Econometric methodology applied to financial systems

Salvador Climent-Serrano; University of Valencia; Salvador.climent@uv.es

In this paper we review the main characteristics of econometric models, focusing on their application in financial systems. The objective is to give an overview of the econometric technique and its practical application. The methodology used is a description of the bases of econometrics and the properties it assumes. The different types of adjustments are developed so that they can be made (logarithmic, linear, etc.), and the goodness-of-fit measurements. Finally, they are analysed by the practical application of the different types of variables and their real application to regression models on Spanish credit institutions.

Introduction: Economic and econometric models

In a market economy, we can assume that the higher the GDP, the greater will be the deposits in credit institutions. At first, we can also assume that this relationship is linear. In this case, an example of the relationship between two variables can be seen in Figure 1. In this case we are dealing with an economic model.

The representation will be: $D = \beta_1 + \beta_2 GDP \qquad 0 < \beta_2 < 1$

In the representation, D are the bank deposits, GDP the gross domestic product, and β_1, β_2 the parameters of the model.

However, when we look at the real economy, the relationship between GDP and deposits is not exact and may be more like the representation shown in Figure 2.

Figure 1: Linear adjustment

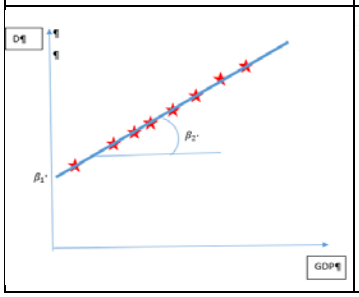
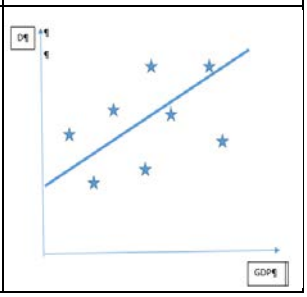


Figure 2: Real situation



The relationships between economic variables are usually not exact, since they depend on multiple factors, such as the propensity to save or to spend, age, ease of access to bank branches, even random factors. To take all these factors into consideration, a random term "u" is included in the model, which we will call error or resid, which is not observable and which represents the factors that are not in the deterministic part of the model. In this case it will be an econometric model. So the model will be:

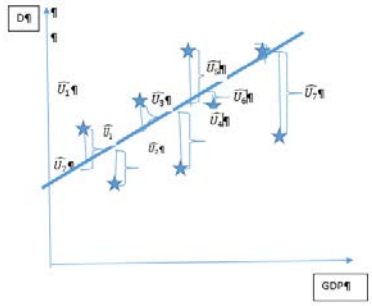
$$D = \beta_1 + \beta_2 GDP + u \qquad 0 < \beta_2 < 1$$

The objective of the regression is to obtain the parameters β_1 y β_2 from an available sample.

Estimates should be sought for $\widehat{\beta_1}$ y $\widehat{\beta_2}$, (as the results obtained are estimates, a circumflex accent is introduced above the betas, or of what we estimate).

If we look at Figure 2, the distance from each point to the line will be $D_i - \widehat{D_i} = \widehat{u_i}$. This error or resid is the difference between the observed value of the endogenous variable (dependent variable) and the adjusted value.

Figure 3: Adjustments



The objective of the regression is to adjust the line to the points so that all the residuals or errors, which are shown in Figure 3, are as small as possible.

Mathematically as in the $\sum_{i=1}^{N} \widehat{u_i}$, we have positive and negative errors; so that they are not compensated, we modify the equation to: $\sum_{i=1}^{N} \hat{u}_i^2$.

