Université d'Ottawa | University of Ottawa

The Pástor-Stambaugh Empirical Model Revisted: Evidence from Robust Instruments

François-Éric Racicot
William F. Rentz

WORKING PAPER
WP.2014.07

October 2014
ISSN 0701-3086

École de gestion TELFER School of Management

AACSB ACCREDITED EST. 1916 · Accredited by Association of MBAs · EFMD EQUIS ACCREDITED

The Pástor-Stambaugh Empirical Model Revisted: Evidence from Robust Instruments

François-Eric Racicot[a,]* and William F. Rentz[b]

[a]*Telfer School of Management, University of Ottawa, ON, Canada; CGA-Canada Accounting and Governance Research Centre; Chaire d'information financière et organisationnelle, ESG-UQAM.*

[b] *Telfer School of Management, University of Ottawa, ON, Canada.*

Abstract

This paper uses a parsimonious and robust instrumental variables technique to minimize the specification errors in the Pástor-Stambaugh (PS) empirical model. In particular, we use an improvement of Hansen's generalized method of moments (GMM) that uses higher moments that are robust instruments. Results with these instruments indicate that the liquidity measure used in the PS empirical model is improperly measured and/or is ill-conceived. Although this article applies a GMM framework to a financial application, this technique is applicable to estimation problems in the presence of specification errors in all areas of quantitative finance.

Keywords: GMM; specification errors; robust instrumental variables; higher moments; Pástor-Stambaugh; liquidity risk.

*Corresponding author. E-mail: Racicot@telfer.uottawa.ca

The Pástor-Stambaugh Empirical Model Revisted: Evidence from Robust Instruments

Abstract

This paper uses a parsimonious and robust instrumental variables technique to minimize the specification errors in the Pástor-Stambaugh (PS) empirical model. In particular, we use an improvement of Hansen's generalized method of moments (GMM) that uses higher moments that are robust instruments. Results with these instruments indicate that the liquidity measure used in the PS empirical model is improperly measured and/or is ill-conceived. Although this article applies a GMM framework to a financial application, this technique is applicable to estimation problems in the presence of specification errors in all areas of quantitative finance.

Keywords: GMM; specification errors; robust instrumental variables; higher moments; Pástor-Stambaugh; liquidity risk.

1. Introduction

Since the seminal work of Frisch (1934), treatment of specification errors, particularly endogeneity, is regarded as a challenging problem in empirical economics. Endogeneity, measurement errors, or more broadly, specification errors may lead to an inconsistent ordinary least squares (OLS) estimator and yield unreliable results. In the econometric literature, specification errors generally lead to non-orthogonality between the regressors and the error term. Spencer and Berk (1981) conjecture that specification errors originate from many sources, such as omission of relevant regressors, errors in variables, inappropriate aggregation over time, simultaneity (endogeneity), and incorrect specification form. Traditionally, a Hausman (1978) test may be used to identify this problem. This paper proposes a modified Hausman test using robust instrumental variables. As is well known in the literature, the use of weak instrumental variables can actually worsen the problem. Greene (2012, p. 249) noted that the use of weak instruments can lead to "perverse and contradictory results[1]." We propose a procedure that generates robust instruments that are able to tackle the weak instrumental variables problem[2].

The Fama and French (1992, 1993, 2000, 2012) or FF model as well as the Pástor and Stambaugh (2003) or PS extension are expressed in terms of unobservable expectations of the explanatory and dependent variables. In fact, however, estimates of these models use *realized* values of the variables. In essence, these realizations are the expectations measured with error. So, *a priori*, using OLS to find the parameters of the FF or PS models would yield incorrect estimation. More precisely, when there are measurement errors[3], endogeneity, or more generally specification errors, the OLS estimator is inconsistent. Thus, a robust instrumental variables approach is strongly recommended when estimating financial models based on expected values.

A concern in the PS model is a possible relation between the PS liquidity measure and the FF small firm anomaly variable (SMB), as small firms tend to be less liquid than large firms. This might create some specification error in the empirical PS model.

The remainder of the paper is organized as follows. Section 2 presents the robust instruments for the Pástor-Stambaugh empirical model. Section 3 describes the empirical models development. Section 4 discusses empirical results. Finally, section 5 contains conclusions and suggestions for further research.

2. Applying the robust instruments to the Pástor-Stambaugh empirical model

2.1 The Pástor and Stambaugh five-factor model

The cost of equity for firm *i*, $E(R_i)$, is given by Equation (1) and follows the well-known convention that now appears in many textbooks such as Copeland, Weston, and Shastri (2005),

$$E\left(R_i\right)-R_f=\sum_{k=1}^{n}E\left(\tilde{\delta}_k\right)\beta_{ik} \tag{1}$$

[1] Nelson and Startz (1990a,b) and Hahn and Hausman (2003) list two important implications using weak instruments. (i) The 2SLS is badly biased towards the OLS estimator, and (ii) The standard first-order asymptotics will not give an appropriate framework for statictical inference.
[2] This paper closely follows the estimation and testing approach described in Racicot and Théoret (2014). We improve on their approach by explicitly identifying and discussing the importance of using a robust instrumental variables method to shed new light on the weak instrumental variables problem.
[3] In his book *Irrational Exurberance*, Shiller (2005) states that speculative bubbles can incur when price increases spur investor enthusiasm. In other words, observed prices are not always equilibrium prices and hence observed prices may be viewed as equilibrium prices with measurement errors. See also Shiller (2014).

where $E(\cdot)$ is the expectation operator, $\tilde{\delta}_k$ is typically an unobservable variable, and β_{ik} is the sensitivity of stock i to the unobservable variable $\tilde{\delta}_k$. For $n = 1$, we obtain the CAPM with $E(\tilde{\delta}_1)$ equal to the market risk premium (expected return on the market minus the risk free rate). One can argue that the observed value of the market return can be used as a proxy for the expected market return. However, as Roll (1977) pointed out, even the traditional market return measures are themselves proxies for the unobservable true market return. Fama and French (1993) proposed a three-factor model that was then extended to a four-factor model by Carhart (1997). Subsequently, Pástor and Stambaugh (2003) further extended this model to include a fifth factor. The five factors are market risk premium, SMB (return on a portfolio of small cap stocks minus the return on a portfolio of large cap stocks), HML (return of a high book-to-market stock portfolio minus return of a low book-to-market stock portfolio), MOM (Carhart momentum factor), and LIQ (Pástor and Stambaugh measure of market liquidity). Just like the unobservable true market return, these additional variables are truly unobservable as constructed variables are used as proxies for the underlying unobservable variables.

