National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 5

Coskewness and its Implications for Testing Asset Pricing Models

Giovanni Barone-Adesi
Patrick Gagliardini
Giovanni Urga

First version: June 2000
Current version: April 2002

This research has been carried out within the NCCR FINRISK project on "Interest Rate and Volatility Risk".

FINRISK
National Centre of Competence in Research
Financial Valuation and Risk Management

Coskewness and its Implications for Testing Asset Pricing Models[1]

Giovanni Barone-Adesi[2] Patrick Gagliardini[3] Giovanni Urga[4]

Abstract: In this paper we investigate portfolio coskewness using a quadratic market model as return generating process. It is shown that portfolios of small (large) firms have negative (positive) coskewness with market. An asset pricing model including coskewness is tested through the restrictions it imposes on the return generating process. Although the model is statistically rejected, we show that the unexplained component in expected excess returns is constant across portfolios in our sample, and modest in magnitude. We investigate the implications of erroneously neglecting coskewness for testing asset pricing models, with particular interest for the empirically detected explanatory power of size.

[1] We wish to thank participants to the 9th Conference on Panel Data (Geneva, 22-23 June, 2000), and to the 10th annual conference of European Financial Management Association (Lugano, June 27-30, 2001), for helpful discussions on a earlier version of this paper. C. R. Harvey, R. Jagannathan, R. Morck, M. Rockinger, T. Wansbeek, and C. Zhang provided very useful comments which helped to improve the paper. We wish to thank K. French, R. Jagannathan and R. Kan for providing us with their data set. This research has been supported by the Swiss NCCR Finrisk. The usual disclaimer applies.

[2] Università della Svizzera Italiana, Lugano, CH, and City University Business School, London, U.K.

[3] Università della Svizzera Italiana, Lugano, CH. Corresponding author: Patrick Gagliardini, Institute of Finance, Università della Svizzera Italiana, Via Buffi 13, Lugano, CH. Tel.: +/41/91/9124723. Fax: +/41/91/9124647.

[4] City University Business School, London, U.K.

Introduction

Asset pricing models generally express expected returns on financial assets as linear functions of covariances of returns with some systematic risk factors. Sharpe (1964), Lintner (1965), Black (1972), Merton (1973), Kraus and Litzenberger (1976), Ross (1976), Breeden (1979), Barone-Adesi (1985), Jagannathan and Wang (1996), Harvey and Siddique (1999,2000), Dittmar (2002) have proposed several formulations of this general paradigm. [See Campbell (2000) for a recent survey on the field of asset pricing.] Unfortunately, most of the empirical tests to date have produced negative or ambiguous results. These findings have spurred renewed interest in the statistical properties of testing methodologies currently available. Among recent studies, Shanken (1992), Kan and Zhang (1999a,b) provide thorough analyses of the statistical methodologies commonly employed and highlight the sources of ambiguity that plague their findings.

Although a full specification of the return generating process is not needed for the formulation of most asset pricing models, it appears that only its preliminary knowledge may lead to the design of reliable tests. Because this condition is never met in practice, researchers are forced to make unpalatable choices between two alternative approaches. On the one hand, powerful tests can be designed in the context of a (fully) specified return generating process, but they are misleading in the presence of possible model misspecifications; on the other hand, more tolerant tests may be considered, but they may lack of power [Kan and Zhou (1999a,b) and Jagannathan and Wang (1998, 2001)]. Notice that the first choice may lead not only to the rejection of correct models, but also to the acceptance of irrelevant factors as sources of systematic risk, as noted by Kan and Zhang (1999a,b).

To complicate the picture, a number of empirical regularities have been detected. Among them, Banz (1981) relates expected returns to firm size, Fama and French (1995) link expected returns to the ratio of book to market value. Some of these anomalies fade over time, other ones seem to be more persistent, and deserve attentive investigation. Pricing anomalies may be related to the possibility that useless factors appear to be priced. Of course it is also possible that pricing anomalies proxy for omitted factors. While statistical tests do not allow us to choose among these two possible explanations of pricing anomalies, Kan and Zhang (1999a,b) suggest that perhaps large increase in R^2 and persistence of sign and size of coefficients over time are most likely to be associated with truly priced factors.

In this paper we consider coskewness and its role in testing asset pricing models in the light of the above considerations by using a data set of monthly returns on 10 stock portfolios. Following Harvey and Siddique (2000), an asset is defined to have "positive coskewness" with the market when the residuals of the regression of its returns on a constant and the market returns are positively correlated with squared market returns. Therefore, an asset with positive (negative) coskewness decreases (increases) the risk of the portfolio to large absolute market returns, and should command a lower (higher) expected return in equilibrium. Kraus and Litzenberger (1976), Barone-Adesi (1985) and Harvey and Siddique (1999, 2000) have studied non-normal[1] asset pricing models related to coskewness. Kraus and Litzenberger (1976) and Harvey and Siddique (2000) formulate expected returns as function of covariance and coskewness with the market portfolio. Further, Dittmar (2002) presents a framework in which agents are also adverse to kurtosis, implying that asset returns are influenced by both coskewness and cokurtosis with the return on aggregate wealth. Their model is very general, since it does not require the specification of an underlying return generating process. However, covariance and coskewness with market are almost perfectly collinear across their portfolios, limiting the power of the test. To remedy that, we adopt the formulation of Barone-Adesi (1985), where expected returns are functions of the coefficients of the quadratic market model, an extension of the traditional market model including squared market returns as an additional factor. This allows us to test an asset pricing model including coskewness by means of the constraints it imposes on the return generating process. By this choice the form of the latter is restricted, but it allows us for the formulation of the model in terms of the market and the squared market return, which are regressors almost orthogonal to each other. Although this does not allow for a more precise estimation of coskewness premia, it provides a powerful test of the linear relationship between risk and expected returns implied by the asset pricing model.

Apart from the interest in asset pricing models including coskewness, it is important to investigate the consequences on asset pricing tests, when coskewness is erroneously neglected. On the one hand, it turns out that portfolios coskewness coefficients tends to be correlated with size. This suggests that size has spurious explanatory power in the cross-section of expected returns since it proxies for omitted coskewness risk. On the other hand, we investigate the effects of misspecifications deriving from neglected coskewness on the power of the return generating process. In the simple example of the market model, we show that its power is seriously compromised when it is used as the alternative in a test of the CAPM.

The remaining of the paper is organized as follows. Section 1 introduces the quadratic market model. An asset pricing model including coskewness is derived from it using arbitrage pricing, and various testing methodologies are discussed. Section 2 reports estimators and test statistics used in the empirical part of the paper. Section 3 describes the data, and reports empirical results. Section 4 provides Monte Carlo simulations for investigating the finite sample properties of the test statistics, and Section 5 concludes.

1 Asset Pricing Models with Coskewness.

Factor models are amongst the most widely used return generating processes in financial econometrics. They explain comovements in asset returns as arising from the common effect of a (small) number of underlying variables, called factors [Gourieroux and Jasiak (2001)]. In this paper, a linear two-factor model, called quadratic market model, is used as return generating process. Market returns and the square of the market returns are its two factors. Specifically, let us denote by R_t the $N \times 1$ vector of returns in period t of N portfolios, and by $R_{M,t}$ the return of the market. If $R_{F,t}$ is the return in period t of a (conditionally) risk free asset, excess returns are defined as: $r_t = R_t - R_{F,t}\iota$, $r_{M,t} = R_{M,t} - R_{F,t}$, $q_{M,t} = R_{M,t}^2 - R_{F,t}$, where ι is a $N \times 1$ vector of ones. The quadratic market model is then specified by:

$$\begin{array}{rl} & r_t = \alpha + \beta r_{M,t} + \gamma q_{M,t} + \varepsilon_t,\ t = 1, ..., T, \\ \mathcal{H}_F : & \gamma \neq 0 \end{array} \tag{1}$$

where α is a $N \times 1$ vector of intercepts, β and γ are $N \times 1$ vectors of sensitivities and ε_t is an $N \times 1$ vector of errors satisfying[2]:

$$E\left[\varepsilon_t \mid \underline{R_{M,t}}, \underline{R_{F,t}}\right] = 0.$$

The quadratic market model is a direct extension of the well-known market model, which corresponds to the restriction $\gamma = 0$ in (1):

$$\begin{array}{rl} & r_t = \alpha + \beta r_{M,t} + \varepsilon_t,\ t = 1, ..., T, \\ \mathcal{H}_F^* : & \gamma = 0 \text{ in } (1). \end{array} \tag{2}$$

The motivation for including the square of the market returns is to fully account for coskewness with the market portfolio. In fact, deviations from the linear relation between asset returns and market returns implied by (2) are empirically observed. Indeed, for some classes of assets, residuals from the regression of returns on a constant and market returns tend to be positively (negatively) correlated with squared market returns. These assets show therefore a tendency to have relatively higher (lower) returns when the market experiences high absolute returns, and are said to have positive (negative) coskewness with the market. To be more specific, we anticipate that in our empirical investigation (see section 3), in accord with the results of Harvey and Siddique (2000), we find that portfolios formed by assets of small firms tend to have a negative coskewness with the market, whereas portfolios formed by assets of large firms have positive market coskewness. In addition to classical beta, market coskewness is therefore another very important risk characteristic: an asset that has positive coskewness with the market diminishes the risk of the portfolio with respect to large absolute market returns, and, everything else being equal, investors should prefer assets with positive market coskewness to those with negative coskewness. The quadratic market model (1) is a specification which provides us with a very simple way to take into account market coskewness. Indeed, we have:

$$\gamma = \frac{1}{V\left[\epsilon_{q,t}\right]} cov\left[\epsilon_t, R_{M,t}^2\right], \tag{3}$$

where ϵ_t ($\epsilon_{q,t}$) are the residuals from a theoretical regression of portfolio returns R_t (market square returns $R_{M,t}^2$) on a constant and market returns $R_{M,t}$ [3]. We use the estimate of γ in model (1) to investigate the properties of the coskewness coefficients of the N portfolios. The statistical (joint) significancy of coskewness γ is assessed by testing the null hypothesis $\mathcal{H}_F^*$ against the alternative $\mathcal{H}_F$.

