THE PROBLEM OF ESTIMATING CAUSAL RELATIONS BY REGRESSING ACCOUNTING (SEMI) IDENTITIES

F. Javier Sánchez-Vidal*

WP-EC 2007-06

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

Primera Edición Julio 2007

Depósito Legal: V-3151-2007

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

---

* Dpto. de Economía Financiera y Contabilidad, Facultad de Ciencias de la Empresa, Universidad Politécnica de Cartagena, Paseo Alfonso XIII, 50, 30203 Cartagena (Murcia), Tel : +34 968325495, Fax: +34 968325782, Email: javier.sanchez@upct.es.

THE PROBLEM OF ESTIMATING CAUSAL RELATIONS BY REGRESSING ACCOUNTING (SEMI) IDENTITIES

F. Javier Sánchez-Vidal

ABSTRACT

Inferences about the coefficient values of a model estimated with a linear regression cannot be made when both the dependent and the independent variable are part of an accounting (semi) identity. The coefficients will no longer indicate a causal relation as they must adapt to satisfy the identity. A good example is an investment-cash flow sensitivity model.

KEYWORDS: Investment-cash flow sensitivities, Accounting identities, Accounting semi-identities.

JEL classification: G32, B4

RESUMEN

Este trabajo habla de la imposibilidad de extraer conclusiones sobre el valor de los coeficientes de un modelo de regresión lineal que intenta estimar una relación causal, cuando tanto la variable dependiente como la variable independiente forman parte de una (semi) identidad contable. Los coeficientes no sirven para explicar la relación causal, ya que su valor se adaptará para cumplir la identidad. Como ejemplo ilustrativo se presenta el modelo de la sensibilidad de la inversión al cash-flow.

PALABRAS CLAVE: Sensibilidad de la inversión al cash flow, identidades contables, semi-identidades contables.

1. Introduction

In the Fazzari, Hubbard and Petersen (1988) (FHP) model, investments that a company makes in a certain year are dependent on the cash flows the company was generating in the same period. FHP (1988) relate these two accounting variables by a linear regression model, calling the coefficient of the explanatory variable "investment-cash flow sensitivity."

For companies with severe information asymmetry, external finance is either too expensive or not available. FHP (1988) argue that the higher the coefficient, the more that company investments depend on company's ability to generate cash flows to finance them. So, these higher coefficients would describe companies with more severe information asymmetry.

Since then, Kaplan and Zingales (1997, 2000), (KZ) have used information in the required 10-k, or annual reports, on dividends, stock repurchase, firm liquidity measures, and the company's own opinions on obtaining funding to classify firms as (1) not financially constrained, (2) likely not financially constrained, (3) possibly financially constrained, (4) likely financially constrained and (5) financially constrained.[1] Estimation of FHP (1988) model in this case reveals that coefficients strongly reject the hypothesis that the more financially constrained a company is, the greater its investment-cash flow sensitivity. The KZ results are robust to different ways of subsampling, various financing restriction definitions, alternative specifications of the regression equation, different definitions of investments and controlling for outliers.

KZ (1997) conclude that the reason behind this paradox may be related either to the nonmonotonic investment-cash flow sensitivity to the degree of financing constraints or to a mispecification of the external finance cost function, in that all the factors causing a company to raise external finance are not fully explained. KZ (1997) also conclude that their findings are not caused by an inappropriate classification scheme or econometric problems.

---

[1] Kaplan and Zingales (2000) is a response to Fazzari, Hubbard and Petersen (2000), which is itself a reply to KZ (1997).

I believe the last conclusion is incorrect. The answer to why investment-cash flow sensitivities are not valid measures is simple. Coefficients do not mean the same when we are regressing causal relationships that are at the same time accounting or accounting semi-identities as when we are regressing only pure causal relations, because the identity has to be mathematically fulfilled.

2. The FHP (1998) equation is an accounting semi-identity

Let us take a look at what it is being regressed in estimation of the FHP (1988) investment-cash flow sensitivity equation. These are simply left-hand side and right-hand side long-term components of the balance sheet of a company. Figure I demonstrates the components.

Figure II shows the long-term elements on both sides of the balance sheets, considering only the increase that has taken place in one period, and taking into account that the increase in retained earnings is the result of the cash flows generated in that year minus dividends and depreciation.

