
Original Article

A new empirical version of the Fama and French model based on the Hausman specification test: An application to hedge funds

Received (in revised form): 4th January 2010

François-Éric Racicot

is Professor of Finance at the Department of Administrative Sciences, University of Quebec – Outaouais (UQO). His research interests focus on the problems of measurement errors and specification errors in financial models of returns. He is also interested in developing new methods used for forecasting financial time series. He is associate researcher at the Chaire d'information financière et organisationnelle, ESG-UQAM and a permanent member at the LRSP.

Raymond Théoret

is Professor of Finance at the Department of Finance, University of Quebec – Montreal (UQAM). His research focuses on banking risk and accruals in relation with stock market forecasts, self-enforcing labour contracts and problems of measurement errors in financial models. He is an associate researcher at the Chaire d'information financière et organisationnelle, ESG-UQAM.

Alain Coën

is Associate Professor of Finance at the Graduate School of Business (ESG) of the University of Quebec in Montreal (UQAM). He obtained his PhD in Finance from the University of Grenoble, and his PhD in Economics from the University of Paris I Panthéon-Sorbonne. He holds a Master of Arts in Economics with major in Macroeconomics from Laval University and an Accreditation to supervise research (HDR) from Paris-Dauphine University. His research interests focus on asset pricing, international finance, hedge funds, business cycles and financial econometrics. He has published in several international journals and has written a book in financial management.

Correspondence: François-Éric Racicot, University of Quebec – Outaouais, (UQO), 101 St-Jean-Bosco Street, Lucien-Brault Building, Gatineau (Hull), Québec, Canada J8X 3X7
E-mail: francoiseric.racicot@uqo.ca

ABSTRACT In this article, we propose a new empirical version of the Fama and French (F&F) model based on a new version of the Hausman (1978) specification test. This empirical model incorporates correction factors for risk exposure. These factors take into account the problem of errors-in-variables. Our model also features another innovation. It uses higher moments of the four factors of the augmented F&F model as instruments to tackle the problem of errors-in-variables. This new method reflects the nonlinear character of risk. We highlight the link between our new version of the Hausman artificial regression and the standard two-stage least-squares method. Our results suggest that it is preferable to account for errors-in-variables and more generally specification errors when estimating the F&F model because it might cause serious biases of important parameters like alpha and beta.

Journal of Derivatives & Hedge Funds (2011) 16, 278–302. doi:10.1057/jdhf.2010.22

Keywords: asset pricing; portfolio selection; errors-in-variables; higher moments; instrumental variables; specification tests

INTRODUCTION

Financial returns, and especially hedge funds returns, are contaminated by errors-in-variables. If neglected, this problem may completely invalidate the results obtained when estimating financial models, like the market model or the Fama and French (F&F)¹⁻³ one or any standard linear factor model. Nevertheless, researchers seldom tackle the problem of errors-in-variables in their empirical works, perhaps because the econometric methods related to the correction of this problem are not so advanced as other fields of financial econometrics. Some papers⁴ have proposed correction techniques for errors-in-variables in a capital asset pricing model (CAPM) setting, but these studies are not numerous and much work remains to be done.

In this article, we develop a new empirical version of the F&F model based on a new version of the Hausman⁵ specification test.⁶ This empirical model incorporates correction factors for risk exposure. These factors deal with the problem of errors-in-variables. Our model also includes another innovation. It uses higher moments as instruments for correcting the problem of errors-in-variables. In fact, the methods of financial econometrics were traditionally well suited to estimate financial models, which, like the CAPM, postulate a linear relationship between return and risk. But this kind of relation is valid only if risk is not very high or under very special assumptions as Gaussian returns or quadratic utility function, which is not very realistic because it implies increasing absolute risk aversion. Investor prudence,⁷ which is associated to the utility

function third derivative, and fat-tail risk, which is related to the fourth derivative, introduce nonlinearities in the relation between return and risk. In this article, we reformulate the problem of correcting errors-in-variables in the setting of a nonlinear relation between return and risk. Higher-order moments and cumulants of returns become very important in this analysis in relation to nonlinear risk (Racicot and Théoret^{8,9}).

According to Campbell *et al.*,¹⁰ the errors-in-variables problem in asset pricing models can be addressed in two ways. The first way (Fama and MacBeth¹¹) is to reduce this problem by pooling stocks into portfolios. The second way (Shanken¹²) is to explicitly adjust the coefficient standard errors to reduce the biases resulting from the errors-in-variables. However, there is a more recent approach proposed by Kandel and Stambaugh¹³ using generalized least squares instead of ordinary least squares (OLS) as estimator. But, in this approach, the covariance matrix required to weight the observations must be estimated, which might be a problem.

Durbin,¹⁴ Pal¹⁵ and more recently Racicot (1993) and Dagenais and Dagenais^{16,17} have proposed an estimation method to tackle the problem of errors-in-variables based on an optimal combination of estimators built on the cumulants of the explanatory variables. These cumulants are used as instruments to reduce the problem of errors-in-variables. In this article, we use a variant of this method to correct the errors-in-variables problem, which might bias the estimation of the well-known augmented F&F model and which is based on higher moments.

As shown in this article, using higher moments to tackle errors-in-variables opens the way to a synthesis between the modern asset pricing theory and the financial econometric treatment of errors-in-variables. Surely, Durbin,¹⁴ Pal,¹⁵ Racicot¹⁸ and Dagenais and Dagenais^{16,17} did not aim at transposing their technique to the asset pricing theory and risk measures. But it is widely recognized now that the first two moments of returns, that is, the mean and the variance, are largely insufficient to measure the risk of a portfolio. Huang and Litzenberger,¹⁹ Ingersoll²⁰ and Levy²¹ have mentioned that portfolio selection paradigm based on the first two moments of the returns distribution maximizes the expected utility of a representative agent only in either of the two following situations: his utility function is quadratic or the returns distribution is normal. Of course, these two postulates are violated in the real world. Following this 'constat', Samuelson,²² Rubinstein,²³ Kraus and Litzenberger,²⁴ Friend and Westerfield²⁵ and Sears and Wei²⁶ have laid the foundations of the approach based on higher moments for pricing financial instruments. These theoretical developments gave birth to the three-moment and the four-moment CAPM.²³

The theoretical developments related to the use of higher moments as measures of financial risk are linked to other sections of risk theory, but a synthesis is yet to do. The theory of stochastic dominance²⁷ has a long history. The stochastic dominance theory considers an increasing amount of higher moments of a return distribution to evaluate whether a portfolio is superior to another. This theory has delivered new risk measures like the risk of shortfall, which is based on the returns probability distribution and transfers of probability masses between

high-wealth states of nature and low-wealth states of nature. In the same line of ideas, Scott and Hovarth²⁸ argue that the odd moments of the returns distribution, like a positive mean and a positive skewness, provide positive marginal utility to an investor. The positive even moments, like variance and kurtosis, entail negative marginal utility. These developments lay the foundations of the modern theory of risk, which is under construction.

The new empirical version of the F&F model, based on the Hausman specification test we propose, is in line with the new risk literature. Our contribution is twofold. First, we add factors adjusting or correcting for risk exposure in the F&F model, this correction being required by a potential problem of errors-in-variables, which will be detected by our new version of the Hausman specification test. Next, we use as instruments the higher moments of the regressors in the F&F equation. These instruments not only serve as technical tools, but also as risk measures that implement our estimation method in the modern theory of risk analysis, which is essentially nonlinear. As we will observe, our procedure has another advantage as the regressors (factors) of the F&F equation are long-short portfolios, which are consequently similar to those of hedge funds. Indeed, these long-short portfolios are mimicking market anomalies, and thus they are long in a category of returns (for example, returns of small firms) and short in the opposite category (for example, returns of big firms). Consequently, they behave as options portfolios that incorporate many nonlinearities. Higher-order moments are therefore required to account for these nonlinearities. Taleb²⁹ even suggests using moments of order higher than four to measure the risk of an option, odd moments

being measures of asymmetry and even moments being measures of convexity.³⁰ The empirical work shows that our choice of instruments is judicious. In fact, the correlation of a regressor with moments higher than two is often much more important than the correlation with conventional inferior moments.

This article is organized as follows. The methodology used to carry our empirical analysis is explained in the next section. First, we discuss the theoretical aspects related to errors-in-variables and to the choice of instruments, which consist in the higher moments of the regressors of the F&F model. Second, the estimation method used to correct the problem of errors-in-variables in the F&F model is presented. That gives rise to the development of our new empirical version of the F&F model, which features a new two-stage least-squares estimator incorporating a Hausman specification test using an artificial regression. The subsequent section describes the data series used in this study and reports the empirical results. Concluding remarks are given in the final section.

METHODOLOGY

The biases caused by errors-in-variables

The errors-in-variables problem is well-known in the econometric literature, but is often overlooked in finance. Assume³¹ a regression model where observed y and x variables are measured with errors. The true values of these variables are y^* and x^* . The relations between the observed variables and their true counterparts are:

$$y = y^* + u \quad (1)$$

$$x = x^* + v \quad (2)$$

where u and v are error vectors. These error vectors have zero means and their variances are: $E(uu') = \sigma_u^2 I$ and $E(vv') = \sigma_v^2 I$. The vectors u and v are assumed orthogonal.

The exact linear relationship between the two unobserved variables is:

$$y^* = x^* \beta \quad (3)$$

As y^* and x^* are unobservable, we substitute equations (1) and (2) in equation (3):

$$y = x\beta + \varepsilon \quad (4)$$

where $\varepsilon = v - u\beta$. Consequently, by equation (2), x is correlated with the error term ε , and this creates a bias. To compute it, we solve equation (4) for $\hat{\beta}$, the estimated value of parameter β , by the method of least squares:

$$\hat{\beta} = (x'x)^{-1} x'y \quad (5)$$

Substituting equation (4) in equation (5), we obtain:

$$\hat{\beta} = \beta + (x'x)^{-1} x'e \quad (6)$$

Equation (6) shows that $\hat{\beta}$ is biased. And in large samples, it is not consistent:

$$p \lim \hat{\beta} = \beta \left[1 - \frac{\sigma_u^2}{\sigma_x^2} \right] \quad (7)$$

where $\sigma_x^2 = \sigma_{x^*}^2 + \sigma_u^2$. We can rewrite expression (7) as:

$$p \lim \hat{\beta} = \beta(1 - \lambda) \quad (8)$$

where $\lambda < 1$.

