TÍTULO DE LA COMUNICACIÓN:  
Market potential and firm-level productivity in Spain

AUTOR 1:  
ADELHEID HOLL
Email:  
a.holl@cchs.csic.es

DEPARTAMENTO:  
IPP

UNIVERSIDAD:  
CSIC

ÁREA TEMÁTICA:  
Transporte, movilidad e infraestructuras

RESUMEN:  The literature has documented a large degree of heterogeneity across firms in terms of productivity. In this paper I focus on market potential as a source of differences in productivity across Spanish manufacturing firms. Market potential is conditioned by the existing transport infrastructure. Transport infrastructure investment improves accessibility to input and output markets and thus increases market potential. Market potential is measured by travel time through the real transport network and takes into account the immense improvements that have taken place in Spain over the last decades. The results from the preferred system GMM approach show a significant positive effect of market potential on firm-level productivity for small and medium sized firms. This indicates an important mechanism of how road infrastructure improvements can generate wider economic impacts.

PALABRAS CLAVE: market potential, transport infrastructure, firm-level productivity
1. Introduction
The concept of market potential has received renewed interest in the New Economic Geography literature. Theoretical contributions in this field show how market potential can be derived from formal spatial models and that the extent of market potential is an important determinant of firms’ productivity (Fujita et al. 1999). Key to new economic geography models are transport costs and the way they influence market potential. Empirical studies in this field provide evidence that market potential matters (Redding and Venables 2004, Hanson 2005, Head and Mayer 2004, 2006). However, few studies have provided empirical evidence for the role of transport costs in shaping market potential. In particular, time series information on transport infrastructure investments that reduce transport costs has generally not been taken into account when calculating market potential.

In contrast, the role of infrastructure investment in aggregate productivity growth has received substantial attention and there is now a large body of aggregate studies that treat public infrastructure as input in production functions contributing to national and regional output and that has led to mixed empirical results. On the other hand, empirical studies at the micro-level have typically been developed within the cost-benefit framework and have tended to look at narrow measures of transport user benefits. Recently, the interest is moving towards the assessment of indirect economic impacts and towards understanding the mechanisms whereby transport improvements benefit firms. This is however yet one of the least developed areas in the field of transport project evaluation, but precisely these effects are of great policy interest.

The literature has emphasised important productivity differences across firms, even within narrowly defined industries (Baily et al. 1992; Olley and Pakes 1996).¹ This literature has analysed a range of sources of productivity growth such as the effect of firm size (Geroski 1998), R&D investment (see, Griliches 1995, for a survey), innovation (Crepon et al. 1998; Huergo and Jaumandreu 2004) and human capital investment (Dearden et al. 2006; Almeida and Carneiro 2009) among others, but the effect of transport infrastructure improvements has not been tested directly at the micro-level. Yet there are several ways in which improvements of transport infrastructure can impact on firm-level productivity (see, Anderson and Lakshmanan 2007, for example, for a review). Here, I focus on effects through market potential. Improvements of

¹ For a recent review of this literature, see, Syverson (2010).
transport infrastructure reduce the cost of doing business over distance primarily through reducing travel times and this can increase economic opportunities for firms through providing greater market potential. Greater market potential can allow firms greater specialisation and to exploit scale economies to a greater degree. At the same time, transport improvements that increase market potential also reduce geographic market segmentation and this way can increase substitutability in differentiated goods markets and consequently raise productivity through increased levels of competition (Syverson 2004). Finally, transport improvements can impact on the externalities of agglomeration (Venables 2007). With market expansion and integration, the geographic scope at which agglomeration economies materialise (Graham 2007a, 2007b) can also increase. Through these different mechanisms transport improvements may influence firm-level productivity.