The empirical version of the cost of equity for stock i may be written as

$$R_i - R_f = \alpha_i + \sum_{k=1}^{n} \delta_k \beta_{ik} + \varepsilon_i \tag{2}$$

where $n = 1$ for CAPM and $n = 5$ for the Pástor and Stambaugh model. The parameter α_i is the abnormal return for stock i known as the Jensen (1968) performance measure, δ_k is a proxy for the unobservable variable $\tilde{\delta}_k$, and ε_i is the error term. The proxy variable δ_k is defined by matrix equation (3).

$$\delta = \tilde{\delta} + u \tag{3}$$

δ is a matrix of dimension $T \times n$ of the n observable proxy factors that contain measurement errors and $\tilde{\delta}$ is a matrix of dimension $T \times n$ of the factors measured with error. u is a matrix of measurement errors, which we assume to be normally distributed. Substituting (3) into the matrix version of (2) yields (4).

$$R_i - R_f = \alpha_i i_T + \delta\beta_i + \varepsilon_i - u\beta_i = \alpha_i i_T + \delta\beta_i + e_i \tag{4}$$

where i_T is a identity vector of dimension $T \times 1$. Estimating (4) by OLS yields inconsistent estimators. This is the classical errors-in-variables problem (Fomby, *et al.,* 1984)[4] and Greene (2012, p. 99).

2.2 Robust instrumental variables for GMM estimation

Here we present an extension of the generalized method of moments (GMM) originally developed by Hansen (1982). This approach is called GMM_d and is based on robust instrumental variable that can be visualized as a distance estimator (Racicot and Théoret, 2012, 2014)[5]. In this paper we show how to incorporate this measure into the GMM framework[6].

The GMM_d formulation of the robust instrumental variable estimator is as follows:

$$\arg\min_{\hat{\beta}} \left\{ n^{-1} \left[d'\left(Y - X\hat{\beta} \right) \right]' W n^{-1} \left[d'\left(Y - X\hat{\beta} \right) \right] \right\} \quad (5)$$

The variables in (5) are defined below in (6) through (9). We start with *W*, which is a weighting matrix that can be estimated using the HAC[7] estimator and Y is defined as

$$Y = X\beta + \varepsilon \quad (6)$$

where X is assumed to be an unobserved matrix of explanatory variables.The observed matrix of observed variables is assumed to be measured with normally distributed error[8], *viz*., $X^* = X + v$.

$\hat{\beta}$ is defined as

$$\hat{\beta} = \hat{\beta}_{TSLS} = \left(X' P_z X \right)^{-1} X' P_z Y \quad (7)$$

P_z is defined as the standard "predicted value maker or projection matrix" used to compute

$$P_z X = Z\left(Z'Z \right)^{-1} Z'X = Z\hat{\theta} = \hat{X} \quad (8)$$

[4] Note that in the classical errors-in-variables problem the assumptions of normally distributed errors is not required but the OLS estimators remain inconsistent even with the normally distributed assumption.
[5] Note that this approach was first developed in Racicot (1993) and later published in Racicot (2014).
[6] The GMM_d estimator first appeared in Racicot and Théoret (2012, 2014).
[7] HAC is the heteroscedasticity and autocorrelation consistent estimator. We used the "Iterate to Convergence" Newey-West (1987) methodology of EViews 8.
[8] The assumption of a normally distributed matrix of errors is used to simplify the mathematical proof of the consistency of the estimators in this paper. This assumption is in no way a limitation in the modeling process of the time series used in this paper. Our proposed GMM_d estimator is based on the higher moments of the observed financial data and is thus able to capture the data's non-linearity, which is one of the important goals of this estimator.

where Z is obtained by optimally combining the Durbin (1954) and Pal (1980) estimators using GLS. The result is based on the Bayesian approach of Theil and Goldberger (1961). This leads to estimators that are more asymptotically efficient or at least as asymptotically efficient as using either only the Durbin or Pal estimators. This approach for obtaining Z is implemented in Equation (11) below in deviation form.

From (8) we extract the matrix of residuals

$$\boldsymbol{d} = X - \hat{X} = X - P_Z X = (I - P_Z)X \tag{9}$$

In (9) the matrix $\boldsymbol{d}$ is a matrix of instruments than can be defined individually in deviation form as

$$d_{it} = x_{it} - \hat{x}_{it} \tag{10}$$

Intuitively, the variable d_{it} is a filtered version of the endogenous variables. It potentially removes non-linearities that might be hidden in x_{it}. The $\hat{x}_{it}$ in (10) are obtained applying OLS on the z instruments.

$$\hat{x}_{it} = \hat{\gamma}_0 + z\hat{\phi} \tag{11}$$

The z instruments are defined as $z = \{z_0, z_1, z_2\}$, where $z_0 = i_T$, $z_1 = x \bullet x$, and $z_2 = x \bullet x \bullet x$ $-3x\,[D(x'x/T)]$. The symbol $\bullet$ is the Hadamard product, $D(x'x/T) = \underset{T\to\infty}{p\lim}(x'x/T) \bullet I_n$ is a diagonal matrix, and I_n is a identity matrix of dimension n x n, where n is the number of independent variables. z_1 contains the instruments used in the Durbin (1954) estimator, and z_2 contains the cumulant instruments used by Pal (1980). These instruments are consistent with Dagenais and Dagenais (1994).

