

Volume 37, Issue 2

Murphy-Topel adjustment of the variance-covariance matrix of a two-step panel data model: Evidence from competition-fragility nexus in banking

Ion Lapteacru

Larefi, Department of Economics, University of Bordeaux

Abstract

We develop the Murphy-Topel adjustment of the variance-covariance matrix for two-step panel data models. We apply it on the competition-fragility nexus in banking with different samples for two equations. Indeed, this issue is often observed in this field of research. A competition measure of banks is constructed for each country (first equation), whereas a risk measure is regressed on the entire sample of countries (second equation). Any statistical adjustment will only provide approximate results for the second equation, because of possible correlations between the results of both models. As a precise adjustment, the Murphy-Topel method seems to be more appropriate.

Citation: Ion Lapteacru, (2017) "Murphy-Topel adjustment of the variance-covariance matrix of a two-step panel data model: Evidence from competition-fragility nexus in banking", *Economics Bulletin*, Volume 37, Issue 2, pages 902-910

Contact: Ion Lapteacru - ion.lapteacru@u-bordeaux.fr.

Submitted: June 21, 2016. **Published:** May 01, 2017.

Introduction

Many papers in economic empirical literature treat a subject with a two-step econometric model. For instance, Buch *et al.* (2013) compute the market power of banks (Lerner index and Boone indicator) from a first step regression and then apply it into a second regression equation in order to determine the impact of the banks' internationalisation on the market power – risk nexus. Tabak *et al.* (2012) use also a two-step econometric model in order to determine whether size and market capitalisation matters in the relationship between banking market power (Boone indicator) and risk-taking. The impact of the market power (Boone indicator) on the interest rate pass-through in the euro area is studied by Van Leuvensteijn *et al.* (2013) within a two-step econometric model too. Schaeck *et al.* 2009 estimate, at the first stage, the level of competition of banks (H-statistic of Panzar and Rosse) and check, at the second stage, its impact on banking systems' stability. And Leon (2015) applies all three bank competition measures (Lerner index, Boone indicator and H-statistic) to study their effect on bank credit constraints in developing countries.

The authors regress a first equation whose results are then applied in the computation of an explanatory variable (Lerner index, Boone indicator and/or H-statistic) that is used in the regression of a second equation. And many of them make these regressions independently of each other, considering therefore that this explanatory variable has no estimation error. Nevertheless, the errors of parameters from the first regression provide to this variable a random character and the values of parameters allow the determination of its expected value, which cannot be considered as an observed data unlike other explanatory variables. Consequently, since it depends on the regression results of the first model, the precision (the standard errors) of parameters of the second model must be adjusted.

Given this embedding of regressions and results, few authors merely bootstrap the standard errors of the second equation (for instance, Schaeck *et al.* 2009, Buch *et al.* 2013). However, this statistical procedure does not take into account the results of the first equation, neither its parameters nor its variance-covariance matrix, and does not consider a possible correlation between these results and those of the second equation, thereby providing approximate results. Murphy and Topel (1985, 2002) propose an adjustment of the variance-covariance matrix for the second model and therefore of its standard errors. The aim of the present paper is to develop this procedure for panel data models with different samples for the first and second equations, and to apply it to the estimation of the competition-fragility relationship in banking.

This subject is widely studied in empirical banking literature, as suggested by few studies cited above, and all papers are made on panel data. However, very few of them bootstrapped the standard errors and none made a precise adjustment with the Murphy-Topel approach. The importance of the subject and the difficulty to estimate correctly the standard errors of the second equation, especially with different samples for regression equations, lead us to develop the Murphy-Topel adjustment of the variance-covariance matrix of two-step panel data models.

1. Two-step econometric model and Murphy-Topel variance-covariance matrix

Two-step econometric model consists of two equations, where the parameters of the first equation will appear in the second equation and the precision of their estimation, that is, the variance-covariance matrix, interferes therefore in the precision of the second equation's parameters.

Let

1st stage (model 1): $y_1 = f_1(\mathbf{x}_1, \boldsymbol{\alpha})$

be the first equation, whose econometric results, that is, coefficients and variance-covariance matrix, are involved in the regression of the following second equation

2nd stage (model 2): $y_2 = f_2(\mathbf{x}_1, \mathbf{x}_2, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta})$,

where \mathbf{x}_{i1} and \mathbf{x}_{i2} are vectors of explanatory variables of the first and second equations, respectively, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of parameters of model 1 and model 2, respectively. Hat symbol means that parameters are estimated. Vector $\boldsymbol{\beta}$ includes also the coefficient $\boldsymbol{\gamma}$ that connects to the second equation the explanatory variable computed with the parameters of the first equation.

