Testing, not modelling, the impact of Cohesion support: a theoretical framework and some preliminary results for the Spanish regions

Angel de la Fuente*
Instituto de Análisis Económico (CSIC)
November 2009

1. Introduction

Many attempts to evaluate the macroeconomic effects of Cohesion policies proceed by inserting into an already existing model some measure of EU support. Impact estimates are typically obtained as the difference between the time paths of the variables of interest (normally output and employment) in a baseline simulation and those corresponding to a counterfactual scenario in which EU subsidies for investment and/or job creation have been removed. Even if we assume that we are working with the correct model, the procedure makes a very strong implicit assumption, namely, that assisted investment projects and other EU-supported measures are "just as good" as other projects and measures of a similar nature. Since this is precisely what many critics of Cohesion policies question, all such exercises can be dismissed as irrelevant, or at least as unconvincing, until the critical underlying assumption can be explicitly tested.

To provide a convincing test of the effectiveness of Cohesion policies a basic falsifiability requirement must be met: data on actual support levels has to be confronted with output indicators in the context of a statistical model which could conceivably lead us to conclude that EU aid has had no effect on output and employment --something which is just not possible within the framework sketched above. In the current paper, I develop a methodology that meets this requirement and apply it, using Spanish regional data, to estimate the "relative effectiveness" of infrastructure projects co-financed by the Structural and Cohesion Funds. The basic idea is to keep track of the accumulated stock of EU-assisted capital and to use a specification of the regional production function that allows this component of the capital stock to be more or less productive than that arising from unassisted projects.

The remainder of the paper is organized as follows. Section 2 sets out the proposed framework in its simplest possible form and briefly discusses the sort of data that would be required to implement it. In Section 3 an extended version of the model is applied to the case of infrastructure investment in the Spanish regions. Section 4 concludes with some reflections on the implications of the results and on the need to collect appropriate data if we really want to estimate the impact of Cohesion policies.

* This paper has been prepared for the Sixth European Conference on the Evaluation of the Structural Funds (Warsaw, November 2009). It draws on preliminary results from a larger project financed by the ERDF and Fundación Caixa Galicia. I would like to thank José Luis Kaiser, David Azcárate and Ángeles Holgado (from the Directorate General for European Funds of the Spanish Ministry of Finance) for their invaluable help in putting together the data on Structural and Cohesion Fund expenditure that are used here.

Address for correspondence: Angel.delaFuente@uab.es
2. A falsifiable test of the impact of cohesion expenditure

Infrastructure and other investment projects that benefit from EU cohesion support are not qualitatively different from many other projects that do not. They may, however, be less efficient if selection criteria or cost control procedures are relaxed when a substantial part of the cost is born by an outside party. On the other hand, it is also possible that EU-assisted projects may be of a higher average quality than other projects, either because Commission involvement helps improve selection and cost-control procedures or because national Governments select for cohesion support “good” projects that are more likely to pass EU filters.

A simple way to formalize these alternative hypotheses and put them into a testable form is to assume that a euro of investment may generate a different amount of effective capital (\(K\)) depending on how it is financed. Letting \(K^a\) and \(K^n\) denote the stocks of EU-assisted and fully nationally financed capital, I will assume that the effective stock of capital is given by the following expression

\[
(1) \quad K^e_t = K^n_t + (1 + \phi)K^a_t = \frac{K^n_t + (1 + \phi)K^a_t}{K^a_t} K_u = \left[\frac{(1 - \omega_u) + (1 + \phi)\omega_u}{K^a_t}\right] K_u = (1 + \phi \omega_u) K_u
\]

where

\[
K_u = K^n_t + K^a_t
\]

is the total stock of capital in region \(i\) at time \(t\) calculated in the usual way (i.e. without taking into account potential differences in quality across projects) and \(\omega_u\) denotes the weight of EU-assisted capital in this aggregate. The parameter \(\phi\) measures the relative effectiveness of capital resulting from projects that have received cohesion support. If \(\phi = 0\), both types of capital are equally productive and it makes no difference whether investment has been partly financed by EU grants or not. If \(\phi < 0\), on the other hand, each euro of EU-assisted investment adds less than one euro to the effective stock of capital, indicating that subsidized projects are, on average, less productive than the rest. The reverse will be true, finally, whenever \(\phi > 0\).

