fit than the GARCH(1,1) model. Note, however, that the linear standard deviation model has three more parameters than the GARCH model, and this improvement in log-likelihood is commensurate with this.

We plot the squared residual, which is a measure of realized volatility, are plotted in Fig. 3. The linear instrumental variable-based volatility is plotted as a solid line and the fitted GARCH volatility is given by the dashed line. Except for a few minor differences, the two models give virtually identical conditional volatility estimates. Both models also seem to fit the squared unexpected market return (the dotted line) rather well.

To compare the two volatility models more formally we apply two information criteria. The Akaiki information criteria (AIC) suggests that the M extra parameters in a more heavily parameterized model are justified if the log-likelihood increase by $2 \times M$, or in our case by 6; while the Bayesian information criteria (BIC) would require the log-likelihood to increase by $M \times log(T) = 18{:}07$. Since the log-likelihood increases by 13.25 with the addition of three parameters, the

AIC would choose the more heavily parameterized linear model, while the BIC, which places a more stringent penalty on more complex models, would choose the more parsimonious GARCH model.

Appendix B. Comparison with the SDF approach

We find support for the three-moment CAPM, while Dittmar (2002) rejected a conditional pricing kernel for the three-moment CAPM. In fact Dittmar (2002) rejected all market return based asset pricing models which incorporate up to, and including, the fourth moment. However, by adding labor growth rates, the empirical performance of the model improves dramatically, increasing the p-value of the Hansen–Jagannathan distance statistic to approximately 23%. We now explore the reason our GMM-based tests find support for the three-moment CAPM while Dittmar's do not.

Virtually all asset pricing theories can be represented as an SDF which is a random variable ξ_t such that all asset prices satisfy the pricing equation

$$E_{t-1}[(1+R_t)\xi_t]=1. \tag{20}$$

The conditionally risk-free asset is R_f: $E_{t-1}[\xi_t]=(1+R_{f,t})^{-1}$. Here we compare two different nonlinear SDFs. The first SDF, which was used by Dittmar (2002), is a p-th order polynomial of market returns:

$$\xi_t=\eta_{0,t}+\sum_{i=1}^{p}\eta_{i,t}r_{m,t}^{i}, \tag{21}$$

with $\eta_{i,t}=s_i(Z_{t-1}^{\top}\eta_i)^2$, and $s_i=(-1)^i$ accounts for the preference for mean and skewness and aversion to variance and kurtosis. This SDF is motivated by a Taylor series expansion of the intertemporal marginal rate of substitution for the representative agent. The second nonlinear SDF is derived by dividing the moment conditions of the three-moment CAPM by the gross conditional risk-free return and rewriting in SDF form, giving:

$$\xi_t=\zeta_{0,t}+\zeta_{1,t}(r_{m,t}-\mu_{m,t})+\zeta_{2,t}(e_{m,t}^{2}-\sigma_{m,t}^{2}), \tag{22}$$

where $\zeta_{0,t}=(1+R_{f,t})^{-1}, \zeta_{1,t}=\dfrac{\mu_{1,t}}{\sigma_{m,t}^{2}(1+R_{f,t})}$ and $\zeta_{2,t}=\dfrac{\mu_{2,t}}{\kappa_{m,t}^{3}(1+R_{f,t})}$. Each term in the SDF includes $1+R_{f,t}$ in the denominator to ensure that it correctly prices the risk-free asset.

Table 9
Stochastic discount factor models

Model	HJ Distance	H_T	p-value
Two-moment CAPM	0.8636	135.8420	0.0047
Three-moment CAPM	0.6714	106.5885	0.0416
First-order SDF	1.2224	186.6006	0.0000
Second-order SDF	0.8434	158.8670	0.0000
Third-order SDF	0.7405	152.7452	0.0000

We report the value of the Hansen–Jagannathan distance statistic (column 1), the minimized criterion function for the multi-moment asset pricing models in SDF using the inverse of the second-moment matrix, and the H_T test of Zhou (1994) (column 2) and its p-value (column 3). We report the SDF version of the two- and three-moment CAPM models and the first-through third-order Taylor series expansion of Dittmar (2002).