From the point of view of financial economics, a linear factor model is only a return generating process, which is not necessarily consistent with notions of economic equilibrium. Constraints on its coefficients are imposed e.g. by arbitrage pricing [Ross (1976), and Chamberlain and Rothschild (1983)]. It implies that expected excess returns of assets following the factor model (1) satisfy the restriction [Barone-Adesi (1985)]:

$$E(r_t) = \beta\lambda_1 + \gamma\lambda_2, \tag{4}$$

where λ_1 and λ_2 are expected excess returns on portfolios whose excess returns are perfectly correlated with $r_{M,t}$ and $q_{M,t}$ respectively. Equation (4) is in the form of a typical linear asset pricing model, which relates expected excess returns to covariances and coskewnesses to market. Asset pricing models of this kind are considered in Kraus and Litzenberger (1976) and Harvey and Siddique (2000). Harvey and Siddique (2000) introduce their specification as a model where the stochastic discount factor is quadratic in market returns. Specifically, in our notation, condition (4) is equivalent to:

$$E\left[r_t m_t\left(\delta\right)\right] = 0, \tag{5}$$

where discount factor $m_t\left(\delta\right)$ is given by: $m_t\left(\delta\right) = 1 - r_{m,t}\delta_1 - q_{m,t}\delta_2$, and $\delta = (\delta_1, \delta_2)$ is a two-dimensional parameter. This quadratic discount factor $m_t\left(\delta\right)$ can be justified as a (formal) second order Taylor expansion of a stochastic discount factor, which is nonlinear in the market returns. Thus, in this approach, derivation and testing of (5) do not require a prior specification of a data generating process. Harvey and Siddique (2000) measure coskewness either by direct measures similar to γ in (3), $\widetilde{\gamma}$ say, or by the portfolios beta with a portfolio containing exclusively assets having positive (negative) coskewness, β_S say. They focus on the explanatory power of coskewness for asset expected returns, adopting a classical two step methodology, and assessing the contribute of coskewness by the increase in R^2 in the cross sectional regression when including (estimated) $\widetilde{\gamma}$ or β_S among the regressors. More recently, in a similar conditional GMM framework Dittmar (2002) uses a stochastic discount factor model embodying both quadratic and cubic terms, and he tests the model by a GMM statistics using the weighting matrix proposed in Jagannathan and Wang (1996) and Hansen and Jagannathan (1997). As mentioned in the introduction, the main feature of our paper, especially with respect the contributions of Harvey and Siddique (2000) and Dittmar (2002), is that we focus on testing model (4), but instead of adopting a methodology using an unspecified alternative (e.g. by a GMM test), we test (4) through the restrictions it imposes on the return generating process (1). Let us now derive these restrictions. Since the excess market return $r_{M,t}$ satisfies (4), it must be that $\lambda_1 = E(r_{M,t})$. A similar restriction doesn't hold for the second factor since it is not a traded asset. However, we expect $\lambda_2 < 0$, since assets with positive coskewness decrease the risk of a portfolio with respect to large absolute market returns, and therefore should command a lower risk premium in an arbitrage equilibrium. Thus, equation (4) implies the cross-equation restriction $\alpha = \vartheta\gamma$, where ϑ is the scalar parameter $\vartheta = \lambda_2 - E(q_{M,t})$, and arbitrage pricing is consistent with the following restricted model:

$$\begin{aligned} & r_t = \beta r_{M,t} + \gamma q_{M,t} + \gamma\vartheta + \varepsilon_t,\ t = 1, ..., T, \\ \mathcal{H}_1 : \quad & \exists \vartheta : \alpha = \vartheta\gamma \text{ in } (1). \end{aligned} \tag{6}$$

Therefore, the asset pricing model (4) is tested by testing $\mathcal{H}_1$ against $\mathcal{H}_F$.

If (4) turn out not to be supported by the data, this implies the existence of an additional component $\widetilde{\alpha}$, a $N \times 1$ vector, in expected excess returns, other than those related to market risk and coskewness risk: $E(r_t) = \beta\lambda_1 + \gamma\lambda_2 + \widetilde{\alpha}$. In this case, the intercepts α of model (1) satisfies the restriction: $\alpha = \vartheta\gamma + \widetilde{\alpha}$. It is crucial to investigate how the additional component $\widetilde{\alpha}$ varies across assets. Indeed, if it arises from an omitted factor, it will provide us with information about the sensitivities of portfolios to this factor. Furthermore, variables representing portfolio characteristics, which turn out to be correlated with $\widetilde{\alpha}$ across portfolios, will have spurious explanatory power for expected excess returns, since they proxy for the sensitivities of the omitted factor. A case of particular interest arises when $\widetilde{\alpha}$ is homogeneous across assets: $\widetilde{\alpha} = \delta\iota$, for a scalar δ, corresponding to the following specification:

$$\begin{aligned} & r_t = \beta r_{M,t} + \gamma q_{M,t} + \gamma\vartheta + \delta\iota + \varepsilon_t,\ t = 1, ..., T, \\ \mathcal{H}_2 : \quad & \exists \vartheta, \delta : \alpha = \vartheta\gamma + \delta\iota \text{ in } (1). \end{aligned} \tag{7}$$

If $\mathcal{H}_2$ is not rejected against $\mathcal{H}_F$, the risk factor omitted in (4) has homogeneous sensitivities across portfolios, and we expect portfolio characteristics not to have additional explanatory power for expected excess returns. In addition, a more powerful test of the asset pricing model (4) should be provided by a test of $\mathcal{H}_1$ against the alternative $\mathcal{H}_2$.

Finally, we can test asset pricing models which neglect coskewness. We have restricted specifications corresponding to $\mathcal{H}_1$ and $\mathcal{H}_2$, when $\gamma = 0$. The former is the Sharpe-Lintner CAPM model, which is consistent with arbitrage pricing if asset returns obey the market model (2) [4]:

$$\begin{aligned} & r_t = \beta r_{M,t} + \varepsilon_t,\ t = 1, ..., T, \\ \mathcal{H}_1^* : \quad & \alpha = \gamma = 0 \text{ in } (1)\ , \end{aligned} \tag{8}$$

while the latter is an extended version of it with δ as an additional component in the expected excess returns, common to all portfolios:

$$\begin{aligned} & r_t = \beta r_{M,t} + \delta\iota + \varepsilon_t,\ t = 1, ..., T, \\ \mathcal{H}_2^* : \quad & \exists \delta : \alpha = \delta\iota,\ \gamma = 0 \text{ in } (1), \end{aligned} \tag{9}$$

In addition to testing $\mathcal{H}_1^*$ and $\mathcal{H}_2^*$ against $\mathcal{H}_F$, it is important to investigate the consequences on tests of asset pricing models of erroneously neglecting coskewness. On the one hand, we consider the possibility that portfolio characteristics such as size are empirically found to explain expected excess returns since a truly priced factor, coskewness, is omitted. Indeed, if market coskewness is correlated with a variable such as size, this variable will have spurious explanatory power for the cross-section of expected returns, since it proxies for omitted coskewness. As we will see in the empirical section of this paper, this is exactly what happens. This suggest that a possible explanation for the empirically observed relation between size and assets excess returns is the omission of a systematic risk factor: market coskewness. As another consequence on inference of neglecting erroneously coskewness, we expect the power of the return generating process to be seriously compromised. As an example, we can compare results from testing the CAPM $\mathcal{H}_1^*$ when the market model $\mathcal{H}_F^*$ is the alternative, and when the quadratic market model $\mathcal{H}_F$ is the alternative. If $\mathcal{H}_F^*$ is misspecified due to omitting the quadratic market returns, its power is presumably low against $\mathcal{H}_1^*$.

To summarize, the various specifications we have introduced are

Eq. (1)	$\mathcal{H}_F$	$r_t = \alpha + \beta r_{M,t} + \gamma q_{M,t} + \varepsilon_t$
Eq. (7)	$\mathcal{H}_2: \quad \exists \vartheta, \delta : \alpha = \vartheta\gamma + \delta\iota$	$r_t = \beta r_{M,t} + \gamma q_{M,t} + \gamma\vartheta + \delta\iota + \varepsilon_t$
Eq. (6)	$\mathcal{H}_1: \quad \exists \vartheta : \alpha = \vartheta\gamma$	$r_t = \beta r_{M,t} + \gamma q_{M,t} + \gamma\vartheta + \varepsilon_t$
Eq. (2)	$\mathcal{H}_F^*: \quad \gamma = 0$	$r_t = \alpha + \beta r_{M,t} + \varepsilon_t$
Eq. (9)	$\mathcal{H}_2^*: \quad \exists \delta : \alpha = \delta\iota,\ \gamma = 0$	$r_t = \beta r_{M,t} + \delta\iota + \varepsilon_t$
Eq. (8)	$\mathcal{H}_1^*: \quad \alpha = \gamma = 0$	$r_t = \beta r_{M,t} + \varepsilon_t$

with their nesting structure

$$\begin{array}{ccc} \mathcal{H}_F & \supset & \mathcal{H}_F^* \\ \cup & & \cup \\ \mathcal{H}_2 & \supset & \mathcal{H}_2^* \\ \cup & & \cup \\ \mathcal{H}_1 & \supset & \mathcal{H}_1^* \end{array}$$

and, finally, the tests we are interested in are:

test	null hypothesis	alternative hypothesis
τ_1	$\mathcal{H}_F^*$	$\mathcal{H}_F$
τ_2	$\mathcal{H}_1^*$	$\mathcal{H}_F$
τ_3	$\mathcal{H}_1$	$\mathcal{H}_F$
τ_4	$\mathcal{H}_2$	$\mathcal{H}_F$
τ_5	$\mathcal{H}_1$	$\mathcal{H}_2$
τ_6	$\mathcal{H}_2^*$	$\mathcal{H}_F$
τ_7	$\mathcal{H}_1^*$	$\mathcal{H}_F^*$

2 Estimators and Test Statistics.

This section derives the estimators and test statistics used in our empirical applications. We report and discuss the various procedures, widely investigated in the literature [Campbell, Lo, MacKinlay (1997) and Gourieroux and Jasiak (2001)], and their properties within the alternative coskewness asset pricing models introduced in the previous section. Full derivations are provided in the Appendices for completeness.

