EMPIRICAL LIKELIHOOD AS AN ALTERNATIVE TO GMM ESTIMATION IN THE AREA OF ASSET PRICING

Angelica Gonzalez

September 2005

Working Paper 05.04
Empirical Likelihood as an alternative to GMM estimation in the area of asset pricing

Angelica Gonzalez*
University of Edinburgh
Management School and Economics
e-mail: Angelica.Gonzalez@ed.ac.uk
September 20, 2005

Abstract
This paper proposes and analyses Owen’s (1998, 1990, 1991) Empirical Likelihood (EL) as an alternative to the General Method of Moments (GMM) within the Capital Asset Pricing Model (CAPM). We concentrate on the finite-sample properties, size and power, of their overidentification tests. Our simulation evidence shows that there are no clear advantages in terms of size when the GMM’s overidentification tests —based on two-step and continuously updated estimators— are compared to that based on the Empirical Likelihood Ratio (ELR) within a Mean-Variance and Three-Moment setting. The three tests have moderate size distortions. However, our findings illustrate that the ELR overidentification statistic is more powerful in detecting deviations from the null under the alternatives that we analyse.

1 Introduction

Over the last 40 decades our progress in understanding why some assets pay higher average returns than others has been tremendous. It was with the development of asset pricing models that economists were able to quantify risk and the reward for bearing it. The CAPM, based on the work of Markowitz (1959) and extended by Sharpe (1964) and Lintner (1965), was the first and probably most widely used model in asset pricing. The Mean-Variance CAPM states that the expected return of an asset is linearly related to the covariance of its return with the return of the market portfolio. Despite our ability to model expected returns using the CAPM, empirical evidence on the Mean-Variance CAPM has not been consistent. The early evidence was mostly positive but\(^1\) in the late 1970s less favourable empirical results for the Mean-Variance CAPM came out.\(^2\) There is still controversy over how these

---

*Thanks are due to Karim Abadir, Francesco Bravo, Patrick Marsh and Peter Phillips.

1Black et al, 1972; Fama and MacBeth, 1973; Blume and Friend, 1973; reported evidence consistent with the model.

discrepancies must be interpreted. Yet despite growing criticism, the CAPM remains widely used in finance.

Seeking a refinement and/or uniformity in results, some authors proposed extensions of the Mean-Variance framework to incorporate higher moments, while others advised alternative estimation procedures. However, even with the incorporation of higher moments empirical evidence is still contradictory. \(^3\) In this paper we mainly concentrate on looking at alternative estimation procedures to test the validity of the CAPM.

At the beginning estimation was carried out using Ordinary Least Squares (OLS) and Maximum Likelihood (ML). However, when deviations from the assumptions that returns are jointly normal and independent through time were accounted for, methods which accommodate non-normality, heteroscedasticity and temporal dependence of returns are to be preferred. Since the development of the GMM by Hansen (1982), this has dominated most of the literature. Within the GMM framework, the distribution of returns is not specified. It can be both serially dependent and conditionally heteroscedastic, the only assumption necessary being that excess asset returns are stationary and ergodic with finite fourth moments (Campbell et al, 1997).

The debate in respect of which approach—OLS, ML or GMM—is the appropriate estimating method to employ in asset pricing models is practically over. Nowadays, most empirical work uses the GMM.

This paper focuses on a new debate. It is devoted to an alternative estimation procedure to the GMM: EL. The ELR, as defined in (12), is a nonparametric analogue of likelihood estimation. It possesses an asymptotic variance that is the same as for the efficient GMM, thus it is asymptotically efficient. The overidentification test based on the ELR (see Section 2.1.3) is similar to that based on the GMM. They are asymptotically first-order equivalent and have the same interpretation. Both tests are distribution-free and their general setting is moment-condition models. But despite all the appealing

\(^3\) Two of the most famous papers, Friend and Westerfield (1980) and Kraus and Litzenberger (1976), lead to different conclusions. The first reports significant coefficients on beta and co-skewness, while the latter does not.
properties of EL statistics, it has had limited diffusion in the area of asset pricing. The usefulness of studying alternatives to GMM estimation are numerous. It has been extensively documented that the asymptotic approximation for GMM confidence intervals and tests can be poor (see, among others, the 1996 special issue of the *Journal of Business and Economic Statistics* on GMM). For example, the asymptotic properties of the GMM test of overidentifying restrictions can be a poor guide to finite-sample behaviour in small data sets often encountered in empirical analyses. Therefore, it is important either to explore new procedures or to improve on the existing ones.

This paper investigates the finite-sample properties, size and power\(^4\), of moment restrictions tests\(^5\) based on GMM and EL within the Mean-Variance and Three-Moment CAPM through simulation evidence.

Authors providing simulation evidence on the finite-sample properties of the GMM overidentification test have tended to consider the Mean-Variance setting and have mainly analysed the size and not the power properties of this statistic.\(^6\) However, the finite-sample size properties of the ELR overidentifying restrictions statistic remain practically unexplored in this context and the power properties more generally\(^7\). This paper considers not only the widely analysed Two-Moment setting but also a Three-Moment model. The practical usefulness of using the latter model is evident since the inclusion of higher moments allows the expected return of an asset to be related not only to the covariance but to the co-skewness and co-kurtosis of its return with the return of the market portfolio.\(^8\)

\(^4\)We define size (level) as estimates of Type I error probabilities. We will refer to Monte Carlo sizes and rejection frequencies, interchangeably. These are obtained by: (1) Drawing \(n\) observations from a DGP that satisfies the moment equations of the model. (2) Calculating the test statistic. (3) Repeating steps 1 to 2 \(m\) times. (4) Reporting the proportion of the simulated statistics that exceeds the asymptotic critical value.

\(^5\)We will refer to moment restrictions tests, overidentifying tests, tests of overidentifying restrictions and J-tests interchangeably.


\(^7\)One of the few studies that compares the asymptotic optimality of EL for testing moment restrictions to tests based on different versions of the GMM: two-step, ten-step and continuously updating; is that of Kitamura (2001). Kitamura examines the Hall and Horowitz (1996) model (refer to Chapter 2) and a simulation at the null hypothesis showed that every method gave confidence regions that undercovered the true parameter. After an adjustment to make coverage 95%, the power was compared in simulations that varied the parameters at 8 places along 4 line segments through the null. Of 32 simulations EL had the greatest power 22 times, two-step updating did this 5 times, 10-step updating 7 times and continuous updating never had the greatest power. EL’s power ranking was best at hypothesis farther from the null. Where any of simulated methods achieved power over 80%, EL had the greatest power.

The rest of the paper is organized as follows: In Section 2 the main theory underlying the EL and CAPM are presented. References are provided for further insight into the GMM.

Section 3 focuses on the Mean- Variance CAPM. We examine the finite-sample properties of tests of overidentifying restrictions based on the ELR, $W_j$; two-step GMM, $J_{2GMM}$; and continuously updated GMM, $J_{CuGMM}$. To study the size properties of these tests, we use Monte Carlo techniques to simulate their finite-sample distribution. We consider two data generated processes (DGPs) that are nested in the Mean-Variance framework: a DGP based on mexican information and a linear market model. We also assess the power properties of overidentification tests. After an adjustment to make coverage 95%, the power is compared in simulations that vary the means and variances through the null. In total, we carry out 96 experiments.

Section 4 studies the size and power properties of overidentication tests using the Three-Moment CAPM. We consider a quadratic market model to simulate data consistent with this setting and report rejection frequencies. We equalize the size of the different tests and examine their power using two different experiments. The first experiment varies the mean of the error term through the null. The second experiment assesses power under the alternative hypothesis of overidentifying structural restrictions, rather than model misspecification.

Section 5 concludes. Proofs are provided in the Appendices.

2 Theoretical Background

2.1 Empirical Likelihood

Qin and Lawless (1994) extend Owen’s (1991, 1990, 1988) formulation by combining the concept of unbiased estimating functions and EL. They assume that $x_1, x_2, ..., x_n$ are i.i.d. random variables from an unknown distribution function $F$, that there is a $q$-dimensional parameter $\theta$ associated with $F$ and that information about $\theta$ and $F$ is available in the form of $r \geq q$ functionally independent unbiased estimating functions. This is functions

$$g_j(x, \theta), \ j = 1, 2, ..., r,$$
such that

\[ E_F \{g_j (x, \theta)\} = 0. \]

The notation \( E_F \) is used to emphasize that expectations are being taken with respect to the EDF. In vector form we have

\[ g (x, \theta) = (g_1 (x, \theta), \ldots, g_r (x, \theta))^\top, \]

\[ E_F \{g (x, \theta)\} = 0. \]

Note that when \( r = q \), estimators of the parameters can be obtained as roots of the corresponding estimating equations. In order to recover an estimate of the probability distribution function from the observed sample \( x_1, x_2, \ldots, x_n \); Stein (1956) approximates it with a multinomial distribution. Owen (1990) applies Stein’s estimate and defines a nonparametric (multinomial type) likelihood function, the EL function

\[ L(F) = \prod_{i=1}^n dF(x_i) = \prod_{i=1}^n p_i, \tag{1} \]

where \( p_i = dF(x_i) = \Pr (X = x_i) \).

To apply EL to this framework we maximize the logarithm of (1) subject to: \( p_i \geq 0, \sum p_i = 1 \) and \( \sum_i p_i g (x_i, \theta) = 0 \) via Lagrange multipliers. Let

\[ \mathcal{L}(p, \lambda, t) = \sum_i \ln (p_i) + \lambda \left( 1 - \sum_i p_i \right) - nt^\top \sum_i p_i g (x_i, \theta), \]

where \( \lambda \) and \( t = (t_1, t_2, \ldots, t_r)^\top \) are Lagrange multipliers.