Unlike many empirical studies that use two-step econometric models and consider the variables with coefficient $\boldsymbol{\gamma}$ as they contain observed data, one might consider them as estimators and adjust therefore the variance-covariance matrix of the second equation. Murphy and Topel (1985, 2002) proposes a full information maximum likelihood estimation with two-step procedure that consists to estimate, firstly, the parameters $\boldsymbol{\alpha}$ of model 1 by maximizing its likelihood function

$$\ln L_1(\boldsymbol{\alpha}) = \sum_i \ln f_1(y_{1i} | \mathbf{x}_{1i}, \boldsymbol{\alpha}) \quad (1)$$

and, secondly, the parameters $\boldsymbol{\beta}$ of model 2 by maximizing the respective likelihood function

$$\ln L_2(\hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}) = \sum_i \ln f_2(y_{2i} | \mathbf{x}_{1i}, \mathbf{x}_{2i}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}), \quad (2)$$

where the embedded consistent estimator $\hat{\boldsymbol{\alpha}}$ is treated as given. Greene (2012, p. 576) explains the reasons why one might proceed with two-step estimations. Among them, the easiness to formulate and to maximize the two separate log-likelihoods leads to choose this approach, instead of the maximisation of a complicated joint distribution function. Thus, Murphy and Topel (1985, 2002) show that if the regression results can be obtained with MLE, then the variance-covariance matrix of the second equation must be adjusted in the following way

$$V_2^* = V_2 + V_2[CV_1C' - RV_1C' - CV_1R']V_2, \quad (3)$$

where

$$V_1 = \left[\sum_i \left(-\frac{\partial^2 \ln f_{1,i}}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} \right) \right]^{-1} = \left[\sum_i \left(\frac{\partial \ln f_{1,i}}{\partial \boldsymbol{\alpha}} \right) \left(\frac{\partial \ln f_{1,i}}{\partial \boldsymbol{\alpha}'} \right) \right]^{-1} \quad (4)$$

is the variance-covariance matrix of model 1,

$$V_2 = \left[\sum_i \left(-\frac{\partial^2 \ln f_{2,i}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) \right]^{-1} = \left[\sum_i \left(\frac{\partial \ln f_{2,i}}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \ln f_{2,i}}{\partial \boldsymbol{\beta}'} \right) \right]^{-1} \quad (5)$$

is the variance-covariance matrix of model 2 when no adjustment is made, that is, when the constructed explanatory variable is considered with given observed data instead of computed econometric results, and

$$C = \sum_i \left(-\frac{\partial^2 \ln f_{2,i}}{\partial \boldsymbol{\beta} \partial \hat{\boldsymbol{\alpha}}'} \right) = \sum_i \left(\frac{\partial \ln f_{2,i}}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \ln f_{2,i}}{\partial \hat{\boldsymbol{\alpha}}'} \right), \quad (6)$$

$$R = \sum_i \left(\frac{\partial \ln f_{2,i}}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \ln f_{1,i}}{\partial \hat{\boldsymbol{\alpha}}'} \right) \quad (7)$$

are the matrices of cross products of first derivative vectors. Since the variance of the first derivative vector of the log-likelihood is the negative of the expected second derivative matrix, the two formulas for matrices V_1 , V_2 and C provide the same result. As the covariance of the two first derivative vectors, it is very likely that the matrix R has all elements very close to zero.

2. Application: competition-fragility nexus in banking

There is some difficulty to develop the Murphy-Topel adjustment for panel data models with different samples. As we explained in Introduction section, there are a lot of studies in

banking literature that apply a two-step panel data models, including to study the competition-fragility relationship. This is why we apply our development on competition-fragility nexus in banking but it can be employed in any other two-step panel data model.

2.1 Two-step panel data model

At the first stage, one might estimate the bank competition measure. Let us take the following examples for the first equation, which are widely applied in banking empirical literature.