According to equation (8), the problem of errors-in-variables tends to underestimate $\hat{\beta}$, the degree of underestimation being related to λ . The more λ is near 1, the more serious is the problem of errors-in-variables. At the limit, λ is 1 and $p \lim \hat{\beta} = 0$. When only y is plagued with errors of measurement in equation (1), there is

no bias because x remains uncorrelated with the innovation of the equation. The problem appears when x is measured with error. This creates a correlation between x and the innovation term and therefore a bias.

Unfortunately, if there is more than one explanatory variable in a model, we cannot know *a priori* the relative impact of errors-in-variables on the estimation process. Some parameters will be overstated and others understated. But as seen in another section, the Hausman test, and more precisely the version of this test based on artificial regressions, will not only help us to detect errors in variables, but also give us more information about the incidence of this problem on estimated parameters.

The choice of instrumental variables to estimate the augmented F&F model

The augmented F&F (1992, 1993 and 1997) model is a purely empirical model that may be written as:

$$R_{pt} - R_{ft} = \alpha + \beta_1(R_{mt} - R_{ft}) + \beta_2SMB_t + \beta_3HML_t + \beta_4UMD_t + \varepsilon_t \quad (9)$$

with:

- $R_{pt} - R_{ft}$ the excess return of a portfolio, R_{ft} being the risk-free return;
- $R_{mt} - R_{ft}$ the market risk premium;
- SMB* a portfolio that mimics the ‘small firm anomaly’, which is long in the returns of selected small firms and short in the returns of selected big firms;
- HML* a portfolio that mimics the ‘income stock anomaly’, which is long in returns of stocks of selected firms having a high (book value/market

value) ratio (income stocks) and short in selected stocks having a low (book value/market value) ratio (growth stocks);

UMD a portfolio that mimics the ‘momentum anomaly’, which is long in returns of selected stocks having a persistent upper trend and short in stocks having a persistent downwards trend.

To explain the return of a stock or a portfolio of stocks, the F&F model adds to the unique factor retained by the CAPM, the market risk premium, three other factors that are assumed to represent market anomalies: the small firm anomaly, the book value to market value anomaly and the momentum anomaly.³²

We postulate that the three mimicking portfolios *SMB*, *HML* and *UMD* might be measured with errors. They are thus possibly correlated with the innovation term in equation (9), and the estimators of the parameters of this equation obtained by OLS are consequently biased and not consistent. To purge these coefficients from these biases, we must regress in a first pass the independent variables on instrumental variables. The estimated method used in this article, which is based on the Hausman test, will be explained below. The problem lies in the choice of these instruments.

As said previously, it is difficult to find valuable instruments for the excess returns of the mimicking portfolios. Being long in some stocks and short in others, their cash flows are similar to those of hedge funds. Higher moments of returns, as asymmetry and kurtosis, might have a great influence on these returns. This suggests the use of higher moments of the variables on the RHS of equation (9) as instrumental

variables. An econometric theory is indeed in construction on this subject. Following Durbin¹⁴ and Pal,¹⁵ Dagenais and Dagenais^{16,17} showed that higher moments³³ of independent variables of a regression might be valid instruments to remove errors-in-variables. But instead of defining higher moments as in these papers, we will adopt a method more akin to asset pricing theory, which defines higher moments of returns by powers of these returns.

The method of asset pricing based on higher moments is not new. Rubinstein²³ and Kraus and Litzenberger²⁴ put the foundations of the three-moment and four-moment CAPM. The three-moment CAPM integrated asymmetry of returns in the analysis while the four-moment CAPM added kurtosis.

Some authors, like Kraus and Litzenberger,²⁴ use a general utility function to derive the moment-CAPM. Others use a Taylor expansion of the utility function, which, following Samuelson²² and Rubinstein,²³ allows expressing utility in terms of the higher moments of returns.

Let us assume that the expected utility of wealth, $E[U(W)]$, is function of the n first moments of the distribution of wealth:

$$E[U(W)] = \varphi[\bar{W}, \sigma_W, skew_W, kur_W, \dots, sm_nW] \quad (10)$$

with \bar{W} the expected value of wealth; σ_W , the volatility of wealth; $skew_W$, its skewness; kur_W , its kurtosis and sm_nW , the n th moment of the distribution of wealth. We incorporate moments of order 5, particularly because we know that they might be important to explain the returns of long-short portfolios like mimicking portfolios that are the foundation of the F&F model.

The expected utility of end-of-period wealth is maximized over the one period horizon subject to the constraint of initial wealth, which is:

$$a_0 + \sum_{i=1}^n a_i = w_0 \quad (11)$$

According to equation (11), the initial wealth w_0 is allocated between the risk-free asset a_0 and n other risky assets designated by a_i . To maximize the utility of end-of-period wealth subject to (11), we form the usual Lagrangian:

$$L = E[U(W)] - \phi \left(a_0 - \sum_{i=1}^n a_i - w_0 \right) \quad (12)$$

Taking the first-order conditions for a maximum, we have:

$$\frac{\partial L}{\partial a_0} = \varphi_{\bar{W}} \times \frac{\partial \bar{W}}{\partial a_0} - \phi = 0 \quad (13)$$

$$\begin{aligned} \frac{\partial L}{\partial a_i, i \neq 0} &= \varphi_{\bar{W}} \times \frac{\partial E(w)}{\partial a_i} + \varphi_{\sigma_W} \times \frac{\partial \sigma_W}{\partial a_i} + \varphi_{skew_W} \\ &\times \frac{\partial skew_W}{\partial a_i} + \varphi_{kur_W} \times \frac{\partial kur_W}{\partial a_i} \\ &+ \dots + \varphi_{sm_nW} \frac{\partial sm_nW}{\partial a_n} - \phi = 0 \quad (14) \end{aligned}$$

with $\varphi_x = (\partial E[U(X)]) / (\partial x)$.

End-of-period expected wealth is equal to:

$$\bar{W} = (1 + R_0)a_0 + \sum_{i=1}^n [(1 + E(R_i))a_i] \quad (15)$$

We thus have:

$$\frac{\partial \bar{W}}{\partial a_0} = 1 + R_0 \quad (16)$$

$$\frac{\partial \bar{W}}{\partial a_i} = 1 + E(R_i) \quad (17)$$

We can therefore express the moments of wealth in terms of the moments of returns of the portfolio as:

$$\sigma_W = \sum_i a_i \beta_{ip} \sigma_p \quad (18)$$

$$skew_W = \sum_i a_i \gamma_{ip} skew_p \quad (19)$$

$$kur_W = \sum_i a_i \theta_{ip} kur_p \quad (20)$$

$$sm_{nW} = \sum_i a_i \omega_{ip} sm_{np} \quad (21)$$

with $\sigma_p = \sqrt{E(R_p - \bar{R}_p)^2}$, the volatility of the return of the risk assets; $\beta_{ip} = (E[(R_i - \bar{R}_i)(R_p - \bar{R}_p)]) / (\sigma_p^2)$, the beta of risk asset i with the investor's portfolio of risk assets.

$skew_p = \sqrt[3]{E[(R_p - \bar{R}_p)^3]}$, the asymmetry of the return of the portfolio and $\gamma_{ip} = (E[(R_i - \bar{R}_i)(R_p - \bar{R}_p)^2]) / (sk_p^3)$, the gamma of risk asset i with the investor's portfolio of risk assets. It is easy to generate all the other variables by following this pattern.

Equating equations (13) and (14) to delete ϕ and taking into account equations (18) to (21), we arrive at an expression for $E(R_i)$ in terms of the moments of the distribution of the return

of an investor's portfolio:

$$\begin{aligned} E(R_i) - R_0 = & - \left(\frac{\varphi_{\sigma_W}}{\varphi_{\bar{W}}} \right) \beta_{ip} \sigma_p \\ & - \left(\frac{\varphi_{skew_W}}{\varphi_{\bar{W}}} \right) \gamma_{ip} skew_p \\ & - \left(\frac{\varphi_{kur_W}}{\varphi_{\bar{W}}} \right) \theta_{ip} kur_p \\ & - \dots - \left(\frac{\varphi_{sm_{nW}}}{\varphi_{\bar{W}}} \right) \omega_{ip} sm_{np} \quad (22) \end{aligned}$$

with $(\varphi_{moment})/(\varphi_{\bar{w}})$ being an investor's marginal rate of substitution between expected wealth and a specific moment. According to Scott and Hovarth,²⁸ these marginal rates of substitution are positive for odd moments, like mean and positive asymmetry, and negative for even moments, like variance and kurtosis. From equation (22), odd moments have, *ceteris paribus*, a negative impact on expected return from the point of view of investors. Even moments have a positive impact because they represent risk.