To complete the specification and see how it may be taken to the data, let us assume for now that there is a single type of capital and that the aggregate production function is of the Cobb-Douglas type,

\[
(2) \quad Y_i = A_i \left( (1 + \phi \omega_u) K_u \right)^{\theta_l} L_i^{\theta_l}
\]

where \(Y\) is output, \(L\) employment, and \(A\) the level of technical efficiency or total factor productivity (TFP). The parameters \(\theta_i\) (with \(i = l, k\)) measure the elasticity of aggregate output with respect to the stocks of the different productive factors. An increase of 1% in the stock of capital, for instance, will increase output by \(\theta_k\)% holding constant employment and the level of technical efficiency.
Taking logs of (2) and noting that

\[ \ln(1 + \phi_{it}) = \phi_{it} \]

provided that \( \phi_{it} \) is small, we obtain a transformation of (2) that is approximately linear in the parameters:

\[
\ln Y_{it} = \ln A_{it} + \theta_i \ln(1 + \phi_{it}) + \theta_i \ln K_{it} + \theta_i \ln L_{it}
\]

Taking first differences of (4) we finally arrive at the following expression

\[
\Delta \ln Y_{it} = \Delta \ln A_{it} + \phi_{it} \Delta \omega_{it} + \theta_i \Delta \ln K_{it} + \theta_i \Delta \ln L_{it}
\]

where the symbol “\( \Delta \)” denotes time differences (i.e. for any variable \( X \), \( \Delta X \) is defined as \( \Delta X_{it} = X_{it+1} - X_{it} \)).

Given data on output, employment and the stocks of assisted and non-assisted capital, equation (5) can be used to estimate the elasticity of output with respect to the different productive factors and the relative effectiveness of EU-assisted investment projects. I will be primarily concerned with the sign of the coefficient of the increase in the share of assisted capital, \( \Delta \omega_{it} \), because this is the parameter that can be used to test the hypothesis set out at the beginning of this section: if, controlling for other things, output growth is a decreasing function of \( \Delta \omega_{it} \), we can conclude that EU-subsidized capital is less productive than non-assisted capital. The size of the parameter will also be of considerable interest for it will provide an estimate of the size of the inefficiencies arising from the availability of cohesion support and will have a direct impact on the return to EU-assisted investment.

The main difficulty one faces when trying to estimate a model of the type I have just sketched is the lack of appropriate data on stocks and flows of EU-assisted capital. In principle, we would need annual series of assisted investment, broken down by function, that adequately capture the moment in which EU grants translate into productive capital. Somewhat surprisingly, data of this sort are practically inexistent. While both the European Commission and the governments of member states publish an enormous amount of material on the subject, the available information on EU grants is disperse, unsystematic, heterogeneous, difficult to interpret and, in most cases, almost useless as an input for statistical analysis. In many cases, the available information corresponds to ex-ante budget appropriations rather than to actual executed expenditure. When data on actual expenditure are available, they are typically not broken down by year. And when such a breakdown does exist, it often reflects payment flows that do not correspond to the execution of the underlying projects. The functional breakdown of expenditure is also far from ideal. The axes, measures and fields of intervention that are used by national governments and by the Commission to classify expenditure change over time, often combine actions of different economic nature and are not always precisely defined.
3. Some preliminary results for the Spanish regions

Drawing on ongoing and still incomplete work, in this section I will present some preliminary results on the effectiveness of EU-assisted infrastructure investment in the Spanish regions. I will start by extending the model developed in the previous section in order to distinguish between infrastructure and non-infrastructure capital and to introduce average schooling as an additional determinant of labor productivity. I will then briefly discuss the data I have used to estimate the model and present the results.

