ESG UQÀM

Cahier de recherche

2008-05

Chaire d'information financière et organisationnelle

Optimal Instrumental Variables Generators Based on Improved Hausman Regression, with an Application to Hedge Fund Returns

François-Éric Racicot[†]

Raymond Théoret[††]

[†] *Associate Professor, Department of administrative sciences, University of Quebec (Outaouais), (UQO), 101 St-Jean-Bosco Street, Gatineau (Hull), Quebec, Canada, J8X 3X7. E-mail: francoiseric.racicot@uqo.ca, Tel. number: (819) 595-3900 ext. 1727.*
[††] *Professor, Department of finance, University of Quebec (Montreal), (UQAM), 315 Ste-Catherine East, Montreal, Quebec, Canada. H3X 2X2. E-mail: theoret.raymond@uqam.ca. Tel. number: (514) 987-3000 ext. 4417.*

Résumé

Dans cet article, nous proposons de nouveaux estimateurs inspirés d'Hausman et qui reposent sur des instruments optimaux construits à partir de cumulants. En utilisant ces nouveaux instruments robustes dans un contexte de GMM, nous obtenons de nouveaux estimateurs GMM que nous appelons GMM-C et GMM-hm, son homologue. Cette procédure représente une amélioration prometteuse de la méthode des moments pour identifier les paramètres d'un modèle. De plus, notre étude développe un nouvel indicateur qui signale la présence d'erreurs de spécification à l'intérieur d'un modèle économique ou financier. Nous appliquons notre batterie de tests et d'estimateurs à un échantillon d'indices de fonds de couverture HFR observés mensuellement au cours de la période 1990-2005. Nos tests révèlent que des erreurs de spécification contaminent l'estimation des paramètres de modèles financiers de rendements. Par conséquent, il n'est pas surprenant que la notation des fonds de couverture soit très sensible au choix des estimateurs. Notre nouvel indicateur des erreurs de spécification se révèle lui-même très puissant pour détecter ces erreurs.

Mots-clés : Modèles d'évaluation des actifs; erreurs de spécification; tests d'Hausman; moments supérieurs; instruments optimaux; GMM.

Abstract

This paper proposes new Hausman-based estimators lying on cumulants optimal instruments. Using these new generated strong instruments in a GMM setting, we obtain new GMM estimators which we call GMM-C and its homologue, the GMM-hm. This procedure improves the method of moments for identifying the parameters of a model. Also, our study gives way to a new indicator signalling the presence of specification errors in financial models. We apply our battery of tests and estimators to a sample of 22 HFR hedge fund indices observed monthly over the period 1990-2005. Our tests reveal that specification errors corrupt parameters estimation of financial models of returns. Therefore, it is not surprising that the ranking of hedge funds is very sensitive to the choice of estimators. Our new indicator of specification errors reveals itself very powerful to detect those errors.

Keywords: Asset pricing models; specification errors; Hausman tests; higher moments; optimal instruments; GMM.

JEL classification: C13; C19; C49; G12; G31.

1. Introduction

The presence of specification errors is an important problem when estimating economic and financial models. Those errors may emanate from many sources. An obvious one is due to the fact that the variables of the theoretical models are often formulated in expected or forecasted values. As only observed values of the variables are available, the explanatory variables of a financial model, like the CAPM, are then measured with error (Cragg, 1994, 1997; Racicot, 2003; Coën and Racicot, 2007). Another source of specification errors in financial models is the neglect of higher moments and co-moments of the variables when estimating those models (Harvey and Siddique, 2000; Chung, Johnson and Schill, 2006). We could lengthen the list of the causes of specification errors such as the omission of an important factor when formulating a financial model like the illiquidity premium (Chan and Faff, 2005). As it is well-known, those errors lead to inconsistent estimators of the factor loadings of the risk variables incorporated in a financial model. Those errors may upset completely the conclusions derived from an OLS estimation of a model (Dagenais and Dagenais, 1997).

The solutions to the problem of specifications errors are yet limited. The Generalized Method of Moments (GMM) is often used in finance to correct specification errors (Hansen and Singleton, 1982; Hansen and Jagannathan, 1997; Cochrane, 2001) but resorting to this procedure requires a judicious choice of instruments. However, the instruments used in the majority of financial studies may be considered as weak (Watson, 2003). Even the Chen-Roll-Ross (1986) instruments derived from the arbitrage pricing theory (APT) are not very reliable.

In this paper, we revisit financial models of returns in the framework of the estimation of hedge fund returns whose distributions display a high degree of skewness and kurtosis . We resort to two new sets of instruments based respectively on higher moments and cumulants of the explanatory variables. Those instruments, which are borrowed from the method of moments (Geary, 1942; Durbin, 1954; Pal, 1981; Dagenais and Dagenais, 1997; Gillard, 2006; Gillard and Iles, 2005) are quite promising as tools to analyse the distribution of returns or of other financial variables for which the asymmetry or kurtosis cannot be disregarded. As we will show in this paper, higher moments and cumulants thus qualify as optimal instrumental variables (IV) to estimate financial models like the CAPM or the Fama and French (F&F) one by two-stage least squares (TSLS) or GMM.

Furthermore, the Hausman (1978) test is often invoked to detect specification errors in an estimated model. It is less known that a specific version of the Hausman test based on an artificial regression may be equivalent to a TSLS procedure (Pindyck and Rubinfeld, 1998; Spencer and Berk, 1981; Wu, 1973). Resorting to this equivalence, we show how we can use our new sets of higher moment or cumulant instruments to generate innovative versions of financial models which give direct information on the severity of the specification errors. We thus propose new procedures to generate strong instruments based on higher moments and cumulants of the explanatory variables. By doing so, we rehabilitate the well-known estimator (GMM), which uses these innovative instruments and we thus construct our new estimator, called the GMM-C, and its homologue, the GMM-hm (Racicot and Théoret 2008a and 2008b). These developments allow us to build new indicators of measurement errors, one for each explanatory variable.

This paper is organised as follows. The next section proposes two new sets of instruments based respectively on higher moments and cumulants and provides the econometric and financial foundations of these instruments. The third section shows how these instruments may be integrated in a Hausman test to give way to indicators of specification errors in the framework of the CAPM or the F&F model. These indicators are related to the spread between the coefficients estimated by ordinary least squares (OLS) and by an IV method. They supply direct information on the degree of overstatement or understatement of the coefficients estimated by the OLS procedure. The fourth section deals with the empirical validity of these Hausman-based estimators for calibrating the F&F model. Our sample consists in a series of 22 monthly Hedge Fund Research (HFR) indices observed over the period 1990-2005, a quite long period for hedge fund returns given the short length of most hedge fund series. The final section concludes.

2. In search of higher moment optimal instruments to detect and correct specification errors[1]

As said previously, the aim of this paper is to propose an integrative method to detect and correct specification errors in a financial model. To formulate this method, we appeal to the econometric theory of measurements errors. According to Gillard and Iles (2005), the word error is an unfortunate choice for designating what should be meant by this word: a disturbance, a departure from equilibrium, a perturbation, a noise or a random component. Actually, the methods used to detect measurement errors are a subset of those used to identify specification errors. Both are concerned with the orthogonality between the explanatory variables and the innovation of a model and the resulting inconsistency of the estimators. That is why we speak indifferently of specification errors and measurement errors in this paper.

Let us assume that a variable, say a return, is measured with error, that is:

$$R = \widetilde{R} + v \qquad (1)$$

with $\nu \sim N(0,1)$. In equation (1), R is the observed return variable and $\widetilde{R}$ is the unobserved or latent variable so that v is the measurement error. Let us introduce this variable in the simple market model, that is:

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it} \qquad (2)$$

---

[1] For this section, see also: Théoret and Racicot (2007); Racicot and Théoret (2006, 2007, 2008); Coën and Racicot (2007), Racicot, Théoret and Coën (2007).

where r_{it} is the excess return[2] of portfolio i; r_{mt}, the excess return of the market portfolio and ε_t, the innovation. Obviously, the variable r_m is measured with error because the financial models of returns are formulated using expected values of the variables and not observed ones[3]. Therefore, there is an absence of orthogonality between r_m and ε, which gives way to an inconsistent estimator of the parameter of r_m. We must thus search for additional information to identify the parameter of r_m.

For the method of moments[4], which was initiated by Geary (1942) for analysing measurement errors, this new information must be found in the higher moments of the variables of the estimated model. When there are no measurement errors, the first and second-order moments are sufficient to estimate equation (2). The resulting ordinary least squares (OLS) estimators of α_i and β_i are then:

$$\hat{\alpha}_i = \bar{r}_i - \hat{\beta}\bar{r}_m \qquad (3)$$

$$\hat{\beta}_i = \left(s^2_{r_m}\right)^{-1} s_{r_i r_m} \qquad (4)$$

with $\bar{r}_i$ and $\bar{r}_m$ being respectively the first-order moments of r_i and r_m and where:

$$\left(s^2_{r_m}, s_{r_i r_m}\right) = (n-1)^{-1} \sum_{t=1}^{n} \left(r_{mt} - \bar{r}_m, r_{it} - \bar{r}_i\right)\left(r_{mt} - \bar{r}_m\right) \qquad (5)$$

with n, the number of observations on the variables. $s^2_{r_m}$ is the second-order moment of r_m and $s_{r_i r_m}$, the second-order co-moment of r_i and r_m.

In the presence of measurement errors, one possibility is to resort to moments of order higher than two to identify the parameters of a model. To do so, the method of moments defines estimators by equating the sample moments to their population equivalents which contain the parameters to be estimated. More specifically, the method of moments creates

[2] That is the spread between the gross return of the portfolio and the risk-free rate.
[3] Actually, there are other causes of measurement errors for this variable.
[4] On the relationship between the method of moments and the identification of the parameters in presence of errors in variables, see: Gillard and Iles (2005) and Gillard (2006).

estimators by equating the sample moments to their expectations (Fuller, 1987; Cheng and Van Ness, 1999), that is:

$$\hat{m} = m(\theta) \qquad (6)$$

where $\hat{m}$ are the sample moments and $m(\theta)$ are the population moments[5] which incorporate the parameters θ to be estimated. Durbin (1954) defined an estimator based on third order-moment and co-moment. In the case of our market model (equation 2), this estimator, designated by $\hat{\beta}_d$, would be:

$$\hat{\beta}_d = \frac{s_{r_m r_m r_i}}{s^3_{r_m}} \qquad (7)$$

where $s_{r_m r_m r_i}$ is the third-order co-moment between r_i and r_m and $s^3_{r_m}$ is the third-order moment of r_m, defined as:

$$\left(s^3_{r_m}, s_{r_m r_m r_i}\right) = (n-1)^{-1}\sum_{i=1}^{n}\left((r_m - \bar{r}_m)\ , r_i - \bar{r}_i\right)(r_m - \bar{r}_m)^2 \qquad (8)$$

Furthermore, Pal (1980) defined an estimator based on the fourth-order moment and co-moment of the variables of a model defined by:

$$\hat{\beta}_p = \frac{s_{r_m r_m r_m r_i} - 3\left(\sum r^2_{mt}/n\right)s_{r_m r_i}}{s^4_{r_m} - 3\left(\sum r^2_{mt}/n\right)s^2_{r_m}} \qquad (9)$$

where the subtracted terms account for the normality of ν in equation (1). This estimator may be seen as a ratio of cumulants because its numerator and denominator combine moments and co-moments of different orders.