It should be emphasized that the 3rd and 4th cross sample moments are used as instruments to estimate the model parameters. This is in line with the work of Mandelbrot (1963) and Fama (1963, 1965) who found that stock returns are not normally distributed. We believe that the assumption of normality is a sufficient condition for the estimators to be consistent once measurement errors are purged using these 3rd and 4th cross sample moments.

2.3 Hausman artificial variables regression test

Alternatively, $\hat{\beta}$ in (5) above is obtained by estimating the following equation using OLS:

$$Y = X\beta + \hat{\omega}\varphi + e \qquad (12)$$

It is a two-stage least squares estimator because $\hat{\omega}$ is also obtained by OLS and (12) can be rewritten as

$$Y = X\hat{\beta}_{TSLS} + \hat{\omega}\varphi + e^* \qquad (13)$$

where $\varphi = \psi - \beta$ measures the under/over estimation of the OLS benchmark estimator. Following Pindyck and Rubinfeld (1998, 195-197) and Racicot and Theoret (2012, 2014), φ can be obtained using the following procedure.

Assume a regression model of the form $Y = X\beta + \varepsilon$ and that X is an unobservable variable that is related to the observable variable X^*, where $X^* = X + v$ and v is matrix of measurement errors that are assumed to be normally distributed. The OLS regression $Y = X^*\beta + \varepsilon^*$ is related to the original regression by noting that $\varepsilon^* = \varepsilon - v\beta$. We can write $X^* = \hat{X}^* + \hat{\omega}$, where $\hat{\omega}$ are the regression residuals from (11). Note that (11) is just another representation of (8). Substituting for X^* in the equation for y yields $Y = \hat{X}^*\beta + \hat{\omega}\beta + \varepsilon^*$. Let ψ represent the coefficient of the variable $\hat{\omega}$. Substituting $\hat{X}^* = X^* - \hat{\omega}$ yields $Y = X^*\beta + \hat{\omega}(\psi - \beta) + \varepsilon^*$, which is analogous to (13).

The resulting t statistics can be analyzed in the usual fashion. That is, if a significant t statistic is obtained for any φ variable, there are significant specification/measurement errors in the model[9].

(13) is a Hausman (1978) artificial regression that can also be obtained using TSLS with the same set of instruments (Spencer and Berk, 1981). To be precise, the estimated 'slope' coefficients of the GMM_d regression should be the same as the corresponding 'slope' coefficients in the Hausman artificial regression.

[9] An F test can be done to see if collectively, none of the coefficients of the variables in the artificial regression are significantly different from zero. This turns out to be unnecessary, since at least one coefficient in every regression is significantly different from zero using t tests on the individual coefficients.

In (13), $\hat{\omega}$ is a matrix of residuals of the regression of each explanatory variable on the instrument set. The notation $\hat{\omega}$ is commonly used in Hausman artificial regressions. It is equivalent to the ***d*** matrix of residuals of (9) that emphasizes the idea of a 'distance' variable.

3. Empirical models development

Equation (14) is the empirical formulation of the Fama-French model as augmented by Carhart (1997) by the momentum factor.

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}\left[R_{Mt} - R_{ft}\right] + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{it} \quad (14)$$

Equation (15) is the GMM$_d$ formulation of equation (14).

$$R_{it} - R_{ft} = \alpha_{GMM_d i} + \beta_{GMM_d 1i}\left[R_{Mt} - R_{ft}\right] + \beta_{GMM_d 2i}SMB_t + \beta_{GMM_d 3i}HML_t + \beta_{GMM_d 4i}MOM_t + \tilde{\varepsilon}_{it} \quad (15)$$

Equation (16) is the Pástor-Stambaugh (2003) empirical model.

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}\left[R_{Mt} - R_{ft}\right] + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \beta_{5i}LIQ_t + \varepsilon_{it} \quad (16)$$

Equation (17) is the GMM$_d$ formulation of equation (16).

$$\begin{aligned} R_{it} - R_{ft} = {} & \alpha_{GMM_d i} + \beta_{GMM_d 1i}\left[R_{Mt} - R_{ft}\right] + \beta_{GMM_d 2i}SMB_t + \beta_{GMM_d 3i}HML_t \\ & + \beta_{GMM_d 4i}MOM_t + \beta_{GMM_d 5i}LIQ_t + \tilde{\varepsilon}_{it} \end{aligned} \quad (17)$$

Equation (18) is the Haus$_d$ formulation of equation (14).

$$\begin{aligned} R_{it} - R_{ft} = {} & \alpha_{Haus_d i} + \beta_{Haus_d 1i}\left[R_{Mt} - R_{ft}\right] + \beta_{Haus_d 2i}SMB_t + \beta_{Haus_d 3i}HML_t + \beta_{Haus_d 4i}MOM_t \\ & + \varphi_{Mi}\hat{\omega}_{Mi} + \varphi_{SMBi}\hat{\omega}_{SMBi} + \varphi_{HMLi}\hat{\omega}_{HMLi} + \varphi_{MOMi}\hat{\omega}_{MOMi} + \tilde{\varepsilon}_{it} \end{aligned} \quad (18)$$

Equation (19) is the Haus$_d$ formulation of equation (15).

$$\begin{aligned} R_{it} - R_{ft} = {} & \alpha_{Haus_d i} + \beta_{Haus_d 1i}\left[R_{Mt} - R_{ft}\right] + \beta_{Haus_d 2i}SMB_t + \beta_{Haus_d 3i}HML_t + \beta_{Haus_d 4i}MOM_t \\ & + \beta_{Haus_d 5i}LIQ_t + \varphi_{Mi}\hat{\omega}_{Mi} + \varphi_{SMBi}\hat{\omega}_{SMBi} + \varphi_{HMLi}\hat{\omega}_{HMLi} + \varphi_{MOMi}\hat{\omega}_{MOMi} + \varphi_{LIQi}\hat{\omega}_{LIQi} + \tilde{\varepsilon}_{it} \end{aligned} \quad (19)$$

In the next section we discuss are empirical results.