Isolating the increase in the fixed assets, i.e., this year's investments, we have:

$$\mathrm{Inv} = \mathrm{CF} + (\Delta\ \mathrm{LTD} + \Delta\mathrm{Capital\ Stock} - \mathrm{Depreciation} - \mathrm{Dividends} - \Delta\mathrm{Working\ capital} - \Delta\mathrm{OFA})$$

FHP (1988) suggest regressing the investments on the cash flows, [2] (see Figure II):

$$\mathrm{Inv} = \alpha + \beta\mathrm{CF} + \varepsilon \qquad (1)$$

where the accounting identity is now an accounting semi-identity. That is, one important part is missing (the elements in the parentheses above, henceforth called the "rest"). The model with the complete accounting identity is given by equation 2:

[2] The complete FHP (1998) equation model also includes the Tobin's q. In pages 7 and 8, before Table II, I explain how the inclusion of a variable that is not comprised in the accounting identity could affect results.

FIGURE I. Balance sheet of a company.

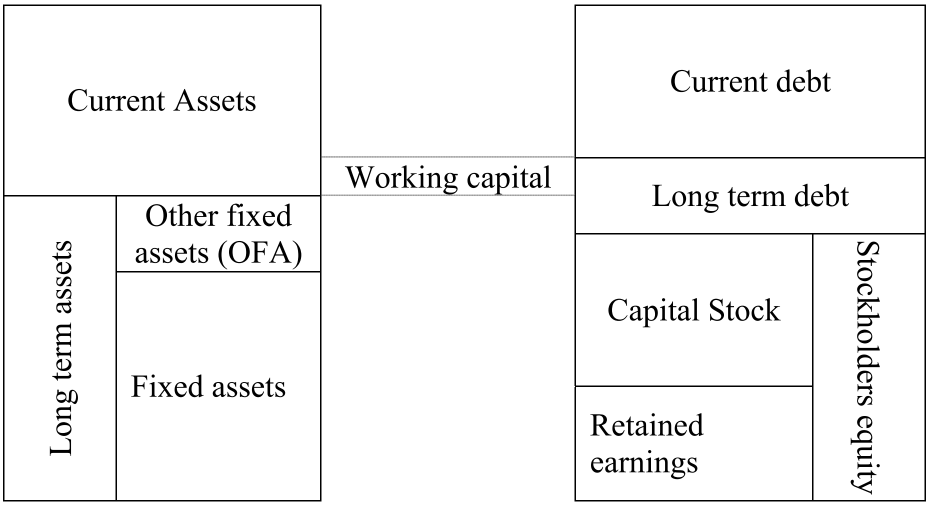
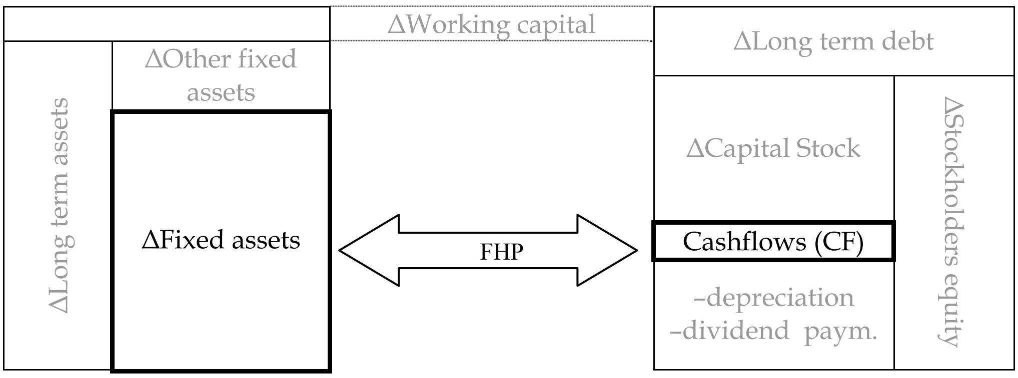


FIGURE II. Increases in long-term elements of the balance sheet.



$$Inv_{it} = \alpha + \beta_1 CF_{it} + \beta_2 Rest_{it} + \varepsilon_{it} \qquad (2)$$

where:

$$Rest_{it} = \Delta LTD_{it} + \Delta Capital\,Stock_{it} - Depreciation_{it} - Dividends_{it} - \Delta Working\,capital_{it} - \Delta OFA_{it}$$

A problem arises when the researcher wants to infer conclusions from the value of the estimated coefficient, β of equation 1, as this value is no longer valid as an expression of the causal relationship. The reason is that both the dependent and the independent variable are part of an accounting identity, and results are not the same as in estimation of pure causal relationships because the coefficients must adapt to satisfy that identity.