Although both of these nonlinear SDFs are polynomial functions of the market return, there are some important differences. Firstly, the conditional CAPM approach is more heavily parameterized (requiring $4\boldsymbol{L}+1$ parameters) than the conditional SDF approach (the quadratic SDF only requires $3\boldsymbol{L}$ parameters). The second major difference is the functional form of the conditional coefficients in the SDF. Dittmar (2002) uses a sign corrected linear specification for the conditional coefficients whereas the three-moment CAPM implies the following nonlinear functional form, obtained by collecting the coefficients in Eq. (22)in terms of $r^i_{m,t}$:

$$\eta_{0,t} = (1+R_{f,t})^{-1} - \mu_{m,t}\frac{\mu_{1,t}}{\sigma^2_{m,t}(1+R_{f,t})} - (\mu^2_{m,t}+\sigma^2_{m,t})\frac{\mu_{2,t}}{\kappa^3_{m,t}(1+R_{f,t})}$$

$$\eta_{1,t} = \frac{\mu_{1,t}}{\sigma^2_{m,t}(R_{f,t})} - 2\mu_{m,t}\frac{\mu_{2,t}}{\kappa^3_{m,t}(1+R_{f,t})}$$

$$\eta_{2,t} = \frac{\mu_{2,t}}{\kappa^3_{m,t}(1+R_{f,t})}.$$

We estimate the model parameters by GMM using the inverse of the second moment matrix following Hansen and Jagannathan (1991) which is model independent. We follow Dittmar and use all the industry portfolios and add the gross risk-free return yielding, in our case, a total of $N=18$ assets to be priced.[20] Including the gross risk-free return in our setup serves to link the conditional mean and variance to the market returns, so we will continue to interpret these parameters in this fashion.[21] We use the six conditioning variables previously introduced.

We report the parameter estimates and the value of the objective function, along with the p-value of the Zhou's (1994) *HT* specification test in Table 9. We find that the three-moment CAPM-based SDF provides a much better fit to the data than either the quadratic or even the cubic SDF. This is true even though the cubic pricing kernel is consistent with a four-moment asset pricing model and requires more parameters. All three of Dittmar's SDFs and the two-moment CAPM-based SDF are rejected at conventional significance levels, whereas the SDF implied by the three-moment CAPM is not. It appears that the nonlinear functional form implied by the three-moment CAPM provides a better pricing fit than the linear coefficient model estimated by Dittmar (2002).

Kan and Zhou (1999) argue that the conditional beta method provides more efficient parameter estimates (i.e., lower standard errors) than the SDF method. Jagannathan and Wang (2001) dispute this claim and demonstrate that when the factors are traded portfolios and returns are multivariate normal with constant parameters then the two methods are asymptotically equivalent. In reply, Kan and Zhou (2001) demonstrate that when the factor is non-normal then the beta method provides more efficient parameter estimates. The reason is that the SDF price of risk is the beta price of risk divided by the variance of the factor. The

[20] Where Dittmar used 20 industry portfolios in his tests, we use 17 industry portfolios.

[21] The gross risk-free return, $(1+R_{f,t})$, cancels with the same term which appears in the denominator of each coefficient in the SDF, leaving the relevant moment condition as $E[(r_{m,t}-\mu_{m,t})]+(e^2_{m,t}-\sigma^2_{m,t})\frac{\mu_{1,t}}{\sigma^2_{m,t}}=0$, which serves to link the conditional mean $\mu_{m,t}$ and variance $\sigma^2_{m,t}$ to the market returns.