Spain is a particular interesting country to study effects of market potential on firm performance. The country has developed an ambitious road building programme, increasing its motorway network from about 1,900 kilometres at the beginning of the 1980s to nearly 14,000 kilometres by the year 2006. Over this period the basic national motorway network has been established. This immense investment has considerably improved accessibility and market potential, particularly for the more peripheral regions of the country as they became linked to the national and international motorway network. It has been argued that in areas where the network is already very dense impacts from transport investment might actually be small and difficult to measure. The large-scale improvements that have taken place in Spain over the last two decades make the country a particular interesting case study.

In this paper, I analyse the effect of market potential on individual firm performance in the Spanish manufacturing sector. I use panel data which allows controlling for unobserved firm-specific heterogeneity, time specific influences, and endogeneity. This is important, because the type of economic activity in a particular area is not independent of the characteristics of the location. Thus, with firm and workers migration, market potential could be determined simultaneously with productivity. To isolate the real productivity effect resulting from improved market potential, I estimate firm-level production functions using the instrumental variable.

The closest related papers are Lall et al. (2004), Graham (2007a, 2007b) and more recently Combes et al. (2010) who all use firm level data to study the impact of market
potential on productivity for India, UK, and France respectively. Only Lall et al. (2004) and Graham (2007b) calculate market potential with travel times based on the real road network, however for a single point in time only. In this paper I contribute to this literature by constructing market potential measures that are based on the real road network and that take into account its evolution over a period of time that has witnessed important investment in road transport infrastructure. Using the real transport network and its evolution allows to assess the impact of transport improvements on productivity and to take into account impacts from integrating and extending the geographical scope of markets. This is an important issue both for policy makers and academics. Given the considerable funds allocated to transport investment, it is important to know more about the range of economic benefits created by such investment.

The paper is organised as follows. The next section provides a review of the related literature. Section 3 covers data and provides some descriptives. Section 4 presents the econometric model and discusses the main estimation issues. Section 5 presents the results of the empirical estimation, Section 6 [work in progress] will deal with the robustness of the results to alternative estimation methods. Section 7 presents preliminary conclusions.

2. Market potential, transport improvements, and productivity: a survey of the related literature

The literature has for a long time shown that firms are more productive in large and denser urban areas. Ciccone and Hall (1996) show for the U.S. that higher employment density increases labour productivity. Ciccone (2002) and Brülhart and Mathy (2008) provide recent evidence for European regions. For reviews of this literature see, Eberts and McMillen (1999), Rosenthal and Strange (2004), Melo et al. (2009), and Puga (2010). Building on Marhall (1920) the basic idea is that having more other firms nearby generates positive externalities in terms of knowledge spillovers, labour market pooling and input sharing. Traditionally, this literature has used city or regions size or density to measure agglomeration economies. Since cities and regions are not islands and hinterlands matter, agglomeration economies are not necessarily limited by administrative boundaries. In this context, transport infrastructure improvements may extend the spatial scope over which these benefits are obtained. Transport infrastructure
studies and the agglomeration literature have however largely developed separately, but as argued by Eberts and McMillen (1999) there are important links between transport provision and agglomeration. Venables (2007) formally models the relation between transport and agglomeration. Transport improvements reduce the cost and increase the potential for interaction and so bring economic agents closer and can therefore enhance the benefits of agglomeration economies and this way induce positive productivity benefits. Rice et al. (2006) estimate the effect of population of working age size in different driving time bands around locations on their productivity. There findings for UK NUTS3 regions show how the productivity effects of agglomeration extend across space. A doubling of population of working age in proximate areas is found to raise productivity by 3.5% with the greatest effects stemming from locations within 40 minutes driving time.

An alternative to own city or region size and density that takes into account spatial externalities is the concept of market potential first introduced by Harris (1954). Harris defined market potential in a given location as the sum of purchasing power in other locations, weighted by a function of distance.

\[
Acc_i = \sum_j \frac{W_j}{d_{ij}^a}
\]

Where \(W_j\) is a measure of the size or economic mass of location \(j\) (usually measured in terms of population, GDP, or employment), and \(d_{ij}\) is the distance between \(i\) and \(j\). The exponent \(a\) refers to the friction of distance and has been typically set to 1. Alternative expressions have also modelled market potential in terms of a negative exponential function. Harris’s (1954) concept of market potential has been used in many empirical studies as proxy for market demand. Fujita et al. (1999) provide theoretical underpinnings of the concept and show how market potential can be given a formal spatial modelling interpretation. Hanson (2005) derives an augmented market potential function directly within the New Economic Geography framework.