We assume that the error term ε_t, $t = 1, ..., T$, is an homoscedastic martingale difference sequence satisfying:

$$\begin{aligned} E\left[\varepsilon_t | \underline{\varepsilon_{t-1}}, \underline{R_{M,t}}, \underline{R_{F,t}}\right] &= 0, \\ E\left[\varepsilon_t \varepsilon_t^{'} | \underline{\varepsilon_{t-1}}, \underline{R_{M,t}}, \underline{R_{F,t}}\right] &= \Sigma, \end{aligned} \tag{10}$$

where Σ is a positive definite $N \times N$ matrix. The factor $f_t = (r_{M,t}, q_{M,t})^{'}$ is supposed to be exogenous in the sense of Engle, Hendry and Richard (1988), and we denote by μ and Σ_f its expectation and variance-covariance matrix, respectively. We conduct estimation and inference in the framework of Pseudo Maximum Likelihood (PML) methods [White (1981), Gourieroux, Monfort and Trognon (1984), Bollerslev and Wooldridge (1992)]. If θ denotes the parameter of interest in the model under consideration, the PML estimator of order 2 is defined by the maximization:

$$\widehat{\theta} = \arg\max_{\theta} L_T(\theta), \tag{11}$$

where the criterium $L_T(\theta)$ is a (conditional) pseudo-loglikelihood, i.e. the (conditional) loglikelihood of the model, assuming a given conditional distribution for ε_t satisfying (10) and such that the resulting pseudo true density of the model is exponential quadratic. Under regularity assumptions, the PML estimator $\widehat{\theta}$ is consistent, for any chosen conditional distribution of ε_t satisfying the above conditions [See Appendix A where we report the asymptotic distribution]. $\widehat{\theta}$ is efficient when the pseudo conditional distribution of ε_t coincides with the true one, being then the PML estimator identical with the maximum likelihood (ML) estimator. Since the PML estimator is based on the maximization of a statistical criterion, hypothesis testing can be conducted by usual general asymptotic tests. In what follows, we will systematically analyze, along these lines, the various specifications introduced in Section 2.

The quadratic market model (1), the market model (2) and the CAPM in (8) are Seemingly Unrelated Regressions (SUR) systems [Zellner (1962)], with the same regressors in each equation. Denoting by θ the parameters[5] of interest in model (1):

$$\theta = (\alpha^{'}, \beta^{'}, \gamma^{'}, vech(\Sigma)^{'})^{'},$$

the PML estimator of θ based on the normal family is obtained by maximizing:

$$L_T(\theta) = -\frac{T}{2}\log\det\Sigma - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(\theta)^{'}\Sigma^{-1}\varepsilon_t(\theta), \tag{12}$$

where

$$\varepsilon_t(\theta) = r_t - \alpha - \beta r_{M,t} - \gamma q_{M,t}, \quad t = 1, ..., T.$$

As is well-known, the PML estimator for $(\alpha^{'}, \beta^{'}, \gamma^{'})^{'}$ is equivalent to the GLS estimator on the SUR system and also to the OLS estimator performed equation by equation in (1). Estimation of models (2) and (8) is similar, after setting $\gamma = 0$ and $\alpha = \gamma = 0$ respectively.

Let $\widehat{\theta}$ denote the PML estimator of θ in model (1). The test of $\mathcal{H}_F^*$ against $\mathcal{H}_F$: $\gamma = 0$ can be easily performed by a Wald statistics [6] [7] [see Appendix B]:

$$\xi_T^{F*} = T\frac{1}{\widehat{\Sigma}_f^{22}}\widehat{\gamma}^{'}\widehat{\Sigma}^{-1}\widehat{\gamma}, \tag{13}$$

which is asymptotically $\chi^2(p)$-distributed, with $p = N$, when $T \longrightarrow \infty$. Similarly, the Wald statistics for testing $\mathcal{H}_1^*$ against $\mathcal{H}_F$, $\alpha = \gamma = 0$, is given by [see Appendix B]:

$$\xi_T^{1*} = T\left(\widehat{\alpha}', \widehat{\gamma}'\right)\left(\widehat{W}^{-1} \otimes \widehat{\Sigma}^{-1}\right)\begin{pmatrix} \widehat{\alpha} \\ \widehat{\gamma} \end{pmatrix} = T \cdot tr\left[\left(\widehat{\alpha}, \widehat{\gamma}\right)' \widehat{\Sigma}^{-1}\left(\widehat{\alpha}, \widehat{\gamma}\right)\widehat{W}^{-1}\right], \tag{14}$$

where:

$$\widehat{W} = \begin{pmatrix} 1 + \widehat{\mu}'\widehat{\Sigma}_f^{-1}\widehat{\mu} & -e_2'\widehat{\Sigma}_f^{-1}\widehat{\mu} \\ -e_2'\widehat{\Sigma}_f^{-1}\widehat{\mu} & \widehat{\Sigma}_f^{22} \end{pmatrix}, \quad e_2' = (0, 1).$$

Models (6), (7), and (9) are more complicated since they entail cross-equation restrictions. We denote by:

$$\theta = \left(\beta', \gamma', \vartheta, \delta, vech\,(\Sigma)'\right)',$$

the vector of parameters of model (7). The PML estimator of order 2 of θ based on a normal pseudo conditional loglikelihood is defined by maximization of:

$$L_T\left(\theta\right) = -\frac{T}{2}\log\det\Sigma - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(\theta)'\Sigma^{-1}\varepsilon_t(\theta), \tag{15}$$

where:

$$\varepsilon_t(\theta) = r_t - \beta r_{M,t} - \gamma q_{M,t} - \gamma\vartheta - \delta\iota, \quad t = 1, ..., T.$$

The PML estimator is given by the following system of implicit equations [see Appendix C]:

$$\left(\widehat{\beta}', \widehat{\gamma}'\right)' = \left(\sum_{t=1}^{T}(r_t - \widehat{\delta}\iota)\widehat{H}_t'\right)\left(\sum_{t=1}^{T}\widehat{H}_t\widehat{H}_t'\right)^{-1}, \tag{16}$$

$$\left(\widehat{\vartheta}, \widehat{\delta}\right)' = (\widehat{Z}'\widehat{\Sigma}^{-1}\widehat{Z})^{-1}\widehat{Z}'\widehat{\Sigma}^{-1}\left(\overline{r} - \widehat{\beta}\overline{r}_M - \widehat{\gamma}\overline{q}_M\right), \tag{17}$$

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_t^{'}, \tag{18}$$

where:

$$\widehat{\varepsilon}_t = r_t - \widehat{\beta}r_{M,t} - \widehat{\gamma}q_{M,t} - \widehat{\gamma}\widehat{\vartheta} - \widehat{\delta}\iota,$$

$$\widehat{H}_t = \left(r_{M,t}, q_{M,t} + \widehat{\vartheta}\right)^{'}, \quad \widehat{Z} = \left(\widehat{\gamma}, \iota\right),$$

and $\overline{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$, $\overline{r}_M = \frac{1}{T}\sum_{t=1}^{T} r_{M,t}$, $\overline{q}_M = \frac{1}{T}\sum_{t=1}^{T} q_{M,t}$. An estimator for λ is simply obtained by:

$$\widehat{\lambda} = \widehat{\mu} + \begin{pmatrix} 0 \\ \widehat{\vartheta} \end{pmatrix}.$$

Note that $\left(\widehat{\beta}^{'}, \widehat{\gamma}^{'}\right)^{'}$ is obtained by (time series) OLS regressions of $r_t - \widehat{\delta}\iota$ on $\widehat{H}_t$ in a SUR system, performed equation by equation, whereas $\left(\widehat{\vartheta}, \widehat{\delta}\right)^{'}$ is obtained by (cross-sectional) GLS regression of $\overline{r} - \widehat{\beta}\overline{r}_M - \widehat{\gamma}\overline{q}_M$ on $\widehat{Z}$. A step of a feasible algorithm consists in: a) starting from old estimates; b) computing $\left(\widehat{\beta}^{'}, \widehat{\gamma}^{'}\right)^{'}$ from (16); c) computing $\left(\widehat{\vartheta}, \widehat{\delta}\right)^{'}$ from (17) using new estimates for $\widehat{\beta}$, $\widehat{\gamma}$ and $\widehat{Z}$; d) computing $\widehat{\Sigma}$ from (18), using new estimates. The procedure is iterated until a convergence criterion is met. The starting values for the parameters δ and ϑ are provided by equation (17), where estimates from (1) are used. The asymptotic distributions of the PML estimator are reported in Appendix C. In particular, it is shown that the asymptotic variance of the estimator of $\left(\beta^{'}, \gamma^{'}, \vartheta, \delta, \lambda_1, \lambda_2\right)$ is independent of the true distribution of the error term ε_t, as long as this satisfies the conditions for PML estimation. The results for constrained PML estimation of models (6) follow by setting $\delta = 0$, $\widehat{Z} = \widehat{\gamma}$, and deleting the vector ι, while for model (9) by setting $\gamma = 0$, $\widehat{Z} = \iota$, and deleting the vector γ.