4. Empirical results

4.1 Data

Our sample is composed of monthly returns of 12 indices classified by FF industrial sectors. The observation periods are from August 1962 to December 2012 for a total 605 observations. The FF risk factors are drawn from French's website[10]. The PS liquidity factor is from Pástor's website[11].

Descriptive statistics

Table 1 and 2 present the descriptive statistics of the dependent and independent variables, respectively.

Insert Table 1 here

The Jarque-Bera (1980) statistic is calculated by equation (20),

$$JB = (n-k)\left(\frac{skew^2}{6} + \frac{(kurt-3)^2}{24}\right) \sim \chi^2(2) \qquad (20)$$

where n is the number of observations, k is the number of regressors which is zero when using the raw data, *skew* is the skewness of the data which is zero for a normal distribution, and *kurt* is the kurtosis which is three for the normal distribution. For all sectors, note that the JB statistic is greater than 5.99, which is the critical value of the chi-square distribution at the 5% level for 2 degrees of freedom. Thus, we reject the null hypothesis of normality for all sector returns. This is consistent with Mandelbrot (1963) and Fama (1963, 1965).

Sector 6 Business Equipment has the highest standard deviation of 6.61, which would indicate that it is the riskiest sector on a standalone basis in the Markowitz (1959)[12] mean-variance framework. However, in the Rubinstein (1973) and Jurczenko and Maillet (2006) higher-moments framework, we note that this sector has the lowest kurtosis, which suggests that maybe this sector is not the most risky.

---

[10] French's website is http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
[11] Pástor's website is http://faculty.chicagobooth.edu/lubos.pastor/research/liq_data_1962_2012.txt
[12] Markowitz (2012) noted that the mean-variance model still works well in the presence of moderate amounts of skewness and kurtosis.

Nine of the 12 sectors show negative skewness, which is an indicator of downside risk. Only Sector 2 Durables, Sector 4 Energy, and Sector 10 Health have the desirable positive skewness, which is an indicator of strong upside potential.

Insert Table 2 here

In Table 2, the JB statistics are even more indicative of non-normality. The variables *HML*, *MOM*, and *LIQ* have extremely high JB statistics, indicating that extreme events occur far more frequently than with the normal distribution. This is a reflection of the kurtosis measuring over 9 for each of these 3 variables, which is over 3 times the kurtosis of a normal distribution. Only the JB statistics of 112.81 and 85.75 for the market risk premium and *SMB* variables, respectively, fall within the range of the JB statistics from Table 1 for the sector returns. Nevertheless, as we previously noted for the sector returns, even these values are well above the critical value of 5.99 that allows us to reject the null hypothesis of normality.

All of these results suggest the logic of our proposed methodology which uses higher moments (cumulants) as instruments for the GMM estimation process. Using OLS when such strong non-normality is present in both the dependent and explanatory variables, may lead to wrong inferences.

4.2 Testing for robust instrumental variables

Relevance test

Weak instruments occur when $\left(\frac{1}{n}\right)Z'X$ is close to zero. We proceed analogously to Stock and Yogo (2005)[13] and Stock and Watson (2011, ch. 12) who proposed using the conventional F statistic for testing that all the coefficients in the regression

$$x_i = z_i'\pi + v_i \tag{21}$$

are zero. This is used to test the hypothesis that the instruments are weak. In other words, this is a test of the relevance of the instruments. Specifically, we test each explanatory variable by running regression (21) on all of the instruments. If the resulting F is below 10 for all of the

[13] See also Staiger and Stock (1997) for a similar test for the case of a large number of instruments.

regressions, this is a signal of a potential weak instruments problem.[14] If a least one of the F values is above 10, then the instruments are robust.

Insert Table 3 here

Note from Table 3 that all F values are well over 10. The coefficients of the instrumental variables represent the partial correlation coefficients of the instruments with the explanatory variables. Starting with the coefficient for z_1 on x_1 and ending with the coefficient of z_5 on x_5, these diagonal coefficients are all close to 1 and have significant t values. This means that each individual instrument is highly related to its respective explanatory variable. The off-diagonal coefficients are all not significantly different from zero according to their low t values.

Exogeneity test

When the instruments are uncorrelated with the error terms, *viz*. corr(z_{1i},ε_i) = 0,…, corr(z_{mi}, ε_i) = 0, the instruments are exogenous. Instead of calculating these individual partial correlation coefficients, we regressed the instruments on the error terms as shown in Equation (22),

$$\hat{\hat{\varepsilon}}_i = c + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i} + \gamma_4 z_{4i} + \gamma_5 z_{5i} + \xi_i \tag{22}$$

where $\hat{\hat{\varepsilon}}_i$ is the estimated residual from the five-factor regression Equation (16). According to the Frisch-Waugh (1933)-Lovell (1963) Theorem, the coefficients of this regression partial out or net out the effect of each regressor with the error term. So, these coefficients are analogous to the partial correlation coefficients.

Insert Table 4 here

Table 4 presents the results for regression Equation (22), which is our instruments exogeniety test. Note that all of the coefficients of the instrumental variables in (22) are close to zero and insignificant as their p-values are substantially greater than any of the typical critical levels. Furthermore, the R^2 is essentially zero. Thus, we conclude that our instruments are indeed exogenous.

[14] In theory, the F values could all be less than 10 and yet the instruments not be weak. Godfrey (1999) proposed a joint test on all of the explanatory variables. This test is not necessary here since F is greater than 10 for all 5 of our regressions.

4.3 Estimating the Fama-French and Pástor-Stambaugh models with specification errors

Insert Table 5 here

Estimates of the parameters of all of the models appear in Table 5.