---

2 The larger the value of \(a\), the greater the distinction between nearby and distant destinations. Ideally, it should be estimated from a destination choice model, but adequate disaggregate information is often not available. The value will crucially depend on the type of activity involved. For example, for accessibility to jobs, education, or hospital, shorter average trip length will imply higher values of \(a\), while for industrial activity patterns lower values are used to reflect their larger interaction space.

3 The concept of market potential is appealing because it relates to the natural phenomenon that the volume of interactions such as, for example, trade between locations is lower the further apart they are. While different versions of the concept have been used as applied tool in the new economic geography...
Recent empirical studies of New Economic Geography show that greater market potential raises factor prices and income (Redding and Venables 2004, Hanson 2005, Head and Mayer 2004, 2006). Few studies have, however, estimated directly the impact of market potential on firm-level productivity. Combes et al. (2010) find a positive elasticity of wages with respect to market potential defined as the sum of density of other areas weighted by the inverse distance to these areas. In Graham (2007a) market potential is based on distances that are calculated as direct Pythagoras distances between U.K. ward centroids. In Graham (2007b) the distance based market potential is compared to market potential based on road travel times. Graham (2007a, 2007b) estimates translog firm level production functions and finds positive effects of market potential for most services industries as well as for manufacturing. He argues that the main benefit of transport infrastructure improvement is that it brings more economic agents within accessible reach and thus changes the ‘effective density’ from which agglomeration economies will be derived. Similarly, Lall et al. (2004) also found in a cross-section analysis that access to markets is an important determinant of plant-level productivity among Indian manufacturing firms. As in Graham (2007b) their market potential measure is based on a single cross-section of network travel time along the real road network.

Most studies have calculated market potential on the basis of Great Circle or crow-fly distances between locations, but have not taken into account the real transport network. Yet, the economically relevant distance between i and j is not simply the straight air distance, but the cost to reach locations through the real transport network. Moreover, transport improvements feed into market potential because they alter $d_{ij}$. Thus, from a longitudinal perspective market potential is also determined by the evolution of the transport networks. In panel data studies that use market potential based on direct air distances, this means however that estimates of market potential functions rely on changes in weights, since air distances do not change over time and, thus in practice ignore any potential effect stemming from transport improvement. In contrast, studies that have calculated market potential based on the real transport network have generally used only one cross section and also ignored changes in the transport network. This

---

footnote:

literature, many studies have relied on Harris’s (1954) reduced form market potential measure (Davis and Weinstein 2001; Crozet et al. 2004; Brülhart et al. 2004; Ottaviano and Pinelli, 2006).
paper adds to this literature by estimating the impact of market potential on firm level productivity using a longitudinal data set for firms as well as for transport infrastructure.

3. Estimation strategy

3.1. The model

The objective of this paper is to analyse the degree to which market potential contributes to firm-level productivity. The empirical strategy is based on the estimation of a firm-level production function. The function follows the general Cobb-Douglas form:

\[
Y_{it} = A_{it} K^\beta_{it} L^\beta_{it}
\]

(2)

where \( Y_{it} \) is valued added of firm \( i \) in period \( t \), \( L_{it} \) represents labour input and \( K_{it} \) the capital stock of firm \( i \) at time \( t \). \( A_{it} \) is TFP of firm \( i \) which depends on a firm-level component, \( U_{it} \) and the market potential \( MA \) experienced by the firm:4

\[
A_{it} = (MA_{it})^\delta U_{it}
\]

(3)

After taking logs on both sides, the production function can be written as:

\[
y_{it} = \delta + \beta_2 l_{it} + \delta m a_{it} + \eta_i + u_{it}
\]

(4)

where lower case letters denote the corresponding logarithmic values in equation (2) and (3). The specification also contains a permanent but unobservable firm-specific effect \( \eta_i \) and an error term \( u_{it} \).