We now consider testing hypotheses $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_2^*$ against the alternative $\mathcal{H}_F$. If θ denotes the parameter of model $\mathcal{H}_F$, these hypotheses can be written in mixed form:

$$\{\theta : \exists a \in A \subset \mathbb{R}^q : g(\theta, a) = 0\}, \tag{19}$$

where g is a vector function with values in $\mathbb{R}^r$. Assuming that the rank conditions:

$$rank\left(\frac{\partial g}{\partial \theta'}\right) = r, \quad rank\left(\frac{\partial g}{\partial a'}\right) = q,$$

are satisfied at the true values θ^0, a^0, a specification test for the hypothesis (19) based on Asymptotic Least Squares (ALS) consists in verifying whether the constraints $g(\widehat{\theta}, a) = 0$ are satisfied, where $\widehat{\theta}$ is an unconstrained estimator of θ (the PML estimator in our case) [Gourieroux, Monfort and Trognon (1985)]. It is based on the following statistics:

$$\xi_T = \arg\min_a T g(\widehat{\theta}, a)' \widehat{S} g(\widehat{\theta}, a),$$

where $\widehat{S}$ is a consistent estimator for

$$S_0 = \left(\frac{\partial g}{\partial \theta'} \Omega_0 \frac{\partial g'}{\partial \theta}\right)^{-1},$$

evaluated at the true values θ^0, a^0, where $\Omega_0 = V_{as}\left[\sqrt{T}\left(\widehat{\theta} - \theta^0\right)\right]$. Under regularity conditions, ξ_T is asymptotically $\chi^2(r-q)$-distributed, and is asymptotically equivalent to the other asymptotic tests[8].

We report the ALS test statistics for testing the hypotheses $\mathcal{H}_2$, $\mathcal{H}_1$, and $\mathcal{H}_2^*$ against the alternative $\mathcal{H}_F$ [They are fully derived in Appendix D]. The hypothesis $\mathcal{H}_2$ against $\mathcal{H}_F$ is tested by the statistics:

$$\xi_T^2 = T \frac{\left(\widehat{\alpha} - \widetilde{\vartheta}\widehat{\gamma} - \widetilde{\delta}\iota\right)' \widehat{\Sigma}^{-1} \left(\widehat{\alpha} - \widetilde{\vartheta}\widehat{\gamma} - \widetilde{\delta}\iota\right)}{1 + \widetilde{\lambda}' \widehat{\Sigma}_f^{-1} \widetilde{\lambda}} \sim \chi^2(p), \tag{20}$$

with $p = N - 2$, where $\widetilde{\lambda} = \widehat{\mu} + \left(0, \widetilde{\vartheta}\right)'$, and:

$$\begin{aligned} \left(\widetilde{\vartheta}, \widetilde{\delta}\right)' &= \arg\min_{\vartheta,\delta} \left(\widehat{\alpha} - \vartheta\widehat{\gamma} - \delta\iota\right)' \widehat{\Sigma}^{-1} \left(\widehat{\alpha} - \vartheta\widehat{\gamma} - \delta\iota\right) \\ &= (\widehat{Z}'\widehat{\Sigma}^{-1}\widehat{Z})^{-1}\widehat{Z}'\widehat{\Sigma}^{-1}\widehat{\alpha}. \end{aligned}$$

After deleting the vector ι, we get the test statistics ξ_T^1 of the hypothesis $\mathcal{H}_1$ against $\mathcal{H}_F$:

$$\xi_T^1 = T \frac{\left(\widehat{\alpha} - \widetilde{\vartheta}\widehat{\gamma}\right)' \widehat{\Sigma}^{-1} \left(\widehat{\alpha} - \widetilde{\vartheta}\widehat{\gamma}\right)}{1 + \widetilde{\lambda}' \widehat{\Sigma}_f^{-1} \widetilde{\lambda}}, \tag{21}$$

where $\widetilde{\vartheta} = (\widehat{\gamma}'\widehat{\Sigma}^{-1}\widehat{\gamma})^{-1}\widehat{\gamma}'\widehat{\Sigma}^{-1}\widehat{\alpha}$, which is asymptotically $\chi^2(p)$-distributed, with $p = N-1$. The test statistics for testing $\mathcal{H}_2^*$ against $\mathcal{H}_F$ is given by:

$$\xi_T^{2*} = T \cdot tr\left[\left(\widehat{\alpha} - \widetilde{\delta}\iota, \widehat{\gamma}\right)' \widehat{\Sigma}^{-1}\left(\widehat{\alpha} - \widetilde{\delta}\iota, \widehat{\gamma}\right)\widehat{W}^{-1}\right] \sim \chi^2(p), \tag{22}$$

with $p = 2N-1$, where:

$$\begin{aligned}\widetilde{\delta} &= \arg\min_{\delta} T tr\left[(\widehat{\alpha} - \delta\iota, \widehat{\gamma})' \widehat{\Sigma}^{-1}(\widehat{\alpha} - \delta\iota, \widehat{\gamma})\widehat{W}^{-1}\right] \\ &= \left[\left(\iota', 0'\right)\left(\widehat{W}^{-1} \otimes \widehat{\Sigma}^{-1}\right)\begin{pmatrix}\iota \\ 0\end{pmatrix}\right]^{-1}\left[\left(\iota', 0'\right)\left(\widehat{W}^{-1} \otimes \widehat{\Sigma}^{-1}\right)\begin{pmatrix}\widehat{\alpha} \\ 0\end{pmatrix}\right].\end{aligned}$$

Finally a test of $\mathcal{H}_1$ against $\mathcal{H}_2$ is simply performed by a t-test for the parameter δ, whereas the test of $\mathcal{H}_1^*$ against $\mathcal{H}_F^*$ is the famous Gibbons-Ross-Shanken test [MacKinlay (1987), and Gibbons, Ross and Shanken (1989)]:

$$\xi_T^{GRS} = T\frac{1}{1 + \frac{\widehat{\mu_1}^2}{\widehat{\Sigma_{f,11}}}}\widehat{\alpha}'\widehat{\Sigma}^{-1}\widehat{\alpha}, \tag{23}$$

where $\widehat{\alpha}$ and $\widehat{\Sigma}$ are the intercepts and the variance-covariance matrix estimated in the market model (2).

3 Empirical Results.

3.1 Data Description.

Our dataset consists of 450 (percentage) monthly returns of the 10 stock portfolios formed by size by French, for the period from July 1963 to December 2000[9]. The portfolios are constructed at the end of each June, using the June market equity and NYSE breakpoints. The portfolios for July of year t to June of $t+1$ include all NYSE, AMEX, and NASDAQ stocks for which we have market equity data for June of year t. Portfolios are ranked by size, with portfolio 1 the smallest, and portfolio 10 the largest.

The market return is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks. The risk free rate is the one-month Treasury bill rate from Ibbotson Associates[10].

3.2 Results.

We begin with the investigation of the coskewness coefficients of the portfolios in our sample, obtained by estimating the quadratic market model (1). PML-SUR estimates of the coefficients α, β, γ and of the variance Σ in model (1) are reported in Tables 1 and 2, respectively.

[**Insert somewhere here Tables 1 and 2**]

As explained in Section 3, these estimates are obtained by OLS regressions, performed equation by equation on the system (1). As expected, the beta coefficients are strongly significant for all portfolios, with smaller portfolios having larger betas in general. From the estimates of the γ parameter, we see that small portfolios have significantly negative market coskewness coefficients, whereas the latter are significantly positive for the two largest portfolios. In particular, we notice that the β and γ coefficients are strongly correlated across portfolios. We can test for joint significancy of the coskewness parameter γ by using the Wald statistics ξ_T^{F*} in (13). It assumes the value:

$$\xi_T^{F*} = 35.34, \tag{24}$$

which is strongly significant at the 5 percent level, being the associated critical value $\chi^2_{0.05}(10) = 18.31$. Finally, from Table 2, we also see that smaller portfolios are characterized by larger variances of the residual error terms.

Comparison of t-statistics in Table 1 computed under (10) with those obtained using a Newey-West estimator with 5 lags [Newey and West (1987)] suggests that errors ε_t do not show significant autocorrelation or heteroscedasticity, in accordance with our assumption (10)[11].

For our analysis, one central result from Table 1 is that the coskewness coefficients are (significantly) different from zero for all portfolios in our sample, except for two of moderate size. Furthermore, coskewness coefficients tend to be correlated with size, with small portfolios having negative coskewness with the market, and the largest portfolios having positive market coskewness. This result is consistent with the findings of Harvey and Siddique (2000). It is worth noticing that the dependence between portfolios returns and market returns deviates from that of a linear specification (as that assumed in the market model), in directions of smaller (larger) returns for small (large) portfolios when the market has a large absolute return. This finding has important consequences for the assessment of risk in various portfolio classes: small portfolios, having negative market coskewness, are exposed to a source of risk additional to market risk, and related to large absolute market returns. In addition, as we have already seen, the market model (2), if tested against the quadratic market model (1), is rejected with a largely significant Wald statistics. In the light of these findings, we conclude that the extension of the return generating process to include squared market returns is valuable.

We now investigate market coskewness in the context of models consistent with arbitrage pricing. This is done by considering constrained PML estimation of specification (6), obtained from the quadratic market model after imposing restrictions from the asset pricing model (4), and of specification (7), where a homogeneous additional constant in expected excess returns is allowed for. These PML estimators are obtained from the algorithm based on equations (16) to (18), as reported in Section 3. As convergence criterium we required the update of each parameter to be smaller than $5 \cdot 10^{-3}$. We obtained convergence of the algorithm to the same estimates over the range of sensible alternative starting points we have tried. The results for model (6) are reported in Table 3 and for model (7) in Table 4.

[**Insert somewhere here Tables 3 and 4**]

The point estimates and standard errors of the parameters β and γ are similar in the two models, and close to those obtained from (1). In particular, the estimates of the parameter γ confirm that small (large) portfolios have significantly negative (positive) coskewness coefficients. The parameter ϑ is found significantly negative in both models, as expected, but the implied estimate for the risk premium for coskewness, $\widehat{\lambda}_2$, is not statistically significant in both models. However, the estimate in model $\mathcal{H}_2$, $\widehat{\lambda}_2 = -7.439$, has at least the expected negative sign. Using it, we deduce that, for a portfolio with coskewness $\gamma = -0.01$ (a moderately small portfolio), its contribution to the expected excess return on a annual percentage basis is approximately 0.9. This contribution raises to 1.5 for the smallest portfolio in our data set.