The first order conditions (FOC) for \( p_i, \lambda \) and \( t \) are

\[ \frac{1}{p_i} = \lambda + nt^\top g (x_i, \theta), \tag{2} \]

\[ \sum_i p_i = 1, \tag{3} \]

\[ \sum_i p_i g (x_i, \theta) = 0. \tag{4} \]

Multiplying (2) by \( p_i \), summing over \( i \) and using (3) and (4) yields

\[ \lambda = n, \]
\[ p_i = p_i(\theta, t) = n^{-1} \{1 + t^\tau g(x_i, \theta)\}^{-1}, \]  

(5)

with the restriction from (4) that

\[ \sum_i p_i g(x_i, \theta) = \sum_i n^{-1} \{1 + t^\tau g(x_i, \theta)\}^{-1} g(x_i, \theta) = 0. \]  

(6)

Qin and Lawless (1994) show that a solution for \( t \) can be determined in terms of \( \theta \) from (6) if:

(i) \( 0 \leq p_i \leq 1 \), which implies that \( t \) and \( \theta \) must satisfy \( 1 + t^\tau g(x_i, \theta) \geq 1/n \) for each \( i \),

(ii) \( 0 \) is inside the convex hull of the \( g(x_i, \theta) \)'s.

By substituting the optimal Lagrange multiplier, \( t(\theta) \), into the expression for the optimal \( p \) weights, \( p_i(\theta, t) \) in (5), the empirical probabilities can be represented in terms of \( \theta \) as

\[ p_i(\theta, t(\theta)) = n^{-1} \{1 + t(\theta)^\tau g(x_i, \theta)\}^{-1}. \]  

(7)

It follows that the profile EL function\(^9\) for \( \theta \) takes the form

\[ L_E(\theta) = \prod_{i=1}^{n} \left\{ \left( \frac{1}{n} \right) \frac{1}{1 + t(\theta)^\tau g(x_i, \theta)} \right\}. \]  

(8)

Since \( p_i = n^{-1} \) in the absence of constraints, the empirical log-likelihood ratio is

\[ l_E(\theta) = \sum_{i=1}^{n} \ln \{1 + t(\theta)^\tau g(x_i, \theta)\}. \]  

(9)

### 2.1.1 Maximum Empirical Likelihood Estimator

The form of \( \tilde{\theta}_{EL} \), the MEL estimator for \( \theta \), is the solution to the minimization of (9).\(^{10}\) Substituting \( \tilde{\theta}_{EL} \) into (7) we find

\[ \tilde{p}_{iEL} = p_{iEL}\left(\tilde{\theta}_{EL}, t(\tilde{\theta}_{EL})\right) = n^{-1} \{1 + t(\tilde{\theta}_{EL})^\tau g(x_i, \tilde{\theta}_{EL})\}^{-1}, \]  

(10)

and an estimator for the distribution function \( F \) is

\[ \tilde{F}_{nEL}(x) = \sum_i \tilde{p}_{iEL} I(x_i < x). \]  

(11)

\(^9\)A profile likelihood function has been partially maximized with respect to a subset of its parameters conditional on given values of the remaining parameters.

\(^{10}\)When \( r > q \) computational issues arise as the best way to obtain \( \tilde{\theta}_{EL} \). We will discuss computational procedures in more detail in the following Section and in Chapter 2.
2.1.2 Empirical Likelihood Ratio

The following Theorem, which was proved by Qin and Lawless (1994), allows us to use the ELR statistic for testing and/or obtaining confidence limits for parameters.

**Theorem 1** Under the assumptions of Theorem 1 in Qin and Lawless 1994\(^{11}\), the ELR statistic for testing \(H_0 : \theta = \theta_0\) is

\[
W_E(\theta_0) = 2l_E(\theta_0) - 2l_E(\bar{\theta}_{EL}),
\]

where \(l_E\) is the empirical log-likelihood function as given in (9).

\[
W_E(\theta_0) \rightarrow \chi^2_q
\]

as \(n \to \infty\) when \(H_0\) is true. \(l_E\) is the empirical log-likelihood function as given in (9).

An asymptotic \(\alpha - \text{level}\) test of \(H_0\) is

\[
\text{reject } H_0 : \theta = \theta_0 \text{ if } W_E(\theta_0) \geq c_{\alpha},
\]

where \(c_{\alpha}\) is defined such that \(\Pr(\chi^2_q \leq c_{\alpha}) = 1 - \alpha\). An asymptotic 100 \((1 - \alpha)\%\) confidence region for \(\theta\) can be obtained in the usual way applying the duality principle to the test procedure given in (13), resulting in the set of \(\theta_0\) values not rejected by the test.

2.1.3 Overidentification Test

The ELR’s test for overidentifying moment conditions requires two values for the probabilities. One in which the overidentifying restrictions holds, \(\hat{\pi}_{EL}\) in (10), and one in which these are removed from the optimization problem, \(p_i = \frac{1}{n}\). After substituting \(\hat{\pi}_{EL}\) and \(p_i = \frac{1}{n}\) in the EL functions, a test of the \(r\) restrictions can be conducted based on the ELR statistic. Corollary 1 formalizes the test and provides the asymptotic distribution of the statistic (this Corollary corresponds to Corollary 4 in Qin and Lawless, 1994).

**Corollary 1** Under the conditions of Theorem 1 in Qin and Lawless (1994), the statistic given by \(W_j\) is asymptotically \(\chi^2_{(r-q)}\) if the estimating equations are unbiased, i.e.

\(\text{Their regularity conditions relate to twice continuous differentiability of } g(x, \theta) \text{ with respect to } \theta \text{ and the boundedness of } g \text{ and its first and second derivatives, all in a neighbourhood of the true value } \theta_0. \text{ They also assume that the row rank of } E \left| \frac{\partial g(x, \theta)}{\partial \theta} \right|_{\theta_0} \text{ equals the number of parameters in the vector } \theta.\)
\[ W_j = 2 \sum_i \log \left[ 1 + \tilde{r}_{EL} g \left( x_i, \tilde{\theta}_{EL} \right) \right] \rightarrow \chi^2_{(r-q)}. \]

An asymptotic \( \alpha \)-level test of the validity of the moment restrictions is then conducted as:

\[ reject \ H_0 : E \left[ g(x, \theta) \right] = 0 \quad if \quad W_j \geq c_{\alpha}, \quad (14) \]

where \( c_{\alpha} \) is defined such that \( \Pr \left( \chi^2_{(r-q)} \leq c_{\alpha} \right) = 1 - \alpha \). An asymptotic \( 100 \(1 - \alpha\) \)% confidence region can be obtained in the usual way applying the duality principle to the test procedure given in (14).

The overidentification tests based on GMM estimators are similar to that based on the ELR, \( i.e. \) Corollary 1. They are asymptotically first-order equivalent and have the same interpretation (refer to Hansen, P.L. \textit{et al} (1996)) We will denote the statistics based on the two-step and the continuously-updated GMM estimators as \( J_{2GMM} \) and \( J_{CuGMM} \); respectively.

### 2.2 CAPM

The aim of this section is to derive the CAPM. This is important because by doing this we will get further insight into the assumptions and implications of the model.


In general terms, the main problem of the CAPM can be stated as that of an investor with a specified utility function facing an investment environment with a riskless asset and \( N \) risky assets. Her aim is to maximize her utility by combining the risky assets and the riskless one in an optimal way. This maximization leads to the expected return of the risky asset being expressed in terms of its relationship with the market.
We follow Hwang and Satchell (1999) in the derivation of the CAPM.

There is a representative investor and all returns are in units of period one consumption. There is a riskless asset whose return is $R_f$ and $N$ risky assets whose $i^{th}$ return is represented as $R_i$. Investment proportions on the riskless asset and $N$ risky assets are $x_0$ and $x_i$ ($i = 1, \ldots, N$), respectively; where:

$$x_0 + \sum_i x_i = 1.$$ 

For the investor, the initial investment is one and the end of period wealth is represented as $\omega$. Hence, her end of period wealth is

$$\omega = x_0 (1 + R_f) + \sum_i x_i (1 + R_i).$$

Consider a portfolio composed of combinations of the risky assets and the riskless one. The return of the portfolio is

$$R_P = x_0 R_f + \sum_i x_i R_i.$$ 

It is sensible to argue that the expected return on a security should be positively related to its risk. That is, individuals will hold risky securities only if its expected return compensates for their risk. According to Sharpe (1964), every investment carries two distinct risks. The systematic risk, which cannot be diversified away, and the unsystematic risk, which is specific to individual securities. Since the latter can be eliminated through appropriate diversification, the expected return hinges not on the asset’s variance, skewness and kurtosis — which are common measures of dispersion — but on the covariances, co-skewnesses and co-kurtosis of the returns. The systematic risk measures are given by beta, systematic skewness and systematic kurtosis$^{12}$, i.e.:

$$\beta_{iP} = \frac{E[(R_i - E(R_i))(R_P - E(R_P))]}{E[(R_P - E(R_P))^2]},$$  (15)

$$\gamma_{iP} = \frac{E[(R_i - E(R_i))(R_P - E(R_P))^2]}{E[(R_P - E(R_P))^3]},$$  (16)

$^{12}$The terms co-skewness and systematic skewness as well as co-kurtosis and systematic kurtosis are interchangeably applied.
\[ \theta_{iP} = \frac{E \left[ (R_i - E(R_i))(R_P - E(R_P))^3 \right]}{E \left[ (R_P - E(R_P))^4 \right]}. \]  

(17)

To link the systematic risk measures to the investor, information about the investor’s preferences must be incorporated. The investor’s expected utility is a function of the expected value of end of period wealth and higher moments: variance, skewness and kurtosis. The standard assumption is that preferences induce the favouring of higher means, smaller variances, higher skewness and smaller kurtosis. The investor is concerned as to the proportions to allocate to the riskless and risky assets and be compensated for bearing risk. Loosely put, the investor will maximize her utility, which depends on her wealth and hence on the combination of risky and riskless assets, by obtaining the optimal proportions of assets to allocate into her portfolio.