This adjustment is called ordinary least squares and the intention is to minimize the expression: $min \sum_{i=1}^{N} \hat{u}_i^2 = min \sum_{i=1}^{N} \left(GDP_i - \widehat{GDP_i}\right)^2$

The goal is to find $\widehat{\beta_1}$ and $\widehat{\beta_2}$) that minimize the expression: $min \sum_{i=1}^{N} \hat{u}_i^2 = min \sum_{i=1}^{N} \left(GDP_i - \widehat{\beta_1} - \widehat{\beta_2 GDP_i}\right)^2$

To minimize, firstly it is derived partially with respect to $\widehat{\beta_1}$ and $\widehat{\beta_2}$, The first order condition is obtained by equating the expressions obtained to zero. Solving the equations, we obtain the normal equations and solving these we arrive at the following expressions:

$\widehat{\beta_1} = \overline{D} - \widehat{\beta_2}\overline{GDP}$ $\widehat{\beta_2} = \frac{Cov\,(GDP,D)}{Var\,(GDP)}$

Dependent Variable: _2_DEPOSITS
Method: Panel Least Squares
Date: 12/12/18 Time: 00:22
Sample: 2004 2017
Periods included: 14
Cross-sections included: 77
Total panel (unbalanced) observations: 569
White diagonal standard errors & covariance (no d.f. correction)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.720250	0.002407	299.2344	0.0000
_6_REAL_GDP	0.296165	0.083927	3.528834	0.0005

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.798783	Mean dependent var	0.724545
Adjusted R-squared	0.767228	S.D. dependent var	0.113928
S.E. of regression	0.054966	Akaike info criterion	-2.837470
Sum squared resid	1.483447	Schwarz criterion	-2.241999
Log likelihood	885.2602	Hannan-Quinn criter.	-2.605117
F-statistic	25.31367	Durbin-Watson stat	0.981463
Prob(F-statistic)	0.000000		

The estimated line will be $D_i = \beta_1 + \beta_2\, GDP_i\, U_i$. This means that $\widehat{D_i} = \beta_1 + \beta_2\, GDP_i$

for example $\widehat{D_i} = 5 + 0.30\, GDP_i$.

Properties:

1.-The sum of the residuals is zero: $\sum_{i=1}^{N} \widehat{U} = 0$

2.-The adjusted line passes through the point of means: $\overline{GDP}, \overline{D}$

3.-The sum of the cross products of the residuals with the explanatory variable is zero: $\sum_{i=1}^{N} \widehat{U}\, GDP = 0$

4.-The sum of the cross products of the resid with $\widehat{D_t}$ is zero: $\sum_{i=1}^{N} \widehat{U}\, \widehat{D} = 0$

Non-linear adjustments

In the event that the variables are not linear, non-linear estimates can be made with alternative functional forms, for example: Logarithmic adjustment: $ln\widehat{D_i} = \beta_1 + \beta_2\, lnGDP_i + U_i$. In this case, the interpretation of the coefficient is related to the elasticity. In other words, if GDP changes 1% (ΔGDP / GDP) = 1%, then D will change $(\Delta D/D) = \beta_2\%$.

As well as this functional form, others can be estimated. The summary of the alternative functional forms, and the interpretation of each of the parameters, is shown in Table 1.

Table 1: Functional forms of the econometric regressions

	MODEL	MARGINAL PROPENSION	ELASTICITY
LINEAR	$D_I = \beta_1 + \beta_2 * GDP_i + U_i$	$\hat{\beta}_2$	$\hat{\beta}_2 * \frac{\overline{GDP}}{\overline{D}}$
INVERSE	$D_I = \beta_1 + \beta_2 * \frac{1}{GDP_I} + U_i$	$\widehat{-\beta}_2 * \frac{1}{\overline{GDP^2}}$	$\widehat{-\beta}_2 * \frac{1}{\overline{GDP} * \overline{D}}$
LINEAR-LOG	$D_I = \beta_1 + \beta_2 * \ln GDP_i + U_i$	$\hat{\beta}_2 * \frac{\overline{D}}{\overline{GDP}}$	$\hat{\beta}_2 * \frac{1}{\overline{D}}$
LOG - LOG	$lnD_I = \beta_1 + \beta_2 * \ln GDP_i + U_i$	$\hat{\beta}_2 * \frac{1}{\overline{GDP}}$	$\hat{\beta}_2$
LOG-LINEAR	$lnD_I = \beta_1 + \beta_2 * GDP_i + U_i$	$\hat{\beta}_2 * \overline{D}$	$\hat{\beta}_2 * \overline{GDP}$
LOG-INVERSE	$lnD_I = \beta_1 + \beta_2 * \frac{1}{GDP_{I_i}} + U_i$	$\widehat{-\beta}_2 * \frac{\overline{D}}{\overline{GDP^2}}$	$\widehat{-\beta}_2 * \frac{1}{\overline{GDP}}$