We test the empirical validity of the asset pricing model (4) in our sample by testing hypothesis $\mathcal{H}_1$ against the alternative $\mathcal{H}_F$. The ALS test statistics ξ_T^1, obtained from (20) deleting the vector ι, assumes the value:

$$\xi_T^1 = 16.27,$$

which is not significant at the 5 percent level, but is very close to the critical value $\chi^2_{0.05}(9) = 16.90$. Thus, there is a modest evidence that asset pricing model (4) could not be satisfied in our sample. In other words, an additional component, other than covariance and coskewness to market, could be present in expected excess returns. In order to test for the homogeneity of this component across assets, we test $\mathcal{H}_2$ against $\mathcal{H}_F$. The test statistics ξ_T^2 in (20) assumes the value:

$$\xi_T^2 = 5.32,$$

largely below the critical value $\chi^2_{0.05}(8) = 15.51$. A more powerful test of the asset pricing model (4) should be provided by testing $\mathcal{H}_1$ against the alternative $\mathcal{H}_2$. This test is performed by the simple t-test of significancy of δ, and from Table 4, we see that $\mathcal{H}_1$ is quite clearly rejected. This confirms our previous impression that asset pricing model (4) is not satisfied in our sample. However, since $\mathcal{H}_2$ is not rejected, this implies that, if the additional component unexplained by (4) comes from an omitted factor, at least its sensitivities are homogeneous across portfolios in our sample, and characteristics such as size and book to market should not have explanatory power for expected excess returns. Moreover, the contribution to expected excess returns of the unexplained component, deduced from the estimate of parameter δ, is quite modest, approximately 0.4 on a annual percentage basis. Notice in particular that this is less than the half of the contribution due to coskewness for portfolios of modest size.

We can compare these findings with those obtained from asset pricing models not including coskewness. For the CAPM model, if $\mathcal{H}_1^*$ is tested against the alternative of the quadratic market model $\mathcal{H}_F$, it is strongly rejected with a statistics ξ_T^{1*} [equation (14) in Section 2] which assumes the value:

$$\xi_T^{1*} = 51.84, \tag{25}$$

well above the critical value $\chi^2_{0.05}(20) = 31.41$. With reference to the extended version of CAPM with additional homogeneous component, the hypothesis $\mathcal{H}_2^*$ is also rejected against $\mathcal{H}_F$, with a test statistics ξ_T^{2*} in (22) given by:

$$\xi_T^{2*} = 42.85,$$

which is strongly significant at the 5 percent level, being the associated critical value $\chi^2_{0.05}(19) = 30.14$. These results were expected, since we have already seen that even the market model (2) is rejected against (1). In the light of these findings, we conclude that the quadratic market model, used as return generating process, has enough power to strongly reject all specifications not including coskewness.

As explained in Section 3, we are interested in the consequences on asset pricing tests of erroneously neglecting coskewness. The results above suggest that the market model (2), not taking into account quadratic market returns, is misspecified. As already reported in (24), if tested against the quadratic market model (1), it is strongly rejected. For comparison, we report the estimates of the parameter α and β in the market model (2) in Table 5.

[**Insert somewhere here Table 5**]

We notice that the β coefficients are close to those obtained in the quadratic market model in Table 1. Therefore, neglecting the quadratic market returns does not seem to have dramatic consequences for the estimation of the model parameters. However, we expect that the consequences of this misspecification to be serious for inference. Indeed, we have seen above that the coskewness coefficients are correlated with size, small portfolios having negative market coskewness and large portfolios positive market coskewness. Although the parameter λ_2 is not found to be statistically significant, this suggests that size can have spurious explanatory power in the cross-section of asset expected excess returns since it proxies for omitted coskewness. Therefore, as anticipated in Section 2, the empirically observed ability of size to explain expected excess returns could be due to misspecification of models neglecting coskewness risk. It is interesting to compare these findings with those reported in Barone Adesi (1985), whose investigation runs over the period 1931-1975. We see that the sign of the premium for coskewness has not changed over time, with assets having negative coskewness commanding higher expected returns, as expected. On the contrary, both the sign of the premium for size and consequently the link between coskewness and size are inverted. While it appears difficult to discriminate statistically between a structural size effect and reward for coskewness, Kan and Zhang (1999a,b) suggest that perhaps persistence of sign and size of coefficients over time are most likely to be associated with truly priced factors. In this case, the explanation of the size effect as arising from neglected coskewness seems to be favored.

Misleading inference can also arise if the power of the return generating process is seriously compromised by erroneously neglected coskewness. To provide an example, we consider testing $\mathcal{H}_1^*$ of the CAPM model. We have already seen that it is strongly rejected if tested against the alternative of the quadratic market model $\mathcal{H}_F$ [see (25)]. However, if the CAPM hypothesis $\mathcal{H}_1^*$ is tested against the alternative of the market model $\mathcal{H}_F^*$, it is not rejected. Indeed, from the estimates in Table 5, we see that each intercept in the market model is separately not significant. Moreover, at a joint level, the classical Gibbons-Ross-Shanken test does not reject the CAPM hypothesis $\alpha = 0$, with a statistics ξ_T^{GRS} in (23) given by:

$$\xi_T^{GRS} = 16.49,$$

below, albeit not dramatically, the critical value $\chi^2_{0.05}(10) = 18.31$. What is likely to occur is that the market model, being itself misspecified, has insufficient power to reject the CAPM.

4 Monte Carlo simulations.

In this section we perform Monte Carlo simulations to investigate the finite sample properties of the ALS statistics ξ_T^1 in (21) proposed in this paper to test the coskewness asset pricing model (4). In particular, we compare the performance of ξ_T^1 with that of the Hansen (1982) GMM test statistics ξ_T^{GMM}, and investigate the effects of possible misspecification in the return generating process.

4.1 Experiment 1.

The data generating process used in Experiment 1 is given by:

$$r_t = \alpha + \beta r_{m,t} + \gamma q_{m,t} + \varepsilon_t,\ t = 1, ..., 450.$$

where $r_{m,t} = R_{M,t} - r_{f,t}$, $q_{m,t} = R^2_{M,t} - r_{f,t}$, with

$$\begin{aligned} R_{m,t} &\sim iidN(\mu_m, \sigma_m^2), \\ \varepsilon_t &\sim iidN(0, \Sigma),\ (\varepsilon_t) \text{ independent of } \left(R_{m,t}\right), \\ r_{f,t} &= r_f, \text{ a constant,} \end{aligned}$$

and

$$\alpha = \vartheta\gamma + \delta\iota.$$

The values of the parameters are chosen to be equal to the estimates obtained in the empirical analysis reported in the previous section. Specifically, β and γ are the third and fourth columns respectively in Table 1, the matrix Σ is taken from Table 2, ϑ is taken from Table 3, $\mu_m = 0.52$, $\sigma_m = 4.41$, and $r_f = 0.4$, corresponding to the average of the risk free return in our data set. We will refer to this data generating process as DGP1. Under DGP1, when $\delta = 0$, the quadratic equilibrium model (4) is satisfied. When $\delta \neq 0$, the equilibrium model (4) is not correct, due to an additional component homogeneous across portfolios. However, the quadratic model (1) is in any case well-specified.

We perform Monte Carlo simulation (10000 repetitions), for different values of δ, and report the rejection frequencies of the two test at the nominal size of 0.05 in Table 6.

[**Insert somewhere here Table 6**]

The second row, $\delta = 0$, reports the empirical size of the two tests statistics, ξ_T^{GMM} and ξ_T^1. Both statistics control the size quite well in finite sample, at least for sample size $T = 450$. The subsequent rows, corresponding to $\delta \neq 0$, report the power of the two test statistics against alternatives corresponding to unexplained components in expected excess returns, which are homogeneous across portfolios. Note that such additional components, with $\delta = 0.033$, were found in the data in the empirical analysis. From Table 6 the power of the ALS statistics ξ_T^1 is considerably higher than that of the GMM statistics ξ_T^{GMM}. This is due to the fact that the ALS statistics ξ_T^1 uses a well-specified alternative for testing (1), whereas the alternative for the GMM statistics ξ_T^{GMM} is left unspecified.

4.2 Experiment 2.

Under DGP1, the residuals ε_t are normal. When the ε_t are not normal, the alternative used by the ALS statistics ξ_T^1, that is model (1), is still correctly specified, since PML estimators are used to construct ξ_T^1. However, these estimators are less efficient. In this section we investigate the effect on the ALS test statistics of non-normality of the residuals ε_t. The data generating process used in this experiment, called DGP2, is equal to DGP1 but the residuals ε_t follow a multivariate t-distribution with $df = 5$ degrees of freedom, and a correlation matrix such that variance of ε_t is the same as under DGP1. The rejection frequencies of the Monte Carlo simulation (10000 replications) for the ALS statistics ξ_T^1 are reported in Table 7.

[Insert somewhere here Table 7]

The ALS statistics appears to be only slightly oversized. As expected, the power is reduced compared to the case of normality, however the loss of power caused by non-normality is limited. These results suggest that the ALS statistics does not unduly suffer from departures from normality of the residuals.