In the estimation of equation (4) several econometric issues arise and are discussed below.

3.2. Estimation issues

*Unobserved heterogeneity and simultaneity bias in inputs*: First, firm-specific characteristics, such as management ability or organisational capital which are not directly observable may affect productivity. Endogeneity of labour, capital and other variables in the production function may arise because such unobserved time-invariant

---

4 This is the commonly used framework to examine the contribution of external factors arising from a firm’s environment within the production function approach. Alternatively, studies looking at transport infrastructure specifically have also included infrastructure as a separate production input.
factors can have an impact on the explanatory variables as well as on the value added of firms. With panel data, one can apply the within-firms estimator to control for such time-invariant firm-specific factors.

Second, another source of potential endogeneity in inputs arises from simultaneity if, for example, transitory productivity shocks determine firms’ decisions on input variables. This will arise whenever such productivity shocks are anticipated by the firm in their choices of inputs. This will result in a positive correlation between the right-hand variables and the error term. In this context, OLS, as well as the within-firm estimator will yield inconsistent results. Usual approaches to deal with this problem are the GMM method proposed by Arellano and Bond (1991) or the control function approach by Olley and Pakes (1996).

Spatial sorting and the potential endogeneity of market potential: A further practical problem faced by empirical studies analysing spatial differences in productive performances is that the type of economic activity in a particular area is not independent of the characteristics of the location. For instance, firms in higher value added functions have been documented to be predominantly located in urban and metropolitan areas which are also high market potential areas (Rice et al. 2006). In a new economic geography framework, Baldwin and Okubo (2006) show how location in the largest markets is most attractive for the most productive firms. Nocke (2006) shows in a theoretical model how more efficient entrepreneurs self select into larger markets. Thus, high market potential areas are likely to attract more productive firms and cross-section analysis could therefore produce upwards biased estimates of productivity simply because the higher value added activities typically located in high market potential areas tend to be more productive. This implies that the choice of location may be related to unobservable characteristics such as management ability or attitude towards risk that influence at the same time firms’ productivity. In this case, the market potential measure is partly determined by the initial location decision and thus potentially endogenous. The within-firm estimator provides consistent estimates if the simultaneous nature of firm productivity and market potential is due to time in-variant unobserved characteristics.

---

5 Recent empirical evidence on the importance of self-selection is also provided in Saito and Gopinath (2009) for the Chilean food industry.
However, productivity and market potential can also be simultaneously determined. If some locations experience, for example, positive productivity shocks (e.g. improvements in other endowments than transport infrastructure; subsidies, taxes or factors related to the local business climate), this will attract firms and workers to that location and hence increase market potential. Graham et al. (2010) empirically show that higher levels of productivity can induce growth in the scale of local urban and industrial environments. Since market potential is determined by population size through $W_j$ in Eq. (1), a potential endogeneity problem exits.

Yet another source of simultaneity could arise in Eq. (1) through the denominator if roads where specifically improved in areas that experienced greater productivity increases. In the case of the Spanish road building programme, I argue that this has not been the case. Although roads are clearly not built randomly, road network design is primarily conditioned by the existing spatial pattern of economic activity, i.e. the transport network is designed to link important economic centres. In the case of Spain, the government decided at the beginning of the 1980’s for a doubling of lanes of the existing principal trunk roads that connected the major cities of the country (M.O.P.T., 1993). These were major connections that existed well before the study period. As a consequence, I regard market potential endogenous through firms’ initial location decision, but changes in the road network as exogenous to firm level productivity. Since the GMM estimator of Arellano and Bond (1991) allows for any correlation between explanatory variables and transitory shocks it provides a way to handle both the endogeneity of inputs and market potential. This estimator applies first differences to the equation of interest and instruments with lagged values of the endogenous explanatory variables in levels. This is a suitable estimation method even in the presence of endogenous regressors resulting from simultaneity bias. However, when explanatory variables are persistent over time, lagged levels are weak instruments for the regression in differences. Company variables from micro data tend to be highly persistent as documented in Blundell and Bond (2000). Arellano and Bover (1995) and Blundell and Bond (1998) have developed a GMM system estimator that combines in a system the regressions in differences with the regression in levels. The second part of the system requires additional moment conditions based on no correlation between the