Figure 4: Measures of goodness of the estimation by econometric regressions

Dependent Variable: _2_DEPOSITS
Method: Panel Least Squares
Date: 12/12/18 Time: 00:22
Sample: 2004 2017
Periods included: 14
Cross-sections included: 77
Total panel (unbalanced) observations: 569
White diagonal standard errors & covariance (no d.f. correction)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.720250	0.002407	299.2344	0.0000
_6_REAL_GDP	0.296165	0.083927	3.528834	0.0005

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.798783	Mean dependent var	0.724545
Adjusted R-squared	0.767228	S.D. dependent var	0.113928
S.E. of regression	0.054966	Akaike info criterion	-2.837470
Sum squared resid	1.483447	Schwarz criterion	-2.241999
Log likelihood	885.2602	Hannan-Quinn criter.	-2.605117
F-statistic	25.31367	Durbin-Watson stat	0.981463
Prob(F-statistic)	0.000000		

Coefficient of determination or R^2 being $R^2 = \frac{\sum_{i=1}^{N}(\widehat{D_i} - \widehat{D})^2}{\sum_{i=1}^{N}(D_i - \overline{D})^2}$; therefore, R^2 (R squared) is the proportion of the sample variance of D explained by the regression; the greater R squared is, the better the adjustment.

Multiple linear regression

A usual extension of the simple linear regression model is to include more than one explanatory variable in the model.

For example, that bank deposits depend on the GDP and the interest rate (I). In this case, the econometric model would be the following:

$$D_I = \beta_1 + \beta_2 GDP_i + \beta_3 I_i + U_i$$

Beta 2 will be expected to be positive, with more GDP plus bank deposits. Beta three is expected to be negative; the higher the interest rate, the lower the incentive to have the money in bank deposits and the higher in other financial assets.

For the estimation of the coefficients, beta, with a simple linear adjustment, with two variables it has been possible to intuit on the two-dimensional Cartesian axis. In the same way as with two variables, it can also be done with three variables, so one can intuit on a three-dimensional axis. With n variables you cannot intuit visually, since an n dimensional axis will be used, but the process is the same. In this case, the procedure for estimating betas is the same as that developed with two variables, but with matrix notation.

Basic assumptions that must comply with residuals or errors, u_i

$E\ (u_i) = 0$ residuals are balanced between positive and negative.

Homoscedasticity: $Var\ (u_i) = E\ (u_i^2) = \sigma^2$ The variance is constant. This is verified by the White test; if homoscedasticity persists, the Generalized Moment Method (MCM) procedure can be used since this procedure allows the presence of the homoscedasticity to estimate the coefficients of the model when using panel data. (Arellano y Bold, 1991)

No autocorrelation: $cov\ (u_{i,}\ u_j) = E(u_{i,}\ u_j) = 0$. This is checked by the Durbin Watson test.

The explanatory variables are not related to the residual: $cov\ (GDP_{i,}\ u_i) = 0$

No multicollinearity: The independent variables are linearly independent of each other.

The vector of the residuals "u" has a **normal distribution**. This is checked by the Jarque-Bera Test.

The model is well specified, in the form and in the variables.