4.3 Experiment 3.

In the experiments conducted so far, the alternative used by the ALS statistics was well-specified. In this last experiment we investigate the effect of a misspecification in the alternative in the form of conditional heteroscedasticity. We thus consider two data generating processes having the same unconditional variance of the residuals ε_t, but such that the residuals ε_t are conditionally heteroscedastic in one case, and conditionally homoscedatic in the other. Specifically, DGP3 is the same as DGP1, but the innovations ε_t follow a conditionally normal, multivariate ARCH(1) process without cross effects:

$$cov\left(\varepsilon_{i,t}, \varepsilon_{j,t} \mid \underline{\varepsilon_{t-1}}\right) = \begin{cases} \omega_{ii} + \rho\varepsilon_{i,t-1}^2, & i = j \\ \omega_{ij}, & i \neq j \end{cases}.$$

The matrix $\Omega = [\omega_{ij}]$ is chosen as in Table 2, and $\rho = 0.2$. DGP4 is the same as DGP1, with i.i.d. normal innovations whose unconditional variance matrix is the same as the unconditional variance of ε_t in DGP 3. Thus under DGP4 the alternative of the ALS statistics is well-specified, but not under DGP3. The rejection ratios of the ALS statistics under DGP3 and DGP4 are reported in Table 8.

[Insert somewhere here Table 8]

The misspecification in form of conditional heteroscedasticity has no effect on the empirical size of the statistics in these simulations. The power of the test is reduced, but not dramatically.

5 Conclusions

In this paper we consider coskewness and its implications for testing asset pricing models. We use a quadratic market model as return generating process, with market returns and the square of market returns as the two factors. It is shown that portfolios of small (large) firms have negative (positive) coskewness with market. This implies that small portfolios are subject to a further source of risk other than covariance with market, that is market coskewness, which arises from (negative) covariance with large absolute market returns. Coskewness coefficients of the portfolios in our sample are shown to be jointly significant, rejecting the usual market model. These findings imply that the quadratic market model, used as a return generating process, is a valuable extension of the market model.

In order to obtain methodologies of superior power, we propose to test an asset pricing model, including coskewness, through the restrictions it imposes on the return generating process. We use an asymptotic test statistics whose finite sample properties are validated via a series of Monte Carlo simulations. Although the model is statistically rejected, we show that the unexplained component in expected excess returns is constant across portfolios and modest in magnitude. This implies that additional variables representing portfolios characteristics have no explanatory power for expected excess returns when coskewness is taken in account. This result cannot be obtained if coskewness is neglected.

In addition to that, our results have implications for testing methodologies, since they show that neglecting coskewness risk can cause misleading inference. Indeed, we find that coskewness is positively correlated with size. This suggests that a possible justification for the anomalous explanatory power of size in the cross-section of expected returns, is that it proxies for omitted coskewness risk. This view is supported by the fact that the sign of the premium for coskewness, contrary to that of size, has not changed over time. Furthermore, misspecifications due to neglected coskewness are shown to seriously compromise the power of the return generating process.

APPENDIX

A: Asymptotic distribution of the Pseudo Maximum Likelihood (PML) estimator.

Under general regularity conditions (see general references in the test), the asymptotic distribution of the PML estimator $\widehat{\theta}$ defined in (11) is:

$$\sqrt{T}\left(\widehat{\theta}-\theta^0\right) \xrightarrow{d} N\left(0, J_0^{-1} I_0 J_0^{-1}\right),$$

where J_0 (the so called information matrix), and I_0 are symmetric, positive defined matrices defined by:

$$J_0=\lim_{T\rightarrow\infty} E\left[-\frac{1}{T}\frac{\partial^2 L_T}{\partial\theta\partial\theta'}(\theta^0)\right], \quad I_0=\lim_{T\rightarrow\infty} E\left[\frac{1}{T}\frac{\partial L_T}{\partial\theta}(\theta^0)\frac{\partial L_T}{\partial\theta'}(\theta^0)\right].$$

B: PML in Seemingly Unrelated Regression systems (1), (2) and (8).

The PML estimator of the parameter $B = (\alpha, \beta, \gamma)$ defined by maximization of (12) is the well-known SUR estimator:

$$\widehat{B} = \left(\sum_{t=1}^{T} r_t F_t^{'} \right) \left(\sum_{t=1}^{T} F_t F_t^{'} \right)^{-1},$$

where $F_t = (1, r_{M,t}, q_{M,t})^{'}$. By usual asymptotic arguments, or using the results in Appendix A, it is shown that its asymptotic distribution is given by:

$$\sqrt{T} \left(\widehat{B} - B \right) \xrightarrow{d} N(0, \Sigma \otimes E \left[F_t F_t^{'} \right]^{-1}).$$

We see that the asymptotic variance of $\widehat{\gamma}$ is given by:

$$V_{as} \left[\sqrt{T} \left(\widehat{\gamma} - \gamma \right) \right] = E \left[F_t F_t^{'} \right]^{33} \Sigma = \Sigma_f^{22} \Sigma.$$

We deduce the Wald statistics for testing $\mathcal{H}_F^*$ against $\mathcal{H}_F$, $\gamma = 0$:

$$\xi_T^{F*} = T \frac{1}{\widehat{\Sigma}_f^{22}} \widehat{\gamma}^{'} \widehat{\Sigma}^{-1} \widehat{\gamma}.$$

Similarly, the asymptotic variance of $\left(\widehat{\alpha}^{'}, \widehat{\gamma}^{'} \right)^{'}$ is given by:

$$V_{as} \left[\sqrt{T} \left(\left(\begin{array}{c} \widehat{\alpha} \\ \widehat{\gamma} \end{array} \right) - \begin{array}{c} \alpha \\ \gamma \end{array} \right) \right] = W \otimes \Sigma,$$

where W is the 2×2 matrix formed by the first and third rows and columns of $E \left[F_t F_t^{'} \right]^{-1}$. Straightforward calculations give:

$$W = \left(\begin{array}{cc} 1 + \mu^{'} \Sigma_f^{-1} \mu & -e_2^{'} \Sigma_f^{-1} \mu \\ -e_2^{'} \Sigma_f^{-1} \mu & \Sigma_f^{22} \end{array} \right), \ e_2^{'} = (0, 1).$$

Then the test statistics for testing $\mathcal{H}_1^*$ against $\mathcal{H}_F$, $(\alpha = \gamma = 0)$, is:

$$\xi_T^{1*} = T \left(\widehat{\alpha}^{'}, \widehat{\gamma}^{'} \right) \left(\widehat{W}^{-1} \otimes \widehat{\Sigma}^{-1} \right) \left(\widehat{\alpha}^{'}, \widehat{\gamma}^{'} \right)^{'} = T \cdot tr \left[\left(\widehat{\alpha}, \widehat{\gamma} \right)^{'} \widehat{\Sigma}^{-1} \left(\widehat{\alpha}, \widehat{\gamma} \right) \widehat{W}^{-1} \right],$$

where $\widehat{W}$ is obtained by replacing unknown quantities by unconstrained estimators.

C: PML in model (7).

In this appendix we consider the Pseudo Maximum Likelihood (PML) estimator of model (7) defined by maximization of (15). The score vector is given by:

$$
\begin{aligned}
\frac{\partial L_T}{\partial \left(\beta', \gamma'\right)'} &= \sum_{t=1}^{T} H_t \otimes \Sigma^{-1} \varepsilon_t, \\
\frac{\partial L_T}{\partial \left(\vartheta, \delta\right)'} &= \sum_{t=1}^{T} Z' \Sigma^{-1} \varepsilon_t, \\
\frac{\partial L_T}{\partial vech(\Sigma)} &= \frac{1}{2} P^T \Sigma^{-1} \otimes \Sigma^{-1} P vech \left[\sum_{t=1}^{T} \left(\varepsilon_t \varepsilon_t' - \Sigma \right) \right],
\end{aligned}
$$

where $H_t = \left(r_{M,t}, q_{M,t} + \vartheta\right)'$, $\varepsilon_t = r_t - \beta r_{M,t} - \gamma q_{M,t} - \gamma\vartheta - \delta\iota$, $Z = (\delta, \iota)$ and P is such that $vec\,(\Sigma) = Pvech\,(\Sigma)$. By equating the score to 0, we immediately find the equations (16) to (18). The second derivatives are given by:

$$
\begin{aligned}
\frac{\partial^2 L_T}{\partial \left(\beta', \gamma'\right)' \partial \left(\beta', \gamma'\right)} &= -\sum_{t=1}^{T} H_t H_t' \otimes \Sigma^{-1}, \\
\frac{\partial^2 L_T}{\partial \left(\beta', \gamma'\right)' \partial \left(\vartheta, \delta\right)'} &= -\sum_{t=1}^{T} H_t \otimes \Sigma^{-1} Z, \\
\frac{\partial^2 L_T}{\partial \left(\vartheta, \delta\right) \partial \left(\vartheta, \delta\right)'} &= -T Z' \Sigma^{-1} Z, \\
\frac{\partial L_T}{\partial vech(\Sigma) \partial vech(\Sigma)'} &= \frac{T}{2} P^T \Sigma^{-1} \otimes \Sigma^{-1} P - \frac{T}{2} P^T \Sigma^{-1} \otimes \Sigma^{-1} \left(\sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right) \Sigma^{-1} P \\
&\quad - \frac{T}{2} P^T \Sigma^{-1} \left(\sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right) \Sigma^{-1} \otimes \Sigma^{-1} P,
\end{aligned}
$$

with the other ones vanishing in expectation. It results that matrices J_0 and I_0 (see Appendix A) are block diagonal in $\left(\beta', \gamma', \vartheta, \delta\right)'$ and $vech\,(\Sigma)$:

$$
J_0 = \begin{bmatrix} J_0^* & \\ & \widetilde{J}_0 \end{bmatrix}, \qquad I_0 = \begin{bmatrix} J_0^* & \eta S \widetilde{J}_0 \\ \widetilde{J}_0 S' \eta' & \widetilde{J}_0 K \widetilde{J}_0 \end{bmatrix},
$$

where, in the block form corresponding to $\left(\beta^{'}, \gamma^{'}\right)^{'}$, $(\vartheta, \delta)^{'}$, we have:

$$J_0^* = \left[\begin{array}{cc} E\left[H_t H_t\right] \otimes \Sigma^{-1} & \lambda \otimes \Sigma^{-1} Z \\ \lambda^{'} \otimes Z^{'} \Sigma^{-1} & Z^{'} \Sigma^{-1} Z \end{array} \right], \quad \widetilde{J}_0 = \frac{1}{2} \left(P^T \Sigma^{-1} \otimes \Sigma^{-1} P \right),$$

$$\eta = \left[\begin{array}{c} \lambda \otimes \Sigma^{-1} \\ Z^{'} \Sigma^{-1} \end{array} \right]$$

$$S = cov \left[\varepsilon_t, vech \left(\varepsilon_t \varepsilon_t^{'} \right)^{'} \right], \quad K = Var \left[vech \left(\varepsilon_t \varepsilon_t^{'} \right) \right],$$

and all parameters are evaluated at the true value. Therefore, the asymptotic variance-covariance matrix of the PML estimator $\widehat{\theta}$ is:

$$V_{as} \left[\sqrt{T} \left(\widehat{\theta} - \theta_0 \right) \right] = J_0^{-1} I_0 J_0^{-1} = \left[\begin{array}{cc} J_0^{*-1} & J_0^{*-1} \eta S \\ S^{'} \eta^{'} J_0^{*-1} & K \end{array} \right].$$