1st stage: Lerner index

The Lerner index is the mark-up of price over the bank's marginal cost:

$$L_{it} = \frac{p_{it} - \frac{\partial C_{it}}{\partial y_{it}}}{p_{it}}, \quad (8a)$$

where $p_{it} = TR_{it}/(y_{1,it} + y_{2,it} + y_{3,it})$ is the average price of a bank i 's output at time t , C_{it} is the bank's total cost and y_{it} is its output. Higher is the Lerner index, higher is the market power and less competitive banking institutions are.

The econometric task is thus to estimate the marginal costs of banks, $\partial C_{it}/\partial y_{it}$. As in the most literature, we take three inputs, prices of which are w_j , and three outputs y_m :

$$\begin{aligned} \ln(C_{it}/w_{3,it}) = & c_0 + \alpha_i + \mu_t + \sum_{j=1}^2 \beta_j \ln(w_{j,it}/w_{3,it}) + \sum_{m=1}^3 \gamma_m \ln y_{m,it} \\ & + \frac{1}{2} \sum_{k=1}^2 \sum_{l=1}^2 \varphi_{kl} \ln(w_{k,it}/w_{3,it}) \ln(w_{l,it}/w_{3,it}) \\ & + \sum_{m=1}^3 \sum_{k=1}^2 \chi_{mk} \ln(w_{k,it}/w_{3,it}) \ln y_{m,it} + \frac{1}{2} \sum_{m=1}^3 \sum_{s=1}^3 \rho_{ms} \ln y_{m,it} \ln y_{s,it} \\ & + \sum_{m=1}^3 \omega_{1m} \ln y_{m,it} \ln gdp_t + \sum_{m=1}^3 \omega_{2m} \ln y_{m,it} \ln grgdp_t + \varepsilon_{it}, \end{aligned} \quad (8b)$$

To control the effects of the business cycle, the real GDP, gdp_t , and/or the real GDP growth, $grgdp_t$, can be considered in levels and/or in cross-product terms. The estimated coefficients of this equation are applied to compute the marginal cost:

$$\begin{aligned} mc_{it} = \frac{\partial C_{it}}{\partial (y_{1,it} + y_{2,it} + y_{3,it})} = & \left[\sum_{m=1}^3 \gamma_m + \sum_{m=1}^3 \sum_{k=1}^2 \chi_{mk} \ln(w_{k,it}/w_{3,it}) + \sum_{m=1}^3 \sum_{s=1}^3 \rho_{ms} \ln y_s \right. \\ & \left. + \sum_{m=1}^3 \omega_{1m} \ln gdp_t + \sum_{m=1}^3 \omega_{2m} \ln grgdp_t \right] \times \frac{C_{it}}{y_{1,it} + y_{2,it} + y_{3,it}}. \end{aligned} \quad (8c)$$

1st stage: Boone indicator

The Boone indicator, proposed by Boone (2008), is the elasticity of market share to marginal cost. Based on the idea that more efficient banks, that is, with lower marginal costs, gain higher market shares, this effect will be stronger in more competitive banking market. One needs to run the following econometric model:

$$\begin{aligned} \ln ms_{it} = & \pi_0 + \pi_{1i} + \nu_t + \theta_1 \ln mc_{it} + \theta_2 \ln mc_{it} \times \ln mc_{it} + \theta_3 \ln mc_{it} \times \ln gdp_t \\ & + \theta_4 \ln mc_{it} \times \ln grgdp_t + \varepsilon_{it}, \end{aligned} \quad (9)$$

where ms_{it} stands for earning assets share for bank i in year t . As the elasticity of market share to marginal cost, the Boone indicator is calculated as following:

$$Boone_{it} = \theta_1 + 2\theta_2 \ln mc_{it} + \theta_3 \ln gdp_t + \theta_4 \ln grgdp_t. \quad (10)$$

Since *Boone* variable might take negative values, higher values (or lower absolute values) imply less competitive behaviour.

1st stage: H-statistic

The H-statistic of Panzar and Rosse (1987) is the aggregate elasticity of banks' revenue to input prices. It can therefore be computed regressing the following translog revenue function of banks:

$$\begin{aligned} \ln TR_{it} = & a_0 + \vartheta_i + v_t + \sum_{j=1}^3 \lambda_j \ln w_{j,it} + \psi_1 \ln TA_{it} + \psi_2 \ln Loan_{it} + \frac{1}{2} \sum_{k=1}^3 \sum_{l=1}^3 \kappa_{kl} \ln w_{k,it} \ln w_{l,it} \\ & + \sum_{k=1}^3 \phi_{1k} \ln w_{k,it} \ln TA_{it} + \sum_{k=1}^3 \phi_{2k} \ln w_{k,it} \ln Loan_{it} + \sum_{k=1}^3 \phi_{3k} \ln w_{k,it} \ln gdp_t \\ & + \sum_{k=1}^3 \phi_{4k} \ln w_{k,it} \ln grgdp_t + \varepsilon_{it}, \end{aligned} \quad (11)$$