Model specification

I will assume that regional output is given by a Cobb-Douglas function with constant returns to scale in labor, the effective stock of infrastructure and the stock of other capital for a given level of educational attainment,

\[ Y_{it} = A_{it} K_{it}^\theta (P_{it})^\phi L_{it}^\eta H_{it}^{\theta_p} \]

where the constant returns assumption implies that

\[ \theta_i = 1 - \theta_k - \theta_p \]

The function is similar to the one I have used in the previous section but includes two new inputs: average years of schooling \((H)\) and the effective stock of infrastructure, \(P_e\). Proceeding as in the previous section, I will write \(P_e\) as a function of the total stock of infrastructures, \((P)\), and the weight of EU-assisted infrastructure capital in this aggregate \((\omega)\):

\[ P_{it}^e = P_{it}^n + (1 + \phi) P_{it}^a = (1 + \phi \omega_{it}) P_{it}^a \]

Using the constant-returns assumption given in (7), we can define a per capita production function that will relate average labor productivity to average schooling and to the stocks of infrastructures and other capital per worker. Letting \(Q = Y/L\) denote output per worker and \(X = P/L\) and \(Z = K/L\) the per worker stocks of infrastructures and other capital, and dividing both sides of (6) by employment, \(L\), we have:

\[ Q_{it} = A_{it} Z_{it}^\eta (1 + \phi \omega_{it}) X_{it}^\phi H_{it}^{\theta_p} \]

Proceeding as in the previous section (i.e. taking logs and then first differences of (9) and approximating \(\ln (1 + \phi \omega_{it})\)), it is easy to derive the following expression,

\[ \Delta q_{it} = \Delta a_{it} + \theta_s \Delta a_{it} + \theta_p \Delta X_{it} + \phi \theta_p \Delta \omega_{it} + \theta_h \Delta h_{it} \]

where I have used lower-case letters to indicate that the variables are measured in logarithms.

To estimate equation (10), I will further assume that the rate of technical progress is given by

\[ \Delta a_{it} = \eta_t + \lambda h_{it} \]
where \( \eta_t \) is a fixed period effect that varies over time but not across regions and should help control for technical progress and cyclical factors and \( h_{it} \) is a technological gap measure. This last variable enters the equation as a determinant of the rate of technical progress in order to allow for a catch-up effect whose intensity is measured by the parameter \( \lambda \). The technological gap will be calculated as the log difference in TFP between Madrid \((M)\) and each other region at the beginning of each of the biennial subperiods into which I have divided the sample period. Inverting the per capita production function given in (9), this quantity can be written in the form

\[
(12) \quad h_{it} = a_{Mt} - a_t = (q_{Mt} - \theta_p k_{Mt} - \theta_p x_{Mt} - \phi \omega_{Mt} - \theta_p h_{Mt}) - (q_{at} - \theta_p k_{at} - \theta_p x_{at} - \phi \omega_{at} - \theta_p h_{at})
\]

To estimate the model, I substitute (11) and (12) into (10) and use non-linear least squares with data on both factor stocks and their growth rates to estimate the parameters of interest.

**Data**

The model described in the previous section is estimated using a panel of Spanish regional data that covers the period 1965-2005 at biennial intervals. Data on regional gross value added (GVA) at constant prices and employment (number of jobs) are taken from de la Fuente (2009a) and are constructed (essentially) by linking the historical series constructed by Fundación BBV (1999) with those available in the Regional Accounts of the National Statistical Institute from 1995 onward (Contabilidad Regional de España, base 2000, in INE, 2009). Estimates of average years of schooling are taken from de la Fuente and Doménech (2006a), who estimate the average educational attainment of the adult population using data from the census and municipal registers.

I have constructed series of total stocks of productive infrastructures \((P)\) and other non-residential capital \((K)\) applying a perpetual inventory procedure with geometric depreciation to the data on investment flows, measured at constant prices and disaggregated by function, provided by Mas et al (2007) and Fundación BBVA (2009). (See the Appendix for further details). The stock of productive infrastructure includes publicly financed transportation networks (roads and highways, ports, airports, subways and railways), water works, sewage, urban structures and privately-financed toll highways. The stock of non-infrastructure capital includes private capital, net of the stock of residential housing, and the stock of public capital associated with the provision of education, health and general administrative services. These last three items are aggregated with the capital stock of the private sector because my output measure includes government-provided services. For shortness, I may at times speak of private and public capital to refer to the infrastructure and non-infrastructure components of the stock of physical capital but it should be kept in mind that this is not entirely accurate.