2.1 The higher moment instrumental variables and estimation methods

We would now define our first class of estimators which we call the higher moment estimators. There is obviously a mapping which may be done from the method of moments to the instrumental variables (IV) estimation procedures. Let us consider equation (7), the

[5] That is the expected values of the sample moments.

Durbin estimator obtained by the method of moments. We know that the corresponding two-stages least squares estimator is equal to:

$$\hat{\beta}_d = \left(s_{z,r_m}\right)^{-1}\left(s_{z,r_i}\right) \qquad (10)$$

where z is the instrumental variable of r_m, which is measured with error. By equating equations (7) and (10), it is obvious in this case that the instrument is $\left(r_m - \bar{r}\right)^2$, which is the squared value of r_m expressed in deviation from its mean, an empirical measure of the second-order moment of r_m. And we could do the same exercise for the Pal's beta (Racicot, 1993, 2003).

There is another easy way to show that higher moments are relevant instrumental variables just by resorting to the definition of such variables (Fuller, 1987). If, in the framework of our market model, a variable z_t satisfies the following conditions:

$$E\left\{n^{-1}\sum_{t=1}^{n}\left(z_t - \bar{z}\right)\left(\varepsilon_t, \nu_t\right)\right\} = (0,0);\ E\left\{n^{-1}\sum_{t=1}^{n}\left(z_t - \bar{z}\right)r_{mt}\right\} \neq 0 \qquad (11)$$

where $\bar{z} = n^{-1}\sum_{t=1}^{n} z_t$, then z_t is called an instrumental variable for r_{mt} of the market model. Moreover, we know that the distributions of returns are not normal as it is assumed in the classical econometric model. Two stylised facts of asset returns, especially hedge fund returns which are the object of this paper, are that their distributions are asymmetric and leptokurtic. It is precisely in this situation that we may resort to moments and co-moments of order higher than two to identify the coefficients of explanatory variables contaminated with measurement errors (Cheng and Van Ness, 1999). As the distribution of r_m is asymmetric, we may write:

$$E\left\{\left(r_{mt} - \mu_{r_m}\right)^3\right\} \neq 0 \qquad (12)$$

with μ_{r_m}, the expected value of r_m. This knowledge allows defining an instrumental variable for r_m (Fuller, 1987). If we set $z_t = \left(r_{mt} - \bar{r}_m\right)^2$, then:

$$E\left\{\left(r_{mt} - \mu_{r_m}\right)\left(z_t - \mu_z\right)\right\} = \left(1 - n^{-1}\right)^2 E\left\{\left(r_{mt} - \mu_{r_m}\right)^3\right\} \neq 0 \qquad (13)$$

and by the properties of the normal distribution:

$$E\{z_t \varepsilon_t\} = E\{z_t \nu_t\} = 0 \qquad (14)$$

Thus, the second-order moment, defined by $(r_{mt} - \bar{r}_m)^2$, qualifies as an instrumental variable for r_{mt}. We can follow the same reasoning to show that the third-order moment defined by $(r_{mt} - \bar{r}_m)^3$ may be a relevant instrumental variable for r_{mt} because the distribution of returns is leptokurtic, that is:

$$E\{(r_{m_t} - \mu_{r_m})^4\} \neq 0 \qquad (15)$$

According to Fuller (1987), the co-moment $(r_{it} - \bar{r}_i)(r_{mt} - \bar{r}_m)$ and the second-order moment of the dependent variable, here $(r_{it} - \bar{r}_i)^2$, may also be used as instruments.

Lewbel (1997) and Cragg (1994, 1997) have generalized the transposition of the method of moments to instrumental variables as exposed by Fuller (1987). Lewbel said that his paper is an extension of the method proposed by Dagenais and Dagenais (1997) to the generation of optimal instrumental variables but we only partially agree with this allegation because the Dagenais' method is the combination of previous estimators built in the context of the method of moments, especially the estimators of Durbin and Pal (equations 7 and 9). Thus, we will deal with the Dagenais' method in the next section, which constitutes a separate class of estimators.

Lewbel (1997) assumes that there exists an instrumental variable, say *z*, to estimate a one-variable model, say equation (2). This variable is different from the explanatory variable, here r_m, the excess return of the market portfolio. Lewbel also postulates the following function: $G = G(z)$, G being a linear or non-linear function of z[6]. Lewbel shows that the TSLS procedure is consistent when it uses as instruments the following moments and co-moments:

$$z_1 = (G_t - \overline{G}) \qquad (16)$$

[6] For instance, *G* may be the logarithm of *z* or even *z* itself.

$$z_2 = \left(G_t - \overline{G}\right)\left(r_{it} - \overline{r}_i\right) \qquad (17)$$

$$z_3 = \left(G_t - \overline{G}\right)\left(r_{mt} - \overline{r}_m\right) \qquad (18)$$

$$z_4 = \left(r_{it} - \overline{r}_i\right)\left(r_{mt} - \overline{r}_m\right) \qquad (19)$$

$$z_5 = \left(r_{it} - \overline{r}_i\right)^2 \qquad (20)$$

$$z_6 = \left(r_{mt} - \overline{r}_m\right)^2 \qquad (21)$$

We can resort to an IV method other than the TSLS to estimate the parameters θ of a model. Cragg (1997) shows that the parameters can be estimated by the following procedure:

$$\tilde{\theta} = \underset{\theta}{arg\,min}\left\{\left[\hat{m} - m(\theta)\right]' W \left[\hat{m} - m(\theta)\right]\right\} \qquad (22)$$

with $\hat{m}_{ij} = \sum_{t=1}^{T} \frac{\left(y_t - \overline{y}\right)^i \left(x_t - \overline{x}\right)^j}{T}$ and W, a weighting matrix. To weight their instrumental variables, Dagenais and Dagenais [1997] resorted to an IV estimator suggested by Fuller (1987). As we will see in the section dedicated to the Dagenais and Dagenais (1997) estimator, which will be also used in this paper to estimate the hedge fund returns, it is very important to weight the higher moments because these moments might be erratic due to the presence of outliers in the financial return series. For our part, we will resort to the GMM method to weight the moment conditions related to the instrumental variables. We call this estimator: GMM-hm, using this notation to indicate that we resort to higher moments as instruments for running the GMM.

Using higher moments as risk factors has solid foundations in the financial literature (Malevergne and Sornette, 2005). Samuelson (1970), Rubinstein (1973) and Kraus and Litzenberger (1976) have developed an asset pricing model which incorporates higher moments of the market risk premium to take into account the non-normality of the distribution of returns, at least over short periods. This model was named the n-factor CAPM or alternatively, the n-moment CAPM. Rubinstein (1973) proposed the following version of the n-moment CAPM by assuming that the investors' expectations are identical or homogenous:

$$E(R_i) = R_f + \sum_{j=2}^{n} \lambda_j b_{ij} \qquad (23)$$

with $E(R_i)$ the expected return of a stock or a portfolio and R_f, the riskless rate.

In equation (23), b_{ij} is the systematic co-moment of order j between R_i and R_m, the return of the market portfolio. The parameter λ_j is a market measure of the degree of aversion towards the co-moment of order j. According to the CAPM, only systematic risk is priced by the market, which is measured by the co-moment of order 2 $\left(s_{r_i, r_m}\right)$, that is the covariance between the return of portfolio i and the market one. The non-systematic risk is not priced because it is diversifiable. Scott and Hovarth (1981) have shown that odd moments, like mean and positive asymmetry, provide positive utility to investors while even moments, like variance and kurtosis, provide negative utility.

The works done on the n-CAPM at the beginnings of the 70's have only been revived in the second half of the 90's[7]. Furthermore, the researchers were at this time more preoccupied by the effect of the third moment on risk even if the fourth one seems to be more important when the distribution of returns is not normal. But these gaps in the theory of the n-CAPM have been filled since.

Several authors contributed to the elaboration of the n-moment CAPM. Rubinstein (1973), Ingersoll (1975, 1987), Kraus and Litzenberger (1976), Lim (1989) and Harvey and Siddique (2000) built the three-moment CAPM while Hwang and Satchell (1999) and Dittmar (2002) formulated the four-moment CAPM. There are many empirical studies on the three and four-moment CAPM even though these studies still seem to have a predilection for the third moment in spite of the importance of kurtosis in the specification of non-normal returns.

[7] On that matter, see the comments of Rubinstein in: Jurczenko and Maillet (2006).

These papers gave way to the quadratic and the cubic CAPM, respectively the empirical versions of the third and fourth moment CAPM. The empirical version of the fourth moment CAPM, also called the cubic CAPM, may be written as follows:

$$R_{it} - R_{ft} = \alpha_p + \beta_p\left(R_{mt} - \overline{R}_m\right) + \gamma_p\left(R_{mt} - \overline{R}_m\right)^2 + \delta_p\left(R_{mt} - \overline{R}_m\right)^3 + \varepsilon_t \quad (24)$$

The coefficient γ_p is the exposition or loading of a portfolio to co-asymmetry while the coefficient δ_p is its exposition to co-kurtosis. According to the empirical specification of the CAPM given by equation (24), the higher moments are powers of returns as suggested by the method of moments. The empirical literature has found that higher moments are very relevant to explain non-Gaussian returns[8]. It is thus completely in line with our previous discussions about the use of higher moments of R_{mt} as instrumental variables to correct the first moment of the return of the market portfolio expressed as deviation from its mean for its measurement or other specification errors, that is its correlation with ε_t. In view of our theoretical developments on the use of higher moments as instrumental variables, equation (24) does not however exhaust the list of the possible higher moment instruments.