At this point it is useful to establish the relationship among the moments of the end of period wealth 

\[ -E(\omega), \sigma(\omega)^2, \gamma(\omega)^3 \text{ and } \vartheta(\omega)^4 \] 

—and the measures of systematic risk \(-\beta_{iP}, \gamma_{iP} \text{ and } \theta_{iP}\). After some algebraic manipulations (see Appendix 6.1) we obtain

\[ \sigma(\omega) = \sum_i x_i \beta_{iP} \sigma(R_P), \]  

(18)

\[ \gamma(\omega) = \sum_i x_i \gamma_{iP} \gamma(R_P), \]  

(19)

\[ \vartheta(\omega) = \sum_i x_i \vartheta_{iP} \vartheta(R_P), \]  

(20)

where

\[ \sigma(z) = \left[ E(z - E(z))^2 \right]^{1/2}, \]

\[ \gamma(z) = \left[ E(z - E(z))^3 \right]^{1/3}, \]

\[ \vartheta(z) = \left[ E(z - E(z))^4 \right]^{1/4}, \]

and \(z\) is a random variable.

We now define a constrained optimization problem

\[ \text{Max } E[U(\omega)] = f(E(\omega), \sigma(\omega), \gamma(\omega), \vartheta(\omega)), \]
subject to
\[ x_0 + \sum_i x_i = 1. \]

This optimization may be solved through Lagrange Multipliers. Let
\[ \mathcal{L} = f(E(\omega), \sigma(\omega), \gamma(\omega), \vartheta(\omega)) - \lambda \left( x_0 + \sum_i x_i - 1 \right). \] (21)

Using the relations stated in (18), (19) and (20), the Lagrange Multiplier problem in (21) can be rewritten as:
\[ \mathcal{L} = f \left( E(\omega), \sum_i x_i \beta_{iP} \sigma(R_P), \sum_i x_i \gamma_{iP} \gamma(R_P), \sum_i x_i \vartheta_{iP} \vartheta(R_P) \right) - \lambda \left( x_0 + \sum_i x_i - 1 \right). \] (22)

Then using
\[
\frac{\partial E(\omega)}{\partial x_i} = 1 + E(R_i),
\]
\[
\frac{\partial \sigma(\omega)}{\partial x_i} = \beta_{iP} \sigma(R_P),
\]
\[
\frac{\partial \gamma(\omega)}{\partial x_i} = \gamma_{iP} \gamma(R_P),
\]
\[
\frac{\partial \vartheta(\omega)}{\partial x_i} = \vartheta_{iP} \vartheta(R_P),
\]
we can write the FOC as
\[
\frac{\partial \mathcal{L}}{\partial x_0} = \frac{\partial E[U(\omega)]}{\partial E(\omega)} (1 + R_f) - \lambda = 0,
\]
\[
\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial E[U(\omega)]}{\partial E(\omega)} (1 + E(R_i)) + \frac{\partial E[U(\omega)]}{\partial \sigma(\omega)} \beta_{iP} \sigma(R_P) + \frac{\partial E[U(\omega)]}{\partial \gamma(\omega)} \gamma_{iP} \gamma(R_P)
\]
\[
+ \frac{\partial E[U(\omega)]}{\partial \vartheta(\omega)} \vartheta_{iP} \vartheta(R_P) - \lambda = 0.
\]

Rearranging the FOC we obtain
\[
E(R_i) - R_f = - \left[ \frac{\partial E[U(\omega)]}{\partial \sigma(\omega)} \right] \beta_{iP} \sigma(R_P) - \left[ \frac{\partial E[U(\omega)]}{\partial \gamma(\omega)} \right] \gamma_{iP} \gamma(R_P)
\]
\[
- \left[ \frac{\partial E[U(\omega)]}{\partial \vartheta(\omega)} \right] \vartheta_{iP} \vartheta(R_P). \] (23)

At the maximum, the expected utility is constant and the changes in expected return and variance are zero for a given level of skewness and kurtosis, i.e.
\[ dE[U(\omega)] = \frac{\partial E[U(\omega)]}{\partial E(\omega)} dE(\omega) + \frac{\partial E[U(\omega)]}{\partial \sigma(\omega)} d\sigma(\omega) = 0, \]
\[ dE[U(\omega)] = \frac{\partial E[U(\omega)]}{\partial E(\omega)} dE(\omega) + \frac{\partial E[U(\omega)]}{\partial \gamma(\omega)} d\gamma(\omega) = 0, \]
\[ dE[U(\omega)] = \frac{\partial E[U(\omega)]}{\partial E(\omega)} dE(\omega) + \frac{\partial E[U(\omega)]}{\partial \theta(\omega)} d\theta(\omega) = 0. \]

Incorporating these results into Equation (23) gives:

\[
E(R_i) - R_f = \left[ \frac{dE(\omega)}{d\sigma(\omega)} \right] \beta_{iP}\sigma(R_P) + \left[ \frac{dE(\omega)}{d\gamma(\omega)} \right] \gamma_{iP}\gamma(R_P) + \left[ \frac{dE(\omega)}{d\theta(\omega)} \right] \vartheta_{iP}\vartheta(R_P). \tag{24}
\]

It is important to stress that Equation (24) is defined in terms of a risky asset and a portfolio denoted by the subindex \( P \); thus the pricing results denote an individual equilibrium. To move from an individual equilibrium to a market one—to derive the CAPM—it must be the case that an investor’s choice of a risky investment portfolio is separate from her attitude towards risk. This property is often referred to as a portfolio separation principle.\(^{13}\) Before we formally introduce this principle, its main assumptions are summarized:

(i) Each investor chooses a portfolio with the objective of maximizing a derived utility function,

\[ f(E(\omega), \sigma(\omega), \gamma(\omega), \vartheta(\omega)) \], where the utility function is concave and preferences induce the favouring of higher means, smaller variances, higher skewness and smaller kurtosis.

(ii) All investors have a common time horizon and homogeneous beliefs about \( E(\omega), \sigma(\omega), \gamma(\omega) \) and \( \vartheta(\omega) \).

(iii) Each asset is infinitely divisible.

(iv) The riskless asset can be bought or sold in unlimited amounts.

**Theorem 2** If assumptions (i) to (iv) hold, the optimal combination of risky assets for an investor can be determined without any knowledge of the investor’s preferences towards risk and return.

Theorem 2 is the so-called portfolio separation theorem. Under this theorem the investor makes two separate decisions:

\(^{13}\)This principle is also known as portfolio separation theorem or mutual fund theorem.
1. After estimating the expected returns, variances, covariances, skewnesses, co-skewnesses, kurtosis and co-kurtosis of securities; the investor calculates an efficient set of risky assets. This is a set formed by the combination of assets that for a given level of variance, covariance, kurtosis, co-kurtosis, skewness and co-skewness yield the highest return. No personal characteristics, such as degree of risk aversion, are needed in this step. Intuitively, no other portfolio could be optimal since all investors working with the same inputs, sketch out the same efficient set of risky assets. If all investors choose the same portfolio of risky assets it is possible to determine what that portfolio is. Common sense points to it being a market valued-weighted portfolio of all existing securities: the market portfolio.

2. The investor must now determine how to combine the portfolio of risky assets with the riskless one. This allocation is determined by her tolerance towards risk.

Theorem 2 is fundamental to understanding the CAPM. It ensures that all individual investors maximize their utility with two funds: a riskless asset and the market portfolio.\(^{14}\)

Therefore, after evoking the portfolio separation theorem, Equation (24) can be rewritten in terms of the market portfolio

\[
E(R_i) - R_f = \left[ \frac{dE(\omega)}{d\sigma(\omega)} \right] \beta_{im} \sigma(R_m) + \left[ \frac{dE(\omega)}{d\gamma(\omega)} \right] \gamma_{im} \gamma(R_m) + \left[ \frac{dE(\omega)}{d\vartheta(\omega)} \right] \vartheta_{im} \vartheta(R_m). \tag{25}
\]

Note that (25) is identical to (24) except for \(R_m\), the rate of return of the market portfolio, which is substituted for \(R_P\). The main theoretical difference between both is that (25) is a market equilibrium whereas (24) is an individual equilibrium. From this point onwards, the subindex \(m\) labels the variables and parameters specific to the market portfolio.

Equation (25) is an extension of the Kraus and Litzenberger (1976) Three-Moment CAPM (henceforth, K-L CAPM). Following their notation, (25) can be rewritten as

\[
E(R_i) - R_f = b_1 \beta_{im} + b_2 \gamma_{im} + b_3 \vartheta_{im}, \tag{26}
\]

where

\[ b_1 = \left[ \frac{dE(\omega)}{d\sigma(\omega)} \right] \sigma(R_m), \]  
\[ b_2 = \left[ \frac{dE(\omega)}{d\gamma(\omega)} \right] \gamma(R_m), \]  
\[ b_3 = \left[ \frac{dE(\omega)}{d\theta(\omega)} \right] \theta(R_m). \]  

Equation (26) is the Four-Moment CAPM.

Note that:

(a) \( b_1 > 0 \),

since \( \left[ \frac{dE(\omega)}{d\sigma(\omega)} \right] > 0 \) and \( \sigma(R_m) > 0 \).

(b) \( b_2 > 0 \) if \( \gamma(R_m) < 0 \),

\( b_2 < 0 \) if \( \gamma(R_m) > 0 \),

since \( \frac{dE(\omega)}{d\gamma(\omega)} = -\frac{\partial E[U(\omega)]/\partial \gamma(\omega)}{\partial E[U(\omega)]/\partial E(\omega)} < 0 \).

(c) \( b_3 > 0 \),

since \( \frac{dE(\omega)}{d\theta(\omega)} = -\frac{\partial E[U(\omega)]/\partial \theta(\omega)}{\partial E[U(\omega)]/\partial E(\omega)} > 0 \) and \( \theta(R_m) > 0 \).

**Multiperiod Framework**

Due to the fact that the CAPM is a single period model, all the previous equations do not have a time dimension. For econometric analysis of the CAPM, it is sufficient to assume i.i.d. returns to estimate the model over time (Campbell et al, 1997).