Data on EU-assisted investment in productive infrastructure have been provided by the Directorate General for European Funds of the Spanish Ministry of Finance (MEH) and, for the time being, refer only to the 2000-06 planning period.\(^1\) During this period, assisted investment

\(^1\) See de la Fuente (2009b) for further details on these data.
in productive infrastructures amounted to 35 billion euros (including national confinancing), absorbing 44% of total EU-assisted expenditure. Three fourths of assisted infrastructure investment were co-financed by the Structural Funds (mostly by the ERDF) and the remaining 25% by the Cohesion Fund. The data correspond to what the Spanish government calls expenditure certifications and should be considerably closer to the execution of actual investment projects on the ground than the more standard data on EU commitments or reimbursement that are usually available.

Deflated flows of assisted expenditure are accumulated using the same perpetual inventory procedure described above to construct stocks of assisted infrastructure capital and to recover the corresponding values of $\omega_s$. Since the available data on assisted expenditure start only in 2000, the stock of assisted capital will be set to zero in earlier years. This poses some obvious problems. Since we have only a limited number of observations for $\omega_s$ and the variation of this variable will be relatively small, it may be difficult to estimate its coefficient with precision. In addition, the hypothesis we will be able to test in practice is somewhat less ambitious than the desired one. Given the available data, the only question we may hope to answer is whether or not assisted investment corresponding to 2000-06 is as productive as all other investment, including projects that have received EU support under earlier financial frameworks. Both of these factors can be expected to make it more difficult to reject the null hypothesis that assisted investment is just as effective as non-assisted investment even when this is not the case. As a result, the reports resulted below should be seen as rather preliminary. I hope to be able to extend the series on assisted expenditure back at least to 1994, but progress on this front has been rather slow so far.

**Results**

Table 1 shows the results of the estimation of several variants of the model developed above. Equations [1a] and [1b] correspond to a standard Cobb-Douglas specification without the technological catch-up term. This term is added in the remaining specifications. In equations [2a] and [2b] the TFP gap with Madrid is calculated using output data measured at constant prices of 2000, while in [3a] and [3b] output is valued at average prices, calculated over the entire sample period.\(^2\) For each specification, the model is estimated twice: with and without controlling for the share of assisted capital in the total capital stock.

On the whole, the estimated parameter values seem reasonable and are in line with those of previous studies of the determinants of regional output in Spain. All productive inputs enter the equation with positive and statistically significant coefficients. The coefficient of schooling ($h$) is rather high but is consistent with the findings of de la Fuente and Doménech (2006b) for a sample of OECD countries after correcting for measurement error (which is likely to be a smaller problem with regional data). The infrastructures parameter ($\theta_p$) is somewhat smaller

\(^2\) Measures of the relative productivity of some Spanish regions are quite sensitive to the choice of base year for the price deflator due to the interaction between patterns of specialization and the evolution of relative sectoral prices. To try to mitigate this problem, I have constructed series of GVA valued at average prices computed over the entire sample period. See de la Fuente (2009a) for further details.
than those obtained in some previous studies that used Spanish regional data but remains highly significant and implies rather respectable rates of return on infrastructure investment. The coefficient on private capital \(k\) is a bit on the low side, but the sum of the coefficients on private and public physical capital is not far from the average share of capital in national income, which is precisely what we would expect under perfect competition and suggests there are no significant biases. Finally, the catch-up term is highly significant, indicating that technological diffusion across regions contributes to the convergence of the more backward territories.