Moving from equation (22) to the condition of market equilibrium for $E(R_i)$ requires making, according to Kraus and Litzenberger,²⁴ the strong assumption of homogeneous expectations for investors. Following this assumption, equation (22) becomes, assuming that p is the market portfolio:

$$\begin{aligned} E(R_i) - R_0 = & \left[\frac{d\bar{W}}{d\sigma_W} \right] \sigma_m \beta_{im} + \left[\frac{d\bar{W}}{dskew_W} \right] skew_m \gamma_{im} \\ & + \left[\frac{d\bar{W}}{dkur_w} \right] kur_m \theta_{im} + \dots \\ & + \left[\frac{d\bar{W}}{dsm_{nW}} \right] sm_{im} \omega_{im} \quad (23) \end{aligned}$$

The terms in brackets in expression (23) are the slopes of the efficient frontiers whose arguments are expected wealth and the

respective moment. We obtain finally the n-moment CAPM:

$$E(R_i) - R_f = \psi_1\beta_{im} + \psi_2\gamma_{im} + \psi_3\theta_{im} + \dots + \psi_n sm_{nm} \quad (24)$$

We might use directly expression (24) to define our instruments for removing errors-in-variables by the methods of higher moments in the F&F model. Assume we want to correct the mimicking portfolio *SMB* for errors-in-variables. In the first pass of our regressions, we would regress this variable on the co-moments of the lagged excess return of the market portfolio. The variable *SMB* corrected for errors-in-variables would be:

$$\hat{SMB}_t = \kappa_0 + \kappa_1\beta_{im,t-1} + \kappa_2\gamma_{im,t-1} + \kappa_3\theta_{im,t-1} + \dots + \kappa_n sm_{nm,t-1} \quad (25)$$

In fact, we would also have to introduce the co-moments of the mimicking portfolios. This approach would be laborious and would require computing rolling windows of co-moments. But there is a procedure for simplifying equation (25). Kraus and Lintzenberger^{24,34} have shown that a three-moment CAPM is consistent with the following quadratic form:

$$R_i - R_0 = \alpha_0 + \alpha_1(R_m - R_0) + \alpha_2(R_m - R_0)^2 \quad (26)$$

and consequently a n-moment CAPM can be written as:

$$R_i - R_0 = \alpha_0 + \alpha_1(R_m - R_0) + \alpha_2(R_m - R_0)^2 + \alpha_3(R_m - R_0)^3 + \dots + \alpha_{n-1}(R_m - R_0)^{n-1} \quad (27)$$

A test on α_2 is a test on skewness preferences in asset pricing and a test on α_3 , a test on kurtosis

preferences, and so on. The higher moments are consequently powers of returns in this approach. We therefore use a financial theory, the n-moment CAPM, to give an object to the method of Dagenais and Dagenais¹⁷ for correcting errors-in-variables. Let us return to the variable *SMB*, which we want to correct for the problem of errors-in-variables. In the first pass of the regression, this variable will be regressed on:

$$SMB_t = f(F_{it-1}, F_{it-1}^2, F_{it-1}^3, \dots, F_{it-1}^5, \dots) \quad (28)$$

where F_i are the variables in the RHS of the F&F equation (equation (9)) including *SMB*. They stand for the higher moments of these variables. F_{it-1}^2 stands for the skewness of factor F_i ; F_{it-1}^3 , for its kurtosis, and so on. The variables appearing on the RHS of equation (28) will serve as instrumental variables in the first pass of the Hausman test, as explained in the following section.

Hausman specification test and errors-in-variables

To detect errors in variables in our sample of hedge funds, we could use the original Hausman *h* test.³⁵ To explain this test, let us suppose the following classical model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (29)$$

with \mathbf{Y} a $(n \times 1)$ vector representing the dependent variable; \mathbf{X} , a $(n \times k)$ matrix of the explicative variables; $\boldsymbol{\beta}$, a $(k \times 1)$ vector of the estimates of the parameters and $\boldsymbol{\varepsilon} \sim iid(0, \sigma^2)$.

Hausman compares two sets of estimates of the parameters vector, say, $\boldsymbol{\beta}_{OLS}$, the least-squares estimator (OLS), a $\boldsymbol{\beta}_A$, and alternative estimator that can take a variety of forms, but which for our purposes is the

instrumental variables estimator which we designate by β_{IV} . The hypotheses to test are H_0 , being in our case the absence of errors in variables and H_1 , being the presence of errors in variables. The vector of estimates β_{IV} is consistent under both H_0 and H_1 but β_{OLS} is consistent under H_0 but inconsistent under H_1 . Under H_0 , β_{IV} is indeed less efficient than β_{OLS} .

Hausman wants to verify whether the ‘endogeneity’ of some variables,³⁶ the variables measured with errors in our case, has any significant effect on the estimation of the vector of parameters. To do so, he defines the following vector of contrasts: $\beta_{IV} - \beta_{OLS}$. The test statistic may be written as follows:

$$h = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})^T [\text{Var}(\hat{\beta}_{IV}) - \text{Var}(\hat{\beta}_{OLS})]^{-1} \times (\hat{\beta}_{IV} - \hat{\beta}_{OLS}) \sim \chi^2(g) \quad (30)$$

with $\text{Var}(\hat{\beta}_{IV})$ and $\text{Var}(\hat{\beta}_{OLS})$ being consistent estimates of the covariance matrices of $\hat{\beta}_{IV}$ and $\hat{\beta}_{OLS}$. g is the number of potentially endogenous regressors, that is the variables measured with errors in our case. H_0 will be rejected if the P -value of this test is less than α , with α being the critical threshold of the test, say 5 per cent.

According to MacKinnon,³⁷ this test might run into difficulties if the matrix $[\text{Var}(\hat{\beta}_{IV}) - \text{Var}(\hat{\beta}_{OLS})]$ that weights the vector of contrasts is not positive definite. Fortunately, there is an alternative way to do the Hausman test which is much easier. This test goes as follows.

Assume a two-variable linear model:

$$y_t = \beta_0 + \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \varepsilon_t \quad (31)$$

with $\varepsilon \sim N(0, \sigma^2)$.

The variables x_{1t}^* and x_{2t}^* ³⁸ are observed with errors, that is:

$$x_{1t} = x_{1t}^* + v_{1t} \quad (32)$$

$$x_{2t} = x_{2t}^* + v_{2t} \quad (33)$$

with x_{1t} and x_{2t} , the corresponding observed variables which are measured with errors. By substituting equations (32) and (33) in equation (31), we have:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t^* \quad (34)$$

with $\varepsilon_t^* = \varepsilon_t - \beta_1 v_{1t} - \beta_2 v_{2t}$. As seen before, estimating coefficients of equation (34) by the OLS method gives way to biased and inconsistent coefficients because the explanatory variables are correlated with the innovation.

Consistent estimators can be found if we can identify an instruments vector z_t that is correlated with every explanatory variable but not with the innovation of equation (31). Then we regress these two explanatory variables on z_t . We have:

$$x_{1t} = \hat{x}_{1t} + \hat{w}_{1t} = \hat{\gamma}_1 z_t + \hat{w}_{1t} \quad (35)$$

$$x_{2t} = \hat{x}_{2t} + \hat{w}_{2t} = \hat{\gamma}_2 z_t + \hat{w}_{2t} \quad (36)$$

with \hat{x}_{it} , the value of x_{it} estimated with the vector of instruments and \hat{w}_{it} , the residuals of the regression of x_{it} on \hat{x}_{it} . Substituting equations (35) and (36) into equation (34), we have³⁹:

$$y_t = \beta_0 + \beta_1 \hat{x}_{1t} + \beta_2 \hat{x}_{2t} + \beta_1 \hat{w}_{1t} + \beta_2 \hat{w}_{2t} + \varepsilon_t^* \quad (37)$$

The explanatory variables of this equation are, on the one hand, the estimated values of x_{1t} and x_{2t} , obtained by regressing these two variables on the vector of instruments z_t , and on the other hand, the respective residuals of these regressions. Equation (37) is therefore an augmented version of equation (34), which might be qualified of auxiliary or artificial regression.

We can show that:

$$\begin{aligned}
 p \lim \left[\frac{\sum \hat{w}_{1t} \varepsilon_t^*}{N} \right] &= p \lim \left[\frac{-\beta_1 \sum x_{1t} v_{1t}}{N} \right] \\
 &= -\beta_1 \sigma_{v_1}^2 \quad (38)
 \end{aligned}$$

and the same for w_{2t} . When there is no measurement error, $\sigma_{v_1}^2 = 0$ and OLS gives way to a consistent estimator for the parameter of \hat{w}_{1t} in equation (37), that is β_1 . When there are measurement errors, $\sigma_{v_1}^2 \neq 0$ and therefore this estimator is not consistent.

We can therefore write the following test to detect the presence of errors in variables. As we do not know *a priori* whether there are errors in variables, we replace the coefficients of \hat{w}_{1t} and \hat{w}_{2t} in equation (37) by γ_1 and γ_2 . We have:

$$\begin{aligned}
 \gamma_t &= \beta_0 + \beta_1 \hat{x}_{1t} + \beta_2 \hat{x}_{2t} + \gamma_1 \hat{w}_{1t} \\
 &\quad + \gamma_2 \hat{w}_{2t} + \varepsilon_t^* \quad (39)
 \end{aligned}$$

But following equations (35) and (36) $\hat{x}_{1t} = x_{1t} - \hat{w}_{1t}$ and $\hat{x}_{2t} = x_{2t} - \hat{w}_{2t}$. We can therefore rewrite equation (39) as follows:

$$\begin{aligned}
 \gamma_t &= \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + (\gamma_1 - \beta_1) \hat{w}_{1t} \\
 &\quad + (\gamma_2 - \beta_2) \hat{w}_{2t} + \varepsilon_t^* \quad (40)
 \end{aligned}$$

If there is no measurement error for both variables x_{1t} and x_{2t} , then $\gamma_1 = \beta_1$ and $\gamma_2 = \beta_2$. If there are measurement errors, $\gamma_i \neq \beta_i$ and the coefficients of the residuals terms w_{it} will not be zero.

There is more information that we can draw from equation (40). Indeed, if the estimated coefficient $(\gamma_i - \beta_i)$ is significantly positive, that indicates that the estimated coefficient of the corresponding explanatory variable x_{it} is overstated in the OLS run. Therefore, the estimated coefficient for this variable will

decrease in equation (40). On the other hand, if the estimated coefficient $(\gamma_i - \beta_i)$ is significantly negative, that indicates that the estimated coefficient of the corresponding explanatory variable x_{it} is understated in the OLS run. Therefore, the estimated coefficient for this variable will increase in equation (40). These effects of errors in variables produced by equation (40) are very informative. In the next section, we will transpose these results to the F&F model.