Higher moment estimators will be used in this paper to estimate the augmented version of the Fama and French (1992, 1993, 1997) model on a sample of hedge fund returns. This model may be written as follows:

$$R_{pt} - R_{ft} = \alpha + \beta_1\left(R_{mt} - R_{ft}\right) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \mu_t \quad (25)$$

with $(R_{pt} - R_{ft})$ being the excess return of a portfolio over the risk-free rate (R_{ft}) and $(R_{mt} - R_{ft})$ being the market risk premium. The risk factors related to small cap, growth and momentum exposures are measured by their corresponding F&F (1992, 1993, 1997) mimicking portfolios[9]. *SMB* is a portfolio which mimics the "small firm anomaly", which is long in the

[8] See, for example: Chung, Johnson and Schill (2006).

[9] The original F&F model contained only the first two "anomalies". The momentum anomaly, which is due to Carhart (1997) and Jegadeesh and Titman (1993), was introduced subsequently, to form the augmented Fama and French model.

returns of selected small firms and short in the returns of selected big firms; *HML* is a portfolio which mimics the "income stock anomaly", which is long in returns of stocks of selected firms having a high (book value/ market value) ratio (value stocks) and short in selected stocks having a low (book value/ market value) ratio (growth stocks) and *UMD*, a portfolio which mimics the "momentum anomaly", which is long in returns of selected stocks having a persistent upper trend and short in stocks having a persistent downward trend.

We postulate that equation (25) contains specification errors which might be due to many causes. The most obvious is that the factors appearing on the RHS of equation (25) should be theoretically expected values and not observed ones. These errors may lead to biases in the estimation of the constant of this equation named Jensen alpha, a statistic which is very important for stock-picking activities or for the analysis of the performance of a portfolio manager, or can result in an understatement or an overstatement of the factor loadings of the basic factor model. These biases may thus upset the conclusions derived from the OLS estimation of equation (25) although the sign and the importance of the biases are not known theoretically because they are mixes of attenuation and contamination effects (Cragg, 1997, 1998). We will suggest in the section dedicated to the Hausman instrumental variable test a new indicator to evaluate the empirical biases caused by specification errors.

In choosing higher moment instrumental variables, we must thus restrict ourselves to a subset of them because there is an infinity of such instruments. We saw earlier that the instrument related to the higher moment of order two is linked to the skewness of the corresponding variable and the instrument of order three, to its kurtosis. As we are concerned above all with the third and fourth moments of the distribution of hedge fund returns, we will not exceed the third order moment while building our instruments. We will also neglect the co-moments because they are too numerous.

Let us suppose that x is one of the explanatory variables of the F&F model and that y is the dependent variable, here an excess hedge fund return. These variables are expressed in deviations from their mean. The retained instruments for x are thus the following subset of all possible instruments: $\left\{x_{t-1}, x_t^2, x_t^3, y_t^2\right\}$. To this subset of instruments, we will add other exogenous variables like the Chen-Roll-Ross (1986) variables.

2.2 The cumulant instrumental variables

In a paper published in 1997, Dagenais and Dagenais proposed a new method to purge a model from its measurement errors. This method was not well understood by the academics and that is perhaps why it was not popularized. We will present the instrumental variables constructed by Dagenais and Dagenais (1997) as cumulants to distinguish them from the higher moment instruments defined earlier. Really, there is no distinction between central moments and cumulants up to the fourth order[10].

The purpose of the article of Dagenais and Dagenais (1997) was to combine the higher moment or cumulant estimators developed especially by Durbin (1954) and Pal (1980). It is well-known that the higher moments of the explanatory variables may be quite erratic. Incidentally, according to Cheng and Van Ness (1999), the estimates of third-order moments have much larger variances than the estimates of second-order moments in moderate samples. Smoothing these moments by combining them may be a way to reduce these fluctuations. That was the approach adopted by Dagenais and Dagenais (1997).

We saw earlier that the instrumental variables associated to β_d, that is the estimator developed by Durbin (1954), are: **x*x**, where **x** stands for the explanatory variables expressed in deviation from their mean and where the symbol ***** designates the Hadamard element by element matrix multiplication operator. Moreover, the instrumental variables

[10] Some instruments used by Dagenais and Dagenais (1997), such as the Pal's ones, are ratios of cumulants.

associated to β_p, that is the estimator developed by Pal (1980), are (Racicot, 1993, 2003) : $\mathbf{x*x*x}$ -3$\mathbf{x[Plim(x'x/N)*I_k]}$, where plim ($\mathbf{x'x}$/N) may be replaced by its sample equivalent, that is ($\mathbf{x'x}$/N), when computing the instrumental variables. Dagenais and Dagenais add to these two instruments other cumulants and co-cumulants which were also used previously as instruments by Durbin and Pal for identifying the parameters of a model contaminated by measurement errors. The complete list of the cumulant instrumental variables proposed by Dagenais and Dagenais appears at table 1.

Insert table 1 about here

There are many ways to combine these instrumental variables so they could become optimal instruments. An obvious one is to use those instruments as inputs to the GMM, a procedure which we will adopt and in doing so we thus extend the methodology developed by Dagenais and Dagenais , these authors having chosen to use the Fuller's (1987) IV estimator in their 1997 paper, which is a standard generalized least squares.

Dagenais and Dagenais noticed in their article that the list of instrumental variables may be reduced to the triplet $\{z_0, z_1, z_4\}$ because their results seem better when using this subset of instruments instead of the whole set reported in table 1. We will follow this procedure in the empirical section of our article to reduce the number of instrumental variables. We have indeed four explanatory variables in our model. The number of instruments related to cumulants thus amounts to eight, that is one z_1 and one z_4 for each explanatory variable of the F&F model (equation 25). The set of z variables is therefore in the case of our model: $\{z_0, z_{11}, \ldots z_{14}, z_{41}, \ldots z_{44}\}$, the first index indicating the nature of the cumulant and the second one designating the explanatory variable.

Incidentally, those instruments are very relevant in the context of our sample which is composed of hedge fund returns. As said previously, z_1, that is the Durbin's instruments, is associated to the degree of asymmetry of the distribution of our returns and z_4, the Pal's instruments, are for their part related to the degree of kurtosis of the return distribution. Those instruments are therefore likely to convey much information on our sample of hedge funds whose returns are particularly asymmetric and leptokurtic. On the other hand, we will construct in the following section dedicated to the Hausman's instrumental variable test new instruments which combine in another way z_1 and z_4 and which are based on the distance between an observed variable x and its fitted value. As we will also use those instruments in the GMM estimation of our model, which is a new way to estimate this model, we will call this method: GMM-z.

3. Test of specification errors in the framework of higher moments

To detect specification errors in our sample of hedge funds, we could use the original Hausman h test[11]. To explain this test, let us suppose the following classical model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (26)$$

with $\mathbf{Y}$ a $(n \times 1)$ vector representing the dependent variable; $\mathbf{X}$, a $(n \times k)$ matrix of the explanatory variables; $\boldsymbol{\beta}$, a $(k \times 1)$ vector 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), and $\boldsymbol{\beta}_{A}$, an alternative estimator which can take a variety of forms but which for our purposes is the instrumental variable estimator designated by $\boldsymbol{\beta}_{IV}$. The hypotheses to test are H_0, being in our case the absence of specification errors and H_1, being the presence of specification errors. The vector of estimates $\boldsymbol{\beta}_{IV}$ is consistent under both H_0 and H_1 but $\boldsymbol{\beta}_{OLS}$ is consistent under H_0 but inconsistent under H_1. Under H_0, $\boldsymbol{\beta}_{IV}$ is obviously less efficient than $\boldsymbol{\beta}_{OLS}$.

Hausman wants to verify if "endogeneity" of some variables[12], the variables measured with errors in our case, has any significant effect on the estimation of the vector

[11] On the Hausman test, see : Hausman (1978), Wu (1973), MacKinnon (1992) and Pindyck and Rubinfeld (1998). 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 for one explanatory variable. We have generalized it to a multivariate model.

of parameters. To do so, he defines the following vector of contrasts or distances: $\boldsymbol{\beta}_{\mathbf{IV}} - \boldsymbol{\beta}_{\mathbf{OLS}}$. The test statistic may be written as follows:

$$h = \left(\hat{\boldsymbol{\beta}}_{\mathbf{IV}} - \hat{\boldsymbol{\beta}}_{\mathbf{OLS}}\right)^T \left[\mathbf{Var}\left(\hat{\boldsymbol{\beta}}_{\mathbf{IV}}\right) - \mathbf{Var}\left(\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}\right)\right]^{-1} \left(\hat{\boldsymbol{\beta}}_{\mathbf{IV}} - \hat{\boldsymbol{\beta}}_{\mathbf{OLS}}\right) \sim \chi^2(g) \qquad (27)$$

with $\mathbf{Var}\left(\hat{\boldsymbol{\beta}}_{\mathbf{IV}}\right)$ and $\mathbf{Var}\left(\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}\right)$ being respectively consistent estimates of the covariance matrices of $\hat{\boldsymbol{\beta}}_{\mathbf{IV}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}$. g is the number of potentially endogenous regressors. 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%.

According to MacKinnon (1992), this test might run into difficulties if the matrix $\left[\mathbf{Var}\left(\hat{\boldsymbol{\beta}}_{\mathbf{IV}}\right) - \mathbf{Var}\left(\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}\right)\right]$, which weights the vector of contrasts, is not positive definite. Fortunately, there is an alternative way to perform the Hausman test which is much easier and which was used in its rudimentary form by Dagenais and Dagenais (1997). However, these authors did not realize that the resulting estimator is equivalent to a TSLS, a result which increases its robustness. That is the matter of the following developments.

3.1 Tests based on higher moments[13]

Assume a four-variable linear model which is in this paper the general form of the F&F model with four factors:

$$y_t = \beta_0 + \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \beta_3 x_{3t}^* + \beta_4 x_{4t}^* + \varepsilon_t \qquad (28)$$

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

The variables x_{it}^* [14] are measured with errors, that is:

$$x_{it} = x_{it}^* + \upsilon_{it} \qquad (29)$$

[12] Therefore, the Hausman test is an orthogonality test, that is it aims to verify if plim (1/T) **X'ε** = 0 in large samples (equation 26).

[13] For this section, see also: Théoret and Racicot (2007); Racicot and Théoret (2007); Coën and Racicot (2007), Racicot, Théoret and Coën (2007) and Racicot & Théoret (2006).

[14] As done usually in econometrics, we use the asterisks for the unobserved variables.

with x_{it}, the corresponding observed variables which are measured with errors. By substituting equation (29) in equation (28), we have:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \varepsilon_t^* \qquad (30)$$

with $\varepsilon_t^* = \varepsilon_t - \beta_1 \upsilon_{1t} - \beta_2 \upsilon_{2t} - \beta_3 \upsilon_{3t} - \beta_4 \upsilon_{4t}$. As seen before, estimating coefficients of equation (30) by the OLS method results in biased and inconsistent coefficients because the explanatory variables are correlated with the innovation.