where TR_{it} stands for total revenue of bank i in year t , TA_{it} for total assets and $Loan_{it}$ for ratio of loans on total assets. As for the Lerner index, we consider three inputs for bank intermediation function. Usually in banking empirical literature, the use of TA and $Loan$ variables is to take into account the differences in size and business model, respectively, akin to De Bandt and Davis (2000), Bikker and Haaf (2002), Weill (2013). And as for previous two measures, the economic state variables are considered to control the business cycles. Thus, the H-statistic is equal to

$$\begin{aligned} H_{it} = & \sum_{k=1}^3 \frac{\partial TR_{it}}{\partial w_{k,it}} \times \frac{w_{k,it}}{TR_{it}} = \sum_{k=1}^3 \lambda_j + \sum_{k=1}^3 \sum_{l=1}^3 \kappa_{kl} \ln w_{l,it} + \sum_{k=1}^3 \phi_{1k} \ln TA_{it} + \sum_{k=1}^3 \phi_{2k} \ln Loan_{it} \\ & + \sum_{k=1}^3 \phi_{3k} \ln gdp_t + \sum_{k=1}^3 \phi_{4k} \ln grgdp_t \end{aligned} \quad (12)$$

and it takes values between zero (monopoly or collusive oligopoly situation) and one (perfectly competitive behaviour).

All regressions can be made individually on each country k of the sample. However, since the risk measure cannot be computed for many banks, namely for non-listed entities if a market-based measure is employed, the second equation is often regressed on entire sample of N_c countries.

At the second stage, in order to estimate the relationship between bank competition and risk-taking, two different panel data models are considered in the banking literature:

2nd stage: linear relationship

$$Risk_{i,j,t} = \beta_0 + \gamma_1 BC_{i,j,t} + \sum_{k=1}^n \beta_k X_{k,i,j,t} + \sum_{l=n+1}^m \beta_l Y_{l,j,t} + \mu_i + \pi_t + \varepsilon_{i,j,t} \quad (13a)$$

2nd stage: non-linear relationship

$$Risk_{i,j,t} = \beta_0 + \gamma_1 BC_{i,j,t} + \gamma_2 BC_{i,j,t}^2 + \sum_{k=1}^n \beta_k X_{k,i,j,t} + \sum_{l=n+1}^m \beta_l Y_{l,j,t} + \mu_i + \pi_t + \varepsilon_{i,j,t} \quad (13b)$$

where $Risk_{i,j,t}$ is a risk measure, n is the number of bank-specific variables, $m-n$ is the number of country-specific variables, $BC_{i,j,t}$ and $X_{k,i,j,t}$ are, respectively, bank competition measure and bank-specific factor that controls for the k -th characteristic of the bank i at year t in country j , and $Y_{l,j,t}$ is the country-specific factor that controls for the l -th characteristic of the country j .

In order to determine the adjusted variance-covariance matrix V_2^* of our risk equation, the variance-covariance matrices of the first and second equations V_1 and V_2 , respectively, are extracted from the respective regressions and we must compute only the matrices C and R .

The matrix V_1 is of dimension $l_1 N_c \times l_1 N_c$ and is composed on the diagonal by variance-covariance matrices $V_{1,k}$ for each country k ($k = 1, \dots, N_c$), all other elements being zeros:

$$V_1 = \begin{pmatrix} V_{1,1} & 0 & \dots & 0 \\ 0 & V_{1,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V_{1,N_c} \end{pmatrix}.$$

The estimations of competition measures for all countries are therefore taken into consideration in the adjustment of the variance-covariance matrix V_2 , which is of dimension $l_2 \times l_2$. l_2 and l_1 are the size of vectors \mathbf{x}_2 and \mathbf{x}_1 , respectively. These vectors of explanatory variables include also the intercept and the time fixed effects.

The difficulty to compute the matrices C and R is the difference between samples of two equations. They are determined as cross “products”, for the matrix R , or cross derivations, for the matrix C , between the single sample of the second equation and the bank sample of each country. The computation of the elements $r[i, j]$ and $c[i, j]$ of matrices R and C , respectively, are explained in the next sub-section.