We must note that the coefficients β_i estimated by the equation (40) are identical to those produced by a two-stage least squares (TSLS) procedure using the same instruments. Equation (40) is therefore another way to do a TSLS. But in view of the useful information produced by equation (40), this equation opens the doors to new financial models. We will therefore prefer this formulation to that one represented by TSLS to estimate the augmented F&F model. And we thus have a new empirical formulation for the F&F model.

We therefore proceed as follows to test for errors in variables. First, we regress the observed explanatory variables x_{it} on the instruments vector to obtain the residuals \hat{w}_{it} . Then, we regress γ_t on the observed explanatory variables x_{it} and on the residuals \hat{w}_{it} . This is an auxiliary or artificial regression. If the coefficient of the residuals of an explanatory variable is significantly different from 0, we may conclude that there is a measurement error on this explanatory variable. We may use the Wald test (*F* test) to determine whether the whole set of $(\gamma_i - \beta_i)$ coefficients is significantly different from zero.

We can generalize the preceding procedure to the case of k explanatory variables, which are potentially suffering from the problem of errors

in variables. Let \mathbf{X} be a $(n \times k)$ matrix of explanatory variables that are potentially suffering from the disease of errors in variables and let \mathbf{Z} be a $(n \times s)$ matrix of instruments ($s > k$). To perform the Hausman test based on an artificial regression, we first regress \mathbf{X} on \mathbf{Z} to obtain $\hat{\mathbf{X}}$, that is:

$$\hat{\mathbf{X}} = \mathbf{Z}\hat{\theta} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} = \mathbf{P}_Z\mathbf{X} \quad (41)$$

where \mathbf{P}_Z is the 'predicted value maker'. Having performed this regression, we compute the matrix of residuals $\hat{\mathbf{w}}$:

$$\hat{\mathbf{w}} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X} - \mathbf{P}_Z\mathbf{X} = (\mathbf{I} - \mathbf{P}_Z)\mathbf{X} \quad (42)$$

Then we perform the following artificial regression:

$$\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \lambda\hat{\mathbf{w}} \quad (43)$$

An F test on the λ coefficients will indicate whether they are significant as a group. A t test on individual coefficients will indicate whether the corresponding β is understated or overstated, as discussed previously.

The vector of $\boldsymbol{\beta}$ estimated by equation (43) is identical to the TSLS estimates, that is:

$$\boldsymbol{\beta} = \boldsymbol{\beta}_{IV} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y} \quad (44)$$

To detect errors in variables in the augmented F&F model, we will run two sets of regressions. First, we will run the OLS one, that is:

$$R_{pt} - R_{ft} = \alpha + \beta_1(R_{mt} - R_{ft}) + \beta_2SMB_t + \beta_3HML_t + \beta_4UMD_t + \varepsilon_t \quad (45)$$

Then, we will run the following artificial regression explained previously:

$$R_{pt} - R_{ft} = \alpha^* + \beta_1^*(R_{mt} - R_{ft}) + \beta_2^*SMB_t + \beta_3^*HML_t + \beta_4^*UMD_t + \sum_{i=1}^4 \varphi_i \hat{w}_{it} + \varepsilon_t^* \quad (46)$$

The estimated coefficients φ_i will allow detecting errors in variables, and their signs will indicate whether the corresponding variable is overstated or understated in the OLS regression.

As said previously, the $\boldsymbol{\beta}^*$ estimated by equation (46) are equivalent to the TSLS estimates. But we prefer equation (46) because it gives more information on the problem of errors in variables. Equation (46) is thus our new empirical version of the augmented F&F model. The φ_i are really factors of correction of the risk exposure of a Fund to the i th factor of risk. If φ_i is positive, that means that the exposure to the i th risk factor is overstated in the OLS regression. The β associated with this factor will thus decrease in the artificial regression. And vice-versa if φ_i is negative. Moreover, according to our previous developments, we expect a high positive correlation between $(\hat{\beta}_i - \hat{\beta}_i^*)$, that is the estimated error on the coefficient of factor i , and $\hat{\varphi}_i$, the estimated coefficient of the corresponding artificial variable (\hat{w}_i).

EMPIRICAL RESULTS AND ANALYSIS

Our sample of hedge funds returns comprises the monthly returns of 20 Greenwich-Van US American hedge funds indexes classified by categories or groups of categories. The appendix gives the enumeration of these funds with their chosen symbol. The observation period runs from January 1995 to November 2005, for a total of 131 observations. The risk factors that appear in the F&F equation – that is the market risk premium and the three mimicking portfolios: SMB , HML and UMD – are for their part drawn from the French's website.⁴⁰

Table 1 gives a first glance of our hedge fund sample, and reports the descriptive statistics of

Table 1: Descriptive statistics of the sample of 20 hedge funds returns

<i>Statistics</i>	<i>Average</i>	<i>Std. dev</i>
Mean yield	0.0097	0.003
Median yield	0.0098	0.004
Maximum yield	0.1066	—
Minimum yield	-0.0879	—
Skewness	0.0959	0.817
Kurtosis	6.3812	3.304
Jarque–Bera	133.048	284.206
<i>P</i> -value	0.042	0.121
Number of funds	18	—
with kurtosis > 3		
Number of funds with	17	—
non-Gaussian yields		

the ‘average’ fund. At 11.6 per cent, the annualized mean return of this sample is quite high and its standard deviation is relatively moderate. But it is well known that standard deviation is a reliable measure of risk only if risk is small or under very special conditions. Otherwise, we must consider the higher moments of the distributions of excess returns. The skewness of the average fund is close to 0 but the mean level of kurtosis, at 6.4, is quite high, the level of kurtosis associated with a normal distribution being 3. Obviously, a large number of equity-oriented hedge funds strategies exhibit payoffs that are similar to those of a short position in a put option written on the market index,⁴¹ and therefore bears significant left-tail risk, risk that is ignored by the commonly used mean-variance framework. This may imply that rare events are more frequent

than in a normal distribution, and that nonlinearities of payoffs are quite important. Incidentally, there are 18 funds over 20 that have a kurtosis level exceeding three and 17 that have non-Gaussian returns according to the Jarque–Bera test.

We now define the instruments necessary to perform the Hausman test. Table 2 gives the correlations of the F&F factors with themselves and their instruments from January 1995 to November 2005. In addition to the instruments discussed before, we add other macroeconomic variables: the monthly and annual American inflation rate, *IPC_MENS* and *IPC_ANN*, and the monthly and annual growth rate of the American industrial production, *PROD_MENS* and *PROD_ANN*.

According to Table 2, the F&F factors are more or less correlated with conventional instruments as the first lag of the factor or with the macroeconomic variables. Regarding the macroeconomic variables, we observe that their correlation with the risk factors located at the top of the columns is quite low. Only the industrial production annual growth has a moderate correlation with three of these factors.⁴²

With respect to the risk factors of equation (45), we noted before that there are many nonlinearities in these mimicking portfolios that are similar to hedge fund portfolio. We also noted before that these nonlinearities might be captured by the higher moments of these factors. Corroborating this assumption, we observe from Table 2 that the risk factors are usually more correlated or cross-correlated with the higher moments of the first lag of a risk factor than to the first lag itself. For instance, the market risk premium is more related to the higher moments of *UMD*(-1) than to *UMD*(-1) itself. Indeed, the

Table 2: Correlation between the four factors and their instruments

	<i>RM_RF</i>	<i>SMB</i>	<i>HML</i>	<i>UMD</i>
<i>RM_RF</i>	1.00	—	—	—
<i>SMB</i>	0.23	1.00	—	—
<i>HML</i>	-0.30	-0.37	1.00	—
<i>UMD</i>	-0.23	0.04	-0.62	1.00
<i>RM_RF</i> (-1)	0.03	0.17	0.16	-0.11
<i>RM_RF</i> (-1) ²	0.08	0.07	-0.02	-0.07
<i>RM_RF</i> (-1) ³	-0.07	0.10	0.10	-0.02
<i>RM_RF</i> (-1) ⁴	0.11	0.01	-0.05	-0.05
<i>RM_RF</i> (-1) ⁵	-0.10	0.04	0.08	0.01
<i>HML</i> (-1)	-0.16	0.06	0.11	0.01
<i>HML</i> (-1) ²	-0.06	-0.23	0.17	-0.01
<i>HML</i> (-1) ³	-0.14	0.18	-0.06	0.09
<i>HML</i> (-1) ⁴	0.01	-0.29	0.19	-0.10
<i>HML</i> (-1) ⁵	-0.12	0.24	-0.11	0.11
<i>SMB</i> (-1)	0.08	-0.02	-0.04	0.11
<i>SMB</i> (-1) ²	-0.01	-0.19	0.20	-0.10
<i>SMB</i> (-1) ³	0.12	-0.17	0.03	0.01
<i>SMB</i> (-1) ⁴	0.04	-0.29	0.24	-0.15
<i>SMB</i> (-1) ⁵	0.13	-0.23	0.06	-0.06
<i>UMD</i> (-1)	0.03	-0.15	-0.07	-0.07
<i>UMD</i> (-1) ²	-0.20	-0.08	0.18	0.07
<i>UMD</i> (-1) ³	0.19	-0.06	-0.01	-0.22
<i>UMD</i> (-1) ⁴	-0.21	-0.12	0.16	0.15
<i>UMD</i> (-1) ⁵	0.22	-0.01	-0.05	-0.22
<i>IPC_ANN</i>	-0.03	0.00	0.10	-0.03
<i>IPC_MENS</i>	-0.11	-0.06	-0.01	0.06
<i>PROD_ANN</i>	0.12	-0.23	-0.11	0.05
<i>PROD_MENS</i>	-0.02	-0.04	-0.07	0.02

correlation between *RM_RF* and *UMD*(-1) is quite low: 0.03, but it is equal to 0.22 for *UMD*(-1),⁷ that is the higher moment of *UMD*(-1) of order 5. The same is true for the factor *SMB* and the higher moments of

SMB(-1). Consequently, higher moments of lagged variables may constitute quite good instruments.⁹

Before discussing the results, let us note that we performed a Wald test over the whole set of the four coefficients associated with the risk factors of equation (46). For this test, the null hypothesis H_0 is:

$$\varphi_1 = \varphi_2 = \varphi_3 = \varphi_4 = 0$$

If this hypothesis is not rejected, we cannot detect errors-in-variables for the four factors considered as a group. This hypothesis was not rejected for any fund, and consequently it seems that there is no problem of errors-in-variables in our sample of hedge funds when we consider the four factors as a group. But as we will see, individual *t* tests reveal that in some cases, the bias caused by the presence of errors in variables may be quite serious.