Insert Table 6 here

For all estimation methods, the coefficient for the market factor is significant for all 12 FF sectors as shown in Tables 5 and 6. The coefficient for the SMB factor is significant for 9 of the 12 sectors using OLS. However, this coefficient is significant for only 2 sectors using GMM_d. This suggests that the SMB factor may contain measurement errors. The $Haus_d$ artificial regression further suggests that there are errors, as there are 2 sectors for the estimated ω_{SMB} coefficient that have significant t values. The results for the HML factor are even more strongly suggestive of measurement errors. The coefficient for the HML factor is significant for 5 of the 12 sectors using OLS. This coefficient is **NOT** significant for any sector using GMM_d! The HML factor is regarded as a value return premium. That is, high book to market is indicative of value firms and low book to market is indicative of growth firms. The $Haus_d$ method also suggests that there are errors in this variable, as there are 4 sectors for the estimated ω_{HML} coefficient that have significant t values.

The coefficient for Carhart's MOM factor is significant for 11 of the 12 sectors using OLS. However, this coefficient is **NOT** significant for any sector using GMM_d![15] $Haus_d$ suggests that there are errors, as there are 2 sectors for the estimated ω_{MOM} coefficient that have significant t values. We note that the MOM variable is really a behavioral finance variable, not a risk factor. *So, the fact that the coefficient for the MOM variable for all 12 FF sectors is insignificant, is actually an argument for market efficiency.*

When dealing with the FF sectors, the sector volatility of returns can itself be volatile. Furthermore, the heating and cooling of these volatilities does not always happen simultaneously. Thus, the risk-return profile of the industrial sectors may differ substantially and be quite dynamic[16].

---

[15] This result is in line with Kothari, Shanken, and Sloan (1995).

[16] The authors would like to thank William Ferrell of Ferrell Capital Management for making this observation.

The idea of Pástor-Stambaugh to include a liquidity variable is intuitively appealing. In particular, investors require an illiquidity risk premium for ***less*** liquid assets if this illiquidity risk can **NOT** be diversified away. Although the liquidity variable is not directly tradeable, it can in principle be synthetically created just as one can create a synthetic option.

It appears, however, that the Pástor-Stambaugh liquidity variable LIQ at best contains significant measurement errors or at worst is ill-conceived, as it does not seem to be useful in our sample. As evidence of this, the addition of the liquidity variable actually reduces the adjusted R^2 from 0.69 to 0.67 in the GMM_d estimation in Table 5. Furthermore, the $Haus_d$ artificial regression coefficient for the ω_{LIQ} variable in Table 5 shows 4 sectors with significant measurement error. In Tables 5 and 6 we note that the number of significant sectors for the LIQ factor is 3 for both OLS and GMM_d. However, only Sector 3 Manuf is in common between the two methods, with Sector 5 Chems and Sector 12 Other being significant for OLS and Sector 9 Shops and Sector 11 Money being significant for GMM_d. The adjusted R^2, ω_{LIQ} variable, and OLS versus GMM_d results all tend to imply that the LIQ variable is improperly measured and/or is ill-conceived in its construction.

The results for the *t* tests for the coefficients of the PS empirical model should not be surprising. Pagan (1984, 1986) shows that constructed variables may increase the variance of the OLS estimator but the estimator remains unbiased. Further evidence of the unreliability of the constructed variables used by FF and PS is provided by Harvey, Liu, and Zhu (2013). They present a convincing argument that unless a *t*-ratio for a factor is greater than 3, any claimed research finding for a factor is likely to be false and the result of data mining. Cochrane (2011) expresses his doubts about the importance of the plethora of new factors discovered in the last ten years by referring to them as a “zoo of new factors”.

5. Conclusions

Using our new robust instruments, we were able to improve the GMM estimator. We call this improved estimator GMM_d. Our approach has the virtue of being parsimonious as the researcher does not have to search for instruments. Accordingly, the problem of weak instruments might well be alleviated using GMM_d approach.

We used GMM_d to provide some insight into the Pástor-Stambaugh model. Our results show that there are significant measurement errors in several of the Fama-French sectors. We also note there is a significant reduction in adjusted R^2 in the GMM_d approach compared to the OLS approach in both the Fama-French and Pástor-Stambaugh models. Furthermore, in the GMM_d approach, the adjusted R^2 for the Pástor-Stambaugh model is lower than the value for Fama-French. This suggests that the added regressor LIQ is not very useful in explaining returns in our sample. We also find issues with the HML and MOM variables. The coefficients are insignificant for all of the FF sectors when using GMM_d. Although Fama and French (2015) include HML, they don't include either the LIQ factor or the MOM factor in their empirical model.

One concern of the researchers is that the regression coefficients may change over time. Since we are using sectors rather than individual firms, this should reduce the instability of the coefficients. Nevertheless, we think that this is an issue for further research that we are currently conducting.

Another interesting area of research that we are pursing is to use GMM_d to test the CAPM along the lines of two-pass regression approach of Fama and MacBeth (1973).

For researchers in other areas of quantitative finance, the GMM_d estimation approach is suitable for both linear and non-linear parametric models where specification errors, measurement errors, and/or endogeneity may be suspected.

Finally, the approach used in this paper can be extended to a panel data framework along the lines discussed in Racicot (2015).

Acknowledgements

For providing useful comments, we would like to thank our colleagues Alfred L. Kahl, Edwin Neave, Stylianos Perrakis, and the seminar participants at the May 2014 Canadian Operational Research Society (CORS) held in Ottawa, Ontario. The usual disclaimer applies.

References

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52 (1), 57-82.

Cochrane, J. H., 2011. Presidential address: discount rates. Journal of Finance 66 (4), 1047 – 1108.

Copeland, T.E., Weston, J.F., Shastri, K., 2005. Financial Theory and Corporate Policy, 4e. Pearson Education Inc., Boston, MA.

Dagenais, M.G., Dagenais, D.L., 1994. GMM estimators for linear regression models with errors in the variables. Centre de recherche et développement en économique (CRDE), Working Paper 0594, University of Montreal.