SDF price of risk looses asymptotic efficiency because it incorporates sampling variability from the factor. We find a similar situation in our setting with a conditional nonlinear SDF as the standard errors from the three-moment CAPM is about 26% higher.[22]

Appendix C. Omitted variables bias

We analyze the bias in estimating $\mu_{m,t}$ by omitting coskewness using a slightly simplified setup which takes β_{t-1} and γ_{t-1} as given. In this case the parameters are estimated using the moment conditions on both the market and cross-sectional returns in the two-moment CAPM

$$E(g_t) = E\begin{bmatrix} g_{1,t} \\ g_{2,t} \end{bmatrix} = E\begin{bmatrix} Z_{t-1} r_{m,t} - A_t \mu_m \\ r_t \otimes Z_{t-1} - C_t \mu_m \end{bmatrix}$$

and in the three-moment CAPM

$$E(g_t) = E\begin{bmatrix} g_{1,t} \\ g_{2,t} \end{bmatrix} = E\begin{bmatrix} Z_{t-1} r_{m,t} - A_t \mu_1 - B_t \mu_2 \\ r_t \otimes Z_{t-1} - C_t \mu_1 - D_t \mu_2 \end{bmatrix} = 0$$

where

$$A_t = (Z_{t-1} Z_{t-1}^{\top})$$

$$B_t = (Z_{t-1} \xi_{t-1}^{\top})$$

$$C_t = (\beta_{t-1} \otimes (Z_{t-1} Z_{t-1}^{\top}))$$

$$D_t = (\gamma_{t-1} \otimes (Z_{t-1} \xi_{t-1}^{\top})),$$

$\xi_{t-1} = (Z_{t-1}^{\top}, 1_{\{Z_t^{\top}{}_{-1}\kappa_m > 0\}})^{\top}$, and μ_2 is now an $L+1$-vector which includes δ. We use the subscript T to denote a sample average of these terms, e.g. $A_T = \frac{1}{T}\sum_{t=1}^{T} A_t$. The GMM estimate of μ_m, denoted by μ_m, solves the first-order condition $D_T^{\top} W g_T = 0$ where $D_T = \frac{\partial}{\partial \theta^{\top}} g_T$, and g_T is the sample moment conditions. The difference between the two sample estimates equals

$$\hat{\mu}_m - \hat{\mu}_1 = \left([A_T^{\top} C_T^{\top}] W [A_T^{\top} C_T^{\top}]^{\top}\right)^{-1} [A_T^{\top} C_T^{\top}] W [B_T^{\top} D_T^{\top}]^{\top} \hat{\mu}_2, \tag{23}$$

so whenever $\mu_2 \neq 0$ we have an omitted variables bias. Under suitable regulatory conditions the sample estimates of A_T, B_T, C_T, D_T and $\hat{\mu}_2$ will converge asymptotically to well defined limits, and μ_m will be asymptotically biased. The asymptotic bias depends on the magnitude of μ_2 (the probability limit of $\hat{\mu}_2$). This bias resembles the standard omitted variables bias that arises when regressing cross-sectional expected returns on betas while omitting gammas. One of the determinants of the bias in the intercept of μ_m is the first row of ($[A_T^{\top}\ C_T^{\top}]\ W\ [A_T^{\top}\ C_T^{\top}]^{-1}[A_T^{\top}\ C_T^{\top}]$

[22] Note that the two differences between the estimates are 1) the beta method includes moments to link the conditional mean variance and skewness to market returns whereas the SDF method includes the risk-free asset which has a similar albeit indirect linking, and 2) the SDF estimate scales by the conditional risk-free rate of return to ensure the SDF prices the risk-free asset. The efficiency gains may be due to the scaling by the risk-free asset.

$W[B_T^\top D_T^\top]^\top$. We compute this using both the identity and the second-moment weighting matrices and find very similar results: only the first and the last elements are relevant and both are positive (the first element using both matrices is a little more than 1.0262 and the last element is either 0.3359 and 0.4033 and the other elements are many times smaller). This positive relation arises because the unconditional betas and gammas are positively correlated (the correlation coefficient is 0.73). We find a downward bias in the constant term in μ_m because δ is large and negative which dominates the small positive intercept in μ_2. Thus both the time-series market return and the cross-sectional portfolio return moments suggest that the estimate of the price of covariance risk is too low in the two-moment CAPM.