---

6 In related studies, the GMM approach has, for example, also been used to account for the potential endogeneity of agglomeration measures by Brülhart and Mathy (2008) and Brülhart and Sbergami (2009).
variables in differences and the fixed effect. In contrast, explanatory variables in levels and fixed effects can be correlated. For a discussion of the applications to the estimation of productions functions see Blundell and Bond (2000). By computing first differences, estimations control for firm unobservable and time invariant characteristics. By using lagged values of inputs to instrument current differences in inputs (together with lagged differences in inputs to instrument current levels), estimations account for any correlation between input choices and transitory productivity or cost shocks.⁷

4. Data, variable definition and descriptive statistics

4.1 Data sets

The primary source of data is the Encuesta sobre Estrategias Empresariales (ESEE) published by the Fundación Empresa Pública. This data set provides a wide range of information on a representative sample of Spanish manufacturing firms with more than 10 employees. The survey is undertaken annually since 1990 and constitutes an unbalanced panel. In 1990, all firms with more than 200 employees (large firms) were asked to participate in the survey, and the rate of participation reached around 70 percent. Firms between 10 and 200 employees (small and medium sized firms) were chosen according to a random sampling scheme. In subsequent years the initial selection criteria have been maintained by incorporating newly created firms selected with the same sampling criteria as in the base year (see Ministry of Industry 1992 and Fariñas and Jaumandreu 1999 for more details). The sample used for this analysis covers the period from 1990 to 2005.

To include the impact of transport investment I use spatially geo-referenced data of the Spanish road network. I focus on roads, since this is by far the dominant transport mode in Spain and it is this mode of transport that has experienced by far the largest investment over the period of analysis.⁸ The data includes annual information on the timing of openings of new motorway segments. To calculate market potential indicators, I combine the road network data with spatially geo-referenced annual

---

⁷ All estimations of GMM are carried out using the xtabond2 command for STATA, developed by Roodman (2006).

⁸ About 90 percent of passenger traffic and 80 percent of freight movements in Spain are by road transport.
municipality data. By relating road network data and municipality data in GIS, detailed market potential indicators are calculated over the period of analysis.

To link market potential measures to the firm-level data one would ideally use detailed information regarding the location of each company. Unfortunately, for reasons of confidentiality, such detailed spatial disaggregation is not provided in the ESEE. The ESEE provides location information at the Autonomous Community level together with information on the size (5 size categories) of the municipality where the firms’ main establishment is located.\(^9\) Taking advantage of this information, I have calculated for each Autonomous Community the weighted average of market potential levels for each of the five municipality-size categories in order to relate each ESEE company to the accessibility data. Using regional market potential averages by municipality size categories instead of firm-specific market potential generates, however, an additional component of the error term which, by its own definition is heteroscedastic and orthogonal to the regressors. Therefore, standard errors robust to heteroscedasticity have been computed.

4.2. Variables and summary statistics

*Firm value added:* Value added is computed subtracting the value of inputs from output (sales plus variation of inventories). Inputs are the sum of purchases and external services minus the variation of intermediate inventories. Recent studies have emphasised the importance of using individual firm-level price information instead of industry price indices to calculate real output (see, for example, Klette and Griliches 1996; Ornaghi 2006b).\(^{10}\) The ESEE reports the percentage change in the selling price applied by the firm as well as information on firms’ variation in the cost of raw materials. Following previous studies that have estimated firm level production functions with the ESEE data, such as, for example, Huergo and Jaumandreu (2004) Ornaghi (2006a, 2006b), and, Escribano and Stucchi (2008), I use information on the former to deflate nominal output and on the latter to calculate deflated nominal inputs.\(^{11}\)

\(^9\) Unfortunately, there are no alternative firm level panel data sets for Spanish firms available that would provide more detailed spatial disaggregation.