If we regress the complete accounting identity for a sample of firms, we should obtain that:

$$Inv_{it} = 0 + 1 * CF_{it} + 1 * Rest_{it} + 0$$

To see this, let us examine a sample of 20.000 companies for the year 2000 and regress equation 2. Results are shown in Table I. Coefficients of cash flows and the "rest" become unity, and the constant turns out to be 0 to accomplish the accounting identity.

TABLE I. Results of OLS regression of equation 2: $Inv_{it} = \alpha + \beta_1 CF_{it} + \beta_2 rest_{it} + \varepsilon_{it}$ **for a sample of 20.000 companies in 2000. All the terms are scaled by net fixed assets.**

	Coef.	t
Cash flows	1,0000	∞***
Rest	1,0000	∞***
Constant	0,0000	0,000
Adjusted-R2	1,0000	

*** denote significance at 1% level.

3. When parts of the accounting identity are missing

In the FHP (1988) equation, the "rest" of the accounting identity does not enter the model. Consider a company that has invested \$5 million in a year, has generated \$7 million of cash flows, and the "rest" is equal to -\$2 million. To satisfy the complete identity, equation 2 for this company would lead to these results:

5 = 0 + 1 * 7 + 1 * -2

But, as the -2 is missing in equation 1, the coefficient of the cash flows variable will have to offset this absence:

5 = 0,7143 * 7 , that is, a value lower than 1. [3]

If the “rest” is positive, on the other hand, the coefficient will be greater than one to make the identity. So, the higher coefficients do not correspond to companies with greater information asimmetry, but to companies with a positive “rest,” as can be seen in Table II. In this table, I regress equation 1 for subsamples of positive/negative “rest”: the coefficient above/below the unity confirm my assertions. I have eliminated extreme values for the investment, cash flows and “rest” variables, because high heterogeneity sometimes reflects in a high constant that takes some positivity out of the cash flow coefficient, as is the case of the subsample for the positive “rest” even after controlling for outliers.

TABLE II. Results of regression of equation 1: $Inv_{it} = \alpha + \beta_1 CF_{it} + \varepsilon_{it}$ for the whole sample and for a subsample with a positive "rest" and a subsample with negative "rest", taken from a sample of 20.000 companies in 2000.

	Whole sample		**Positive "rest"**		**Negative "rest"**	
	Coef.	t	Coef.	t	Coef.	t
Cash flows	0,147	64,64***	1,081	99,26***	0,180	113,62***
Constant	0,118	85,83***	13,587	105,03***	0,019	15,28***
Adjusted-R2	0,069		0,392		0,286	

, * denote significance at 5% and 1% levels, respectively.

When analysing accounting semi identities, the presence of the intercept and the inclusion of other variables alters to some extent my arguments. If the value of the intercept is too high it could cause the coefficient of the positive “rest” subsample to be lower than one. But in spite of this “problem” when I have regressed different samples by changing the percentage of elimination of outliers, the results are stubborn: in every case, the cash flows coefficient of the positive “rest” subsample was always higher than the one of the negative “rest” subsample.

When adding a “foreign” variable, that is, a variable that is not included in the accounting identity, the effect would depend on the sign and value of the product of both the value of the estimated variable coefficient and the variable itself. With respect

[3] I have added an econometric example of these arguments in annex 1 for the skeptical reader.

to the validity of the coefficient of the "foreign" variable, I believe that as we are only using the semi identity, there is space for variability (because the R2 are no longer 100%) and for this variable to play its role in affecting the dependant variable, and thus, I believe its coefficient is a valid exponent of this particular causal relation.