Consistent estimators can be found if we can identify an instrument vector $\mathbf{z_t}$ which is correlated with every explanatory variable but not with the innovation of equation (30). These estimators are the higher moments developed in the preceding section. Then we regress these four explanatory variables on $\mathbf{z_t}$. We have:

$$x_{it} = \hat{x}_{it} + \hat{w}_{it} = \hat{\gamma}_i \mathbf{z_t} + \hat{w}_{it} \qquad (31)$$

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 equation (31) into equation (30), we have:

$$y_t = \beta_0 + \beta_1 \hat{x}_{1t} + \beta_2 \hat{x}_{2t} + \beta_3 \hat{x}_{3t} + \beta_4 \hat{x}_{4t} + \beta_1 \hat{w}_{1t} + \beta_2 \hat{w}_{2t} + \beta_3 \hat{w}_{3t} + \beta_4 \hat{w}_{4t} + \varepsilon_t^* \qquad (32)$$

The explanatory variables of this equation are, on the one hand, the estimated values of x_{it}, obtained by regressing these four variables on the vector of instruments $\mathbf{z_t}$, and on the other hand, the respective residuals of these regressions. Equation (32) is therefore an augmented version of equation (30), which might be qualified of auxiliary or artificial regression.

Racicot (2003) was the first to apply this approach to the market model. He postulates that the *t* test issued from the new variable $\hat{w}$ is distributed asymptotically as the normal distribution. According to Pindyck and Rubinfeld (1998), this test is adequate. Furthermore, when generalized to a multivariate dimension, this approach is different from the one proposed by Dagenais and Dagenais (1997) which suggested to run the test on the whole set of coefficients of $\hat{w}_{it}$ by resorting to an *F* test, the usual procedure in this case. Moreover, the Dagenais and Dagenais method was developed for a sample of cross-section data, what is usually the case when resorting to higher moment instrumental variables. Racicot (2003) transposed these asymptotical results to financial time series. He also postulates in this context that the new model resulting from the addition of the artificial variable to the standard market model may be considered as a new model by itself, so we have a new Jensen alpha for this model.

It can be shown that that:

$$p\,lim\left[\frac{\sum \hat{w}_{it}\varepsilon_t^*}{N}\right] = p\,lim\left[\frac{-\beta_1 \sum x_{it}\upsilon_{it}}{N}\right] = -\beta_i\sigma_{\upsilon i}^2 \qquad (33)$$

When there is no measurement error, $\sigma_{\upsilon i}^2 = 0$ and OLS leads to a consistent estimator for the parameter of $\hat{x}_{it}$ in equation (32), that is β_i. When there are measurement errors, $\sigma_{\upsilon_i}^2 \neq 0$ and therefore this estimator is not consistent.

We can thus build the following test to detect the presence of measurement errors. As we do not know a priori if there are such errors, we replace the coefficients β_i of the $\hat{w}_{it}$ variables in equation (32) by θ_i. We thus have:

$$y_t = \beta_0 + \beta_1\hat{x}_{1t} + \beta_2\hat{x}_{2t} + \beta_3\hat{x}_{3t} + \beta_4\hat{x}_{4t} + \theta_1\hat{w}_{1t} + \theta_2\hat{w}_{2t} + \theta_3\hat{w}_{3t} + \theta_4\hat{w}_{4t} + \varepsilon_t^* \quad (34)$$

But according to equation (31), $\hat{x}_{it} = x_{it} - \hat{w}_{it}$. We can therefore rewrite equation (34) as follows:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + (\theta_1 - \beta_1)\hat{w}_{1t} + (\theta_2 - \beta_2)\hat{w}_{2t} + (\theta_3 - \beta_3)\hat{w}_{3t} + (\theta_4 - \beta_4)\hat{w}_{4t} + \varepsilon_t^*$$

(35)

If there is no measurement error for x_{it}, then $\theta_i = \beta_i$. If there is such error, then $\theta_i \neq \beta_i$ and the coefficients of the residuals terms $\hat{w}_{it}$ will be significantly different from 0.

There is more information which we can draw from equation (35). Indeed, if the estimated coefficient $(\theta_i - \beta_i)$ 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 (35). On the other hand, if the estimated coefficient $(\theta_i - \beta_i)$ 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 (35). The estimated parameters of the w_i variables in equation (35) are thus very informative.

We must notice that the coefficients β_i estimated by the equation (35) are identical to those ones produced by a two-stage least squares (TSLS) procedure using the same instruments. Equation (35) is therefore another way to set up a TSLS. But in view of the useful information produced by equation (35), this equation opens the doors to new financial models. We should therefore prefer this formulation to that one represented by a 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 measurement errors. First, we regress the observed explanatory variables x_{it} on the instruments vector to obtain the residuals $\hat{w}_{it}$. Then, we regress y_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 related to this explanatory variable. We may resort to the Wald test (*F* test) to see if the whole set of $(\theta_i - \beta_i)$ coefficients is significantly different from zero.

We can generalize the former procedure to the case of *k* explanatory variables. Let **X** be a *(n x k)* matrix of explanatory variables which is not orthogonal to the innovation and let **Z** be a *(n x s)* matrix of instruments (*s*> *k*). To perform the Hausman test based on an artificial regression, we first regress **X** on **Z** to obtain $\hat{\mathbf{X}}$, that is:

$$\hat{\mathbf{X}} = \mathbf{Z}\hat{\boldsymbol{\theta}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} = \mathbf{P_Z X} \qquad (36)$$

where $\mathbf{P_Z}$ is the "predicted value maker". Having runned this regression, we compute the matrix of residuals $\hat{\mathbf{w}}$:

$$\hat{\mathbf{w}} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X} - \mathbf{P_Z X} = (\mathbf{I} - \mathbf{P_Z})\mathbf{X} \qquad (37)$$

Then we run the following artificial regression:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \hat{\mathbf{w}}\boldsymbol{\lambda} \qquad (38)$$

A *F* test on the $\boldsymbol{\lambda}$ coefficients will indicate if they are significant as a group. A *t* test on individual coefficients will indicate if the corresponding β is understated or overstated, as discussed previously.

The vector of $\boldsymbol{\beta}$ estimated by equation (38) is identical to the TSLS estimates, that is:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\mathbf{IV}} = (\mathbf{X}'\mathbf{P_Z X})^{-1}\mathbf{X}'\mathbf{P_Z y} \qquad (39)$$

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

$$R_{pt} - R_{ft} = \alpha + \beta_1(R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \varepsilon_t \quad (40)$$

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

$$R_{pt} - R_{ft} = \alpha^* + \beta_1^*(R_{mt} - R_{ft}) + \beta_2^* SMB_t + \beta_3^* HML_t + \beta_4^* UMD_t + \sum_{i=1}^{4}\varphi_i \hat{w}_{it} + \varepsilon_t^* \quad (41)$$

The estimated coefficients φ_i will allow detecting measurement errors and their signs will indicate if the corresponding variable is overstated or understated in the OLS regression. To estimate equation (41), we will use as instruments the higher moments discussed previously. We will call this estimation procedure: HAUS-hm.

As said previously, the $\boldsymbol{\beta}^*$ estimated by equation (41) are equivalent to the TSLS estimates. But we could prefer equation (41) because it gives more information on the problem of measurement errors. Equation (41) 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 ith factor of risk. If φ_i is positive, that means that the exposure to the ith risk factor is overstated in the OLS regression. The β associated to this factor will thus decrease in the artificial regression in comparison with the OLS one. And vice-versa if φ_i is negative. Moreover, according to our previous developments, we expect a high positive correlation between $\left(\hat{\beta}_i - \hat{\beta}_i^*\right)$, that is the estimated error on the coefficient of factor *i*, and $\hat{\varphi}_i$, the estimated coefficient of the corresponding artificial variable $\left(\hat{w}_i\right)$.

We can sum up the former arguments by the following empirical equation:

$$Spread_{is} = \pi_0 + \pi_1 \varphi_{is} + \varsigma_{is} \qquad \text{for } s = 1 \text{ to n} \qquad (42)$$

where $Spread_{is} = \hat{\beta}_{is} - \hat{\beta}_{is}^*$, *s* being here an hedge fund specific strategy, and ς_{is} being the innovation of the estimation. According to equation (42), φ may thus be viewed as an indicator of overstatement or understatement of the OLS estimation for the coefficient associated to the factor *i* for the strategy *s*. We will estimate this equation for the most important risk factors in the empirical section of this paper. That constitutes our variant of the original Hausman test. The goodness of fit of equation (42) will provide information about the severity of the measurement error for an explanatory variable as shown in the empirical section.

3.2 Test based on cumulants[15]

To estimate equation (41), we would also use the triplet of instrumental variables $\{z_0, z_1, z_4\}$ discussed previously, what is an innovative way to proceed. We call this estimator: HAUS-z. And we have also a corresponding GMM, the GMM-z. But there is another way to proceed which will reduce the number of instruments to the number of explanatory variables and which will perform quite well in the case of our sample of hedge fund returns. Indeed, regressing the explanatory variables on the z_i , which are cumulants, amounts to performing a polynomial adjustment on each explanatory variable, that is:

$$\hat{x}_{it} = \hat{\gamma}_0 + \hat{\gamma}_1 z_{11t} + ... + \hat{\gamma}_4 z_{14t} + \hat{\gamma}_5 z_{41t} + ... + \hat{\gamma}_8 z_{44t} \qquad (43)$$

where the z were defined previously. Let us define the following distance variable, which we denote iv_{it}:

$$iv_{it} = x_{it} - \hat{x}_{it} \qquad (44)$$

This variable removes from x_{it} some of its nonlinearities. It is thus a smoothed x_{it} which might be seen as a proxy for its long term expected value, the relevant variable in the F&F model which is theoretically formulated on expected values of the explanatory variables. As shown in the empirical section of this article, this new variable labelled iv_{it} is strongly correlated with x_{it} but not with the innovation of our model of returns. It is thus a very good candidate for an instrumental variable. Incidentally, the variable iv_{it} may be viewed, in the framework of our return model, as a portfolio hedged for some extreme events related to the skewness and the kurtosis of returns but which retains its correlation with the non-hedged portfolio. To illustrate the relevance of the variables iv as instrumental variables, figure 1 plots the relation between the observed market risk premium and the corresponding iv variable, denoted by iv_1. As we may see on this figure, the correlation

[15] For this section, see also: Coën and Racicot (2007) and Kendal and Stuart (1963).

between r_m and iv_1 is very good but iv_1 fluctuates less than r_m. It is in this sense that the iv_i variables might be considered as hedged portfolios.