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

Fama, E.F., 1963. Mandelbrot and the stable Paretian hypothesis. Journal of Business 36 (4), 420-429.

Fama, E.F., 1965. Portfolio analysis in a stable Paretian market. Management Science 11 (3), 404-419.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81 (3), 607-635.

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

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

Fama, E.F., French, K.R., 2000. Forecasting profitability and earnings. Journal of Business 73, 161-175.

Fama, E.F., French, K.R., 2012. Size, value, and momentum in international stock returns. Journal of Financial Economics 105, 457-472.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116(1), 1-22.

Fomby, T., Hill, C., Johnson, S., 1984. Advanced Econometric Methods. Springer-Verlag, Needham, MA.

Frisch, R., 1934. Statistical confluence analysis by means of complete regression system. University Institute of Economics, Oslo.

Frisch, R., Waugh, F., 1933. Partial time regressions as compared with individual trends. Econometrica 1, 387-401.

Godfrey, L., 1999. Instrument relevance in multivariate linear models. Review of Economics and Statistics 81, 550-552.

Greene, W.H., 2012. Econometric Analysis, 7e. Pearson, Boston, MA.

Hahn, J., Hausman, J., 2003. Weak instruments: diagnosis and cures in empirical economics. American Economic Review 93, 118-125.

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

Harvey, C.R., Liu, Y., Zhu, H., 2014. … and the cross-section of expected returns. Working Paper SSRN-id2249314, Fuqua School, Duke University.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251-1271.

Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters 6, 255-259.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. Journal of Finance 23, 389-416.

Jurczenko, E., Maillet, B., 2006. The four-moment capital asset pricing model: Between asset pricing and asset allocation. In: Multi-Moment Asset Allocation and Pricing Models (eds. Jurczenko, E., Maillet, B.), Ch. 6, John Wiley & Sons, Chichester, England.

Kothari, S.P., Shanken, J., Sloan, R., 1995. Another look at the cross-section of expected stock returns. Journal of Finance 50 (1), 185-224.

Lovell, M., 1963. Seasonal adjustment of economic time series and multiple regression analysis. Journal of the American Statistical Association 58, 993-1010.

Mandelbrodt, B., 1963. The variation of certain speculative prices. Journal of Business 36, 394-419.

Markowitz, H., 1959. Portfolio Selection: Efficient Diversification of Investments. John Wiley & Sons, New York.

Markowitz, H., 2012. The "Great Confusion" concerning MPT. Aestimatio, The IEB International Journal of Finance 4, 8-27.

Nelson, C., Starz, R. 1990a. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one, Journal of Business, 63(1), 125-140.

Nelson, C., Starz, R. 1990b. Some further results on the exact small sample properties of the instrumental variable estimator, Econometrica, 58(4), 967-976.

Newey, W., West, K., 1987. A simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

Pagan, A.R., 1984. Econometric issues in the analysis of regressions with generated regressors. International Economic Review 25, 221-247.

Pagan, A.R., 1986. Two stage and related estimators and their applications. Review of Economic Studies 53, 517-538.

Pal, M., 1980. Consistent moment estimators of regression coefficients in the presence of errors in variables. Journal of Econometrics 14, 349-364.

Pástor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642-685.

Pindyck, R.S., Rubinfeld, D.L., 1998. Econometric Models and Economic Forecasts, 4ed. Irwin/McGraw-Hill, Boston, MA.

Racicot, F.E., 2015. Engineering robust instruments for panel data regression models with errors in variables: a note. Applied Economics 47 (10), 981-989.

Racicot, F.E., 2014. Erreurs de mesure sur les variables économiques et financières. La Revue des Sciences de Gestion 267-268(3-4): 79-103.

Racicot, F.E., Théoret, R., 2014. Cumulant instrument estimators for hedge fund return models with errors in variables. Applied Economics 46 (10), 1134-1149.

Racicot, F.E., Théoret, R., 2012. Optimally weighting higher-moment instruments to deal with measurement errors in financial return models. Applied Financial Economics 22(14): 1135-1146.

Racicot, F.E., 1993. Techiniques alternatives d'estimation et tests en présence d'erreurs de mesure sur les variables explicatives, M.Sc. thesis, Department of Economics, University of Montreal.

Roll, R., 1977. A critique of the asset pricing theory's tests. Journal of Financial Economics 4, 129-176.

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

Shiller, R.J., 2005. Irrational Exurberance, 2e. Princeton University Press, Princeton, NJ.

Shiller, R.J., 2014. Speculative asset prices. Cowles Foundation Discussion Paper No. 1936, Yale University.

Spencer, D.E., Berk, K.N., 1981. Limited information specification test. Econometrica 49, 1079-1085.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557-586.

Stock, J.H., Watson, M.W., 2011. Introduction to Econometrics, 3e, Pearson, Boston, MA.

Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Identification and Infererence in Econometric Models: Essays in Honor of Thomas J. Rothenberg (eds. Andrews, W.K., Stock, J.H.), Ch. 5, Cambridge University Press, Cambridge, England.

Theil, H., Goldberger, A., 1961. On pure and mixed estimation in economics. International Economic Review 2, 65-78.

Tables

Table 1

Descriptive statistics for the Fama-French 12 sector factors 1962m08 – 2012m12

	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
1 Nodur	1.08	1.08	18.73	-21.03	4.34	-0.24	5.07	114.48
2 Durbl	0.88	0.85	42.78	-32.86	6.35	0.13	7.90	607.02
3 Manuf	0.99	1.37	21.16	-28.58	5.35	-0.49	5.68	205.31
4 Enrgy	1.10	1.04	24.29	-18.39	5.40	0.02	4.46	53.97
5 Chems	0.94	1.05	20.19	-24.58	4.67	-0.21	5.17	122.99
6 Buseq	0.97	0.92	20.46	-26.20	6.61	-0.20	4.17	38.89
7 Telcm	0.85	1.00	22.12	-15.56	4.66	-0.17	4.33	47.43
8 Utils	0.84	0.88	18.80	-12.65	4.03	-0.08	4.18	35.86
9 Shops	1.03	1.03	25.80	-28.24	5.24	-0.24	5.40	150.85
10 Hlth	1.06	1.08	29.58	-20.47	4.95	0.12	5.48	156.80
11 Money	0.95	1.15	21.02	-21.97	5.51	-0.36	4.76	90.84
12 Other	0.86	1.15	19.32	-29.32	5.53	-0.46	5.04	125.65

Note: A total of 605 observations is used to compute the descriptive statistics.