Notice that the asymptotic variance-covariance of $\left(\widehat{\beta}^{'}, \widehat{\gamma}^{'}, \widehat{\vartheta}, \widehat{\delta}\right)^{'}$, J_0^{*-1}, does not depend on the distribution of the error term ε_t, and in particular it coincides with the asymptotic variance-covariance matrix of the maximum likelihood (ML) estimator of $\left(\widehat{\beta}^{'}, \widehat{\gamma}^{'}, \widehat{\vartheta}, \widehat{\delta}\right)^{'}$ when ε_t is normal. On the contrary, asymmetries and kurtosis of the distribution of ε_t influence the asymptotic variance-covariance matrix of $vech(\Sigma)$ and the asymptotic covariance of $\left(\widehat{\beta}^{'}, \widehat{\gamma}^{'}, \widehat{\vartheta}, \widehat{\delta}\right)^{'}$ and $vech(\widehat{\Sigma})$, through matrices S and K.

The asymptotic variance-covariance of $\left(\widehat{\beta}^{'}, \widehat{\gamma}^{'}\right)^{'}$ and $\left(\widehat{\vartheta}, \widehat{\delta}\right)^{'}$ is given explicitly in block form by:

$$J_0^{*-1} = \left[\begin{array}{cc} J_0^{*11} & J_0^{*12} \\ J_0^{*21} & J_0^{*22} \end{array} \right],$$

where:

$$J_0^{*11} = \left(\Sigma_f + \lambda \lambda^{'} \right)^{-1} \otimes \Sigma + \left[\Sigma_f^{-1} \lambda \lambda^{'} \left(\Sigma_f + \lambda \lambda^{'} \right)^{-1} \right] \otimes Z \left(Z^{'} \Sigma^{-1} Z \right)^{-1} Z^{'},$$

$$J_0^{*12} = -\Sigma_f^{-1}\lambda \otimes Z \left(Z'\Sigma^{-1}Z\right)^{-1},$$

$$J_0^{*21} = J_0^{*12'}$$

$$J_0^{*22} = \left(1 + \lambda'\Sigma_f^{-1}\lambda\right)\left(Z'\Sigma^{-1}Z\right)^{-1}.$$

The estimator

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} f_t,$$

where $f_t = (r_{M,t}, q_{M,t})'$, can be seen as PML estimator on the extended pseudo-likelihood:

$$L_T(\theta, \mu, \Sigma_f) = L_T(\theta) - \frac{T}{2}\log\det\Sigma_f - \frac{1}{2}\sum_{t=1}^{T}(f_t - \mu)'\,\Sigma^{-1}(f_t - \mu),$$

where $L_T(\theta)$ is given in (15), and it is easily seen that θ, and (μ, Σ_f) are asymptotically independent. As a result:

$$V_{as}\left[\sqrt{T}\left(\widehat{\lambda}_2 - \lambda_{2,0}\right)\right] = \Sigma_{f,22} + V_{as}\left[\sqrt{T}\left(\widehat{\vartheta} - \vartheta_0\right)\right].$$

D: Asymptotic Least Squares.

In this appendix we derive the ALS statistics ξ_T^1, ξ_T^2 and ξ_T^{2*}. It is sufficient to consider the case where the restrictions are such that:

$$g(\theta, a) = A_1(a) vec(B) + A_2(a),$$

with $B = (\alpha, \beta, \gamma)$ and $A_1(a)$ such that:

$$A_1(a) = A_1^*(a) \otimes I_N,$$

and $A_1^*(a)$ a $\widetilde{r} \times 3$ matrix. In this case we have:

$$\frac{\partial g}{\partial \theta'} \Omega_0 \frac{\partial g'}{\partial \theta} = A_1^* E\left[F_t F_t'\right]^{-1} A_1^{*'} \otimes \Sigma,$$

and the test statistics is:

$$\begin{aligned} \xi_T &= \min_a T g(\widehat{\theta}, a)' \left(A_1^* \widehat{E\left[F_t F_t'\right]}^{-1} A_1^{*'} \right)^{-1} \otimes \widehat{\Sigma}^{-1} g(\widehat{\theta}, a) \\ &= \min_a T \cdot Tr \left[G(\widehat{\theta}, a)' \widehat{\Sigma}^{-1} G(\widehat{\theta}, a) \left(A_1^* \widehat{E\left[F_t F_t'\right]}^{-1} A_1^{*'} \right)^{-1} \right], \end{aligned}$$

where $G(\theta, a)$ is an $N \times \widetilde{r}$ matrix such that $g(\theta, a) = vec(G(\theta, a))$, and a consistent estimator for a is replaced in A_1^*.

We consider the specific cases ξ_T^1, ξ_T^2 and ξ_T^{2*}. For testing $\mathcal{H}_2$ against $\mathcal{H}_F$, $\exists \vartheta, \delta : \alpha = \vartheta\gamma + \delta\iota$, we have:

$$A_1^*(a) = (1, 0, -\vartheta),$$

and

$$A_1^* E\left[F_t F_t'\right]^{-1} A_1^{*'} = 1 + \lambda' \Sigma_f^{-1} \lambda,$$

resulting in the statistics ξ_T^2 in (20). In particular, notice that we can estimate $\widetilde{\vartheta}$ without knowing λ, since the factor $1 + \lambda' \Sigma_f^{-1} \lambda$ does not effect the minimization, and then use $\widetilde{\vartheta}$ to estimate λ. For testing $\mathcal{H}_1$ against $\mathcal{H}_F$,

$\exists \vartheta : \alpha = \vartheta \gamma$, we have the same $A_1^*(a)$, and we get the same statistics as before, if ι is deleted. Finally, the test of $\mathcal{H}_2^*$ against $\mathcal{H}_F$ can be written as:

$$\exists \delta : \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \delta \begin{pmatrix} \iota \\ 0 \end{pmatrix},$$

and we have:

$$A_1^*(a) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$A_1^* E \left[F_t F_t^{'} \right]^{-1} A_1^{*'} = W,$$

resulting in the statistics ξ_T^{2*} given in (22).

Notes

[1]Coskewness cannot be explained e.g. in the framework of MacKinlay and Pastor (2000), because their assumption of multivariate normality implies independence of the unexplained returns to the tested factor.

[2]for a time series $(Y_t, t \in \mathbb{Z})$, $\underline{Y_t}$ denotes all present and past values Y_s, $s \leq t$.

[3]Note that γ can equivalently be written as:

$$\gamma = \frac{1}{V\left[\epsilon_{q,t}\right]} cov\left[\epsilon_t, \epsilon_{q,t}\right].$$

The numerator is a third-order cross moment of the residuals in the regressions of R_t and $R^2_{m,t}$ on $R_{m,t}$. This is slightly different from the measure of coskewness of Krauss and Litzenberger (1976)

[4]Note that $\mathcal{H}^*_1$ is actually a joint hypothesis that $\gamma = 0$ and $E\left(r_t\right) = \beta E\left(r_{M,t}\right)$. We could also consider the single hypothesis $E\left(r_t\right) = \beta E\left(r_{M,t}\right)$, resulting in the restriction $\alpha + \gamma E\left(q_{M,t}\right) = 0$.

[5]For a $n \times n$ symmetric matrix A, $vech(A)$ denotes the $\frac{(n+1)n}{2} \times 1$ vector representation of A, where only elements on and above the main diagonal appear.

[6]It should be noted that exact tests (under normality) exist for testing hypotheses $\mathcal{H}^*_F$, $\mathcal{H}^*_1$ against $\mathcal{H}_F$, as well as $\mathcal{H}^*_1$ against $\mathcal{H}^*_F$ [see Anderson (1984)]. These tests are asymptotically equivalent to the Wald tests proposed in the paper. Future research will be devoted to compare the small sample performance of the two classes of test statistics.

[7]Upper indexes in a matrix denote elements of the inverse.

[8]It should be noted that again exact tests (under normality) can be constructed for testing hypotheses $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}^*_2$ against $\mathcal{H}_F$ [see e.g. Zhou (1995), and Velu and Zhou (1999)]. These tests are asymptotically equivalent to the Asymptotic Least Squares tests, which are proposed in the paper for their computational simplicity. A first assessement of the finite sample properties of the ALS test statistics is presented in section 4.

[9]Data are available from the site
http://web.mit.edu/kfrench/www/data_library.html, in the file "Portfolios Formed on Size".

[10]The market return and risk free return are available from the site
http://web.mit.edu/kfrench/www/data_library.html, in the files "Fama-French Benchmark Factors" and "Fama-French Factors".

[11]The full set of misspecification tests are available from the authors on request.

REFERENCES

Banz, R. W., 1981. "The relationship between Return and Market Value of Common Stocks," *Journal of Financial Economics*, 9, 3-18.

Barone-Adesi, G., 1985. "Arbitrage Equilibrium with Skewed Asset Returns," *Journal of Financial and Quantitative Analysis*, 20, 299-313.