Insert figure 1 about here

Resorting to the *iv* variables allows to weight optimally the vector *z* of instrumental variables to build another instrumental variable because in fact, as mentioned earlier, it is based on a GLS. Indeed, as revealed by our empirical section, the vector of estimated coefficients $\{\hat{\gamma}_1,...,\hat{\gamma}_8\}$ (equation 43) will be such that the extreme events are taken into account by these coefficients.

But the choice between the vector *z* or the vector *iv* as optimal instruments is really an empirical matter. To build the Hausman equation, called HAUS-C, we resort to the following two-step procedure to compute the artificial variables. As the *iv* variables are here the instruments, we regress first the explanatory variables of the F&F model on the iv_i. We obtain the following estimates:

$$x_{it} = c + \sum_{j=1}^{4} \kappa_j iv_{jt} + \zeta_{it} \qquad (45)$$

where x_{it} is an explanatory variable of the model; *iv* are the instruments and ζ is the innovation. The estimated residuals ζ_{it} are the $\hat{w}_{it}$ of equation (41). The resulting regression is called HAUS-C. Table 2 sums up the list of estimators used in this study to correct measurement errors.

Insert table 2 about here

4. Empirical results

4.1 Description of the sample

Our sample of hedge funds comprises the monthly returns of 22 HFR indices classified by categories or groups of categories. The observation period runs from January

1990 to December 2005, for a total of 192 observations. The risk factors which 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[16]. We used as instruments, among others, the Chen-Roll-Ross (1986) factors: the industrial production, the consumer price index, the spread between long and short term bonds, the spread between BBB and AAA corporate bonds and the dividend yield of the S&P500. These factors are drawn from the database of the Federal Reserve Bulletin and the Federal Reserve Bank of St-Louis.

4.2 A first glance at the sample of hedge fund returns

We can get a first glance at our sample of hedge fund returns by looking at table 3, which gives the descriptive statistics of the 22 HFR selected indices. The period of analysis runs from January 1990 to December 2005. The hedge fund indices are sorted by the R^2 resulting from the OLS estimation of the F&F model over the period 1990-2005. We see that the F&F model performs poorly for very specialized hedge fund strategies, like the fixed income arbitrage, convertibles and macro ones. But it seems relevant to explain the returns of the strategies which dominate the hedge fund industry, that is the equity hedge[17], fund of funds and equity non hedge strategies.

Insert table 3 about here

At an annualized value of 14.5% over the 1990-2005 period, the mean return of the hedge fund composite index was higher than the 11.5% realized by the S&P500. However, there is a great diversity of returns over the strategies. The return of the short selling index was a meagre 4% while the equity hedge index, associated to the most important strategy in the hedge fund industry, displayed a return as high as 17.5%.

[16] The address of the French's website is : http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[17] The equity hedge strategy is often called "long-short strategy".

A stylised fact of the hedge fund returns is the high degree of kurtosis of their distribution. Actually, the kurtosis of the returns of the hedge fund composite index was 5.30 over the 1990-2005 period compared to 3.73 for the S&P500. At table 3, kurtosis ranges from a high of 14.71 for the merger arbitrage index to a low of 2.46 for the market timing one. Incidentally, the equity hedge strategy, the most important one in the hedge fund industry, had a kurtosis of 3.92 over the period 1990-2005, a level comparable to the S&P500. Furthermore, the skewness of hedge fund indices is quite high for some strategies like merger arbitrage, event driven and fixed income arbitrage.

4.3 The choice of instruments

This article distinguishes four types of instruments: the classical instruments, the higher moments, the *z* instruments and the *iv* ones. The classical instruments include the traditional predetermined variables and the Chen-Roll-Ross (1986) factors. The higher moments instruments add to the classical ones the following set of variables discussed previously: $\left\{x_t^2, x_t^3, y_t^2\right\}$. The derivation of the cumulant instruments, denoted by *z* and *iv* in table 4, was explained in the preceding section.

Insert table 4 about here

Table 4 gives the adjusted R^2 of the regressions of the risk factors of the F&F model (endogenous variables) on the categories of instruments considered in this paper. As we can see, the classical instruments are weak, the adjusted R^2 being under 0.15 for each endogenous variable. Resorting only to the classical instruments is inappropriate for explaining the market risk premium and the returns of the mimicking portfolios which display a high degree of kurtosis. Nonlinear instruments like higher moments and cumulants are required. Table 4 confirms this allegation for higher moment instruments. The adjusted R^2 for the regressions of the endogenous variables on the set of these instruments excluding[18] y^2 are in a range from 0.55 to 0.64. The adjusted R^2 of the *iv* set of instrumental variables are even higher, being in a range from 0.62 to 0.80. However, the adjusted R^2 of the *z* set of instruments are quite moderate, being in a range from 0.21 to 0.35.

Table 5 gives the regressions of the F&F factors on the four *iv* variables. As we can see in this table, each risk factor has its own instrument. For instance, iv_1 is the instrument of the market risk premium. When regressing the market risk premium on the iv_i set, the variable iv_1 has a coefficient near 1 and the other iv_i have coefficients of 0. iv_2 is the instrument of the *SMB* factor, the regression of *SMB* on the iv_i giving a coefficient of 1 to iv_2 and 0 coefficients for the other iv_i, and so on. Therefore, the *iv* instruments are orthogonal variables. They are thus very appealing in view of their characteristics.

Insert table 5 about here

A good instrument must also be uncorrelated with the innovation term of the OLS estimation of the F&F model. Table 6 gives such information for the four sets of higher moment or cumulant instruments in the case of each hedge fund strategy. The *iv* set has no correlation with the innovation term. They thus qualify as very good, indeed optimal, instruments. The other sets have also low correlation with the innovation term except the higher moment set including y^2. It is for this reason that we will exclude this variable from the higher moment instrumental variables set in our following estimations.

Insert table 6 about here

It is instructive to look at table 7 to get a better grasp of our methodology to construct instrumental variables with cumulants. We know that the *SMB* portfolio incorporates many nonlinearities, being actually a long-short portfolio. Its payoffs are option-like. At table 7, we first show the four instruments related to the Durbin's estimator and thereafter the four

[18] We will explain later why we decided to exclude y^2 from the set of higher moment instruments.

instruments related to the Pal's estimator. As explained previously, the former are related to the asymmetry of a return distribution and the latter, to its kurtosis. As revealed by this table, the *SMB* factor is very sensitive to these groups of instruments. This observation suggests that nonlinearities are very present in this mimicking portfolio. To build the corresponding *iv* instrumental variable, we purge *SMB* from these nonlinearities and we obtain what we called a distance variable, which is more related to the expected value of *SMB*.

Insert table 7 about here

4.4 Comparison of estimation methods for the sample of 22 hedge fund indices

For the following evaluation of the specification errors, the OLS method will serve as benchmark to estimate equation (25) for each hedge fund strategy. The higher moment and cumulant estimation methods used to estimate this equation were previously reported in table 2. For the GMM regressions, we resort to the Newey-West matrix as weighting matrix. Table 7 provides the mean results computed over the hedge fund strategies for the estimation of equation (25) according to the chosen estimation methods. We will firstly discuss the OLS estimation and we will then compare the OLS results with those of the higher moment IV methods.

Insert table 8 about here

At 0.45, the average R^2 associated to the OLS estimation is quite moderate but as we noticed in table 3, it varies greatly from one strategy to the other. Only the TSLS-z and the GMM-z estimations, which are those resorting to the z cumulants set, involve a much lower R^2 while the other are close to the R^2 average.

Let us examine the coefficients averaged over the strategies for each estimation method. The alpha of Jensen measured by the constant of the regressions is quite stable from one estimation method to the other except for the methods based on the z instruments which have a much lower alpha at the expense of a much higher loading for r_m. Excepting the methods based on the *z* instruments, we also note that the factor loadings for the risk factors of the F&F model are quite close from one estimation method to the other. Therefore, resorting to weak instruments, which are the *z* ones here, may give way to biased coefficients.

Compared to the OLS benchmark, the results indicate that measurement errors are quite small when averaging the results. The ranking of the risk factors is the same over the methods. It is the market risk premium which impacts the most on hedge funds returns, followed by the *SMB* and *HML* factors. The *UMD* factor is not very significant as corroborated by other studies on hedge fund returns (Agarwal and Naik, 2000).

It is interesting to look at the Hausman estimations using two kinds of instruments: the higher moments and the cumulants. As explained in the section related to the Hausman tests, we notice that the estimated coefficients are the same for the HAUS-C, TSLS-C and GMM-C because the parameters of these estimations are exactly identified. They are also the same for the HAUS-hm and TSLS-hm estimations. Without being equal, the t-statistics associated to the estimated coefficients are quite close. Their differences may be explained by the weighting matrix which differs from one estimation method to the other. In the case of the TSLS, this matrix is the White one and in the case of the GMM-C, it is the Newey-West one.

Using the higher moment instruments (hm) suggests that hedge funds are riskier than using the cumulant ones except for the *UMD* factor which has a higher loading in the HAUS-C equation. Indeed, the *SMB* and *HML* factor loadings are especially more important in the HAUS-hm estimation than in the HAUS-C one. As we will see in the next section, the two methods tend to identify the same hedge funds for which the F&F model is subject to specification errors but differ in the direction of the correction of these errors. The coefficients resulting from the methods using the *iv* variables as instruments (TSLS-C, GMM-C) are closer to the OLS than the method using the higher moment instruments (TSLS-hm, GMM-hm). We can conclude from those observations that the correction of specification errors depends on the nature of the instruments used. As the cumulant instruments seem superior to the higher moment ones in this study, we could infer that the corrections performed by the cumulant instruments are better[19].

4.5 Hausman tests for the beta and the SMB loadings

We explained previously how to build a modified version of the Hausman test by using an artificial Hausman regression. Those regressions are useful because the estimated coefficients of the *w* variables in equation (41) are indicators of the degree of understatement or overstatement of the coefficient of the corresponding explanatory (endogenous) variable. They are all the more useful since the estimated factor loadings of these regressions are the same as the corresponding TSLS ones, a result which was ignored by Dagenais and Dagenais (1997). The spread between the OLS coefficient for a variable and its corresponding coefficient in the artificial Hausman regression is thus really an indicator of measurement error, -- overstatement or understatement--, on this variable as revealed by equation (42).

Tables 9 and 10 give the results of the estimations of the artificial regressions (41) for the market risk premium. Table 9 resorts to the *iv* variables to compute the residuals included in the artificial regression while table 10 uses the higher moments to do so. The strategies appearing in bold are those for which the estimated φ is significant at the 10% level.