2.2 Murphy-Topel adjustment of the variance-covariance matrix of the risk equation

For the first panel data model (eq. 8b, for the Lerner index, eq. 9 for the Boone indicator, and eq. 11 for the H-statistic), the log-likelihood function is formulated for each country k of the sample.

$$\ln f_{1,k,i} = -\frac{1}{2} \left\{ \frac{1}{\sigma e_{1,k}^2} \left[\sum_{t=1}^{T_{1,k,i}} \varepsilon_{1,k,it}^2 - \frac{\sigma e_{1,k}^2}{\sigma e_{1,k}^2 + T_{1,k,i} \sigma u_{1,k}^2} \left(\sum_{t=1}^{T_{1,k,i}} \varepsilon_{1,k,it} \right)^2 \right] + \ln \left(1 + T_{1,k,i} \frac{\sigma u_{1,k}^2}{\sigma e_{1,k}^2} \right) \right\} - \frac{1}{2} \{ T_{1,k,i} \ln(2\pi \sigma e_{1,k}^2) \} \quad (14)$$

k ranges from 1 to N_c . $\sigma e_{1,k}^2$ and $\sigma u_{1,k}^2$ are the variances of residuals and individual effects, respectively, for each panel data regression k . $\varepsilon_{1,k,it}$ are residuals of this regression and $T_{1,k,i}$ is the number of observations for the bank i . The log-likelihood of the second model takes therefore the following form

$$\ln f_{2,i} = -\frac{1}{2} \left\{ \frac{1}{\sigma e_2^2} \left[\sum_{t=1}^{T_{2,i}} \varepsilon_{2,it}^2 - \frac{\sigma e_2^2}{\sigma e_2^2 + T_{2,i} \sigma u_2^2} \left(\sum_{t=1}^{T_{2,i}} \varepsilon_{2,it} \right)^2 \right] + \ln \left(1 + T_{2,i} \frac{\sigma u_2^2}{\sigma e_2^2} \right) \right\} - \frac{1}{2} \{ T_{2,i} \ln(2\pi \sigma e_2^2) \}, \quad (15)$$

where σe_2^2 and σu_2^2 are the variances of residuals and of individual effects, respectively. $\varepsilon_{2,it}$ are residuals of this regression and $T_{2,i}$ is the number of observations for the bank i .

To compute the elements of matrices R and C , we determine the following derivatives:

For the first equation:

$$\frac{\partial \ln f_{1,k,i}}{\partial \hat{\alpha}_k} = \frac{1}{\sigma e_{1,k}^2} \left[\sum_{t=1}^{T_{1,k,i}} \varepsilon_{1,k,it} \mathbf{x}_{1,k,it} - \frac{\sigma e_{1,k}^2}{\sigma e_{1,k}^2 + T_{1,k,i} \sigma u_{1,k}^2} \left(\sum_{t=1}^{T_{1,k,i}} \varepsilon_{1,k,it} \right) \left(\sum_{t=1}^{T_{1,k,i}} \mathbf{x}_{1,k,it} \right) \right] \quad (16)$$

For the second equation:

$$\frac{\partial \ln f_{2,i}}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma e_2^2} \left[\sum_{t=1}^{T_{2,i}} \varepsilon_{2,it} \mathbf{x}_{2,it} - \frac{\sigma e_2^2}{\sigma e_2^2 + T_{2,i} \sigma u_2^2} \left(\sum_{t=1}^{T_{2,i}} \varepsilon_{2,it} \right) \left(\sum_{t=1}^{T_{2,i}} \mathbf{x}_{2,it} \right) \right] \quad (17)$$

Consequently, the elements of the matrix R for each country k are

$$r0[k, l_2, l_1] = \sum_{i=1}^{N_2} \left\{ \frac{1}{\sigma e_2^2} \left[\sum_{t=1}^{T_{2,i}} \varepsilon_{2,k,it} \mathbf{x}[l_2]_{2,k,it} - \frac{\sigma e_2^2}{\sigma e_2^2 + T_{2,i} \sigma u_2^2} \left(\sum_{t=1}^{T_{2,i}} \varepsilon_{2,k,it} \right) \left(\sum_{t=1}^{T_{2,i}} \mathbf{x}[l_2]_{2,k,it} \right) \right] \right. \\ \left. \times \frac{1}{\sigma e_{1,k}^2} \left[\sum_{t=1}^{T_{1,k,i}} \varepsilon_{1,k,it} \mathbf{x}[l_1]_{1,k,it} - \frac{\sigma e_{1,k}^2}{\sigma e_{1,k}^2 + T_{1,k,i} \sigma u_{1,k}^2} \left(\sum_{t=1}^{T_{1,k,i}} \varepsilon_{1,k,it} \right) \left(\sum_{t=1}^{T_{1,k,i}} \mathbf{x}[l_1]_{1,k,it} \right) \right] \right\}, \quad (18)$$