Lim (1989) tests the validity of the Three-Moment CAPM through the GMM by defining the CAPM in terms of orthogonality conditions. This specification is convenient since the EL is also a moments-based model. The extension of Lim’s (1989) analysis to a Four-Moment framework arises naturally.

Following his work, first define the deflated excess returns for the \( i^{th} \) asset and the market portfolio as:

\( \frac{dE(\omega)}{d\sigma(\omega)} > 0 \) and \( \sigma(R_m) > 0 \).

---

\( \text{\textsuperscript{15}} \) The rates of return on the riskless asset are not constant through time. Thus, the deflated excess returns are used to make moments of the rate of returns intertemporal constants under a changing riskless interest rate (Fama, 1970).
\[ r_{it} = \left( \frac{R_{it} - R_{it}^{*}}{R_{it}^{*}} \right), \]
\[ r_{mt} = \left( \frac{R_{mt} - R_{it}^{*}}{R_{it}^{*}} \right). \]

We now define the moment conditions, \( E [g (r_{it}, r_{mt}, \theta) = 0, \)
for estimating the Four-Moment CAPM:

\[ E \left[ r_{it} \left( b_1 \beta_{im} + b_2 \gamma_{im} + b_3 \vartheta_{im} \right) \right] = 0 \quad i = 1, \ldots, N, \]

\( i = 1, \ldots, N, \)

\[ E \left[ r_{it} r_{mt} - \mu (r_m) r_{it} - \beta_{im} (r_{mt} - \mu (r_m))^2 \right] = 0 \quad i = 1, \ldots, N, \]

\[ E \left[ r_{it} r_{mt}^2 - 2 \mu (r_m) r_{it} r_{mt} + \mu (r_m)^2 r_{it} - \sigma (r_m)^2 r_{it} \right] - \gamma_{im} (r_{mt} - \mu (r_m))^3 \]

\[ i = 1, \ldots, N, \]

\[ E \left[ (r_{mt} - \mu (r_m))^3 r_{it} - \gamma_{im} (r_{mt} - \mu (r_m))^4 \right] = 0 \quad i = 1, \ldots, N, \]

\[ E [r_{mt} - \mu (r_m)] = 0, \]

\[ E \left[ (r_{mt} - \mu (r_m))^2 - \sigma (r_m)^2 \right] = 0, \]

\[ E \left[ (r_{mt} - \mu (r_m))^3 - \gamma (r_m)^3 \right] = 0, \]

\[ E \left[ (r_{mt} - \mu (r_m))^4 - \vartheta (r_m)^4 \right] = 0. \]

These equations are better analysed by dividing them into two groups.

The first group, Equations (30) – (33), specify the relationship between the returns of the risky asset and the market. The \( N \) moment conditions in (30) come from the Four-Moment CAPM as defined in (26). The following \( 3N \) orthogonality conditions, Equations (31)–(33), are \( N \) conditions for beta, \( N \) conditions for co-skewness and \( N \) conditions for co-kurtosis.

The second group, Equations (34) – (37), are particular to the market and they denote common measures for the mean, variance, skewness and kurtosis; respectively.

In total, there are \( 4N+4 \) equations and \( 3N+7 \) parameters to be estimated, \( \theta = (b_1, b_2, b_3, \mu (r_m), \sigma (r_m), \gamma (r_m), \vartheta (r_m), \beta_{im}, \gamma_{im}, \vartheta_{im})^T. \)
For simplicity it is convenient in what follows to make the assumption that $N = 1$. We will denote $\beta_{im}, \gamma_{im}, \vartheta_{im}$, and $r_{it}$ as $\beta_m, \gamma_m, \vartheta_m$, and $r_t$; respectively.

### 3 Mean-Variance CAPM

Markowitz (1959) set down the basis for the CAPM. He formulated the investor’s portfolio selection problem in terms of expected return and variance of return. He showed that investor’s would optimally hold a portfolio with the highest expected value for a given level of variance, i.e. a Mean-Variance efficient portfolio. Sharpe (1964) and Lintner (1965) extended the work of Markowitz (1959) to develop a general equilibrium model, the CAPM. They showed that if investors have homogeneous expectations and optimally hold Mean-Variance efficient portfolios then, in the absence of market frictions, the market portfolio will itself be a Mean-Variance portfolio. The Mean-Variance CAPM states that the expected return of an asset must be linear in the covariance of its return with the return of the market portfolio.

In this section, a comparison of the EL and GMM, in the context of the Mean-Variance CAPM, is carried out. Essentially what we do is to assess the finite-sample properties, size and power, of their moment restrictions tests.

First, we formally introduce the Mean-Variance CAPM as a particular case of the general model, given in Equation (26).

#### 3.1 Moment Equations

The moment equations for estimating the Mean-Variance CAPM are Equations (30), (31) and (34); where $b_1 = E(r_{mt})$, $b_2 = 0$ and $b_3 = 0$ in (30).

Hence, there are 3 equations and 2 parameters, $\theta = (\mu (r_m), \beta_m)^\top$, to be estimated.

#### 3.2 Finite-Sample Properties of Overidentification Tests

The tests of overidentifying restrictions studied in this section have as their null hypothesis that there is a value of $\theta$ consistent with $E [g (r_t, r_{mt}, \theta)] = 0$. We analyse three tests of overidentifying restrictions
in what follows: $W_j$, $J_{2GMM}$ and $J_{CuGMM}$. The three tests have an asymptotic $\chi^2_{(1)}$ distribution under the null.

### 3.2.1 The Data Generating Process

We consider two DGPs. Hwang and Satchell (1999) examine the CAPM for the case of emerging markets. The following figures are obtained from their survey, specifically for the case of Mexico:

\[
\begin{align*}
\sigma (r_{MEX}) &= 14.41, \\
\sigma (r_m) &= 4.11, \\
E (r_{mt}) &= .73, \\
\rho_{rMEX, r_m} &= .65,
\end{align*}
\] (38)

where

\(\sigma (r_{MEX})\) and \(\sigma (r_m)\) are the standard deviations of Mexico and the market,

\(\rho_{rMEX, r_m}\) denotes the correlation coefficient between the returns of Mexico and the market.

The Mean-Variance CAPM predicts

\[
E (r_{MEX}) = \beta_m E (r_{mt}).
\] (39)

Our aim is to use a DGP consistent with the Mean-Variance CAPM.\(^\text{16}\) We consider two processes.

We obtain from substituting (38) into (39):

\[
E (r_{MEX}) = \frac{(.65) (14.41) (4.11)}{(4.11)^2} (.73)
= 1.68.
\]

Hence, a DGP of the form

\[
\begin{bmatrix}
  r_{MEX} \\
  r_{mt}
\end{bmatrix} \sim N (\mu, \Sigma),
\] (40)

where

\(^{16}\text{Consistent in the sense that there is a value of } \beta_m \text{ such that the null hypothesis holds.}\)
\[ \mu = \begin{pmatrix} 1.68 + \Delta_{11} \\ .73 + \Delta_{21} \end{pmatrix}, \]
\[ \Sigma = \begin{pmatrix} \frac{(14.41 + \Delta_{12})^2}{.65} & (14.41 + \Delta_{12})(4.11 + \Delta_{22}) \\ (14.41 + \Delta_{12})(4.11 + \Delta_{22}) & (4.11 + \Delta_{22})^2 \end{pmatrix}, \]

and \( \Delta_{ij} \) for \( i, j = \{1, 2\} \) are constants that allow parameters to vary;
is consistent with the Mean-Variance CAPM if \( \Delta_{ij} = 0 \) \( \forall i, j \). In other words, if \( \Delta_{ij} \neq 0 \) then Equations (30), (31) and (34) do not hold.

The second DGP that we consider, a linear market model, has the following general form

\[ r_t = a_1 \ r_{mt} + \varepsilon_t, \quad (41) \]

where

(i) \( E(\varepsilon_t) = 0 \),

(ii) \( r_{mt} \) and \( \varepsilon_t \) are uncorrelated.

Note that only if (i) and (ii) hold then the linear market model in (41) satisfies the Mean-Variance CAPM, e.g. \( a_1 = \beta_m \).

### 3.2.2 Size of Overidentification Tests

This Section focuses on whether the asymptotic (or nominal) size is a good approximation to that in finite-samples. The following experiments employ either a DGP of the form given in (40) with \( \Delta_{ij} = 0 \) \( \forall i, j \) or a DGP of the form in (41) with \( E(\varepsilon_t) = 0 \) and no correlation between \( r_{mt} \) and \( \varepsilon_t \) so that the Mean-Variance CAPM holds.

**Results** We first consider the DGP related to the mexican figures. We report rejection frequencies, with particular interest being in cases where these probabilities are poorly approximated by the nominal size. Our experiments use 5000 replications. We consider two samples sizes: \( n = 50 \) and 100. Results are summarized in Table 1.
Empirical Levels of J-Tests
Mean-Variance CAPM

<table>
<thead>
<tr>
<th>Level</th>
<th>(n = 50)</th>
<th>(W_1)</th>
<th>(J_{CuGMM})</th>
<th>(J_{2GMM})</th>
<th>(n = 100)</th>
<th>(W_1)</th>
<th>(J_{CuGMM})</th>
<th>(J_{2GMM})</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>.1109</td>
<td>.1098</td>
<td>.1124</td>
<td></td>
<td>.1070</td>
<td>.1140</td>
<td>.0940</td>
<td></td>
</tr>
<tr>
<td>.05</td>
<td>.0593</td>
<td>.0592</td>
<td>.0561</td>
<td></td>
<td>.0590</td>
<td>.0610</td>
<td>.0520</td>
<td></td>
</tr>
<tr>
<td>.01</td>
<td>.0141</td>
<td>.0144</td>
<td>.0110</td>
<td></td>
<td>.0120</td>
<td>.0110</td>
<td>.0100</td>
<td></td>
</tr>
</tbody>
</table>

\(W_1, J_{CuGMM}\) and \(J_{2GMM}\) are J tests based on the ELR, continuously updated and two-step GMM estimators; respectively. \(n\) is the sample size.