References

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

Black, F., Jensen, M., Scholes, M., 1972. The capital asset pricing model: some empirical tests. In: Jensen, M. (Ed.), Studies in the Theory of Capital Markets. Praeger New York, NY.

Bodurtha, J., Mark, N., 1991. Testing the CAPM with time varying risks and returns. Journal of Finance 46, 1485–1505.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance: a review of the theory and empirical evidence. Journal of Econometrics 52, 5–59.

Bollerslev, T., Engle, R., Wooldridge, J., 1988. A capital asset pricing model with time varying covariances. Journal of Political Economy 96, 116–131.

Bos, T., Newbold, P., 1984. An empirical investigation of the possibility of systematic stochastic risk in the market model. Journal of Business 57, 35–41.

Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–400.

Campbell, J.Y., 2000. Asset pricing at the millennium. Journal of Finance 60, 1515–1567.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, USA.

Cochrane, J.H., 2001. Asset Pricing. Princeton University Press, Princeton, USA.

Conrad, J., Cooper, M., Kaul, G., 2003. Value versus glamour. Journal of Finance 68, 1969–1996.

Davidian, M., Carrol, R., 1987. Variance function estimation. Journal of American Statistical Association 82, 1079–1091.

Dittmar, R.F., 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross-section of equity returns. Journal of Finance 51, 369–403.

Fabozzi, F., Francis, J., 1978. Beta as a random coefficient. Journal of Financial and Quantitative Analysis 13, 101–115.

Fama, E., French, K., 1992. The cross-section of expected returns. Journal of Finance 47, 427–465.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E., French, K., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55–84.

Fama, E., MacBeth, J., 1973. Risk, return and equilibrium: empirical tests. Journal of Political Economy 96, 607–636.

Fang, H., Lai, T.-Y., 1997. Co-kurtosis and capital asset pricing. The Financial Review 32, 293–307.

Ferson, W.E., Siegel, A.F., Xu, P.T., in press. Mimicking portfolios with conditioning information. Journal of Financial and Quantitative Analysis.

Ferson, W.E., Foerster, S.R., 1994. Finite sample properties of the generalized method of moments in testes of conditional asset pricing models. Journal of Financial Economics 36, 29–55.

Ferson, W.E., Harvey, C.R., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385–415.

Ferson, W.E., Harvey, C.R., 1993a. Explaining the Predictability of Asset Returns. In Research in Finance, vol. 11. JAI Press Inc.

Ferson, W.E., Harvey, C.R., 1993b. The risk and predictability of international equity returns. Review of Financial Studies 6, 527–566.

Ferson, W.E., Harvey, C.R., 1999. Conditioning variables and the cross section of stock returns. Journal of Finance 54, 1325–1360.

French, K., Schwert, G.W., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–30.

Friend, I., Westerfield, R., 1980. Co-skewness and capital asset pricing. Journal of Finance 35, 1085–1100.

Ghysels, E., 1998. On stable factor structures in the pricing of risk: do time-varying betas help or hurt? Journal of Finance 53, 549–573.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return tradeoff after all. Journal of Financial Economics 76, 509–548.

Gibbons, M.R., Ferson, W.E., 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. Journal of Financial Economics 14, 217–236.

Grundy, B.D., Martin, L.S., 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. Review of Financial Studies 13, 29–78.

Guo, H., Whitelaw, R., 2006. Uncovering the risk-return relation in the stock-market. Journal of Finance 61, 1433–1463.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L.P., Jagannathan, R., 1991. Implications of security market data for models of dynamic economies. Journal of Political Economy 99, 225–262.

Hansen, P.R., Lunde, A., 2001. A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? Unpublished working paper. Department of Economics, Brown University.