where N_2 is the number of banks of the second model. Finally, the matrix R has elements $r[i, j] = r0[k, l_2, l_1]$, where $i = l_2$ and $j = k \times l_1$.

Besides the matrix R , the results of the first equation are also considered in the second one with the matrix C . Since we have two versions of the risk equation, this second derivative considers only BC variable for the linear equation (eq. 13a) and both BC and BC^2 variables for the non-linear equation (eq. 13b) as follows:

For a liner model:

- The derivative of the BC variable with respect to parameters of the first model

$$c0[k, 1, l_1] = - \sum_{i=1}^{N_2} \left\{ \frac{1}{\sigma e_2^2} \left[\sum_{t=1}^{T_{2,i}} (\varepsilon_{2,k,it} - \gamma_1 BC_{k,it}) \mathbf{x}'[l_1]_{2,k,it} \right. \right. \\ \left. \left. - \frac{\sigma e_2^2}{\sigma e_2^2 + T_{2,i} \sigma u_2^2} \left(\sum_{t=1}^{T_{2,i}} \varepsilon_{2,k,it} - \gamma_1 \sum_{t=1}^{T_{2,i}} BC_{k,it} \right) \left(\sum_{t=1}^{T_{2,i}} \mathbf{x}'[l_1]_{2,k,it} \right) \right] \right\}, \quad (19a)$$

- The derivative of other variables of the second equation with respect to parameters of the first model

$$c0[k, l_2, l_1] = - \sum_{i=1}^{N_2} \left\{ \frac{1}{\sigma e_2^2} \left[\sum_{t=1}^{T_{2,i}} -\gamma_1 \mathbf{x}'[l_1]_{2,k,it} \mathbf{x}[l_2]_{2,k,it} \right. \right. \\ \left. \left. - \frac{\sigma e_2^2}{\sigma e_2^2 + T_{2,i} \sigma u_2^2} \left(\sum_{t=1}^{T_{2,i}} -\gamma_1 \mathbf{x}'[l_1]_{2,k,it} \right) \left(\sum_{t=1}^{T_{2,i}} \mathbf{x}[l_2]_{2,k,it} \right) \right] \right\}, \quad (19b)$$

For a non-linear model:

- The derivative of the BC variable with respect to parameters of the first model

$$c0[k, 1, l_1] = - \sum_{i=1}^{N_2} \left\{ \frac{1}{\sigma e_2^2} \left[\sum_{t=1}^{T_{2,i}} (\varepsilon_{2,k,it} + (-\gamma_1 - 2\gamma_2 BC_{k,it}) BC_{k,it}) \mathbf{x}'[l_1]_{2,k,it} \right. \right. \\ \left. \left. - \frac{\sigma e_2^2}{\sigma e_2^2 + T_{2,i} \sigma u_2^2} \left(\left(\sum_{t=1}^{T_{2,i}} (-\gamma_1 - 2\gamma_2 BC_{k,it}) \mathbf{x}'[l_1]_{2,k,it} \right) \left(\sum_{t=1}^{T_{2,i}} BC_{k,it} \right) \right. \right. \right. \\ \left. \left. \left. + \left(\sum_{t=1}^{T_{2,i}} \varepsilon_{2,k,it} \right) \left(\sum_{t=1}^{T_{2,i}} \mathbf{x}'[l_1]_{2,k,it} \right) \right) \right] \right\}, \quad (20a)$$