Table 1: Finite-Sample Size Properties - CAPM mexican data

Table 1 summarizes the rejection frequencies for the tests at the .10, .05 and .01 critical values. Our findings show that the nominal size is a reasonable approximation to finite-sample sizes for the three tests. The asymptotic \(\chi^2_{(1)}\) approximation of the finite-sample distribution improves for \(W_1\) and \(J_{2GMM}\) as \(n\) increases.

Our experiments illustrate that for these sample sizes (and a well behaved DGP), the nominal critical values of the overidentification tests can be a useful guide to finite-sample behaviour.

Consider the second DGP, Equation (41). We generate pseudorandom samples with the following characteristics:

\[
\begin{bmatrix}
    r_{mt} \\
    \varepsilon_t
\end{bmatrix} \sim N\left[\begin{bmatrix} 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right]
\]

and we arbitrarily set \(a_1 = 1.5\) in (41), \(i.e.:\)

\[r_t = 1.5 \ r_{mt} + \varepsilon_t.\]

The empirical levels for 5000 replications and for two sample sizes: \(n = \{50, 100\}\); are summarized in Table 2. The ELR test is more oversized\(^{17}\) than its GMM counterparts for the three critical values. However, differences among the three tests are small. We note the familiar decrease in the size distortions as \(n\) increases.

\(^{17}\)Oversized (undersized) tests refer to tests whose Monte Carlo sizes are larger (smaller) than their nominal counterparts.
### Empirical Levels of J-Tests

#### Mean-Variance CAPM

<table>
<thead>
<tr>
<th>Level</th>
<th>( n = 50 )</th>
<th>( n = 100 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.1250</td>
<td>0.1180</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0706</td>
<td>0.0598</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0182</td>
<td>0.0146</td>
</tr>
</tbody>
</table>

\( W_j, J_{CuGMM} \) and \( J_{2GMM} \) are J tests constructed through the ELR, continuously updated and two-step GMM estimators; respectively.

Table 2: Finite-Sample Size Properties - CAPM linear market model

#### 3.2.3 Power of Overidentification Tests

When drawing inferences using a given test statistic it is important to consider its power. This is the probability that the null hypothesis will be rejected given that the alternative hypothesis is true. Low power suggests that the test is not useful to discriminate between the alternative and the null hypothesis.

To document the power of a test it is necessary to specify the alternative DGP and the size of the test. In what follows we consider a DGP of the form given in (40). Under the null hypothesis \( \Delta_{ij} = 0 \) for all \( i, j \).

The experiments reported in this section set \( \Delta_{ij} \neq 0 \) in so that the moment conditions are invalid. We use the rejection frequencies as estimates of one minus the probability of Type II error.

Two main experiments are carried out. The first one considers variations in the means of the returns by setting \( \Delta_{11} \neq 0 \) for \( i = 1, 2 \). The second experiment deals with fluctuations of the variances of the returns by letting \( \Delta_{22} \neq 0 \) for \( i = 1, 2 \).

To separate the effect of size distortions we report the rejection frequencies for the cases where the critical values are given by the (estimated) .10, .05 and .01 critical values of the finite-sample null distribution.

**Size Correction** To obtain the .10, .05 and .01 finite-sample critical values we perform a Monte Carlo experiment with \( \Delta_{ij} = 0 \) for all \( i, j \). After ordering the simulated values of the overidentification tests from the largest to the smallest we find the 500\(^{th} \), 250\(^{th} \) and 50\(^{th} \) values (since 5000 replications
were performed). These values are the corrected critical values. Results for \( n = 100 \) are summarized in Table 3.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Asymptotic critical value</th>
<th>Corrected Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>2.7055</td>
<td>2.8091</td>
</tr>
<tr>
<td>.05</td>
<td>3.8414</td>
<td>4.1485</td>
</tr>
<tr>
<td>.01</td>
<td>6.6348</td>
<td>6.8828</td>
</tr>
</tbody>
</table>

\( W_j, J_{CuGMM} \text{ and } J_{2GMM} \) are \( J \) tests constructed through the ELR, continuously updated and two-step GMM estimators; respectively.

Table 3: Size Correction Mean-Variance CAPM - mexican data

Results

Experiment 1: Variations in the Means

We set \( \Delta_{i1} \neq 0 \) in (40) for \( i = 1, 2 \). It is easy to see that deviations from the null hypothesis are given by

\[
\Delta_{21} \frac{38.56}{(4.11)^2} - \Delta_{11}.
\]

If \( \Delta_{21} \) and \( \Delta_{11} \) are both positive (negative), it is ambiguous if these increments (decrements) lead to departures from the null because both effects might cancel each other. Moreover, larger \( \Delta_{i1} \)'s are not necessarily interpreted as larger deviations from the null. Hence, we concentrate on the cases in which the means vary in opposite directions: \( i.e. \) \( \Delta_{11} > 0 (\Delta_{11} < 0) \) and \( \Delta_{21} < 0 (\Delta_{21} > 0) \). Forty eight different cases are studied. The ranges of the variations are between -2 and +2: \( \Delta_{i1} = \{-2, ..., +2\} \) for \( i = 1, 2 \). Results for a significance level of 5% are shown in Table 4.

Each coordinate in Table 4 represents \( (\Delta_{11}, \Delta_{21}) \), where \( \Delta_{11} \) are changes induced to the mean of \( r_{MEX} \) and \( \Delta_{21} \) are changes induced to the mean of \( r_{mt} \). We do not experiment with the coordinate \((0,0)\) since the null holds.
### Power of Moment Restrictions Tests

Nominal Size= .05  
Mean-Variance CAPM

<table>
<thead>
<tr>
<th>$\Delta_{11}$</th>
<th>$\Delta_{21}$</th>
<th>-2</th>
<th>-1</th>
<th>-.5</th>
<th>-.2</th>
<th>0</th>
<th>.2</th>
<th>.5</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_j$</td>
<td>2</td>
<td>1</td>
<td>.973</td>
<td>.818</td>
<td>.628</td>
<td>.458</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td>1</td>
<td>.970</td>
<td>.806</td>
<td>.625</td>
<td>.437</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td>1</td>
<td>.972</td>
<td>.803</td>
<td>.595</td>
<td>.415</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td>.998</td>
<td>.852</td>
<td>.518</td>
<td>.285</td>
<td>.159</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td>.998</td>
<td>.848</td>
<td>.511</td>
<td>.277</td>
<td>.142</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td>.995</td>
<td>.833</td>
<td>.471</td>
<td>.236</td>
<td>.146</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td>.992</td>
<td>.708</td>
<td>.323</td>
<td>.148</td>
<td>.086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td>.989</td>
<td>.690</td>
<td>.310</td>
<td>.121</td>
<td>.061</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td>.986</td>
<td>.673</td>
<td>.277</td>
<td>.130</td>
<td>.054</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td>.982</td>
<td>.634</td>
<td>.242</td>
<td>.099</td>
<td>.070</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td>.980</td>
<td>.619</td>
<td>.213</td>
<td>.097</td>
<td>.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td>.978</td>
<td>.610</td>
<td>.221</td>
<td>.075</td>
<td>.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td>.975</td>
<td>.538</td>
<td>.172</td>
<td>.063</td>
<td>.084</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td>.972</td>
<td>.515</td>
<td>.163</td>
<td>.062</td>
<td>.071</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td>.970</td>
<td>.524</td>
<td>.135</td>
<td>.051</td>
<td>.067</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.061</td>
<td>.112</td>
<td>.218</td>
<td>.533</td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.051</td>
<td>.093</td>
<td>.214</td>
<td>.489</td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.044</td>
<td>.082</td>
<td>.207</td>
<td>.492</td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.094</td>
<td>.149</td>
<td>.320</td>
<td>.665</td>
<td>.964</td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.058</td>
<td>.140</td>
<td>.294</td>
<td>.631</td>
<td>.964</td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.070</td>
<td>.129</td>
<td>.283</td>
<td>.618</td>
<td>.964</td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.158</td>
<td>.207</td>
<td>.472</td>
<td>.787</td>
<td>.985</td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.148</td>
<td>.251</td>
<td>.450</td>
<td>.783</td>
<td>.979</td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.139</td>
<td>.241</td>
<td>.468</td>
<td>.763</td>
<td>.978</td>
</tr>
<tr>
<td>$W_j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.446</td>
<td>.604</td>
<td>.792</td>
<td>.953</td>
<td>1</td>
</tr>
<tr>
<td>$J_{CuGMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.436</td>
<td>.572</td>
<td>.782</td>
<td>.933</td>
<td>.997</td>
</tr>
<tr>
<td>$J_{2GMM}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.440</td>
<td>.594</td>
<td>.761</td>
<td>.947</td>
<td>.997</td>
</tr>
</tbody>
</table>

$W_j$, $J_{CuGMM}$ and $J_{2GMM}$ are J-tests based on the ELR, continuously updated and two-step GMM estimators; respectively.

Table 4: Power Properties Mean-Variance CAPM - variations in means

Our results suggest that the ELR overidentification test is more able to detect deviations from the null than tests based on the GMM. Of 48 experiments EL has the greatest power in all 48 of the cases and the two-step and continuously updated GMM are as powerful as the EL in 2 cases.

As expected, power increases as the variations in the means increase and it is also noteworthy that there are no important differences between positive and negative values of $\Delta_{11}$.