Harvey, C., 1989. Time varying conditional covariances in tests of asset pricing models. Journal of Financial Economics 24, 289–317.

Harvey, C., 1991. The world price of covariance risk. Journal of Finance 46, 111–157.

Harvey, C., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34, 465–488.

Harvey, C., Siddique, A., 2000a. Conditional skewness in asset pricing models tests. Journal of Finance 65, 1263–1295.

Harvey, C., Siddique, A., 2000b. Time-varying Conditional Skewness and the Market Risk Premium. In Research in Banking and Finance, vol. 1, pp. 25–58.

He, J., Kan, R., Ng, L., Zhang, C., 1996. Tests of the relations among marketwide factors, firm-specific variables, and stock returns using a conditional asset pricing model. Journal of Finance 51, 1891–1908.

Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 800–802.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.

Jagannathan, R., Wang, Z., 2001. Empirical Evaluation of Asset Pricing Models: A Comparison of the SDF and Beta Methods. Unpublished working paper. Kellogg Graduate School of Management, Northwestern University.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.

Kan, R., Zhou, G., 1999. A critique of the stochastic discount factor methodology. Journal of Finance 54, 1221–1248.

Kan, R., Zhou, G., 2001. Empirical Asset Pricing: The Beta Method Versus the Stochastic Discount Factor Method. Unpublished working paper. Rotman School of Management, University of Toronto.

Kraus, A., Litzenberger, R., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31, 1085–1100.

Kraus, A., Litzenberger, R., 1983. On the distributional conditions for a consumption-oriented three moment CAPM. Journal of Finance 38, 1381–1391.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.

Leland, H., 1999. Beyond mean-variance: risk and performance measurement in a nonsym-metrical world. Financial Analysts Journal 55, 27–36.

Lim, K.-G., 1989. A new test of the three-moment CAPM. Journal of Financial and Quantitative Analysis 24, 205–216.

Linter, J., 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Lo, A.W., MacKinlay, A.C., 1990. Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies 3, 431–467.

Ng, L., 1991. Tests of the CAPM with time-varying covariances: a multivariate GARCH approach. Journal of Finance 46, 1521–2507.

Pastor, L., Sinha, M., Swaminathan, B., 2005. Estimating the Intertemporal Risk-Return Tradeoff Using the Implied Cost of Capital. Unpublished working paper. Graduate School of Business, University of Chicago.

Post, T., Levy, H., van Vliet, P., 2003. Risk Aversion and Skewness Preference: A Comment. Unpublished working paper. Rotterdam School of Economics, Erasmus University Rotterdam.

Poti, V., 2005. The Coskewness Puzzle. Unpublished working paper. Dublin City University Business School.

Rubinstein, M.E., 1973. The fundamental theorem of parameter-preference security valuation. Journal of Financial and Quantitative Analysis 8, 61–69.

Schwert, G.W., Seguin, W., 1990. Heteroskedasticity in stock returns. Journal of Finance 45, 1129–1155.

Scott, R.C., Horvath, P.A., 1980. On the direction of preference for moments of higher order than the variance. Journal of Finance 35, 915–1253.

Scruggs, J.T., 1998. Resolving the puzzling intertemporal relationship between the market risk premium and conditional market variance: a two-factor approach. Journal of Finance 53, 575–603.

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

Smith, D.R., 2004. Testing for Structural Breaks in GARCH Models. Unpublished working paper. Faculty of Business Administration, Simon Fraser University.

Smith, D.R., 2005. Risk an Return in Stochastic Volatility Models: Volatility Feedback Matters! Unpublished working paper. Faculty of Business Administration, Simon Fraser University.

Whitelaw, R.F., 1994. Time variations and covariations in the expectation and volatility of stock market returns. Journal of Finance 49, 515–541.

Zhou, G., 1994. Analytical GMM tests: asset pricing with time-varying risk premiums. Review of Financial Studies 7, 687–709.