Tables 3 and 4 give a first grasp of the estimations of equations (45) and (46) over the whole sample of our hedge funds. These

Table 3: Count of significant coefficients for the constant and the four factors for the OLS and the artificial Hausman regression^a

	<i>C</i>	<i>R_m-R_f</i>	<i>SMB</i>	<i>HML</i>	<i>UMD</i>	<i>Line total</i>
OLS	13	18	17	7	9	64
Hausman	15	16	17	5	7	60

^aFor a variable, the table gives the number of significant coefficients for $\alpha=5$ per cent. There are 20 yield indexes of hedge funds in the sample. The estimation goes from January 1995 to November 2005.

Abbreviation: OLS, ordinary least squares.

Table 4: Average level of the t statistics for the constant and the four factors for the OLS and the artificial Hausman regression^a

	C	$R_m - R_f$	SMB	HML	UMD	Line average
OLS	3.16	8.62	5.54	1.72	1.99	4.21
Hausman	2.68	8.64	5.18	1.53	1.80	3.97

^aThe average is computed for a sample of 20 hedge funds indexes.

Abbreviation: OLS, ordinary least squares.

tables contain the total count of significant coefficients at a confidence level of 95 per cent and the mean level of the t statistics computed over the 20 funds for both estimation methods. Table 3 indicates that the constant and the coefficients of regression of the risk premium and SMB are significant for the majority of the funds of our sample. The variable UMD is significant for approximately 50 per cent of the sample while the variable HML is more problematic, being significant for only a minority of funds. Table 4, which produces the average of the t statistics for the constant and the four factors computed over the 20 funds, confirms those preliminary observations. Let us note that equation (46) produces less significant coefficients than equation (45), even if this difference is not quite high. It is well known that the coefficients estimated by TSLS tend to have a larger variance than those resulting from a corresponding OLS regression. As seen previously, the artificial regression given by equation (46) is equivalent to TSLS method. On the other hand, the average R^2 of the two estimation methods are quite comparable, being 0.55 for the OLS estimation and 0.54 for the artificial regression.

Table 5: Mean level of the estimated coefficients for the constant and the four factors for the OLS and the Hausman artificial regression

	C	$R_m - R_f$	SMB	HML	UMD
OLS	0.0036	0.2479	0.1709	0.0321	0.0764
Hausman	0.0039	0.2346	0.2057	-0.0126	0.0745
Spread	-0.0003	0.0133	-0.0348	0.0447	0.0019

Abbreviation: OLS, ordinary least squares.

From Table 5, we find the mean levels of the coefficients estimated by the OLS method and by the artificial regression for the whole set of funds. The mean level of the constant, which corresponds to the alpha, is practically the same for both regression methods. We note that the beta seems to be overstated by the OLS regression that produces biased coefficients if there are errors in variables. The average beta resulting from the OLS estimation is 0.25 and 0.23 for the artificial one. Otherwise, the impact of SMB tends to be understated by the OLS regression, its average coefficient being 0.17 in the OLS regression and 0.21 in the artificial one. The average incidence of the UMD factor is quite similar for the two estimation methods but, as we will see, this situation hides quite a high dispersion. Finally, the influence of HML is quite low in both estimation methods, being moderately overstated by the OLS regression.

As the F&F model is a purely empirical one, there is no theory on the signs of the four factors of this model, except perhaps for the market index whose coefficient is generally positive according to the CAPM. But if we consider the three mimicking portfolios SMB , HML and UMD as risk factors, it is reasonable to expect

generally a positive sign for the estimated coefficients of these factors. In fact, a fund can short a factor of risk, as it might short the market index. A hedge fund that makes short selling will have a negative beta. Another fund that short sells the mimicking portfolio *SMB*, say, will have a negative coefficient for this factor. But this behaviour seems to be the exception rather than the rule. Risk factors in returns equations should generally be preceded by positive signs.

Table 6 gives information on this matter for our sample. For both estimation methods, the estimated coefficients of the four factors have positive signs for the majority of funds. That comforts us in our expectations of considering the mimicking portfolios as risk factors and not as market anomalies as they were viewed in the past.

To conclude these preliminary observations, we may say that the problem of errors in variables does not seem to be very serious for the group of funds surveyed in this study. But average behaviour might hide a great dispersion at the individual level. As we will see in the following paragraphs, the biases caused by errors in variables might be severe for some funds.

In our sample, it was the factor *HML* that seemed to be suffering most from errors in variables. The artificial regression reveals that its

residuals were significant at the 95 per cent confidence level for four funds. For both *SMB* and *UMD*, the residuals are significant for three funds.

We will gain a better grasp of our estimations if we look at the individuals results. Table 7 gives the estimated betas, that is the coefficient of the risk premium ($R_m - R_f$), for each fund and for the two estimation methods. This table also provides the corresponding coefficient φ in the artificial regression (equation (46)). This coefficient is in bold when it is significant at the 95 per cent confidence level.⁴³

We note from Table 7 that for five funds, the spread between the OLS and the Hausman estimate, which is the measurement error on this coefficient, is quite high. For these funds, it is overstated four times and understated one time by the OLS method. When the beta of a fund is overstated, Table 7 reveals that the artificial coefficient φ associated with the residuals of the market risk premium in equation (46) is positive, as it must be. In the case of the ‘aggressive growth’ fund designated by *ag*, the beta is greatly understated by the OLS regression and the corresponding φ is therefore negative. Let us note that the correlation between the spreads (errors) column in our tables and the corresponding φ column is 0.98. Therefore, the association between the level of the measurement error on a coefficient and the corresponding level of φ is positive and almost perfect.⁴⁴ Moreover, the regression of the error of the *i*th fund, that is the spread $(\hat{\beta}_{i,OLS} - \hat{\beta}_{i,HAUS})$, on the corresponding artificial variable $\hat{\varphi}_i$, gives the following result:

$$\left(\hat{\beta}_{i,OLS} - \hat{\beta}_{i,HAUS} \right) = -0.005 + 0.70\hat{\varphi}_i \quad (-2.45) \quad (23.96)$$

Table 6: Count of positive signs for the constant and the four factors for both estimation methods

	<i>C</i>	$R_m - R_f$	<i>SMB</i>	<i>HML</i>	<i>UMD</i>	<i>Line total</i>
OLS	17	18	19	15	20	89
Hausman	17	18	18	13	17	83

Abbreviation: OLS, ordinary least squares.

Table 7: Spread (error) between the OLS and the Hausman beta for each fund of the study^a

	<i>OLS beta</i>	<i>Hausman beta</i>	<i>Error</i>	φ
emn	0.095	0.126	-0.031	-0.039
ed	0.264	0.201	0.063	0.085
ds	0.144	0.082	0.063	0.091
ss	0.322	0.275	0.048	0.063
mna	0.082	0.069	0.014	0.018
shs	-0.908	-0.922	0.014	0.043
va	0.515	0.494	0.020	0.022
f	-0.007	-0.129	0.122	0.189
macro	0.303	0.349	-0.046	-0.036
mt	0.325	0.291	0.034	0.051
dtg	0.155	0.088	0.067	0.110
mng	0.145	0.118	0.026	0.035
lsg	0.431	0.431	0.000	0.000
ag	0.650	0.744	-0.094	-0.125
oi	0.368	0.360	0.008	0.013
specg	0.525	0.532	-0.008	0.009
em	0.678	0.686	-0.008	0.021
inc	0.138	0.131	0.007	0.007
msi	0.388	0.427	-0.039	-0.050
hf	0.345	0.338	0.007	0.012

^aThe spread (error) is the difference between the OLS coefficient and the corresponding Hausman coefficient. For each spread, we produce the coefficient (φ) of the corresponding artificial variable. The Funds having the highest spreads in absolute value are bold-faced. The coefficient φ is bold-faced when significant. Note the high positive relation between the error and its associated φ .
Abbreviation: OLS, ordinary least squares.

For this regression, the adjusted R^2 is 0.96. Therefore, the level of the artificial variable is a very good indicator of the measurement error of

the corresponding coefficient. It thus gives precious information on the measurement error, and it is the reason why we prefer the Hausman version of the F&F model to the equivalent TSLS one. We will not repeat this regression for the other factors of risk of the F&F model because the results are very similar to those obtained for the risk premium.

Table 8 provides the same information as Table 7 concerning the *SMB* factor. There are five funds where the measurement error is high. With regard to these funds, we note that the coefficient of this variable is understated four times and overstated one time by the OLS regression. When the impact of *SMB* is understated, the corresponding φ is negative, and when it is overstated the corresponding φ is positive, as it might be. The *SMB* coefficient is particularly understated for the emerging markets fund (*em*), and its associated φ is therefore negative and very high in absolute value.

Table 9 reveals that the estimation of the *HML* coefficient is quite problematic, the φ coefficients for this variable being very high for some funds. The OLS regression tends to overstate greatly the impact of this variable on the returns of four funds. In this study, this overstatement is explained by a possible serious problem of errors in variables.