We carry out a second experiment to assess the power of overidentification tests.
Experiment 2: Variations in the Variances

We set $\Delta_{i2} \neq 0$ in (40) for $i = 1, 2$. The new expected return for the risky asset implied by the Mean-Variance CAPM is

$$E(r_{MEX}) = \frac{.65 (14.41 + \Delta_{12})}{(4.11 + \Delta_{22})} (.73).$$

However, we generate random numbers considering the original expected value

$$E(r_{MEX}) = \frac{.65 (14.41)}{(4.11)} (.73) = 1.68.$$

Note that if $\Delta_{i2}$ and $\Delta_{22}$ are both positive (negative) it is ambiguous if these increases (decreases) lead to departures from the null hypothesis. Hence, we concentrate on the cases in which the variances vary in opposite directions: $\Delta_{i2} > 0$ ($\Delta_{i2} < 0$) and $\Delta_{22} < 0$ ($\Delta_{22} > 0$).

Forty eight different cases are studied. The ranges of the variations are between -4 and +4: $\Delta_{i2} = \{-4, \ldots, +4\}$ for $i = 1, 2$. We omit $\Delta_{22} = -4$ because we encountered several problems when generating random numbers given that the variance is close to zero: $(4.11 - 4) = .11$.

Results for a significance level of 5% and $n = 100$ are shown in Table 5. We performed 5000 replications.

Our findings shown in Table 5 suggest that $W_j$ is more able to detect false moment conditions than tests based on the GMM. In the 48 experiments $W_j$ has the greatest power in all cases. $J_{2GMM}$ is as powerful as $W_j$ in 2 of the cases and $J_{CuGMM}$ is as powerful as $W_j$ in 2 of the replications. As would be expected, power increases as $\Delta_{i2}$ increases. The latter results are consistent with the findings of our first experiment. The GMM tests fail to detect the invalidity of the moment conditions in a higher extent than the ELR test does.
### Table 5: Power Properties Mean-Variance CAPM - variations in variances

<table>
<thead>
<tr>
<th>( \Delta_{12} )</th>
<th>( \Delta_{22} )</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W_j )</td>
<td>( J_{CuGMM} )</td>
<td>.956</td>
<td>.385</td>
<td>.145</td>
<td>.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( J_{2GMM} )</td>
<td>.951</td>
<td>.382</td>
<td>.121</td>
<td>.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( W_j )</td>
<td>( J_{CuGMM} )</td>
<td>.961</td>
<td>.368</td>
<td>.135</td>
<td>.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( J_{2GMM} )</td>
<td>.957</td>
<td>.382</td>
<td>.121</td>
<td>.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( W_j )</td>
<td>( J_{CuGMM} )</td>
<td>.952</td>
<td>.404</td>
<td>.126</td>
<td>.059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( J_{2GMM} )</td>
<td>.945</td>
<td>.371</td>
<td>.090</td>
<td>.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( W_j )</td>
<td>( J_{CuGMM} )</td>
<td>.936</td>
<td>.327</td>
<td>.126</td>
<td>.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( J_{2GMM} )</td>
<td>.918</td>
<td>.321</td>
<td>.101</td>
<td>.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( W_j, J_{CuGMM}, J_{2GMM} \) are J-tests based on the ELR, continuously updated and two-step GMM estimators; respectively.

### 4 Three-Moment CAPM

When contradictory empirical results for the traditional form of the Sharpe-Lintner model emerged, authors such as Kraus and Litzenberger (1976) extended the Mean-Variance framework to incorporate the effect of skewness on valuation. They argue that prior empirical findings that were interpreted as inconsistent with the traditional theory can be attributed to misspecification of the CAPM by omission of systematic skewness.

By setting \( b_3 = 0 \) in (26), the K-L CAPM follows.
It is crucial to address the fact that the market price of beta reduction, $b_1$, and the market price of gamma, $b_2$, can be expressed in terms of the market’s return. To illustrate this assertion consider the special case in which all investors have logarithmic utility functions.\textsuperscript{18}

A Taylor approximation of the investor’s expected utility of end of period wealth, $E[U(\omega)] = f(E(\omega), \sigma(\omega), \gamma(\omega))$, yields:

$$E[U(\omega)] = \log(E(\omega)) - \frac{\sigma(\omega)^2}{2E(\omega)^2} + \frac{\gamma(\omega)^3}{3E(\omega)^3}. \quad (42)$$

When we differentiate (42) with respect to $E(\omega)$, $\sigma(\omega)$ and $\gamma(\omega)$ we obtain

$$\begin{align*}
\frac{\partial E[U(\omega)]}{\partial E(\omega)} &= \frac{1}{E(\omega)} + \frac{\sigma(\omega)^2}{E(\omega)^2} - \frac{\gamma(\omega)^3}{E(\omega)^3}, \quad (43) \\
\frac{\partial E[U(\omega)]}{\partial \sigma(\omega)} &= -\frac{\sigma(\omega)}{E(\omega)^2}, \quad (44) \\
\frac{\partial E[U(\omega)]}{\partial \gamma(\omega)} &= \frac{\gamma(\omega)^2}{E(\omega)^3}. \quad (45)
\end{align*}$$

Substitution of (43), (44) and (45) into (27) and (28) yields

$$\begin{align*}
b_1 &= \left(\frac{-\frac{\sigma(\omega)}{E(\omega)}}{1 + \frac{\sigma(\omega)^2}{E(\omega)^2} - \frac{\gamma(\omega)^3}{E(\omega)^3}}\right) \sigma(r_m), \quad (46) \\
b_2 &= \left(\frac{-\frac{\gamma(\omega)^2}{E(\omega)^3}}{1 + \frac{\sigma(\omega)^2}{E(\omega)^2} - \frac{\gamma(\omega)^3}{E(\omega)^3}}\right) \gamma(r_m). \quad (47)
\end{align*}$$

Since the initial investment is set to one, the moments of end of period wealth are equivalent to those of the rate of return on the portfolio, in equilibrium the market portfolio. Therefore, we can rewrite $b_1$ and $b_2$ as

$$\begin{align*}
b_1 &= \frac{\sigma(r_m)^2}{E(r_{mt})} \cdot \frac{1 + \sigma(r_m)^2}{E(r_{mt})^2 - \gamma(r_m)^3}, \quad (48) \\
b_2 &= \frac{\gamma(r_m)^3}{E(r_{mt})^2} \cdot \frac{1 + \sigma(r_m)^2}{E(r_{mt})^2 - \gamma(r_m)^3}. \quad (49)
\end{align*}$$

\textsuperscript{18}The logarithmic function is representative of utility functions displaying decreasing absolute risk aversion and constant relative risk aversion.
Note that as soon as information about the investor’s preferences is incorporated, \( b_1 \) and \( b_2 \) can be expressed in terms of the market.

Thus, the special case of K-L CAPM, where all investors have logarithmic utility functions, is

\[
E(r_t) = \left[ \frac{-\sigma(r_m)^2}{E(r_{mt})^2} \right] \beta_m + \left[ \frac{\gamma(r_m)^2}{E(r_{mt})^2} \right] \gamma_m. (50)
\]

We can alternatively use a variant of the K-L CAPM that provides information about the structure of the risk premiums, \( b_1 \) and \( b_2 \), by using the Euler condition for the investor’s utility maximization problem as in Seirs and Wei (1985). This is

\[
E(r_t) = \left( \frac{\phi \sigma(r_m)}{\phi \sigma(r_m) - \gamma(r_m)} \right) \beta_m - \left( \frac{\gamma(r_m)}{\phi \sigma(r_m) - \gamma(r_m)} \right) \gamma_m r_{mt}, (51)
\]

where \( \phi \) is the marginal rate of substitution of \( \gamma \) for \( \sigma \) (refer to Seirs and Wei, 1985). Note that as for the logarithmic utility case, \( b_1 \) and \( b_2 \) are now expressed in terms of the market return.

4.1 Moment Equations

The orthogonality conditions that characterize the Three-Moment CAPM are given by Equations (30), (31), (32), (34), (35) and (36); where we set \( b_3 = 0 \) in (30).

4.2 Finite-Sample Properties of Overidentification Tests

The tests of overidentifying restrictions studied in this section have as their null hypothesis that there is a value of \( \theta \) consistent with \( E[g(r_t, r_{mt}, \theta)] = 0 \); where \( \theta = (b_1, b_2, \beta_m, \gamma_m, \mu(r_m), \sigma(r_m), \gamma(r_m)) \).

We again consider the three tests of overidentifying restrictions studied in the Mean-Variance setting: \( W_j, J_{2GMM} \) and \( J_{CuGMM} \). These tests have a \( \chi^2(r-q) \) distribution under the null, where \( \dim(g) = r \) and \( \dim(\theta) = q \).

4.2.1 The Data Generating Process

Assume the following quadratic market model:

\[
r_t = a_1 r_{mt} + a_2 (r_{mt} - E(r_{mt}))^2 + \varepsilon_t, (52)
\]
where

(i) \( a_i \neq 0 \) for \( i = \{1, 2\} \),
(ii) \( \varepsilon_t \) is independent of \( r_{mt} \) and \( (r_{mt} - E(r_{mt}))^2 \),
(iii) \( E(\varepsilon_t) = 0 + \Delta \) and \( \Delta = 0 \).