Insert table 9 about here

At table 9, where the beta is estimated by the HAUS-C method, we notice that this coefficient is significantly overstated by the OLS method for four strategies and significantly understated for two strategies. We must thus disaggregate by strategy to observe measurement errors because we did not suspect serious measurement errors for the beta averaged over strategies (table 8). As we notice in table 9, the φ coefficient associated to the *w* variable of

[19] Resorting to a priori information on the values of the factor loadings might also help to judge the relevance of the corrections of specification errors.

the risk premium is actually an indicator of the degree of overstatement or understatement of the estimated beta. When φ is positive, the spread between the OLS and HAUS-C coefficients is also positive: the coefficient is overstated by the OLS method in this case. Otherwise, when φ is negative, the spread between the OLS and HAUS-C coefficient is negative: the coefficient is understated in this case. If we regress the spread over the φ appearing in table 9, we obtain the following regression[20]:

$$\begin{array}{c} \text{Spread} = 0.001 + 0.1627\varphi \\ \qquad\qquad\quad (1.04) \quad (29.06) \end{array}$$

The R^2 of this regression is 0.97, what is almost a perfect fit. At figure 2, we can observe the close linear relationship between φ and the spread between the OLS and HAUS-C betas. As said previously, such a good fit suggests that measurement errors are far from being negligible for the betas of hedge funds sorted by strategy.

Insert figure 2 about here

Table 10 provides the same information as table 9 for the beta coefficient except that the HAUS-hm method is used in this case. We notice that this table identifies more strategies with significant measurement errors, that is 13 instead of 6 for the HAUS-C method. And we must also notice that the sign of the measurement error is much related to the choice of instruments to discard those errors.

Insert table 10 about here

According to the HAUS-hm regression, four very specialized strategies seem to suffer from significant measurement errors for the loading of the risk premium, actually the same depicted by the HAUS-C regression: the distressed securities, merger arbitrage, event driven and market timing strategies. According to the HAUS-hm method, the betas of the first three strategies are understated while the beta of the last strategy is overstated. We obtain the inverse correction when using the HAUS-C method for those funds. This situation is all the

[20] Note that the t-statistics of the estimated coefficients are in parentheses.

more problematic since we know that the beta estimated by the HAUS-C method is the same as the TSLS one when using the cumulants as instruments for both and that the beta resulting from the HAUS-hm method is the same as the TSLS one when using the higher moments of the explanatory variables as instruments for each of these last two methods. The betas estimated by the artificial regressions have thus a sound foundation, being related to a well-known IV method (TSLS), and the direction of the correction of specification errors is therefore related to the choice of instruments and not to the version of the Hausman test by itself.

As for the HAUS-C method, we can regress the spread between the OLS and HAUS-hm beta on its corresponding φ appearing in the artificial regression. We get:

$$\text{Spread} = -0.002 + 0.5544\varphi$$
$$(-0.67) \quad (20.25)$$

with a R^2 of 0.95. Once again, the coefficient φ is a good indicator of the degree of misspecification errors. Figure 3 plots the close link between the spread and the HAUS-hm estimated φ.

Insert figure 3 about here

We will not report the results of the artificial regressions for the other risk factors but let us notice that our φ indicator may also signal the absence of measurement errors on an explanatory variable. Figure 4 plots the association between the spread between the *SMB* coefficients estimated by the OLS and HAUS-C methods and the corresponding φ. There is no evidence of a linear relationship between these two variables, the estimated φ being incidentally not significant in the HAUS-C regressions, and we thus conclude that there are no measurement errors for this variable.

Insert figure 4 about here

To sum up, our version of the Hausman test proves very useful for detecting measurement errors for the estimated parameters of the F&F model. Very often, the HAUS-C and the HAUS-hm methods identify the same strategies as candidates to measurement errors for a given risk factor. But a problem lies in the choice of the instruments because the correction of these errors by the IV methods is very much conditioned by this choice. On the basis of our previous developments, the cumulants of the variables of the F&F model seem more relevant than the higher moments to remove specification errors from a model.

5. Conclusion

In this paper, we proposed new versions of the Hausman artificial regression to detect and remove specification errors while estimating financial or economic models. These new estimators constitute also new empirical versions of these models, like the well-known CAPM or the Fama and French one. These estimators are equivalent to a TSLS procedure and lead to a new indicator of the spread between the OLS and corrected coefficient of an explanatory variable, this spread being a measure of the measurement error. Therefore, in addition to detecting the variables contaminated with measurement errors, our Hausman-based artificial regression provide a correction of those errors. It is in this sense that those kinds of regressions are very useful empirical versions of existing models.

To set up these tests, we resorted to two new sets of instruments: higher moments and cumulants. These instruments reveal themselves much more powerful than the classical instruments frequently used in the empirical financial literature, like the Chen-Roll-Ross (1986) ones. Our new instruments take into account the high degree of kurtosis which is present in the distributions of financial returns, especially hedge fund returns, what is not the case for the classical instruments.

As shown in this paper, the correction of specification errors in a financial model is much related to a judicious choice of instruments. Our two sets of instruments, higher moments and cumulants, tend to identify the same strategies for which the Fama and French model seems misspecified. Those strategies are not the dominant strategies of the hedge fund industry but instead very specialised ones, like the distressed securities, the fixed income or the event driven strategies. But it is one thing to identify specification errors in an estimated model. It is another thing to correct them. Incidentally, the two sets of instruments used in our study differ when correcting those errors. Where a set of instruments has identified an overstatement in regard to the exposure of a strategy to a risk factor, the other set tends to diagnose an understatement for the exposure. This result is important because it signals that the correction process done by an IV method is much conditioned by the set of instruments used.

In view of this problem, it appears relevant to choose a set of instruments. On the basis of our empirical works, the choice of cumulants over higher moments seems preferable in view of the strict orthogonality between these instruments and the innovation term of the estimated financial models and to the parsimonious character of these instruments, their number being restricted to the number of explanatory variables of a model. The cumulants are also an optimal combination of two well-known estimators, those of Durbin (1954) and Pal (1981). They are distance variables for which an optimal matrix is used to weight the instruments related to these estimators. That is a way to increase the robustness of higher moment instruments which are considered not very robust and which therefore tend to be neglected. Our study of hedge fund returns shows that such instruments may be very relevant if they are combined optimally. They are also especially relevant when the distributions of the studied variables deviate substantially from the Gaussian one like in this paper.

Summing up, our study reveals that we must account for specification errors when estimating a financial model. The method we suggest to do so looks very promising because it integrates new developments in the theory of financial risk in the estimation process of financial models.

References

Agarwal V, Naik N. On taking the alternative route: risks, rewards and performance persistence of hedge funds. Journal of Alternative Investments 2000; 2; 6-23.

Carhart MM. On persistence in mutual fund performance. Journal of Finance 1997; 52; 57-82.

Chan HW, Faff RW. Asset Pricing and the Illiquidity Premium. Financial Review 2005; 40; 429-458.

Chen NF, Roll R, Ross S. Economic Forces and the Stock Market. Journal of Business 1986; 59; 572-621.

Cheng CL, Van Ness JW. Regression with Measurement Error. Arnold: London; 1999.

Chung PY, Johnson H, Schill MJ. Asset Pricing When Returns are Non-normal: Fama-French Factors vs. Higher-order Systematic Co-moments. Journal of Business 2006; 79; 923-941.

Cochrane J. Asset Pricing. Princeton University Press: New Jersey; 2001.

Coën A, Racicot FE. Capital Asset Pricing Models Revisited: Evidence from Errors in Variables. Economics Letters 2007; 95; 443-450.

Cragg JG. Using higher moments to estimate the simple errors-in-variables model. Rand Journal of Economics 1997; 28; S71-S91.

Cragg JG. Making good inferences from bad data. Canadian Journal of Economics 1994; 27; 776-800.

Dagenais MG, Dagenais DL. Higher moment estimators for linear regression models with errors in the variables. Journal of Econometrics 1997; 76; 193-221.

Dittmar R. Nonlinear Pricing Kernels, Kurtosis Preference and Evidence from the Cross-section of Equity Returns. Journal of Finance 2002; 57; 369-403.

Durbin J. Errors in variables. International Statistical Review 1954; 22; 23-32.

Fama, EF, French KR. The cross-section of expected stock returns. Journal of Finance 1992; 47; 427-465.

Fama, EF, French KR. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 1993; 33; 3-56.

Fama EF, French KR. Industry costs of equity. Journal of Financial Economics 1997; 43; 153-193.

Fuller WA. Measurement error models. John Wiley & Sons: New York; 1987.

Geary RC. Inherent relations between random variables. Proceedings of the Royal Irish Academy 1942; 47; 23-76.

Gillard JW. An historical overview of linear regression with errors in both variables. working paper 2006. Cardiff University School of Mathematics.

Gillard JW, Iles TC. Method of Moments Estimation in Linear Regression with Errors in Both Variables. working paper 2005. Cardiff University School of Mathematics.

Hansen L, Jagannathan R. Assessing specification errors in stochastic discount factor model. Journal of Finance 1997; 52; 557-590.

Hansen L, Singleton K. Generalized instrumental variables estimation of non-linear rational expectations models. Econometrica 1982; 50; 1269-1286.

Harvey C, Siddique A. Conditional skewness in asset pricing, Journal of Finance 2000; 55; 1263-1295.

Hausman JA. Specification Tests in Econometrics. Econometrica 1978; 46; 1251-1271.

Hwang S, Satchell S. Modelling Emerging Market Risk Premia Using Higher Moments. International Journal of Finance and Economics 1999; 4; 271-296.

Ingersoll J. Theory of Financial Decision Making. Rowman and Littlefield Publishers: Maryland; 1987.

Ingersoll J. Multidimensional Security Pricing. Journal of Financial and Quantitative Analysis 1975; 5; 785-798.
Jegadeesh N, Titman S. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 1993; 48; 65-91.
Jurczenko E, Maillet B. Multi-Moment Asset Allocation and Pricing Models. Wiley: New York; 2006.
Kendall MG, Stuart A. The advanced theory of statistics, Vol.1. Charles Griffin: London; 1963.
Kraus A, Litzenberger R. Skewness preference and the valuation of risk assets. Journal of Finance 1976; 31; 1085-1100.
Lewbel A. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. Econometrica 1997; 65; 1201-1213.
Lim KG. A New Test of the Three-Moment Capital Asset Pricing Model. Journal of Financial and Quantitative Analysis 1989; 24; 205-216.
MacKinnon JG. Model Specification Tests and Artificial Regressions. Journal of Economic Literature 1992; 30; 102-146.
Malevergne Y, Sornette D. Higher moments portfolio theory, capitalizing on behavioral anomalies of stock markets. Journal of Portfolio Management 2005; 31; 49-55.
Pal M. Consistent moment estimators of regression coefficients in the presence of errors in variables. Journal of Econometrics 1980; 14; 349-364.
Pindyck RS, Rubinfeld DL. Econometric Models and Economic Forecasts. fourth edition. Irwin-McGraw-Hill: New York; 1998.
Racicot FE 2003**.** On Measurement Errors in Economic and Financial Variables. *In* : Three Essays on the Analysis of Economic and Financial Data. Ph.D. thesis. chap. 3. University of Quebec (Montreal); 2003.
Racicot FE. Techniques alternatives d'estimation et tests en présence d'erreurs de mesure sur les variables explicatives. M.Sc. thesis. University of Montreal: department of economics; 1993.
Racicot FE, Théoret R. The Beta Puzzle Revisited: A Panel Study of Hedge Fund Returns. Journal of Derivatives & Hedge Funds 2007; 13; 125-146 .