\(^{10}\) Mairesse and Jaumandreu (2005) however find little improvement of using firm-level price indices over industry-level indices and conclude that other specification errors may be much more important.

\(^{11}\) Prices are expressed in reference to 1990. For firms that entered the sample after 1990, a hypothetical price increase is calculated for every year between 1990 and the first year the firm appears in the sample based on the average price increase by year and industry.
**Capital stock:** The capital stock evaluated at the current replacement value is computed recursively from an initial estimate and the data on current investment applying the perpetual inventory method following Martin-Marcos and Suarez (1997).\(^\text{12}\) For a better proxy of capital services it has been argued that the variations in the firms’ capacity utilisation should be taken into account (Huergo and Jaumandreu 2004, Mairesse and Jaumandreu 2005, Ornaghi 2006b). In the short run, firms may increase output through variations in capacity utilization of machinery and equipment without realizing new investment. Thus, the yearly average rate of capacity utilization reported by the firm in the ESEE survey is also included in estimations.

**Labour:** Labour input is measured as total effective yearly hours of work.

**Market potential:** I use the reduced form accessibility measure first proposed by Harris (1954) to proxy market potential as in equation (1) where distances are based on travel times along the real road network. Real travel times are economically more relevant than alternative measures such as distances, but most importantly, distance do not significantly vary over time. Transport investments primarily reduce travel times. Travel time distances are calculated for all Spanish mainland municipalities serving as origins and the 438 largest Spanish cities with more than 10,000 inhabitants are used as destinations. These cover over 75 percent of the total Spanish peninsular population. Destination weights are based on population and \(a\) is set to 1 (for further details, see, Holl 2007). The Harris market potential measure is straightforward to implement and has been shown to perform well compared to more complex measures that are directly derived from new economic geography models (Head and Mayer 2004, 2006). In order to capture possible non-linear effects, I also include the squared term of market potential. With increasing market potential benefits may be limited by diseconomies such as commuting and congestion costs.

The final sample consists of an unbalanced panel of 2,688 firms. I have dropped all firms which changed location or industrial sector as such changes imply that their productivity is not comparable in time. I have also dropped multi-plant companies because with the information available from the ESEE survey it is neither possible to know the location of additional plants, nor to assign correctly value added, employment

\(^{12}\) Estimates differ in so far that they are based on industry-specific depreciation rates instead of product-specific depreciation rates as information on the latter is not available. Similarly, the average life of the fixed assets is always based on the average oldness of the fixed assets as information on whether the firm has updated the value of its fixed assets is not available either.
and capital stock data among plants of multi-plant companies. Furthermore, in cleaning
the data set, observations with missing, negative or null value added, employment or
stock of capital have also been dropped. Table A1 in the Appendix, provides
information on the distribution of firms by year in the final sample used for estimation.
Table A2 in the Appendix shows the usual descriptive statistics.

5. Results

First estimation results are presented in Table 1 to Table 3. All estimations include
unreported annual year dummies as well as industry dummy variables to account for
productivity shocks common to all firms and sector-specific factors accounting for
productivity differences. Furthermore, estimations are carried out separately for small
and medium sized firms (up to 200 employees) and large firms (more than 200
employees). This accounts for the characteristics of the data set; in particular, the fact
that sampling proportions are different for small and medium sized firms and large
firms. Furthermore, the relative importance of improvements in market potential may
actually vary across firm size.

As a starting point and benchmark, pooled cross section OLS estimates are reported in
column (1) and (3) of Table 1 for small and medium sized firms and large firms
respectively. These estimates treat all variables as exogenous and are only consistent if
explanatory variables are uncorrelated with both time variant and time-invariant firm-
specific unobservables. Column (2) and (4) show the results of fixed effects estimations
which allow for any correlation structure between the covariates and the firm-specific
time invariant unobservables.