Black, F., 1972. "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business*, 45, 444-55.

Bollerslev, T., and J. M. Wooldridge, 1992. "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143-172.

Breeden, D. T., 1979. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265-96.

Campbell, J. Y. (2000), "Asset Pricing at the Millenium," *Journal of Finance*, 55, 1515-1567.

Campbell, J. Y., Lo, A. W., and A. C. MacKinlay, 1987, *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

Chamberlain, G., and M., Rothschild, 1983."Arbitrage, Factor Structure and Mean Variance Analysis in Large Asset Markets", *Econometrica*, 51, 1281-1301.

Cochrane, J. H., 1996. "A Cross-sectional Test of an Investment-based Asset Pricing Model," *Journal of Political Economy*, 104, 572-621.

Dittmar, R. F., 2002."Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from Cross Section of Equity Returns", *Journal of Finance*, 57, 368-403.

Engle, R. F., Hendry, D. F., and J. F. Richard, 1988. "Exogeneity," *Econometrica*, 51, 277-304.

Fama, E. and K. R. French, 1995. "Size and Book-to-Market Factors in Earnings and Returns," *Journal of Finance*, 50, 131-55.

Gibbons, M., S. Ross, and J. Shanken. 1989. "A Test of the Efficiency of a Given Portfolio", *Econometrica* 57, 1121-1152.

Gourieroux, C., and J. Jasiak, 2001, *Financial Econometrics*, Princeton University Press.

Gourieroux, C., Monfort, A., and A. Trognon, 1984. "Pseudo Maximum Likelihood Methods: Theory," *Econometrica,* 52, 681-700.

Gourieroux, C., Monfort, A., and A. Trognon, 1985. "Moindres Carrés Asymptotiques," *Annales de l'INSEE*, 58, 91-122.

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

Hansen, L. P., and R. Jagannathan, 1997. "Assessing specification errors in stochastic discount factor models, *Journal of Finance,* 52, 557-590.

Harvey, C.R. and A. Siddique, 1999. "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34, 465-487.

Harvey, C.R. and A. Siddique, 2000. "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263-1295.

Jagannathan, R. and Z. Wang, 1996. "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3-53.

Jagannathan, R. and Z. Wang, 1998. "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression," *Journal of Finance*, 53, 1285-1309.

Jagannathan, R. and Z. Wang, 2001. "Empirical Evaluation of Asset Pricing Models: A Comparison of the SDF and Beta Methods", working paper, Northwestern University, Illinois.

Kan, R. and C. Zhang, 1999a. "GMM Test of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54, 103-27.

Kan, R. and C. Zhang, 1999b. "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203-35.

Kan, R. and G. Zhou, 1999. "A Critique of the Stochastic Discount Factor Methodology", *Journal of Finance*, 54, 1221-1248.

Kraus, A. and R. Litzenberger, 1976. "Skewness Preferences and the Valuation of Risk Assets," *Journal of Finance*, 31, 1085-1100.

Lintner, J., 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13-37.

MacKinlay, A. C., 1987. "On Multivariate Tests of the CAPM", *Journal of Financial Economics*,18, 341-372.

MacKinlay, A.C. and L. Pastor, 2000. "Asset Pricing Models: Implications for Expected Returns and Portfolio Selection," *The Review of Financial Studies*, 13, 883-916.

Merton, R. C., 1973. "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867-87.

Newey, W. K., and K. D. West, 1989. "A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-8.

Ross, S. A., 1976. "Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341-60.

Shanken, J., 1992. "On the Estimation of Beta Pricing Models," *Review of Financial Studies*, 5, 1-33.

Sharpe, W. F., 1964. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk ," *Journal of Finance*, 40, 1189-96.

Velu, R., and G. Zhou, 1999. "Testing Multi-Beta Asset Pricing Models," *Journal Of Empirical Finance,* (6)3, 219-241.

White, H., 1981. "Maximum Likelihood Estimation of Misspecified Models," *Econometrica,* 50, 1-25.

Zellner, A., 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of American Statistical Association*, 57, 348-68.

Zhou, G., 1995. "Small Sample Rank Tests with Applications to Asset Pricing," *Journal Empirical Finance,* (2)1, 71-93.

Table 1: Coefficient estimates of model (1).

Portfolio i	α_i	β_i	γ_i
1	0.418 (1.84) [1.70]	1.101 (24.23) [20.24]	−0.017 (−3.32) [−2.94]
2	0.299 (1.65) [1.56]	1.188 (32.62) [27.07]	−0.013 (−3.05) [−2.65]
3	0.288 (1.88) [1.86]	1.182 (38.37) [29.18]	−0.010 (−2.84) [−2.45]
4	0.283 (1.96) [1.83]	1.166 (39.99) [30.98]	−0.010 (−3.00) [−2.82]
5	0.328 (2.73) [2.51]	1.135 (46.94) [34.16]	−0.009 (−3.34) [−2.68]
6	0.162 (1.59) [1.53]	1.110 (54.02) [37.85]	−0.006 (−2.58) [−2.28]
7	0.110 (1.29) [1.24]	1.105 (64.37) [50.66]	−0.002 (−0.88) [−0.84]
8	0.076 (1.02) [0.90]	1.083 (72.59) [56.61]	−0.000 (−0.18) [−0.23]
9	−0.016 (−0.30) [−0.28]	1.017 (92.76) [98.43]	0.003 (2.06) [2.26]
10	−0.057 (−1.10) [−0.99]	0.933 (88.77) [66.71]	0.003 (2.64) [2.73]

Notes: We report t-statistics computed under (10) in round parentheses, while t-statistics calculated with Newey-West heteroscedasticity and autocorrelation consistent estimator with 5 lags are in square parentheses.

Table 2: Variance estimates of model (1).

	1	2	3	4	5	6	7	8	9	10
1	17.94	13.42	10.69	9.41	6.93	5.20	4.02	2.64	0.51	−3.11
2		11.50	9.02	8.27	6.35	4.81	3.69	2.61	0.58	−2.72
3			8.24	7.18	5.65	4.51	3.34	2.39	0.68	−2.40
4				7.39	5.56	4.37	3.40	2.41	0.78	−2.33
5					5.07	3.71	2.82	2.21	0.77	−1.93
6						3.67	2.42	1.85	0.78	−1.59
7							2.56	1.68	0.75	−1.29
8								1.93	0.85	−1.05
9									1.04	−0.50
10										0.96

Notes: Estimate of the variance $\Sigma = E\left[\varepsilon_t\varepsilon_t' \mid \underline{r_{M,t}}, \underline{q_{M,t}}\right]$ of the error ε_t in the model **(1)**.

Table 3: PML estimates of model (6).

Portfolio i	$\widehat{\beta}_i$	$\widehat{\gamma}_i$
1	1.106 (24.50)	−0.017 (−3.25)
2	1.191 (32.97)	−0.012 (−2,99)
3	1.186 (38.79)	−0.009 (−2.74)
4	1.170 (40.41)	−0.009 (−2.90)
5	1.140 (47.38)	−0.009 (−3.14)
6	1.112 (54.56)	−0.006 (−2.50)
7	1.107 (65.07)	−0.001 (−0.76)
8	1.085 (73.37)	−0.001 (−0.05)
9	1.017 (93.66)	0.002 (2.14)
10	0.933 (89.53)	0.003 (2.63)

$\widehat{\vartheta} = -14.955$ (−2.23)	$\widehat{\lambda_2} = 4.850$ (0.70)

Notes: t-statistics in parentheses.

Table 4: PML estimates of model (7).

Portfolio i	$\widehat{\beta}_i$	$\widehat{\gamma}_i$
1	1.100 (24.38)	−0.017 (−3.32)
2	1.187 (32.84)	−0.012 (−3.05)
3	1.183 (38.70)	−0.010 (−2.91)
4	1.167 (40.31)	−0.010 (−3.07)
5	1.137 (47.35)	−0.009 (−3.52)
6	1.110 (54.45)	−0.006 (−2.62)
7	1.107 (65.07)	−0.002 (−1.06)
8	1.085 (73.40)	−0.001 (−0.38)
9	1.018 (93.72)	0.002 (1.90)
10	0.934 (89.60)	0.003 (2.57)

$\widehat{\vartheta} = -27.244$ (−3.73)	$\widehat{\lambda}_2 = -7.439$ (1.01)	$\widehat{\delta} = 0.032$ (3.27)

Notes: t-statistics in parentheses.

Table 5: Estimates of model (8).

Portfolio i	$\widehat{\alpha}_i$	$\widehat{\beta}_i$
1	0.080 (0.39)	1.102 (23.97)
2	0.050 (0.31)	1.188 (32.34)
3	0.092 (0.67)	1.183 (38.09)
4	0.088 (0.67)	1.167 (39.65)
5	0.148 (1.36)	1.135 (46.43)
6	0.044 (0.48)	1.110 (53.69)
7	0.076 (1.00)	1.105 (64.39)
8	0.069 (1.05)	1.083 (72.67)
9	0.034 (0.71)	1.017 (92.41)
10	0.005 (0.10)	0.933 (88.18)

Notes: t-statistics in parentheses.

Table 6: Rejection frequencies in experiment 1

δ	ξ_T^{GMM}	ξ_T^1
0.00	0.0404	0.0559
0.03	0.0505	0.4641
0.06	0.0712	0.9746
0.1	0.1217	0.9924
0.15	0.2307	0.9945

Table 7: Rejection ratios in experiment 2

δ	ξ_T^1
0.00	0.0617
0.03	0.3781
0.06	0.9368
0.1	0.9876
0.15	0.9910

Table 8: Rejection ratios in experiment 3

δ	ξ_T^1 under DGP 4 (cond. homosced.)	ξ_T^1 under DGP 3 (cond. heterosced.)
0.00	0.0587	0.0539
0.03	0.3683	0.1720
0.06	0.9333	0.5791
0.1	0.9855	0.9373