Figure 5: Regression model using the Eviews application

Dependent Variable: _2_DEPOSITS
Method: Panel Least Squares
Date: 12/16/18 Time: 00:13
Sample: 2004 2017
Periods included: 14
Cross-sections included: 77
Total panel (unbalanced) observations: 569
White diagonal standard errors & covariance (no d.f. correction)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.741834	0.005231	141.8253	0.0000
_6_REAL_GDP	0.426065	0.086496	4.925809	0.0000
_6_EURIBOR	-0.703335	0.253538	-2.774075	0.0057
_6_IAPC	-0.290112	0.295553	-0.981591	0.3268

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.808275	Mean dependent var	0.724545
Adjusted R-squared	0.777301	S.D. dependent var	0.113928
S.E. of regression	0.053764	Akaike info criterion	-2.878762
Sum squared resid	1.413468	Schwarz criterion	-2.268023
Log likelihood	899.0077	Hannan-Quinn criter.	-2.640452
F-statistic	26.09531	Durbin-Watson stat	1.012126
Prob(F-statistic)	0.000000		

To validate the model, the level of significance of each of the betas or parameters of the model is used. The level of significance or P-value is used, normally using levels of significance of 1%, 5% or 10%. With these levels of significance the variable is considered valid and for the whole model, the F test is used.

In this case, as shown in Figure 5, the real GDP and the interest rate are significant variables, with a level of significance lower than 1% (confidence level above 99%), while the IPAP is not significant, since its p-Value is 0.3268.

TYPE OR CLASSES OF VARIABLES

The variables that can be used in the regression models can be:

Metric variables: for example: GDP, bank deposits, interest margin, capital, profits, etc.

Variables with ordinal scale: for example: small, medium, large, systemic

Dummy variables: they are artificially constructed variables that collect this type of qualitative information, for example: bank, savings bank. Normally one is coded with zero (0) and the alternative with one (1). The interpretation is the difference between one of the variables and the alternative.

Figure 6

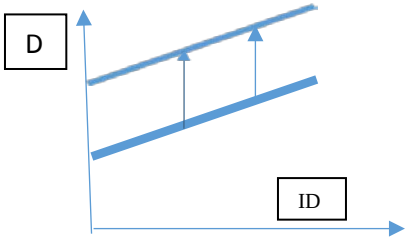


Figure 6 shows the change that would occur if the dummy variable corresponding to a beta were positive. The economic interpretation to be given would be, as an example, if the bank is assigned the 1 and the savings banks the 0, that with the same GDP, the bank gets more deposits than the savings banks.

It can also be used for more than two categories, for example to differentiate sectors or autonomous communities etc.

Figure 7

Equation: Z_DET_BALANCE Workfile: PANEL 2015_TRAB::Untitled\
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: _0_DET_CREDITOS
Method: Panel EGLS (Cross-section weights)
Date: 12/17/18 Time: 20:31
Sample (adjusted): 2005 2015
Periods included: 11
Cross-sections included: 74
Total panel (unbalanced) observations: 455
Linear estimation after one-step weighting matrix
White diagonal standard errors & covariance (no d.f. correction)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011423	0.001487	7.680269	0.0000
_1_BDE	0.041622	0.007776	5.352466	0.0000
_1_CRED_DEP	-0.002492	0.000705	-3.537298	0.0004
_1_FP_ACTIVO	-0.075622	0.013459	-5.618596	0.0000
_1_ADJUDICACIONES	0.167371	0.025980	6.442228	0.0000
_3_DUMMY_12	0.026733	0.002909	9.189745	0.0000
_0_DET_CREDITOS(...	0.163841	0.035502	4.614991	0.0000

Weighted Statistics

R-squared	0.653781	Mean dependent var	0.021166
Adjusted R-squared	0.649144	S.D. dependent var	0.024472
S.E. of regression	0.012634	Sum squared resid	0.071510
F-statistic	140.9966	Durbin-Watson stat	1.721722
Prob(F-statistic)	0.000000		

Figure 7 shows an example of a model with metric variables and a dummy variable. On the left, the results are shown as they appear in the Eviews application. In this case, the model studies the losses by default of the Spanish banks with a sample that goes from 2004 to 2015.