Finally, according to Table 10, the overstatement of the impact of the factor *UMD* is quite high in three cases. It is moderately understated in two cases. Moreover, if we correlate the OLS- and Hausman-estimated coefficients for this factor, we obtain only 0.23. The corresponding coefficients for the three other factors – $R_m - R_f$, *SMB* and *HML* – are 0.99, 0.91 and 0.80, respectively. Consequently, there is a great divergence between the results

Table 8: Spread (error) between the OLS and the Hausman SMB estimated coefficients for each fund of the study^a

	SMB coef. OLS	SMB coef. Hausman	Spread	φ
emn	0.123	0.158	-0.035	-0.036
ed	0.225	0.252	-0.027	-0.052
ds	0.188	0.232	-0.044	-0.074
ss	0.250	0.279	-0.028	-0.056
Mna	0.062	0.066	-0.004	-0.012
shs	-0.422	-0.359	-0.063	-0.077
va	0.331	0.272	0.058	0.101
f	0.021	-0.053	0.073	0.149
macro	0.199	0.381	-0.182	-0.238
mt	0.102	0.070	0.032	0.055
dtg	0.087	0.080	0.007	0.039
Mng	0.133	0.150	-0.017	-0.028
lsg	0.288	0.248	0.040	0.074
ag	0.359	0.368	-0.010	-0.025
oi	0.318	0.256	0.062	0.140
specg	0.276	0.459	-0.183	-0.293
em	0.396	0.659	-0.264	-0.423
inc	0.105	0.145	-0.040	-0.068
msi	0.159	0.214	-0.055	-0.079
hf	0.220	0.236	-0.016	-0.021

^aThe spread (error) is the difference between the OLS coefficient and the corresponding Hausman coefficient. For each spread, we produce the coefficient (φ) of the corresponding artificial variable. The Funds having the highest spreads in absolute value are bold-faced. This coefficient φ is bold-faced when significant. Note the high positive relation between the error and its associated φ .

Abbreviation: OLS, ordinary least squares.

Table 9: Spread (error) between the OLS and the Hausman HML estimated coefficients for each fund of the study^a

	HML coef. OLS	HML coef. Hausman	Spread	φ
emn	0.016	0.021	-0.004	0.009
ed	0.074	0.065	0.009	-0.006
ds	0.071	0.033	0.039	0.052
ss	0.084	0.093	-0.009	-0.037
mna	0.004	0.014	-0.010	-0.023
shs	0.508	0.320	0.187	0.334
va	0.029	0.019	0.010	0.024
f	0.165	-0.137	0.302	0.548
macro	0.025	-0.080	0.106	0.217
mt	-0.175	-0.249	0.073	0.131
dtg	0.037	-0.161	0.198	0.362
mng	0.025	0.026	-0.002	-0.009
lsg	-0.046	-0.072	0.026	0.058
ag	-0.297	-0.250	-0.047	-0.069
oi	-0.086	-0.185	0.099	0.208
specg	0.072	0.098	-0.026	-0.047
em	0.082	0.096	-0.014	-0.021
inc	0.086	0.105	-0.020	-0.046
msi	-0.018	0.012	-0.030	-0.046
hf	-0.014	-0.020	0.006	0.013

^aThe spread (error) is the difference between the OLS coefficient and the corresponding Hausman coefficient. For each spread, we produce the coefficient (φ) of the corresponding artificial variable. The Funds having the highest spreads in absolute value are bold-faced. This coefficient φ is bold-faced when significant. Note the high positive relation between the error and its associated φ .

Abbreviation: OLS, ordinary least squares.

Table 10: Spread (error) between the OLS and the Hausman UMD estimated coefficients for each fund of the study^a

	<i>UMD coef.</i> OLS	<i>UMD coef.</i> Hausman	<i>Spread</i>	φ
emn	0.101	0.160	-0.060	-0.079
ed	0.086	0.083	0.003	-0.001
ds	0.019	0.009	0.009	0.016
ss	0.115	0.120	-0.005	-0.015
mna	0.029	0.036	-0.007	-0.014
shs	0.047	-0.103	0.150	0.256
va	0.045	0.054	-0.009	-0.016
f	0.187	-0.027	0.213	0.352
macro	0.150	0.204	-0.055	-0.037
mt	0.038	-0.027	0.065	0.103
dtg	0.139	0.024	0.115	0.199
mng	0.060	0.078	-0.017	-0.027
lsg	0.064	0.054	0.010	0.017
ag	0.059	0.082	-0.023	-0.023
oi	0.112	0.102	0.010	0.033
specg	0.075	0.176	-0.100	-0.136
em	0.082	0.215	-0.133	-0.173
inc	0.071	0.090	-0.019	-0.031
msi	0.021	0.073	-0.052	-0.072
hf	0.064	0.085	-0.021	-0.028

^aThe spread (error) is the difference between the OLS coefficient and the corresponding Hausman coefficient. For each spread, we produce the coefficient (φ) of the corresponding artificial variable. The Funds having the highest spreads in absolute value are bold-faced. This coefficient φ is bold-faced when significant. Note the high positive relation between the error and its associated φ .

Abbreviation: OLS, ordinary least squares.

obtained by the OLS and Hausman regressions for the *UMD* factor, which is not observed for the other ones.

From Tables 11 and 12, we find, respectively, the OLS and artificial regressions for the funds that seem to suffer the most from a problem of error in variables. The adjusted R^2 varies greatly for these funds, having a low of 0.02 for the futures funds and a high of 0.85 for the short selling funds. The artificial regressions (Table 12) were performed by using the higher moments of the predetermined variables of our model as instruments.

Table 12 reports our preferred empirical version of the F&F model. It is a new version that has not been produced yet. Therefore, it is not only an artificial regression, but a new empirical model. It includes the estimated coefficients of the residuals that produce a great amount of information about the correction of the exposures of the funds to the risk factors, as we have seen before. This correcting process is required because of the problem of errors in variables. In Table 13, we note that the coefficients of the regressions performed in Table 12 are identical to the coefficients obtained by a TSLS using the same instruments as equation (46). But this last estimation form is less informative than the one produced by the artificial regression, which constitutes a new empirical model.

To check the relevance of higher moments as instruments, we repeat the estimations appearing in Table 13, but without using higher moments. The instruments used are therefore standard predetermined variables, that is, exogenous variables or lagged endogenous or exogenous variables not powered. Table 14 provides this estimation. The results are obviously bad compared to those obtained by TSLS using higher moments as instruments. Higher moments are therefore good candidates for instruments. They take into account the

Table 11: Estimation of Funds plagued with errors in variables by the OLS method^a

<i>Fund</i>	<i>c</i>	$R_m - R_f$	<i>SMB</i>	<i>HML</i>	<i>UMD</i>	<i>adj. R²</i>	<i>DW</i>
DTG	0.004	0.155	0.087	0.037	0.139	0.129	1.98
	<i>1.80</i>	<i>2.94</i>	<i>1.55</i>	<i>0.51</i>	<i>2.56</i>	—	—
EM	-0.001	0.678	0.396	0.082	0.082	0.456	1.49
	<i>-0.32</i>	<i>7.38</i>	<i>4.06</i>	<i>0.65</i>	<i>0.87</i>	—	—
F	0.005	-0.007	0.021	0.165	0.687	0.023	1.86
	<i>1.61</i>	<i>-0.08</i>	<i>0.23</i>	<i>1.41</i>	<i>2.13</i>	—	—
MACRO	-0.001	0.303	0.199	0.025	0.150	0.361	1.93
	<i>-0.33</i>	<i>5.41</i>	<i>3.33</i>	<i>0.33</i>	<i>2.60</i>	—	—
OI	0.006	0.368	0.318	-0.086	0.112	0.726	1.56
	<i>3.76</i>	<i>9.60</i>	<i>7.81</i>	<i>-1.64</i>	<i>2.85</i>	—	—
SHS	0.002	-0.908	-0.422	0.508	0.047	0.854	1.91
	<i>0.71</i>	<i>-14.13</i>	<i>-6.44</i>	<i>6.01</i>	<i>0.74</i>	—	—
SPECG	0.000	0.525	0.276	0.072	0.075	0.507	1.54
	<i>-0.09</i>	<i>8.43</i>	<i>4.17</i>	<i>0.85</i>	<i>1.18</i>	—	—

^aIn italic, we find the *t* statistic of the corresponding estimated coefficient.

Abbreviation: OLS, ordinary least squares.

nonlinearities of the payoffs of hedge funds that are neglected by classic instruments.

SUMMARY AND CONCLUSION

In this article, we have developed a new empirical version of the F&F model aimed at testing and simultaneously correcting the problem of errors-in-variables. This model includes variables to account for the biases in the estimated exposures to risk factors. This correction process allows reducing the errors-in-variables biases from the estimated coefficients of the risk factors. These correction factors, based on the artificial regression of the Hausman specification test, are interesting because they give information on the understatement or overstatement of risk resulting from the problem

of errors-in-variables and of the kind of specification errors.