From an econometric point of view, if the complete identity is $y_i = \beta_1 x_{1i} + \beta_2 x_{2i}$, with β_2 and β_2=1, but instead we regress y only on x_1 and a constant:

$$y_i = \alpha + b_1 x_{1i} + \varepsilon_{it},$$

there will be a bias if the included variable, x_1, is correlated with the excluded variable, x_2, that will lead to an specification bias, because: $E(b_1) = \beta_1 + \beta_2 \frac{\sum x_{1i} x_{2i}}{\sum x_{1i}^2}$, see Gujarati (1995).

And this is really the case here, as $x_{1i} = y_i - x_{2i}$, due to the accounting identity. So, this is another way of demonstrating that b_1 depends on the value of x_2. If the rest has, for example, the same sign as the cash flows, the correlation will lead to a positive bias, causing the coefficient to be higher than 1 as both β_1 and β_2 are equal to 1. Results of table II confirm this.

We can now see the problem in the work by KZ (2000). Less constrained companies will likely show higher investment-cash flow sensitivities, because it is easier for such companies to add long-term debt, which would increase the value of the "rest," causing the investment-cash flow sensitivity to rise.

Table III shows regression results for equation 1 in companies where the "rest" minus the variation in long-term debt is approximately 0 (no greater than ±0,5% with respect to net fixed assets), so the variation in the "rest" is driven mainly by the variation in long-term debt. There are two subsamples, where the long-term debt variation is higher than and below 0. The first subsample is clearly less financially constrained than the second one, but shows a higher investment-cash flow coefficient.

TABLE III. ***Results of regression of equation 1:*** $Inv_{it} = \alpha + \beta_1 CF_{it} + \varepsilon_{it}$ **for two subsamples in which the increase in the long term debt is above/below 0 and the sum of the other components of the "rest" is aproximately 0, taken from a sample of 20.000 companies in 2000.**

	Positive variation in long-term debt		**Negative variation in long-term debt**	
	Coef.	t	Coef.	t
Cash flows	1,608	18,75***	0,328	13,25***
Constant	0,181	5,93***	-0,004	-0,78**
Adjusted-R2	0,528		0,344	

, * denote significance at 5% and 1% levels, respectively.

4. Conclusions

Knowledge of financial constraints and their implications for investment, job creation, and the survival of many cash-constrained firms is crucial to both economic research and economic policy. The model developed by Fazzari, Hubbard and Petersen purports to be helpful in classifying firms as more or less financially constrained. In fact, the model is not useful as the values of its coefficients are not valid measures of the degree of financing constraints in a firm. The reason is that the model is an accounting semi-identity, that is, the coefficients of the explanatory variable must satisfy this identity, and thus it cannot serve to identify causal relationships.

My findings can be applied to any model in which the dependent and the independent variable are on both sides of an accounting identity.

REFERENCES

Fazzari, S.; Hubbard, R. G. and Petersen, B. (1988): “Financing Constraints and Corporate Investment,” *Brookings Papers on Economic Activity*, pp. 141-195.

Fazzari, S.; Hubbard, R. G. and Petersen, B. (2000) “Investment-Cash Flow Sensitivities are Useful: A Comment on Kaplan and Zingales,” *Quarterly Journal of Economics*, CXV, pp. 695-705.

Gujarati, (1995) : *Basic Econometrics*, 3rd edition, Ed McGraw Hill, New York.

Kaplan, S. and Zingales, L. (1997): “Do Financing Constraints Explain Why Investment is Correlated with Cash Flow?” *Quarterly Journal of Economics*, CXII, pp. 169-215.

Kaplan, S. and Zingales, L. (2000): “Investment-Cash Flow Sensitivities Are Not Valid Measures of Financing Constraints,” *Quarterly Journal of Economics*, CXV, pp. 707-712.

ANNEXES

Annex I. *Results of regression of equation 1:* $Inv_{it} = \alpha + \beta_1 CF_{it} + \varepsilon_{it}$ for a subsample in which the amount of "rest" is approximately (±0,5%) the 50% of the amount of cash flows, taken from a sample of 20.000 companies in 2000.

	Coef.	**t**
Cash flows	1,500	6.369,16***
Constant	-0,000	-0,47
Adjusted-R2	1,0000	

*** denote significance at 1% level.