Racicot FE, Théoret R. On comparing hedge fund strategies using new Hausman-based estimators. Journal of Derivatives & Hedge Funds 2008; 14; 9-30 .
Racicot FE, Théoret R 2006. On Comparing Hedge Fund Strategies Using Higher Moment Estimators for Correcting Specification Errors in Financial Models. In : Gregoriou GN, Kaiser D(Eds.) , Hedge Funds and Managed Futures: a Handbook for Institutional Investors. Risk Books:London; 2006. p. 63-99 .
Racicot FE, Théoret R, Coën A. A New Empirical Version of the Fama and French Model Based on the Hausman Specification Test: An Application to Hedge Funds. Proceedings of the Global Finance Association. GFC-2007. Best Paper Award.
Rubinstein M. The fundamental theorem of parameter-preference security valuation. Journal of Financial and Quantitative Analysis 1973; 8; 61-69.
Samuelson PA. The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. Review of Economic Studies 1970; 37; 537-542.
Scott RC, Hovarth PA. On the direction of preference for moments of higher order than the variance. Journal of Finance 1980; 35; 915-919.
Spencer DE, Berk KN. A Limited Information Specification Test. Econometrica 1981; 49; 1079-1085.

Théoret R, Racicot FE. Specification errors in financial models of returns: an application to hedge funds. Journal of Wealth Management 2007; 10; 73-86.

Watson CT. GMM and the Fama and French Model: The Role of Instruments. working paper 2003. UCLA: Economics Department.

Wu D. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 1973; 41; 733-750.

Tables

Table 1 *List of the cumulant instrumental variables proposed by Dagenais and Dagenais (1997)*

z_0	ι_N
z_1	**x*x (Durbin)**
z_2	**x*y**
z_3	**y*y**
z_4	**x*x*x-3x[E(x'x/N)*I_k] (Pal)**
z_5	**x*x*y-2x[E(x'y/N)*I_k]-y{ι'_k[E(x'x/N)*I_k]}**
z_6	**x*y*y-x[E(y'y/N)]-2y[E(y'x/N]**
z_7	**y*y*y-3y[E(y'y/N)]**

* ι stands for a vector of one. I_k is the identity matrix of dimension (k x k).

Table 2 *List of higher moment and cumulant estimators used to correct specification errors*[*]

Method	Instruments
Hausman-hm	higher moments: $\{x_{t-1}, x_t^2, x_t^3, y_t^2\}$ and other exogenous variables like the Chen-Roll-Ross (1986) factors
Hausman-C	the distance variables called cumulants: $\{iv_1, \ldots, iv_4\}$
TSLS-hm	higher moments: $\{x_{t-1}, x_t^2, x_t^3, y_t^2\}$ and other exogenous variables like the Chen-Roll-Ross (1986) factors
TSLS-z	$\{z_0, z_{11}, \ldots z_{14}, z_{41}, \ldots, z_{44}\}$, which are the Durbin and Pal instruments
TSLS-C	the distance variables called cumulants: $\{iv_1, \ldots, iv_4\}$
GMM-hm	higher moments: $\{x_{t-1}, x_t^2, x_t^3, y_t^2\}$ and other exogenous variables like the Chen-Roll-Ross (1986) factors
GMM-z	$\{z_0, z_{11}, \ldots z_{14}, z_{41}, \ldots, z_{44}\}$, which are the Durbin and Pal instruments
GMM-C	the distance variables called cumulants: $\{iv_1, \ldots, iv_4\}$

***** The variables which enter in the computation of higher moments and cumulants are expressed in deviation from their mean. For the GMM estimations, the weighting matrix used to weight the moment conditions is the Newey West one.

*Table 3 Descriptive statistics of the HFR indices, 1990-2005**

	Mean return (ann)	Median (ann)	s.d. (ann)	skewness	kurtosis	R^2
Fixed income arbitrage	8.14	8.52	5.22	-1.07	13.93	0.01
Funds of funds (FoF) Market defensive	9.43	8.82	5.89	0.18	4.35	0.06
Convertibles	10.00	11.64	3.74	-1.27	6.61	0.16
Macro	13.51	11.28	8.77	0.43	3.82	0.19
Relative Value Arbitrage	11.70	10.68	3.70	0.08	10.30	0.28
FoF Conservative	8.31	8.94	3.23	-0.47	6.50	0.31
Merger Arbitrage	10.48	12.9	4.59	-2.61	14.71	0.32
Fixed income high yield	8.65	8.82	6.16	-0.44	8.98	0.34
Market Neutral Statistical Arbitrage	8.00	8.94	3.95	-0.26	3.67	0.36
Fixed income total	11.70	11.76	4.14	0.06	6.35	0.40
Equity Market Neutral	9.76	9.12	3.60	-0.10	5.16	0.42
FoF diversified	9.02	8.46	5.94	-0.10	7.30	0.42
Distressed Securities	14.42	13.56	6.14	-0.67	8.46	0.44
FoF total	10.25	9.72	4.60	-0.33	7.13	0.45
FoF strategic	12.85	14.82	8.91	-0.38	6.74	0.49
Market timing	12.94	11.94	7.25	0.14	2.46	0.58
Event Driven	14.72	16.98	7.07	-1.24	7.60	0.73
Sector	19.38	21.66	12.78	0.11	6.19	0.75
Equity hedge	17.46	19.38	9.90	0.17	3.92	0.76
Short selling	4.00	-0.90	23.10	0.26	5.10	0.78
Equity non hedge	16.91	21.84	15.45	-0.42	3.98	0.91
Average	*11.51*	*11.85*	*7.34*	*-0.38*	*6.82*	*0.45*
Hedge fund weighted composite	14.51	18.18	7.01	-0.51	5.30	0.82
S&P500	11.09	15.06	14.31	-0.45	3.73	

* The statistics appearing in this table are computed on the monthly returns of the HFR indices over the period running from January 1990 to December 2005. The weighted composite index is computed over the whole set of the HFR indices. The R^2 are those of the OLS estimations of the F&F model for each strategy. The strategies are sorted by increasing value of R^2.

Table 4 *Adjusted R^2 of the OLS regressions of the explanatory variables of the F&F model on instrument catogories**

	classical	**hm with y^2**	**hm without y^2**	***z***	***iv***
r_m	0.10	0.61	0.56	0.21	0.80
SMB	0.11	0.48	0.55	0.35	0.62
HML	0.09	0.61	0.64	0.23	0.74
UMD	0.05	0.53	0.51	0.32	0.65

* In this table, hm stands for higher moments.

Table 5 *Regressions of the explanatory variables of the F&F model on the cumulant instruments (iv)**

	r_m	*SMB*	*HML*	*UMD*
iv_1	**1.02**	0.01	-0.01	-0.01
	23.93	*0.27*	*-0.15*	*-0.12*
iv_2	0.01	**1.00**	-0.01	-0.01
	0.20	*17.01*	*-0.05*	*-0.04*
iv_3	0.01	0.00	**0.99**	-0.01
	0.17	*0.03*	*19.30*	*-0.03*
iv_4	-0.01	-0.01	0.01	**1.00**
	-0.08	*-0.02*	*0.02*	*18.49*
R^2 adj.	0.80	0.62	0.74	0.65
DW	1.87	2.52	2.40	2.22

* The t-statistics of the coefficients are in italics.

Table 6 *Correlation between the residuals of the OLS estimation of the F&F model and instrument sets*

	z	*iv*	**hm with** y^2	**hm without** y^2
Convertibles	0.10	0.01	0.41	0.08
Distressed	0.14	0.02	0.31	0.04
Equity hedge	0.05	0.01	0.32	0.00
Equity Market Neutral	0.06	0.00	0.42	0.01
MN Stat Arbitrage	0.06	0.00	0.44	0.04
Equity non hedge	0.10	0.03	0.48	0.04
Event Driven	0.12	0.01	0.35	0.06
Fund of Funds	0.08	0.01	0.32	0.05
Macro	0.05	0.01	0.59	0.05
Market timing	0.19	0.02	0.60	0.10
Merger Arbitrage	0.14	0.01	0.45	0.08
Relative Value Arb.	0.11	0.01	0.42	0.05
Short selling	0.05	0.01	0.46	0.01
Sector	0.05	0.01	0.29	0.00
Fixed income tot.	0.09	0.01	0.49	0.06
Fixed income arb.	0.05	0.01	0.42	0.00
Fixed income high yield	0.08	0.01	0.42	0.03
FOF Conservative	0.12	0.01	0.42	0.02
FoF diversified	0.08	0.01	0.36	0.00
FoF Market def.	0.08	0.01	0.47	0.00
FoF strategic	0.09	0.01	0.37	0.02
FWC	0.08	0.01	0.27	0.02

Table *7 Regression of SMB over the instrumental z variables*

		Coefficient	t-Statistic	p-value
Durbin's instruments	z_1	-0.0397	*-3.86*	0.00
	z_3	0.0205	*1.29*	0.20
	z_5	-0.0020	*-2.01*	0.04
	z_7	-0.0035	*-2.13*	0.03
Pal's instruments	z_2	-0.0295	*-3.23*	0.00
	z_4	0.0275	*3.47*	0.00
	z_6	0.0015	*3.37*	0.00
	z_8	0.0008	*2.06*	0.04