Our new empirical version of the F&F model provides another novelty. It uses as instruments the higher moments of the mimicking portfolios returns distribution to correct the F&F model from its errors-in-variables. The fact that these instruments should be highly related to the risk factors is explained by the nonlinearities incorporated in the factor mimicking portfolios. These nonlinearities cannot be captured by a standard CAPM or APT model that postulates a linear relation between the returns to be explained and their risk factors. The F&F model is also linear in its factors. But the risk factors *SMB*, *HML* and *UMD* partly account for the presence of these nonlinearities. The strong relation between these factors and the

Table 12: Estimation of Funds plagued with errors in variables by the Hausman method with higher moments as instruments^a

<i>Fund</i>	<i>c</i>	<i>R_m-R_f</i>	<i>SMB</i>	<i>HML</i>	<i>UMD</i>	<i>Resid_{rm-ff}</i>	<i>Resid_{SMB}</i>	<i>Resid_{HML}</i>	<i>Resid_{UMD}</i>	<i>adj R²</i>	<i>DW</i>
DTG	0.006	0.088	0.080	-0.161	0.024	0.110	0.039	0.362	0.199	0.148	2.03
	<i>2.38</i>	<i>0.89</i>	<i>0.91</i>	<i>-1.46</i>	<i>0.26</i>	<i>0.94</i>	<i>0.33</i>	<i>2.47</i>	<i>1.75</i>	—	—
EM	-0.003	0.666	0.659	0.096	0.215	0.021	-0.423	-0.021	-0.174	0.476	1.66
	<i>-0.63</i>	<i>4.03</i>	<i>4.34</i>	<i>0.51</i>	<i>1.36</i>	<i>0.11</i>	<i>-2.13</i>	<i>-0.08</i>	<i>-0.89</i>	—	—
F	0.009	-0.129	-0.053	-0.137	-0.027	0.190	0.149	0.548	0.352	0.035	1.85
	<i>2.33</i>	<i>-0.80</i>	<i>-0.37</i>	<i>-0.77</i>	<i>-0.18</i>	<i>1.00</i>	<i>0.79</i>	<i>2.30</i>	<i>1.90</i>	—	—
MACRO	-0.002	0.349	0.381	-0.080	0.204	-0.036	-0.238	0.217	-0.037	0.399	2.04
	<i>-0.70</i>	<i>3.42</i>	<i>4.19</i>	<i>-0.71</i>	<i>2.16</i>	<i>-0.30</i>	<i>-2.00</i>	<i>1.44</i>	<i>-0.31</i>	—	—
OI	0.006	0.360	0.256	-0.185	0.102	0.013	0.140	0.208	0.033	0.725	1.55
	<i>3.51</i>	<i>5.00</i>	<i>3.99</i>	<i>-2.32</i>	<i>1.52</i>	<i>0.15</i>	<i>1.67</i>	<i>1.95</i>	<i>0.40</i>	—	—
SHS	0.003	-0.922	-0.359	0.320	-0.103	0.043	-0.077	0.334	0.256	0.854	2.02
	<i>1.16</i>	<i>-7.97</i>	<i>-3.49</i>	<i>2.49</i>	<i>-0.96</i>	<i>0.31</i>	<i>-0.57</i>	<i>1.95</i>	<i>1.93</i>	—	—
SPECG	-0.002	0.532	0.459	0.098	0.176	0.009	-0.294	-0.047	-0.136	0.522	1.7
	<i>-0.52</i>	<i>4.57</i>	<i>4.42</i>	<i>0.76</i>	<i>1.62</i>	<i>0.06</i>	<i>-2.16</i>	<i>-0.27</i>	<i>-1.01</i>	—	—

^aThe variable resid is the residuals of the regression of the attached variable, say *umd*, on the set of instruments.

Note: Statistics *t* are in italics.

Table 13: Estimation of Funds plagued with errors in variables by the TSLS method with higher moments as instruments

<i>Fund</i>	<i>c</i>	<i>R_m-R_f</i>	<i>SMB</i>	<i>HML</i>	<i>UMD</i>	<i>adj. R²</i>	<i>DW</i>
DTG	0.006	0.088	0.080	-0.161	0.024	0.064	2.02
	<i>2.27</i>	<i>0.85</i>	<i>0.87</i>	<i>-1.39</i>	<i>0.25</i>	—	—
EM	-0.003	0.686	0.659	0.096	0.215	0.416	1.76
	<i>-0.59</i>	<i>3.76</i>	<i>4.05</i>	<i>0.47</i>	<i>1.27</i>	—	—
F	0.009	-0.129	-0.053	-0.137	-0.027	0.000	1.89
	<i>2.25</i>	<i>-0.77</i>	<i>-0.35</i>	<i>-0.74</i>	<i>-0.18</i>	—	—
MACRO	-0.002	0.349	0.381	-0.080	0.204	0.161	1.98
	<i>-0.60</i>	<i>3.96</i>	<i>3.54</i>	<i>-0.60</i>	<i>1.82</i>	—	—
OI	0.006	0.360	0.256	-0.185	0.102	0.700	1.60
	<i>3.35</i>	<i>4.78</i>	<i>3.82</i>	<i>-2.22</i>	<i>1.46</i>	—	—
SHS	0.003	-0.922	-0.359	0.320	-0.103	0.833	1.87
	<i>1.09</i>	<i>-7.45</i>	<i>-3.26</i>	<i>2.33</i>	<i>-0.90</i>	—	—
SPECG	-0.002	0.532	0.459	0.098	0.176	0.453	1.78
	<i>-0.49</i>	<i>4.27</i>	<i>4.13</i>	<i>0.71</i>	<i>1.52</i>	—	—

Abbreviation: TSLS, two-stage least squares.

Note: Statistics *t* are in italics.

instruments we proposed in this article, which are the higher moments of those variables, tends to validate these points. Moreover, a TSLS run on hedge funds returns using conventional instruments performed very poorly with respect to a TSLS using higher moments instruments.

Many recent articles criticize the F&F model.⁴⁵ Our article provides instead arguments supporting this model. The three factors, *SMB*, *HML* and *UMD*, of the F&F model have their justification in the explanation of risk, because they aggregate the many nonlinearities that are present in the distribution of returns. They are not market anomalies as they were considered in the past. They are instead ‘réservoirs’ of moments and co-moments risks.

Our study also sheds light on the idiosyncratic behaviour of hedge funds. The beta of hedge

funds is not generally high as it does not exceed 0.25. The short sellers differ on that matter, having a mean beta near -1. For hedge funds, the three factors *SMB*, *HML* and *UMD* are definitively factors of risk, their estimated signs being usually positive and significant in the regressions of the funds returns on these factors. After the market risk premium, the factor *SMB* exerts the most prominent positive impact on the returns of the hedge funds of our sample. This factor allows identifying overvalued and undervalued securities, a useful information for the investor.

Globally, the problem of errors-in-variables does not seem to be too serious in our sample of hedge funds. This problem gives rise to an overstatement of the impact of the risk premium and to an understatement of the influence of

Table 14: Estimation of Funds plagued with errors in variables by the TSLS method with classic instruments

<i>Fund</i>	<i>c</i>	$R_m - R_f$	<i>SMB</i>	<i>HML</i>	<i>UMD</i>	<i>adj. R</i> ²	<i>DW</i>
DTG	0.008	0.018	0.064	-0.285	-0.114	0.000	2.07
	<i>2.12</i>	<i>0.10</i>	<i>0.29</i>	<i>-0.96</i>	<i>-0.54</i>	—	—
EM	-0.011	0.849	1.014	0.483	0.811	0.000	2.05
	<i>-1.35</i>	<i>2.08</i>	<i>2.12</i>	<i>0.74</i>	<i>1.74</i>	—	—
F	0.014	-0.231	-0.053	-0.486	-0.372	0.000	2.00
	<i>2.17</i>	<i>-0.72</i>	<i>-0.14</i>	<i>-0.96</i>	<i>-1.02</i>	—	—
MACRO	-0.003	0.280	0.242	0.144	-0.400	0.185	2.05
	<i>-0.77</i>	<i>1.38</i>	<i>1.02</i>	<i>0.44</i>	<i>1.73</i>	—	—
OI	0.005	0.394	0.116	0.120	0.236	0.557	1.64
	<i>1.46</i>	<i>2.52</i>	<i>0.63</i>	<i>0.48</i>	<i>1.32</i>	—	—
SHS	0.003	-0.930	-0.324	0.224	-0.032	0.818	1.80
	<i>0.64</i>	<i>-4.23</i>	<i>-1.26</i>	<i>0.64</i>	<i>-0.13</i>	—	—
SPECG	-0.007	0.636	0.637	0.422	0.595	0.125	2.03
	<i>-1.34</i>	<i>2.36</i>	<i>2.02</i>	<i>0.98</i>	<i>1.94</i>	—	—

Abbreviation: TSLS, two-stage least squares.

Note: Statistics *t* are in italics.

SMB. While the impact of *HML* on hedge funds returns is quite low, the incidence of *UMD* is more questionable. The correlation between the coefficients adjusted and unadjusted for the problem of errors-in-variables is negligible in the case of *UMD*.

In summary, our new empirical version of the F&F model relying on a new version of the TSLS technique incorporating an Hausman test seems quite promising. Another procedure to account for error-in-variables is to use optimal combinations of cumulants⁴⁶ estimators, in addition to the combination of Durbin and Pal estimators we use in Racicot and Théoret,^{8,9} instead of higher moments to build the instruments. In finance, we rely usually on moments to quantify risks, but cumulants are

quite promising on this matter as they generalize the theory of moments.

ACKNOWLEDGEMENTS

We thank Prof. Manuchehr Shahrokhi for his comments on an earlier version of this article, which was awarded the Best Paper Award at the GFC-2007.

REFERENCES AND NOTES

- 1 Fama, E.F. and French, K.R. (1992) The cross-section of expected stock returns. *Journal of Finance* 47(2): 427–465.
- 2 Fama, E.F. and French, K.R. (1993) Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1): 3–56.