- The derivative of the BC^2 variable with respect to parameters of the first model

$$\begin{aligned}
c0[k, 2, l_1] = & - \sum_{i=1}^{N_2} \left\{ \frac{1}{\sigma e_2^2} \left[\sum_{t=1}^{T_{2,i}} (2BC_{k,it} \varepsilon_{2,k,it} + (-\gamma_1 - 2\gamma_2 BC_{k,it}) BC_{k,it}^2) \mathbf{x}'[l_1]_{2,k,it} \right. \right. \\
& - \frac{\sigma e_2^2}{\sigma e_2^2 + T_{2,i} \sigma u_2^2} \left(\left(\sum_{t=1}^{T_{2,i}} (-\gamma_1 - 2\gamma_2 BC_{k,it}) \mathbf{x}'[l_1]_{2,k,it} \right) \left(\sum_{t=1}^{T_{2,i}} BC_{k,it}^2 \right) \right. \\
& \left. \left. + \left(\sum_{t=1}^{T_{2,i}} \varepsilon_{2,k,it} \right) \left(\sum_{t=1}^{T_{2,i}} 2BC_{k,it} \mathbf{x}'[l_1]_{2,k,it} \right) \right) \right\}, \quad (20b)
\end{aligned}$$

and the derivative of other variables of the second equation with respect to parameters of the first model is the same equation (19b). The matrix C has the elements $c[i, j] = c0[k, l_2, l_1]$, where $i = l_2$ and $j = k \times l_1$, in both cases.

The difficulty to construct the matrices R and C is the difference in sample size between the two equations: the first equation is regressed for each country k and the second equation for entire sample of countries. To overcome it, we transformed the explanatory variables, $\mathbf{x}[l_2]_{2,k,it}$, and the error term, $\varepsilon_{2,k,it}$, of the second equation as that each country of the sample has its own variables. Each k variable (or error term) takes therefore its own values if they correspond to country k and zero otherwise.

The vector $\mathbf{x}'[l_1]_2$ is the derivative of the bank competition measure BC , estimated in the first step, with respect to parameters l_1 of the first equation. The index 2 signifies that this vector is obtained from the derivative of the log-likelihood of the second equation. The derivation is firstly made with respect to parameters of explanatory variables and then with respect to parameters of fixed time effects and intercept. $\mathbf{x}'[l_1]_2$ vector is therefore specific for each of our three non-structural competition measures:

- For the Lerner index, based on equations (8a)-(8c), and the definition of the average price of bank output

$$\begin{aligned}
\mathbf{x}'[l_1]_{2,k,it} = & \left[0, 0, -C_{k,it}/TR_{k,it}, -C_{k,it}/TR_{k,it}, -C_{k,it}/TR_{k,it}, 0, 0, -C_{k,it}/TR_{k,it} \right. \\
& \times \ln(w_{1,k,it}/w_{3,k,it}), -C_{k,it}/TR_{k,it} \times \ln(w_{1,k,it}/w_{3,k,it}), -C_{k,it}/TR_{k,it} \\
& \times \ln(w_{1,k,it}/w_{3,k,it}), 0, -C_{k,it}/TR_{k,it} \times \ln(w_{2,k,it}/w_{3,k,it}), -C_{k,it}/TR_{k,it} \\
& \times \ln(w_{2,k,it}/w_{3,k,it}), -C_{k,it}/TR_{k,it} \times \ln(w_{2,k,it}/w_{3,k,it}), -C_{k,it}/TR_{k,it} \\
& \times (\ln y_{1,k,it} + \ln y_{1,k,it}), -C_{k,it}/TR_{k,it} \times (\ln y_{1,k,it} + \ln y_{2,k,it}), -C_{k,it}/TR_{k,it} \\
& \times (\ln y_{1,k,it} + \ln y_{3,k,it}), -C_{k,it}/TR_{k,it} \times (\ln y_{2,k,it} + \ln y_{2,k,it}), -C_{k,it}/TR_{k,it} \\
& \times (\ln y_{2,k,it} + \ln y_{3,k,it}), -C_{k,it}/TR_{k,it} \times (\ln y_{3,k,it} + \ln y_{3,k,it}), -C_{k,it}/TR_{k,it} \\
& \times \ln gdp_{k,it}, -C_{k,it}/TR_{k,it} \times \ln gdp_{k,it}, -C_{k,it}/TR_{k,it} \\
& \times \ln gdp_{k,it}, -C_{k,it}/TR_{k,it} \times gdpgr_{k,it}, -C_{k,it}/TR_{k,it} \\
& \left. \times gdpgr_{k,it}, -C_{k,it}/TR_{k,it} \times gdpgr_{k,it}, \underbrace{0, \dots, 0}_{te}, 0 \right]
\end{aligned}$$

The size of this vector is the same as that of explanatory variables of equation (8b), that is, $27 + te$, and includes 26 explanatory variables, te time fixed effects and 1 intercept.