Table 8 *F&F model estimated by OLS and several higher moment IV methods over 22 HFR indices, 1990-2005[*]*

	c	r_m	*SMB*	*HML*	*UMD*	R^2	*DW*
OLS	**0.4283**	**0.2016**	**0.1226**	**0.0739**	**0.0378**	**0.45**	**1.60**
	4.22	*10.42*	*5.51*	*2.55*	*2.70*		
TSLS-hm	**0.4030**	**0.2304**	**0.1727**	**0.1346**	**0.0118**	**0.45**	**1.61**
	3.44	*8.15*	*4.26*	*2.23*	*2.22*		
TSLS-z	**0.3713**	**0.2884**	**0.1186**	**0.0912**	**0.0375**	**0.32**	**1.70**
	2.71	*4.70*	*1.75*	*1.30*	*1.47*		
TSLS-C	**0.4241**	**0.1910**	**0.1175**	**0.0734**	**0.0506**	**0.45**	**1.61**
	3.44	*6.98*	*3.55*	*1.90*	*2.01*		
GMM-hm	**0.4309**	**0.2133**	**0.1565**	**0.0972**	**0.0144**	**0.42**	**1.62**
	4.20	*7.50*	*6.60*	*2.80*	*2.93*		
GMM-z	**0.3963**	**0.2799**	**0.1106**	**0.0867**	**0.0380**	**0.32**	**1.70**
	2.53	*4.20*	*2.05*	*1.51*	*1.36*		
GMM-C	**0.4241**	**0.1910**	**0.1175**	**0.0734**	**0.0506**	**0.45**	**1.61**
	3.44	*8.15*	*4.26*	*2.23*	*2.22*		
Average	**0.4090**	**0.2341**	**0.1331**	**0.0928**	**0.0317**	**0.40**	**1.64**
	3.42	*7.19*	*4.07*	*2.10*	*2.15*		
HAUS-hm	**0.4030**	**0.2304**	**0.1727**	**0.1346**	**0.0118**	**0.48**	**1.62**
	3.77	*7.91*	*4.97*	*2.68*	*1.82*		
HAUS-C	**0.4241**	**0.1910**	**0.1175**	**0.0734**	**0.0506**	**0.47**	**1.61**
	4.43	*9.54*	*4.90*	*2.30*	*2.53*		

* This exhibit gives the mean results computed over the 22 hedge fund strategies for each estimation method retained in this paper. There are nine estimation methods appearing in this exhibit: the ordinary least squares (OLS) and the eight methods shown at table 2. To avoid overloading this table, we did not report the J-stat and their corresponding p-value for overidentified GMM, which are the GMM-hm and the GMM-z in this table. The J-tests indicate that the instruments are relevant for each GMM method used. By the way, the GMM-C estimation is exactly identified. The t-statistics of the coefficients are in italics.

Table 9 *HAUS-C test for the market risk premium*[*]

	OLS	**HAUS-C**	**Spread**	**φ**	**$t(\varphi)$**
Macro	**0.28**	**0.33**	**-0.04**	**-0.25**	**-1.91**
Market timing	**0.35**	**0.39**	**-0.04**	**-0.25**	**-3.35**
MN Stat Arbitrage	0.19	0.19	0.00	0.01	0.20
Equity Market Neutral	0.09	0.09	0.00	-0.01	-0.16
FoF Market def.	0.04	0.03	0.00	0.00	-0.01
Equity hedge	0.49	0.48	0.01	0.05	0.58
Fixed income arb.	-0.03	-0.04	0.01	0.03	0.40
FOF Conservative	0.12	0.11	0.01	0.07	1.55
Equity non hedge	0.85	0.84	0.01	0.05	0.69
Fund of Funds	0.18	0.17	0.01	0.07	1.29
FoF diversified	0.22	0.21	0.01	0.07	0.95
Fixed income tot.	0.16	0.15	0.01	0.06	1.16
Convertibles	0.10	0.09	0.01	0.06	1.08
Hedge fund weighted comp.	0.38	0.36	0.02	0.11	2.31
Sector	0.51	0.49	0.02	0.13	1.25
Fixed income high yield	0.24	0.22	0.02	0.11	1.27
Short selling	-1.02	-1.04	0.02	0.08	0.47
FoF strategic	0.34	0.32	0.03	0.16	1.54
Merger Arbitrage	**0.19**	**0.16**	**0.03**	**0.16**	**2.55**
Relative Value Arb.	**0.12**	**0.09**	**0.03**	**0.18**	**3.58**
Event Driven	**0.40**	**0.37**	**0.03**	**0.18**	**3.09**
Distressed	**0.24**	**0.20**	**0.04**	**0.24**	**3.32**

* The spread (measurement error) is the difference between the OLS coefficient and the corresponding Hausman coefficient resulting from the estimation of the Hausman artificial regression (equation 41). For each spread, we provide the coefficient φ of the corresponding artificial variable. The funds having a significant φ at the 10% level are bold-faced. Note the strong positive relationship between the spread and φ, the strategies being reported in increasing order of the spread.

Table 10 *HAUS-hm test for the market risk premium*[*]

	OLS	HAUS-hm	Spread	φ	t(φ)
Distressed	**0.24**	**0.37**	**-0.13**	**-0.19**	**-3.64**
Fixed income high yield	**0.24**	**0.34**	**-0.10**	**-0.14**	**-2.33**
Merger Arbitrage	**0.19**	**0.28**	**-0.09**	**-0.18**	**-3.98**
Event Driven	**0.40**	**0.49**	**-0.09**	**-0.14**	**-3.29**
FoF strategic	**0.34**	**0.41**	**-0.07**	**-0.14**	**-1.77**
Convertibles	**0.10**	**0.17**	**-0.07**	**-0.11**	**-2.80**
Fixed income arb.	-0.03	0.04	-0.06	-0.10	-1.56
FoF diversified	**0.22**	**0.26**	**-0.05**	**-0.10**	**-1.77**
FWC	**0.38**	**0.42**	**-0.04**	**-0.07**	**-1.86**
FOF Conservative	**0.12**	**0.16**	**-0.04**	**-0.08**	**-2.47**
Relative Value Arb.	**0.12**	**0.16**	**-0.04**	**-0.07**	**-1.87**
FoF Market def.	0.04	0.07	-0.03	-0.11	-1.56
Fund of Funds	**0.18**	**0.21**	**-0.03**	**-0.07**	**-1.73**
Equity Market Neutral	**0.09**	**0.11**	**-0.03**	**-0.05**	**-1.61**
Fixed income tot.	0.16	0.19	-0.03	-0.03	-0.67
Equity non hedge	0.85	0.87	-0.02	-0.02	-0.29
MN Stat Arbitrage	0.19	0.19	0.00	0.01	0.38
Equity hedge	0.49	0.48	0.01	0.01	0.14
Macro	0.28	0.26	0.03	0.06	0.65
Sector	0.51	0.46	0.05	0.06	0.81
Short selling	-1.02	-1.08	0.07	0.09	0.71
Market timing	**0.35**	**0.21**	**0.14**	**0.28**	**5.19**

* The spread (measurement error) is the difference between the OLS coefficient and the corresponding Hausman coefficient resulting from the estimation of the Hausman artificial regression (equation 41). For each spread, we provide the coefficient φ of the corresponding artificial variable. The funds having a significant φ at the 10% level are bold-faced. Note the strong positive relationship between the spread and φ, the strategies being reported in increasing order of the spread.

Figures

Figure 1 *Relation between the market risk premium and its corresponding iv variable (iv_1)*

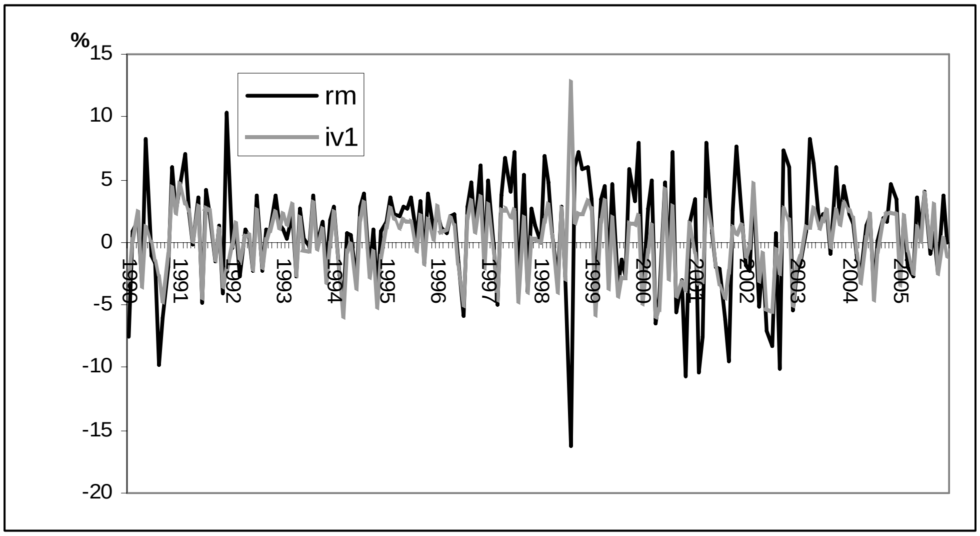
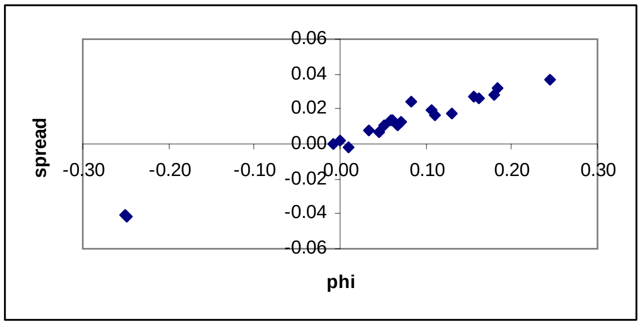
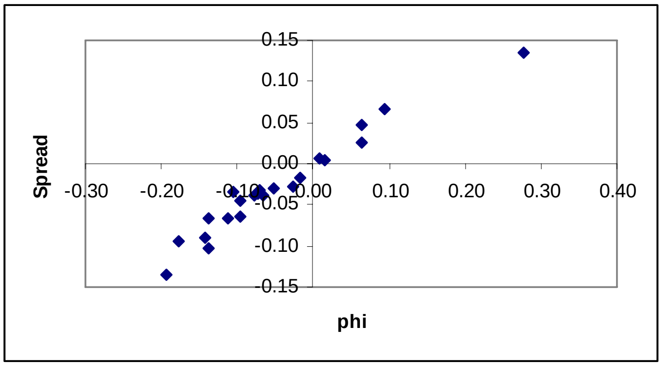


Figure 2 *Relation between the spread and the corresponding φ estimated by HAUS-C for the market risk premium**



* The data used to build this figure are reported at table 9.

Figure 3 *Relation between the spread and the corresponding φ estimated by HAUS-hm for the market risk premium*[*]



* The data used to build this figure are reported at table 10.

Figure 4 *Relation between the spread and the corresponding φ estimated by HAUS-C for the SMB factor*