- 3 Fama, E.F. and French, K.R. (1997) Industry costs of equity. *Journal of Financial Economics* 43(2): 153–193.
- 4 The first study on the treatment of errors-in-variables in the framework of asset pricing models was done by Fama and Macbeth.¹¹ Other posterior papers on this subject are: Shanken (1992); Ferson, W.E. and Locke, D.H. (1998) Estimating the cost of capital through time: An analysis of the sources of error. *Management Science* 44(4): 485–500, and Pastor, L. and Stambaugh, R. (1999) Costs of equity capital and model mispricing. *Journal of Finance* 54(6): 67–121.
- 5 Hausman, J.A. (1978) Specification tests in econometrics. *Econometrica* 46(6): 1251–1271.
- 6 In his PhD thesis, Racicot (2003) transposed a similar form of this model to the CAPM. Instead of using higher moments as instruments, he used cumulants. See: Racicot, F.É. (2003) Trois essais sur l'analyse des données économiques et financières. PhD thesis, chapter 3, University of Quebec, Montreal, Quebec.
- 7 Prudence means 'aversion to downside risk' in finance, which seems to be a universal virtue.
- 8 Racicot, F.É. and Théoret, R. (2008) On comparing hedge fund strategies using new Hausman-based estimators. *Journal of Derivatives & Hedge Funds* 14(1): 9–30.
- 9 In Racicot and Théoret (2010), we show that we can get better instruments by not lagging the higher moment instruments. Racicot, F.É. and Théoret, R. (2010) Optimal instrumental variables generators based on improved Hausman regression, with an application to hedge fund returns. *Journal of Wealth Management* 13(1): 103–123.
- 10 Campbell, J.Y., Lo, A.W. and MacKinlay, C. (1997) *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.
- 11 Fama, E.F. and MacBeth, J.D. (1973) Risk, return, and equilibrium: empirical tests. *Journal of Political Economy* 81(2): 607–636.
- 12 Shanken, J. (1992) On the estimation of beta pricing models. *Review of Financial Studies* 5(1): 1–33.
- 13 Kandel, S. and Stambaugh, R. (1995) Portfolio inefficiency and the cross-section of expected returns. *Journal of Finance* 50(1): 157–184.
- 14 Durbin, J. (1954) Errors in variables. *International Statistical Review* 22(1/3): 23–32.
- 15 Pal, M. (1980) Consistent moment estimators of regression coefficients in the presence of errors in variables. *Journal of Econometrics* 14(3): 349–364.
- 16 See also: Dagenais, M.G. (1994) Parameter estimation in regression models with errors in the variables and autocorrelated disturbances. *Journal of Econometrics* 64(1–2): 145–163, and Racicot (2003)⁶.
- 17 Dagenais, M.G. and Dagenais, D.L. (1997) Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics* 76(1–2): 193–221.
- 18 Racicot, F.É. (1993) Techniques alternatives d'estimation et tests en présence d'erreurs de mesure sur les variables explicatives. MSc thesis, Département de sciences économiques, Université de Montréal.
- 19 Huang, C.F. and Litzenberger, R.H. (1988) *Foundations for Financial Economics*. New York: Prentice-Hall.
- 20 Ingersoll Jr, J. (1987) *Theory of Financial Decision Making*. New York: Rowman and Littlefield Publishers.
- 21 Levy, H. (1992) Stochastic dominance and expected utility: Survey and analysis. *Management Science* 38(4): 555–574.
- 22 Samuelson, P.A. (1970) The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. *Review of Economic Studies* 37(4): 537–542.
- 23 On the three-moment and four-moment CAPM, see: Lhabitant, F.S. (2004) *Hedge Funds: Quantitative Insights*, chapter 8. New York: Wiley; Harvey, C.R. and Siddique, A. (2000) Conditional skewness in asset pricing tests. *Journal of Finance* 55(2): 1263–1295; Lim, K.-G. (1989) A new test of the three-moment capital asset pricing model. *Journal of Financial and Quantitative Analysis* 24(2): 205–216; Kraus, A. and Litzenberger, R. (1976) Skewness preference and the valuation of risk assets. *Journal of Finance* 31(3): 1085–1100 and Rubinstein, M. (1973) The fundamental theorem of parameter-preference security valuation. *Journal of Financial and Quantitative Analysis* 8(1): 61–69.
- 24 Kraus, A. and Litzenberger, R. (1976) Skewness preference and the valuation of risk assets. *Journal of Finance* 31(3): 1085–1100.
- 25 Friend, I. and Westerfield, R. (1980) Co-skewness and capital asset pricing. *Journal of Finance* 35(4): 897–914.
- 26 Sears, R.S. and Wei, K.C.J. (1985) Asset pricing, higher moments, and the market risk premium: A note. *Journal of Finance* 40(4): 1251–1253.
- 27 See the survey of Levy¹⁹ on stochastic dominance.
- 28 Scott, R.C. and Hovarth, P.A. (1980) On the direction of preference for moments of higher order than the variance. *Journal of Finance* 35(4): 915–919.
- 29 For example, the fifth moment is the asymmetry sensitivity of the fourth one. The seventh moment is the sign of the convexity change as the underlying asset moves up or down. For Taleb, moments higher than the fourth order are especially useful for compound options. See: Taleb, N. (1997) *Dynamic Hedging*. New York: Wiley, pp. 202–204.
- 30 For instance, kurtosis measures negative convexity, which is bad from the point of view of a risk averter.
- 31 This presentation follows quite closely Judge, G.G. Griffiths, W.E. Carter Hill, R. Lütkepohl, H. and

- Lee, T.C. (1985) *The Theory and Practice of Econometrics*. New York: Wiley.
- 32 The original F&F model contained only the first two 'anomalies'. The momentum anomaly, which is due to Carhart, M.M. (1997) On persistence in mutual fund performance. *Journal of Finance* 52(1): 57–82, and Jegadeesh, N. and Titman S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48(5): 65–91, was introduced subsequently, to form the augmented Fama and French model.
- 33 The reader will excuse us for confounding at this stage higher moments and cumulants for the sake of the presentation. We will come back on the distinction between these two concepts later.
- 34 See also L'Habitant (2004) on this point, chapter 8. Sometimes, the higher order moments are expressed in deviations from the mean. According to this formulation, the four-moment CAPM would be: $E(R_i) - R_0 = \beta_1[E(R_m) - R_0] + \beta_2[R_m - E(R_m)]^2 + \beta_3[R_m - E(R_m)]^3$.
- 35 On the Hausman test, see: Hausman (1978)⁵; Wu, D. (1973) Alternative tests of independence between stochastic regressors and disturbances. *Econometrica* 41(3): 733–750; MacKinnon, J.G. (1992) Model specification tests and artificial regressions. *Journal of Economic Literature* 30(1): 102–146; Johnston, J. and Dinardo J. (1997). *Econometric Methods*, 4th edn. New York: McGraw Hill; and Pindyck, R.S. and Rubinfeld, D.L. (1998). *Econometric Models and Economic Forecasts*, 4th edn. New York: Irwin-McGraw-Hill. A very good presentation of the version of the Hausman test using an artificial regression in the context of correction of errors in variables may be found in Pindyck and Rubinfeld (1998). This presentation is done only for one explanatory variable.
- 36 Therefore, the Hausman test is an orthogonality test, that is it aims at verifying if $p \lim (1/T) X'\varepsilon = 0$ in large samples. But, to paraphrase Spencer, D.E. and Berk, K.N. (1981) A limited information specification test. *Econometrica* 49(4): 1079–1085; and Taleb, N. (1997) *Dynamic Hedging*. New York: Wiley, the alternative hypothesis of this test is broad and diffuse. Absence of orthogonality might be due to many causes: omission of relevant explanatory variables, errors in variables, inappropriate aggregation over time, simultaneity and incorrect functional form. But, according to these authors, it is true for all existing tests of the orthogonality assumption and it seems that the solution of this problem requires some *a priori* information. For our purpose, this *a priori* information is the potential presence of errors in variables. We want to verify if the endogeneity of some variables, the variables measured with errors, has any significant effect on the parameters estimation.
- 37 MacKinnon, J.G. (1992) Model specification tests and artificial regressions. *Journal of Economic Literature* 30(1): 102–146.
- 38 As it is standard in econometrics, we use the asterisks for unobserved variables.
- 39 Racicot (2003)⁶ transposes this approach to the market model. He assumes that the t test associated to the new variable \hat{u} is distributed asymptotically as the normal distribution. According to Pindyck and Rubinfeld (1998),³³ this test is adequate. Racicot (2003)⁶ also assumes in this context that the new model resulting from the addition of the artificial variable may be viewed as a new model by itself, so we have a new alpha for this model. We follow a similar approach in this article.
- 40 The French's website is: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
- 41 A short put option carries the risk of rare but large losses.
- 42 We note that the correlation between *SMB* and the annual growth of industrial production is quite high, being equal to -0.23 . This correlation suggests that the *SMB* factor would have a positive impact on returns in recessions.
- 43 It seems relevant to consider also the φ which are significant at the 10 per cent level.
- 44 Note that usually, the spread is less than φ .
- 45 For example, Chung, Y.P., Johnson, H. and Schill, M.J. (2001) Asset Pricing When Returns are Nonnormal: Fama French Factors vs. Higher Order Systematic Co-moments. Working Paper, SSRN, noted that *SMB* and *HML* become not significant or less significant when moments or co-moments are taken into account. In that respect, F&F factors might simply be good proxies for the higher moments of the return distribution. But our regression technique, which uses co-moments as instruments, suggests that they capture these co-moments and also other influences.
- 46 On this subject, see: Malevergne, Y. and Sornette, D. (2005) Higher moments portfolio theory, capitalizing on behavioral anomalies of stock markets. *Journal of Portfolio Management* 31(4): 49–55.

APPENDIX

List of funds included in the study

- *USEMN*: Equity market neutral
- *USED*: Event driven

- *USDS*: Distressed securities
- *USSS*: special situations
- *USMMNA*: Market neutral arbitrage
- *USSHS*: short selling index
- *USVA*: Value index
- *USF*: futures index
- *USMACRO*: Macro
- *USMT*: Market timing
- *USDTG*: Directional trading group
- *USMNG*: Market neutral group
- *USLSG*: long short group
- *USAG*: aggressive growth
- *USOI*: opportunity index
- *USSPECG*: special strategies
- *USEM*: emerging markets
- *USINC*: fixed income
- *USMSI*: multi strategy index
- *USHF*: hedge funds index