Then applying the definitions of \( \beta_m \) and \( \gamma_m \) to the quadratic market model we obtain:

\[
\beta_m = a_1 + a_2 \frac{\gamma (r_m)^3}{\sigma (r_m)^2},
\]

\[
\gamma_m = a_1 + a_2 \frac{(\sigma (r_m)^4 - \sigma (r_m)^4)}{\gamma (r_m)^3}.
\]

It is helpful to seek to express \( a_1 \) and \( a_2 \) in terms of \( \beta_m \) and \( \gamma_m \). Solving (54) and (55) for \( a_1 \) and \( a_2 \) yields

\[
a_1 = \frac{-\beta_m \sigma (r_m)^2 + \beta_m \sigma (r_m)^6 + \gamma (r_m)^6 \gamma_m}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6},
\]

\[
a_2 = \frac{\sigma (r_m)^2 \gamma (r_m)^3 (-\gamma_m + \beta_m) + \gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6}.
\]

Thus, the DGP in (52) can be rewritten as:

\[
r_t = \left[ \frac{-\beta_m \sigma (r_m)^2 + \beta_m \sigma (r_m)^6 + \gamma (r_m)^6 \gamma_m}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} \right] r_{mt}
\]

\[
+ \left[ \frac{\sigma (r_m)^2 \gamma (r_m)^3 (-\gamma_m + \beta_m)}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} \right] (r_{mt} - E(r_{mt}))^2 + \varepsilon_t.
\]

Factorizing \( \beta_m \) and \( \gamma_m \) yields

\[
r_t = A_1 \beta_m + A_2 \gamma_m + \varepsilon_t,
\]

where

\[
A_1 = \frac{-\vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} r_{mt}
\]

\[
+ \frac{\sigma (r_m)^2 \gamma (r_m)^3}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} (r_{mt} - E(r_{mt}))^2,
\]

\footnote{The proofs of (54) and (55) are in Appendix 6.2.}
\[ A_2 = \frac{\gamma (r_m)^6}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} r_{mt} \]

\[ - \frac{\sigma (r_m)^2 \gamma (r_m)^3}{\gamma (r_m)^6 - \vartheta (r_m)^4 \sigma (r_m)^2 + \sigma (r_m)^6} (r_{mt} - E(r_{mt}))^2. \]

Note that the proposed quadratic market model, Equation (52), has been rewritten as (59). Here, we have two terms: one which factorizes beta and one which factorizes systematic kurtosis.

For simplicity, we consider a specification of the K-L CAPM that provides information about the structure of the risk premiums\(^{20}\). From this point onwards we focus on the model given in (51). We introduce a normalization variable to generate data consistent with this framework. Let \( \phi \) in (51) be a normalization variable such that

\[ \int \int \left\{ \left( r_t - \left[ \left\{ \frac{\phi \sigma (r_m)}{\phi \sigma (r_m) - \gamma (r_m)} \right\} \beta_m \right. \right. \right. \]

\[ \left. \left. \left. - \left\{ \frac{\sigma (r_m)^2 \gamma (r_m)^3}{\phi \sigma (r_m) - \gamma (r_m)} \right\} \gamma_m \right\} r_{mt} \right\} f_{r_t, r_{mt}} \right\} dr_{mt} dr_t = 0, \]

where \( f_{r_t, r_{mt}} \) is the joint density function of the risky and market returns.

Consider a DGP of the form given in (52) where the conditions in (53) hold. Let \( r_{mt} \sim \chi^2_k \) and \( \varepsilon_t \sim N(0,1) \). Then:

\[ E(r_{mt}) = k, \]
\[ \sigma^2(r_m) = 2k, \]
\[ \gamma^3(r_m) = 8k, \]
\[ \vartheta^4(r_m) = 12k(k + 4). \]

Substitution of these values into (54) and (55) yields:

\[ \beta_m = a_1 + 4a_2, \]
\[ \gamma_m = a_1 + a_2(6 + k). \]

\(^{20}\) The advantage of expressing \( b_1 \) and \( b_2 \) in terms of the market is that these are no longer parameters to be estimated.
Hence, we can rewrite (62) as:

\[
\int \int \left\{ \left( \frac{d}{dr_{mt}} \left[ \left( a_1 r_{mt} + a_2 (r_{mt} - k)^2 \right)^2 + \varepsilon_t \right] - \left[ \frac{\phi(2k)^{1/2}}{\phi(2k)^{1/2} - (8k)^{1/3}} \right] (a_1 + 4a_2) \right] \right\} dr_{mt} d\varepsilon_t = 0,
\]

where \( f_{r_{mt}} \) is the chi-square marginal density of \( r_{mt} \) and \( f_{\varepsilon_t} \) is the standard normal marginal density function of \( \varepsilon_t \).

After some simplification we obtain:

\[
\phi = \frac{(.177 a_1 + 1.06 a_2 - .177 k (a_1 + a_2))}{k^{1/6} \left( .516 \times 10^{-11} a_2 k^2 + .125 a_1 + .501 a_2 - .125 ka_1 - .25 a_2 k \right)}.
\]

To generate \( r_t \) as in (52) we must specify the degrees of freedom for \( r_{mt} \) and set values for \( a_1 \) and \( a_2 \). Before doing this we review a key point. In Section 2 we define the market price of beta, \( b_1 \), and the systematic skewness, \( b_2 \), as:

(i) \( b_1 = \left[ \frac{dE(\omega)}{d\sigma(\omega)} \right] \sigma (r_m) \) where \( b_1 > 0 \),

(ii) \( b_2 = \left[ \frac{dE(\omega)}{d\gamma(\omega)} \right] \gamma (r_m) \) where \( b_2 < 0 \) if \( \gamma (r_m) > 0 \) and \( b_2 > 0 \) if \( \gamma (r_m) < 0 \).

It is easy to see from (26) and (51) that

\[
b_1 = \frac{\sigma(r_m)}{\phi(r_m) - \gamma(r_m)} \quad \text{and} \quad b_2 = -\frac{\gamma(r_m)}{\phi(r_m) - \gamma(r_m)}.
\]

Substituting \( a_1 = 1.5, a_2 = .5 \) and \( k = 1 \) into (64) yields \( \phi = 3.53 \). Since \( \gamma (r_m) > 0 \) we only need that \( \phi \sigma (r_m) - \gamma (r_m) > 0 \) so that (i) and (ii) hold, which is actually true for these values.

### 4.2.2 Size of Overidentification Tests

In this section we examine whether the asymptotic (or nominal) size is a good approximation to the size in finite-samples. The following experiment uses the already defined DGP; i.e.:

\[
r_t = 1.5 r_{mt} + .5 (r_{mt} - E(r_{mt}))^2 + \varepsilon_t,
\]

where \( r_{mt} \sim \chi^2(1) \) and \( \varepsilon_t \sim N(0, 1) \).

Note that for the K-L CAPM specification that we study, (51) replaces (30) so that the moment equations are (51), (31), (32), (34), (35) and (36) with \( \phi = 3.53 \). Thus \( W_j, J_{2GMM} \) and \( J_{CuGMM} \)
have an asymptotic $\chi^2_{(1)}$ distribution under the null. We compute 5000 replications for two sample sizes: $n = \{50, 100\}$. The empirical levels of the J-tests are reported in Table 6.

<table>
<thead>
<tr>
<th>Level</th>
<th>$W_j$</th>
<th>$J_{CuGMM}$</th>
<th>$J_{2GMM}$</th>
<th>$W_j$</th>
<th>$J_{CuGMM}$</th>
<th>$J_{2GMM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>.1160</td>
<td>.1114</td>
<td>.1225</td>
<td>.1020</td>
<td>.0932</td>
<td>.1110</td>
</tr>
<tr>
<td>.05</td>
<td>.0640</td>
<td>.0658</td>
<td>.0705</td>
<td>.0510</td>
<td>.0436</td>
<td>.0620</td>
</tr>
<tr>
<td>.01</td>
<td>.0210</td>
<td>.0238</td>
<td>.0222</td>
<td>.0180</td>
<td>.0096</td>
<td>.0114</td>
</tr>
</tbody>
</table>

$W_j$, $J_{CuGMM}$ and $J_{2GMM}$ are J-tests based on the ELR, continuously updated and two-step GMM estimators, respectively.

Table 6: Finite-Sample Size Properties - Three-Moment CAPM

**Results** Table 6 shows that for both sample sizes the rejection probabilities of the three tests are close to their nominal levels. As $n$ increases size distortions tend to decrease.

Given these results, the size properties cannot be used as a criterion for choosing among the overidentification tests that we study. At this point, it is natural to prefer the test whose power is closer to unity. We investigate power properties of J-tests in the following section.

4.2.3 Power of Overidentification Tests

To make the power (percentage of rejections under the alternative hypothesis) of different test procedures comparable we calculate exact 10%, 5% and 1% critical values from the experiment conducted in the previous Section. These size corrected critical values are used, thus making the power of different test procedures comparable.

To examine power, we concentrate on the following cases:

1. Let $\Delta \neq 0$ in (53), so that $E(\varepsilon_1) \neq 0$.

2. Let $a_2 = 0$ in (52), so that the model is overidentified\(^{21}\).

\(^{21}\)Note that a linear market model is consistent with the Mean-Variance framework whereas a quadratic market model is consistent with the Three-Moment CAPM.
Size Correction  To obtain the .10, .05 and .01 finite-sample critical values we use Monte Carlo simulations. After ordering the simulated values of the overidentification tests from the largest to the smallest we find the 500th, 250th and 50th values (since 5000 replications were performed). These values are the corrected critical values. Results for \( n = 50 \) and \( n = 100 \) are summarized in Table 7.

From our results in Table 7 we note that for \( n = 100 \) the \( \chi^2_{(1)} \) is a good approximation to the finite-sample distribution of the three test statistics. Hence, using asymptotic critical values for this sample size to assess power seems a safe undertake.

<table>
<thead>
<tr>
<th>Level</th>
<th>Asymptotic critical value</th>
<th>Corrected Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>2.7055</td>
<td>3.0104 2.9204 3.2271</td>
</tr>
<tr>
<td>.05</td>
<td>3.8414</td>
<td>4.6525 4.4869 4.9780</td>
</tr>
</tbody>
</table>

\( W_j, J_{CuGMM} \) and \( J_{2GMM} \) are J-tests based on the ELR, continuously updated and two-step GMM, respectively. \( n \) is the sample size.

Table 7: Size Correction - Three-Moment CAPM

Results

Experiment 1: Variations in the error term

We set \( \Delta \neq 0 \) in (53) so that the moment conditions of the Three-Moment CAPM are invalid. Eight departures from the null are considered and the ranges of the variations are between -1 and +1: \( \Delta = \{ -1, ..., +1 \} \).

We calculate the rejection frequencies as estimates of one minus the probability of Type II error at the
nominal .10, .05 and .01 critical values. 5000 replications were used and two sample sizes considered: 

\( n = \{50, 100\} \). Results are reported in Table 8.