Table 1: Estimation results

<table>
<thead>
<tr>
<th></th>
<th>[1a]</th>
<th>[1b]</th>
<th>[2a]</th>
<th>[2b]</th>
<th>[3a]</th>
<th>[3b]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta_p)</td>
<td>0.0793</td>
<td>0.0794</td>
<td>0.0764</td>
<td>0.0765</td>
<td>0.0868</td>
<td>0.0868</td>
</tr>
<tr>
<td></td>
<td>(4.06)</td>
<td>(4.06)</td>
<td>(4.26)</td>
<td>(4.26)</td>
<td>(4.89)</td>
<td>(4.87)</td>
</tr>
<tr>
<td>(\theta_L)</td>
<td>0.2513</td>
<td>0.2519</td>
<td>0.2201</td>
<td>0.2199</td>
<td>0.2249</td>
<td>0.2250</td>
</tr>
<tr>
<td></td>
<td>(6.52)</td>
<td>(6.51)</td>
<td>(5.73)</td>
<td>(5.70)</td>
<td>(5.83)</td>
<td>(5.81)</td>
</tr>
<tr>
<td>(\theta_k)</td>
<td>0.9200</td>
<td>0.8871</td>
<td>0.8669</td>
<td>0.8715</td>
<td>0.8614</td>
<td>0.8604</td>
</tr>
<tr>
<td></td>
<td>(3.70)</td>
<td>(3.25)</td>
<td>(4.02)</td>
<td>(3.94)</td>
<td>(4.40)</td>
<td>(4.33)</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>0.0281</td>
<td>0.0281</td>
<td>0.0281</td>
<td>0.0281</td>
<td>0.0341</td>
<td>0.0341</td>
</tr>
<tr>
<td></td>
<td>(7.32)</td>
<td>(7.25)</td>
<td></td>
<td></td>
<td>(6.87)</td>
<td>(6.81)</td>
</tr>
<tr>
<td>(\phi)</td>
<td>0.940</td>
<td>-0.299</td>
<td>-0.299</td>
<td>-0.299</td>
<td>0.093</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.911</td>
<td>0.911</td>
<td>0.926</td>
<td>0.926</td>
<td>0.925</td>
<td>0.925</td>
</tr>
</tbody>
</table>

- Note: all equations contain a full set of period fixed effects. \(t\) values are shown in parentheses below coefficient estimates.

Having checked that the estimation of the model yields reasonable results, the most interesting question has to do with the value of \(\phi\), the parameter that measures the relative effectiveness of assisted investment. The good news is that the null hypothesis that \(\phi=0\) cannot be rejected in any of the specifications, as indicated by the extremely low values of the corresponding \(t\) ratios shown below the estimated coefficients. That is, the data are consistent with the hypothesis that assisted investment projects are, on average, just as good as non-assisted projects. The bad news is that, as expected, estimates of \(\phi\) are rather imprecise (contain rather large standard errors). This increases the probability that we may be unable to reject the hypothesis that \(\phi=0\) even when we should.


4 For some estimates of rates of return based on a similar model, see de la Fuente and Doménech (2006a).

5 For those readers who may be unfamiliar with the technicalities of regression analysis, \(t\) statistics for the estimates of \(\phi\) given in columns [2b] and [3b] of Table 1 imply that the hypothesis that \(\phi = 0\) can only be rejected with a confidence level of 8.6% and 3.1% respectively. As a reference, the level of confidence generally required to conclude that a hypothesis has been conclusively rejected is 95%, which would require a \(t\) ratio of around 2.
If we focus on the point estimates of the parameter of interest, the results show a considerable dispersion. In the standard Cobb-Douglas specification, the estimated value of $\phi$ is positive and rather large, indicating that assisted projects are roughly twice as good as non-assisted projects - a result which seems rather implausible. A likely explanation for this finding is an omitted variable bias. The significance of the catch-up term in specifications [2] and [3] tells us that poorer regions tend to grow somewhat faster than one would expect on the basis of their rates of factor accumulation alone. Since the weight of assisted capital grows faster in poorer regions, $\Delta \omega$, may be picking up the effects of the omitted gap variable. When the gap is introduced into the equation, the estimated value of $\phi$ becomes considerable smaller. On the other hand, its sign depends on the details of the construction of the gap variable and the coefficient remains statistically indistinguishable from zero in both specifications.