Table 2

Descriptive statistics for the Pástor-Stambaugh factors 1962m08 – 2012m12

	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
R_m-R_f	0.50	0.82	16.01	-23.00	4.48	-0.50	4.86	112.81
SMB	0.25	0.10	14.62	-11.60	2.99	0.33	4.72	85.75
HML	0.33	0.36	19.72	-20.79	3.22	-0.14	10.18	1302.36
MOM	0.70	0.80	18.39	-34.74	4.28	-1.41	13.86	3172.44
LIQ	0.00	0.01	0.29	-0.38	0.06	-1.21	9.43	1189.90

Note: A total of 605 observations is used to compute the descriptive statistics.

Table 3

Relevance Test for Robust Instruments

	c	z_1	z_2	z_3	z_4	z_5	F
x_1	-0.0496	1.0057	-0.0003	-0.0091	0.0112	0.0002	1037.00
	-68.78	*67.18*	*-0.01*	*-0.32*	*0.74*	*0.77*	
x_2	0.0019	0.0008	1.0000	-0.0013	0.0016	0.0000	685.42
	3.99	*0.08*	*56.69*	*-0.07*	*0.16*	*0.16*	
x_3	0.0067	-0.0044	0.0002	1.0069	-0.0086	-0.0001	344.34
	9.73	*-0.31*	*0.01*	*37.92*	*-0.60*	*-0.62*	
x_4	-0.0067	0.0088	-0.0005	-0.0140	1.0173	0.0003	394.94
	-5.89	*0.37*	*-0.01*	*-0.32*	*42.69*	*0.76*	
x_5	0.1974	0.6610	-0.0371	-1.0515	1.2974	1.0193	312.92
	2.10	*0.34*	*-0.01*	*-0.29*	*0.66*	*36.17*	

Note: A total of 605 observations is used to compute the descriptive statistics. The data presented in this table are from a representative sector.

Table 4

Exogeneity Test for Robust Instruments

	c	z_1	z_2	z_3	z_4	z_5
Coef	0.0000	-0.0038	0.0370	-0.0209	-0.0033	-0.0001
P-value	*0.9971*	*0.8148*	*0.1945*	*0.4893*	*0.8394*	*0.6663*
R^2	0.0043					

Note: A total of 605 observations is used to compute the descriptive statistics. The data presented in this table are from a representative sector.

Table 5

OLS versus GMM_d and $Haus_d$ estimation methods for Fama-French and Pástor-Stambaugh models

	c	R_n-R_f	*SMB*	*HML*	*MOM*	*LIQ*	$\hat{\omega}_{R_m-R_f}$	$\hat{\omega}_{SMB}$	$\hat{\omega}_{HML}$	$\hat{\omega}_{MOM}$	$\hat{\omega}_{LIQ}$	R^2	*DW*
Fama-French													
OLS	0.0013	0.9679	0.1594	0.0722	0.0000							0.79	1.91
Abs *t*-mean	*2.92*	*46.96*	*6.39*	*5.57*	*2.24*								
GMM_d	-0.0007	0.9594	0.1375	0.2991	0.0013							0.69	1.82
$Haus_d$	0.0013	0.9594	0.1375	0.2991	0.0013		0.0098	0.0416	-0.2454	-0.0013		0.79	1.91
Abs *t*-mean	*2.84*	*11.47*	*1.53*	*1.81*	*1.52*		*1.49*	*1.08*	*1.57*	*1.31*			
Abs *t*-min	*0.28*	*6.17*	*0.48*	*0.33*	*0.45*		*0.30*	*0.16*	*0.08*	*0.07*			
Abs *t*-max	*6.30*	*19.74*	*3.79*	*4.73*	*4.22*		*3.26*	*2.18*	*4.59*	*4.05*			
# of signif. indices	7	12	4	4	4		4	2	4	2			
Pastor-Stambaugh													
OLS	0.0011	0.9648	0.1568	0.0718	0.0000	0.0110						0.79	1.90
Abs *t*-mean	*2.15*	*11.13*	*1.86*	*1.91*	*1.38*	*0.01*							
GMM_d	-0.0011	0.9475	0.1351	0.2860	0.0012	0.0150						0.67	1.84
$Haus_d$	0.0012	0.9475	0.1351	0.2860	0.0012	0.0150	0.0195	0.0421	-0.2333	-0.0011	-0.0059	0.79	1.90
Abs *t*-mean	*0.80*	*7.28*	*1.08*	*0.95*	*0.81*	*1.28*	*1.37*	*1.08*	*1.55*	*1.36*	*1.60*		
Abs *t*-min	*0.09*	*3.29*	*0.04*	*0.01*	*0.03*	*0.10*	*0.11*	*0.10*	*0.05*	*0.32*	*0.04*		
Abs *t*-max	*1.69*	*17.61*	*2.23*	*1.71*	*1.31*	*2.66*	*3.69*	*2.26*	*4.11*	*3.38*	*3.18*		
# of signif. indices	5	12	3	5	5	6	4	2	5	3	4		

Note: Results in this table are the averages of the 12 Fama-French sectors. The t-statistics are in italics. DW represents the Durbin-Watson statistics. $\overline{R}^2$ represents the adjusted R squared.