<table>
<thead>
<tr>
<th>( \Delta )</th>
<th>( W_j ) Levels</th>
<th>( J_{CuGMM} ) Levels</th>
<th>( J_{2GMM} ) Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>.878</td>
<td>.374</td>
</tr>
<tr>
<td>-.5</td>
<td>.902</td>
<td>.701</td>
<td>.231</td>
</tr>
<tr>
<td>-.2</td>
<td>.444</td>
<td>.201</td>
<td>.080</td>
</tr>
<tr>
<td>-.1</td>
<td>.200</td>
<td>.073</td>
<td>.041</td>
</tr>
<tr>
<td>.1</td>
<td>.358</td>
<td>.088</td>
<td>.021</td>
</tr>
<tr>
<td>.2</td>
<td>.182</td>
<td>.111</td>
<td>.020</td>
</tr>
<tr>
<td>.5</td>
<td>.301</td>
<td>.300</td>
<td>.201</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.706</td>
<td>.489</td>
</tr>
</tbody>
</table>

| \( n = 50 \) |
|-----|--------|--------|--------|
| -1  | 1      | 1      | 1      | .965   | .941   | .803   | .990   | .950   | .921   |
| -.5 | .926   | .768   | .412   | .779   | .649   | .318   | .879   | .701   | .361   |
| -.2 | .575   | .351   | .118   | .288   | .171   | .037   | .442   | .272   | .069   |
| -.1 | .183   | .107   | .039   | .151   | .081   | .016   | .166   | .093   | .021   |
| .1  | .127   | .073   | .027   | .103   | .058   | .020   | .100   | .052   | .018   |
| .2  | .149   | .121   | .058   | .156   | .111   | .049   | .162   | .131   | .041   |
| .5  | .311   | .309   | .367   | .433   | .375   | .250   | .301   | .300   | .190   |
| 1   | 1      | 1      | 1      | .780   | .744   | .643   | .711   | .691   | .611   |

**Table 8:** Power Properties Three-Moment CAPM - variation in the error term

In most of the cases power increases as the departures from the null increase. We observe that for all the cases that we examine, the ELR test performs better than the GMM tests. Intriguingly, power is higher for negative departures from the null hypothesis than for positive deviations.

These results are new in this kind of literature. Kitamura (2001) found that the power of ELR was greater than that of GMM tests when power was already high. Here, we find that the power of ELR is uniformly better.
Experiment 2: Three-Moment CAPM versus Mean-Variance CAPM

The null and alternative hypothesis that we consider are:

\[ H_0 : \text{The Three – Moment CAPM is valid} \]

\[ H_a : \text{The Mean – Variance CAPM is valid.} \]

We have already shown that while linear market characteristic lines are consistent with the Mean-Variance CAPM, quadratic market lines characterize the Three-Moment CAPM. Hence, if we set \( \alpha_2 = 0 \) in (52) then the model characterized by (51), (31), (32), (34), (35) and (36) is overidentified.

We perform 5000 replications and calculate the rejection frequencies as estimates of one minus the probability of Type II error at the nominal .10, .05 and .01 critical values for two sample sizes: \( n = 50 \) and \( n = 100 \). Results are reported in Table 9.

<table>
<thead>
<tr>
<th>Power of Moment Restriction Tests</th>
<th>Three-Moment CAPM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n=50 )</td>
</tr>
<tr>
<td>( W_j )</td>
<td>.4912</td>
</tr>
<tr>
<td>( J_{CuGMM} )</td>
<td>.3416</td>
</tr>
<tr>
<td>( J_{2GMM} )</td>
<td>.3754</td>
</tr>
</tbody>
</table>

\( W_j \), \( J_{CuGMM} \) and \( J_{2GMM} \) are J tests based on the ELR, continuously updated and two-step GMM; respectively.

Table 9: Power Properties Three-Moment CAPM vs. Mean-Variance CAPM

The ranking among tests show that the EL performs better than GMM tests. Throughout this experiment design, \( W_j \) has the highest power for both sample sizes. For most of the cases that we study, \( J_{CuGMM} \) has the lowest power.

5 Conclusions

We compared the finite-sample size properties of overidentification tests based on EL and GMM within two variants of the CAPM. While there is a large amount of literature on the GMM that uses a Two-Moment framework to examine size and power of its overidentifying restrictions tests, there are no
studies which use a higher moment setting. The finite-sample properties of the J-test based on the EL has not been previously assessed in the asset pricing literature. In addition, little is known about its power properties in general.

Our experiments show that there are no clear advantages in terms of size when the GMM overidentification tests are compared to those based on EL within a Two-Moment and Three-Moment setting. The three tests have moderate size distortions. However, our findings illustrate that the ELR overidentification statistic is more powerful in detecting deviations from the null under the alternatives that we analysed. We also found some evidence that this statistic has uniformly greater power than tests based on GMM whereas Kitamura (2001) shows that the ELR J-test has better power when this is already high.

When we compared the power of overidentification tests within the Three-Moment framework, we tested against the alternative of overidentifying orthogonality conditions instead of false moment equations (model misspecification). We are not aware of any other study, at least in the asset pricing literature, which assesses the power properties of tests of overidentifying restrictions using the former interpretation of these tests.

Our results illustrate that EL may be a good alternative to GMM estimation within the area of asset pricing.

6 Appendices

6.1 Appendix 1

Since the initial investment is set to one, the moments of end of period wealth are equivalent to those of the rate of return on the portfolio, i.e.:

\[ \sigma(\omega) = \sigma(R_p), \]
\[ \gamma(\omega) = \gamma(R_p), \]
\[ \theta(\omega) = \theta(R_p). \]
Using $\sum x_i R_i = R_p - x_0 R_f$ and $\sum x_i E(R_i) = E(R_p) - x_0 R_f$ gives

$$\sum x_i \beta_{ip} = \sum x_i \frac{E \left[ \{ R_i - E(R_i) \} \{ R_p - E(R_p) \} \right]}{E \left[ \{ R_p - E(R_p) \}^2 \right]}$$

$$= \frac{E \left[ \left\{ \sum x_i R_i - \sum x_i E(R_i) \right\} \{ R_p - E(R_p) \} \right]}{E \left[ \{ R_p - E(R_p) \}^2 \right]}$$

$$= 1.$$  

Therefore,

$$\sigma(\omega) = \sum x_i \beta_{ip} \sigma(R_p).$$

Following the same procedure we obtain:

$$\sum x_i \gamma_{ip} = 1,$$
$$\sum x_i \theta_{ip} = 1,$$

which leads to

$$\gamma(\omega) = \sum x_i \gamma_{ip} \gamma(R_p) \text{ and } \theta(\omega) = \sum x_i \theta_{ip} \theta(R_p).$$

6.2 Appendix 2

We define $\beta_m$ as

$$\beta_m = \frac{\text{Cov}(r_t, r_{mt})}{\text{Var}(r_{mt})}.$$  

Substituting the proposed DGP into $\beta_m$ yields

$$\beta_m = \frac{\text{Cov}(a_1 r_{mt} + a_2 (r_{mt} - E(r_{mt}))^2 + \varepsilon_t , r_{mt})}{\text{Var}(r_{mt})}$$

$$= \frac{a_1 \text{Cov}(r_{mt}, r_{mt})}{\text{Var}(r_{mt})} + \frac{a_2 \left[ \text{Cov}(r_{mt}^2, r_{mt}) - 2 E(r_{mt}) \text{Cov}(r_{mt}, r_{mt}) \right]}{\text{Var}(r_{mt})}$$

$$= a_1 + a_2 \frac{E \left( r_{mt}^3 - 3E(r_{mt}) E(r_{mt}^2) + 2(E(r_{mt}))^3 \right)}{\text{Var}(r_{mt})}$$

$$= a_1 + a_2 \frac{\gamma(r_m)^3}{\sigma(r_m)^2}.$$  

We define $\gamma_m$ as

$$\gamma_m = \frac{\text{Cov}(r_t, r_{mt}) - 2 E(r_{mt}) \text{Cov}(r_t, r_{mt})}{\gamma(r_{mt})^3}.$$  

35
Substituting the proposed DGP into $\gamma_m$ yields

$$
\gamma_m = \frac{\text{Cov} \left( a_1 r_{mt} + a_2 (r_{mt} - E(r_{mt})^2 + \varepsilon_t, r_{mt} \right) - 2E(r_{mt}) [a_1 (E(r_{mt}) - E(r_{mt})) + a_2] \gamma (r_{mt})^3}{\gamma (r_{mt})^3} = a_1 + a_2 \frac{E(r_{mt}^4) - (E(r_{mt}))^2 - 2E(r_{mt}) [E(r_{mt}^3) - (E(r_{mt}))(E(r_{mt})^2)]}{\gamma (r_{mt})^3} - a_2 \frac{2E(r_{mt}) E(r_{mt}^3) + 6 (E(r_{mt}))^2 E(r_{mt}^2) - 4 (E(r_{mt}))^4}{\gamma (r_{mt})^3}.
$$

After rearranging:

$$
\gamma_m = a_1 + a_2 \frac{E(r_{mt}^4) - (E(r_{mt}))^2 - 4E(r_{mt}) E(r_{mt}^3)}{\gamma (r_{mt})^3} + a_2 \frac{8 (E(r_{mt}))^2 E(r_{mt}^2) - 4 (E(r_{mt}))^4}{\gamma (r_{mt})^3}.
$$

To simplify this expression we note that

$$
E \left[ (r_{mt} - E(r_{mt}))^4 \right] - E \left[ (r_{mt} - E(r_{mt}))^2 \right]^2 = E \left( r_{mt}^4 \right) - (E(r_{mt}))^2 - 4E(r_{mt}) E(r_{mt}^3) + 8 (E(r_{mt}))^2 E(r_{mt}^2) - 4 (E(r_{mt}))^4.
$$

Hence,

$$
\gamma_m = a_1 + a_2 \frac{\varphi(r_{mt})^4 - \sigma(r_{mt})^2}{\gamma(r_{mt})^3}.
$$

References