4. Conclusion

This paper develops a simple model that can be used to estimate the effectiveness of Cohesion expenditure relative to similar but unsubsidized projects, thereby making it possible to explicitly test an important assumption that is often implicit in estimates of the impact of Cohesion policies. The model is then applied to the case of infrastructure investment using regional data from Spain.

The results should be considered rather tentative due mainly to the limitations of the available data but they are also encouraging. On a first look, the data suggest that the availability of Cohesion support has not led to the funding of low-quality projects or translated into increases in unit costs that would drive the returns to such projects significantly below those that have not received such support. The estimated elasticity of output with respect to the stock of infrastructures, moreover, is highly significant and large enough to imply rather respectable social returns to infrastructure investment in Spain.

To conclude, I would like to emphasize the importance of collecting the right kind of data for evaluation purposes. If we do not know at what point in time EU grants translate into productive capital on the ground or into other measures that affect the behavior of economic agents, I do not see how we can hope to estimate the impact of Cohesion policies on economic aggregates or to identify effective and ineffective measures. If the Commission and the member states of the EU really want to make progress in this area, they need to build appropriate data collection procedures into program design and into the management and reporting process for the Structural and Cohesion Funds. Worrying a bit more about getting good quality data for evaluation studies and a bit less about already cumbersome accounting and administrative controls would, in my view, be an excellent investment. It would help us get better answers to the question of whether or not EU resources are being used productively and it would make it possible to direct the available resources into the most effective programs.
Appendix

As noted in the text, a perpetual inventory method with geometric depreciation has been used to construct series of total stocks of private and public capital and the stock of EU-assisted infrastructures. Annual series on total regional investment, disaggregated by types of assets and measured at constant 2000 prices, have been taken from Mas et al (2007) for all years between 1965 and 2004 and from Fundación BBVA (2009), which is an update of the previous work, for the last years in the sample. These sources also provide series of net stocks of capital and other related aggregates. I have not used these data directly because the authors do not provide sufficient details to allow me to replicate exactly their procedure using the data on EU-assisted investment and because the depreciation rates implicit in some of these series seem suspiciously low. I have, however, taken Mas et al’s estimates of 1965 stocks as the starting point for the construction of my series. I have also used their price index for infrastructure investment to deflate the series of EU-assisted investment. The depreciation rates I have used, shown in the last column of Table A.1, are taken mostly from Timmer et al (2003 and 2007) and are similar to the BEA rates given in Fraumeni (1997).

<table>
<thead>
<tr>
<th>Table A.1: Depreciation rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Residential structures</td>
</tr>
<tr>
<td>Non-residential structures</td>
</tr>
<tr>
<td>Infrastructures</td>
</tr>
<tr>
<td>Transport equipment</td>
</tr>
<tr>
<td>Computing equipment</td>
</tr>
<tr>
<td>Communications equipment</td>
</tr>
<tr>
<td>Other machinery and equipment</td>
</tr>
<tr>
<td>Agricult. and forestry products</td>
</tr>
<tr>
<td>Other products</td>
</tr>
<tr>
<td>Software</td>
</tr>
<tr>
<td>Other intangibles</td>
</tr>
</tbody>
</table>

The perpetual inventory method has been applied in a slightly different way to different types of investment. In the case of infrastructures, the capital stock at the end of period $t$, $PE_t$, will be given by

$$(A.1) PE_t = (1 - \delta_p)PE_{t-1} + IP_t$$

where $\delta_p$ is the depreciation rate and $IP_t$ the flow of infrastructure investment during period $t$. Hence, I am assuming that investment in long-lived structures becomes operative and begins to depreciate only at the end of the period during which it is executed. In coherence with this
assumption, the stock variable that has been used in the regressions described in the text (what I will call the *usable stock of public capital,* $PU$) will be the beginning of period stock:

$$PU_t = PE_{t-1}$$

In the case of “private” capital, however, I have assumed that investment becomes operative at half-year and begins to depreciate at that time. Hence, end of period stock for each asset of this type will be given by

$$KE_t = (1 - \delta_t)KE_{t-1} + (1 - 0.5 * \delta_t)IK_t$$

and the corresponding usable stock will be

$$KU_t = KE_{t-1} + 0.5 * IK_t$$
References