Table 6

OLS versus GMM_d estimation methods for Pástor-Stambaugh model by Fama-French sectors

		c	*R_m-R_f*	*SMB*	*HML*	*MOM*	*LIQ*	$\bar{R}^2$	*DW*
Sector	**Pastor-Stambaugh**								
1 NoDur	OLS	-0.0025	0.8626	0.0996	0.0767	0.0003	0.0278	0.80	1.75
	t-stat	*-1.90*	*44.49*	*2.97*	*2.17*	*1.30*	*1.57*		
	GMM_d	-0.0045	0.9071	0.1118	0.5644	0.0038	0.0213	0.72	1.68
	t-stat	*-0.50*	*6.15*	*0.42*	*1.31*	*1.21*	*0.32*		
2 Durbl	OLS	0.0053	1.1039	0.3777	0.4663	-0.0008	0.0387	0.81	1.97
	t-stat	*2.96*	*41.28*	*8.17*	*9.56*	*-2.25*	*1.58*		
	GMM_d	-0.0122	0.8310	0.9171	1.1571	0.0013	0.0938	0.65	1.95
	t-stat	*-0.95*	*3.77*	*2.23*	*1.71*	*0.27*	*1.04*		
3 Manuf	OLS	0.0054	1.0702	0.3174	0.1370	0.0002	0.0505	0.92	2.00
	t-stat	*5.48*	*72.66*	*12.47*	*5.10*	*1.16*	*3.74*		
	GMM_d	0.0060	1.0805	0.2177	0.2329	0.0000	0.1000	0.91	2.04
	t-stat	*1.32*	*17.61*	*1.75*	*0.96*	*-0.03*	*2.66*		
4 Enrgy	OLS	0.0000	0.9399	-0.0883	0.2520	0.0021	0.0418	0.62	1.87
	t-stat	*0.00*	*29.11*	*-1.58*	*4.28*	*4.65*	*1.41*		
	GMM_d	0.0224	1.4261	-0.4388	0.2256	0.0029	-0.1503	0.44	1.86
	t-stat	*1.69*	*7.42*	*-0.94*	*0.36*	*0.68*	*-1.63*		
5 Chems	OLS	0.0011	0.9613	-0.0054	0.0703	0.0002	0.0373	0.84	2.00
	t-stat	*0.86*	*50.37*	*-0.16*	*2.02*	*0.65*	*2.13*		
	GMM_d	-0.0105	0.8892	0.3167	1.0474	0.0062	0.0729	0.56	1.76
	t-stat	*-0.74*	*3.94*	*0.71*	*1.47*	*1.17*	*0.78*		
6 BusEq	OLS	0.0097	1.0921	0.4644	-0.5701	-0.0020	-0.0368	0.84	1.95
	t-stat	*5.97*	*44.84*	*11.03*	*-12.83*	*-5.86*	*-1.65*		
	GMM_d	-0.0024	0.7924	0.6595	-0.7455	-0.0040	-0.0087	0.79	1.88
	t-stat	*-0.25*	*4.93*	*2.15*	*-1.37*	*-1.03*	*-0.10*		
7 Telcm	OLS	-0.0038	0.8616	-0.0540	0.0910	-0.0005	-0.0433	0.69	1.95
	t-stat	*-2.27*	*34.02*	*-1.23*	*1.97*	*-1.39*	*-1.87*		
	GMM_d	-0.0099	0.5862	-0.2313	-0.5906	-0.0057	0.0805	0.50	1.83
	t-stat	*-0.97*	*3.29*	*-0.69*	*-1.01*	*-1.31*	*0.89*		
8 Utils	OLS	-0.0110	0.7529	-0.1083	0.3590	0.0014	-0.0207	0.60	1.87
	t-stat	*-6.26*	*28.62*	*-2.38*	*7.48*	*3.87*	*-0.86*		
	GMM_d	-0.0116	0.7717	-0.1767	0.4772	0.0031	-0.1176	0.56	1.86
	t-stat	*-1.14*	*5.45*	*-0.51*	*0.90*	*0.84*	*-1.73*		
9 Shops	OLS	0.0019	0.9510	0.3241	-0.0882	-0.0009	0.0121	0.82	1.87
	t-stat	*1.32*	*44.42*	*8.76*	*-2.26*	*-2.90*	*0.62*		
	GMM_d	0.0007	0.9365	0.1290	0.0883	-0.0002	0.1395	0.79	1.84
	t-stat	*0.09*	*8.12*	*0.44*	*0.25*	*-0.08*	*2.25*		
10 Hlth	OLS	0.0011	0.8947	-0.1081	-0.3381	-0.0002	-0.0420	0.74	1.95
	t-stat	*0.69*	*36.51*	*-2.55*	*-7.57*	*-0.66*	*-1.87*		
	GMM_d	-0.0032	0.9662	-0.0152	0.5575	0.0059	-0.1694	0.50	1.89
	t-stat	*-0.24*	*4.22*	*-0.04*	*0.73*	*1.05*	*-1.49*		
11 Money	OLS	0.0041	1.0494	0.1399	0.3265	-0.0001	0.0279	0.84	1.78
	t-stat	*2.97*	*50.68*	*3.91*	*8.65*	*-0.27*	*1.47*		
	GMM_d	0.0069	1.0363	-0.2465	-0.0038	-0.0020	0.1045	0.79	1.72
	t-stat	*0.96*	*8.75*	*-1.16*	*-0.01*	*-0.82*	*2.12*		
12 Other	OLS	0.0023	1.0382	0.5226	0.0794	0.0003	0.0389	0.92	1.89
	t-stat	*2.19*	*66.93*	*19.50*	*2.81*	*1.61*	*2.74*		
	GMM_d	0.0046	1.1466	0.3776	0.4217	0.0029	0.0138	0.87	1.80
	t-stat	*0.81*	*13.73*	*2.00*	*1.32*	*1.25*	*0.34*		

Note: Results in this table are each of the 12 Fama-French sectors. The t-statistics are in italics. DW represents the Durbin-Watson statistics. $\bar{R}^2$ represents the adjusted R squared.