- For the Boone indicator, based on equations (9) and (10)

$$\mathbf{x}'[l_1]_{2,k,it} = \left[1, 2 \ln mc_{k,it}, \ln gdp_{k,it}, gdpgr_{k,it}, \underbrace{0, \dots, 0}_{te}, 0 \right]$$

The size of this vector is the same as that of explanatory variables of equation (9), that is, $5 + te$, and includes 4 explanatory variables, te time fixed effects and 1 intercept.

- For the H-statistic, based on equations (11) and (12)

$$\mathbf{x}'[l_1]_{2,k,it} = \left[1, 1, 1, 0, 0, \ln w_{1,k,it} + \ln w_{1,k,it}, \ln w_{1,k,it} + \ln w_{2,k,it}, \ln w_{1,k,it} + \ln w_{3,k,it}, \ln w_{2,k,it} + \ln w_{2,k,it}, \ln w_{2,k,it} + \ln w_{3,k,it}, \ln w_{3,k,it} + \ln w_{3,k,it}, \ln TA_{k,it}, \ln TA_{k,it}, \ln TA_{k,it}, Loan_{k,it}, Loan_{k,it}, Loan_{k,it}, \ln gdp_{k,it}, \ln gdp_{k,it}, \ln gdp_{k,it}, gdpgr_{k,it}, gdpgr_{k,it}, gdpgr_{k,it}, \underbrace{0, \dots, 0}_{te}, 0 \right]$$

The size of this vector is the same as that of explanatory variables of equation (11), that is, 24+ te , and includes 23 explanatory variables, te time fixed effects and 1 intercept.

3. Conclusion

In this paper, we developed the Murphy-Topel adjustment of the variance-covariance matrix of a two-step panel data model and for different samples of two equations. Since this type of model is often employed in banking literature, especially in the study of the competition-stability relationship, we developed the Murphy-Topel adjustment to examine this subject. However, it can be applied for any economic empirical study that use a two-step panel data model. Because of possible correlations between the results of both models, the Murphy-Topel method eventually seems to be more appropriate than any statistical (numerical adjustment) as the bootstrapping procedure.

References

- Bikker, J. A. and K. Haaf (2002) "Competition, concentration and their relationship: an empirical analysis of the banking industry", *Journal of Banking and Finance* 26, 2191-2214.
- Boone, J. (2008) "A new way to measure competition", *The Economic Journal* 118, 1245-1261.
- Buch, C. M., Koch, C. T. and M. Koetter (2013) "Do Banks Benefit from Internationalization? Revisiting the Market Power–Risk Nexus", *Review of Finance* 17, 1401–1435.
- De Bandt, O. and E. Davis (2000) "Competition, contestability and market structure in European banking sector on the eve of EMU", *Journal of Banking and Finance* 24, 1045-1066.
- Greene, W. H. (2012) "Econometric Analysis". Seventh edition. International edition. Pearson.
- Leon, F. (2015) "Does bank competition alleviate credit constraints in developing countries?", *Journal of Banking and Finance* 57, 130-142.
- Murphy, K. and R. Topel (1985) "Estimation and Inference in Two Step Econometric Models", *Journal of Business and Economic Statistics* 3, 370–379.
- Murphy, K. and R. Topel (2002) "Estimation and Inference in Two Step Econometric Models", *Journal of Business and Economic Statistics* 20, 88–97.
- Panzar, J. and J. Rosse, (1987) "Testing for "monopoly" equilibrium", *Journal of Industrial Economics* 35, 443-456.
- Schaeck, K., Cihak, M. and S. Wolfe (2009) "Are competitive banking systems more stable?", *Journal of Money, Credit and Banking* 41, 711–734.
- Tabak B.M., Fazio D.M. and D.O. Cajueiro (2012) "The relationship between banking market competition and risk-taking: Do size and capitalization matter?", *Journal of Banking and Finance* 36, 3366-3381.
- Van Leuvensteijn M., Sørensen C.K., Bikker J.A. and A.A.R.J.M. van Rixtel (2013) "Impact of bank competition on the interest rate pass-through in the euro area", *Applied Economics* 45, 1359-1380.
- Weill, L. (2013) "Bank competition in the EU: How has it evolved?", *Journal of International Financial Markets, Institutions and Money* 26, 100-112.