The explanatory variables are all significant. This is verified with the p-value (Prob.), which is obtained from the t-statistic.

The different types of variables are:

1.- Variable with positive sign. BDE, the money that the banks have requested from the ECB. In this case, the results indicate that the greater the need for central bank funds, the greater the losses due to late payments.

2.- Variable with negative sign. The relationship of equity assets; in this case the results indicate that the credit entities that have more equity funds have fewer losses due to delinquency.

3.- Control variable. To give robustness to the model, control variables are introduced, which are variables whose sign is predicted in advance. In this case they are the adjudications that the credit entities have. Here it is obvious that the result has to be positive and with a significant value. In this case we see that it is positive and with a coefficient of 0.16, much higher than the rest of the coefficients.

4. Dummy variable. Dummy_12. This variable represents the change in regulations that occurred in 2012 with respect to the allocation of provisions for impairments by what are colloquially known as "Guindos" decrees. The results indicate that the impairment losses in 2012 were 2.67% of the total assets as a consequence of this change. The way to include it in the model is to give the value zero to all the years of the sample, except for the year 2012 which is assigned the value one.

5. Variables with delay. In some cases, the dependent variable may have inertia, that is, it depends on the same variable from the previous year. This case occurs in delinquency, where the results of the variable DET_CREDITOS (-1) indicate that the losses due to delinquency depend 16% on the delinquency of the credit institution of the previous year. The way to include this variable is by introducing the dependent variable as explicative, with a delay of one year; if it is significant, this indicates that there is inertia. When entering the variable with a delay, one year of the sample is lost.

TYPE OR CLASSES OF DATA

Time series data: data of the same variable at different moments of time; for example: the profits of Banco Santander in 2010, 2011, 2012 etc.

Cross-section data: data of the same variable in the same period of time; for example: the profits in 2015 of Banco Santander, BBVA, Bankia etc.

Panel data: Combination of the two previous ones, which is usually used in the studies of the financial entities when the data is available; for example: the profits of Banco Santander, BBVA, Bankia, etc. in 2010, 2011, 2012 etc.

In addition, the form of the series can be modified; for example, if it is used to see how interest rates affect non-performing loans and loan defaults, the relationship may not be linear but quadratic, so the variable of interest rates is introduced to the table, etc.

It can also be introduced in a multiplicative way with fictitious variables, for example, interest rates multiplied by bank (1) and savings bank (0). In this case, the information it provides is how a variation in interest rates affects banks and savings banks in a different way.

OTHER CHECKS

The Chow test (Chow, 1960) is used to test that there is no structural break in the model, that is, that the linearity does not break. But it can also be used to the contrary, to show that when a certain effect occurs, the variables do not behave in the same way. For example, if you study how GDP affects delinquency, the effect may not be the same during stages of growth as in recession and it can be contrasted that in year X (moment of change) there is a structural change, so that the coefficients of the parameters during growth stages will be different to the same parameter, but during recession stages.

Acknowledgement: The authors wishes to thank the support of the Chair of International Finance-Banco Santander.

REFERENCES

Arellano M. and Bond S. (1991) Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. The Review of Economic Studies, Vol. 58, No. 2 (Apr., 1991), pp. 277-297 Stable URL: https://www.jstor.org/stable/2297968

Chow G.C. (1960) Tests of Equality Between Sets of Coefficients in Two Linear Regressions Econometrica, 28, (3) pp. 591-605 DOI: 10.2307/1910133

Granger C. W. J. (1969) Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. Econometrica, 37, (3) pp 424-438 DOI: 10.2307/1912791

Uriel Jiménez, E. and Gea Rosat. I. (1967) Econometría aplicada. Madrid : AC