Starting with the pooled OLS results for the small and medium sized firms in column
(1), coefficient estimates of capital, labour are significantly different from zero at the
one-percent level and have the expected sign. Output elasticity with respect to labour
amounts to about 80%, output elasticity with respect to capital is about 27%. These are
reasonable estimates that reflect input shares fairly well. Capacity utilization shows a
small positive and significant coefficient as expected.13 Most importantly to this
analysis, the point estimate of the market potential coefficient is positive and significant

13 All estimations have also been carried out without inclusion of the capacity utilization variable and
results remain basically unchanged.
while the coefficient for the square term is negative and significant. This indicates an inverted U-shaped relation where market potential increases productivity, but the effect gets smaller as market potential increases. Thus, there appear to be diminishing returns to increasing market potential.

Column (2) shows fixed effects estimations for small and medium sized firms. These estimates control for time-invariant firm unobservable factors and thus the potential endogeneity of market potential that may result from such factors affecting firms’ initial location decision. Point estimates for capital and labour are substantially lower than values obtained from simple pooled OLS regressions. Market potential is again positive and significant and the square term is negative and significant.

Column (3) and (4) show the corresponding results for the large firm sample. Market potential turns out significant in the pooled OLS regression. As with small and medium sized firms, there appears to be a non-linear relationship between market potential and valued added. There is no significant effect in the fixed effects estimates.

In Table 2, results of GMM estimates for small and medium sized firms are presented. I start with results of first-difference GMM (GMM-FD) which account for endogeneity. Contemporaneous inputs in differences are instrumented with the corresponding values in the past. The corresponding results from the two-step estimation are presented in column (1). The capital coefficient is similar to the fixed effects result, but the labour coefficient is now markedly lower. The market potential coefficient turns out insignificant, as well as its quadratic term. This may be due to the weak power of the instruments. This is a well known problem with this type of estimator when variables show a high degree of persistence over time. Market potential does not change drastically even in the case of a large case infrastructure programme as the one carried out in Spain. This problem of weak instruments can lead to substantial bias in finite samples.

To address this problem, in column (2) and (3) results of the system GMM estimation are presented. I treat all time dependent regressors as potentially endogenous and instrument first differences with lagged levels and their current values in the level equation with lagged first differences. This approach allows for contemporaneous correlation between these variables and shocks to the production function equation, as well as correlation with unobserved firm-specific effects. In column (2), I use lagged
values two periods and earlier to instrument the difference equation. These instruments are valid as long as input decisions two periods earlier are made without knowledge of the current transitory shocks in the production. This is to say that firms cannot forecast future shocks. The assumption will hold if there are no strong serial correlations in the transitory shocks. In this sense, the test for second-order correlation is satisfactory. The main results are confirmed. Market potential is now again strongly significant and positive while the square term is again negative and significant.

The marginal effect of market potential is given by \( \delta_1 + 2\delta_2 ma_i \), and amounts to 0.347 indicating that a 10% increase in market potential leads on average to approximately 3.5 percentage increase in value added.

In column (3), I present results from a dynamic specifications of the production function. Dynamics can arise, for example in the presence of adjustment costs causing value added to respond sluggishly to changes in independent variables. If, for example, benefits from increased market potential take time to materialise then one would expect to see also delayed responses in productivity to changes in accessibility. Results indicate a high degree of persistence. The coefficient on lagged value added is positive and highly significant and amounts to 0.534. Market potential is again statistically significant and positive and the square term is as before negative and significant. The contemporaneous marginal effect of market potential is markedly lower with 0.139. The long run elasticity is given by \( \frac{\delta_1 + 2\delta_2 ma_i}{1 - \alpha} \) and amounts to 0.298.

Note that this dynamic specification uses instruments dated \( t-3 \) and before in the difference equation. This is because there are signs of significant second-order serial correlation which invalidates instruments dated \( t-2 \).\(^{14}\) To further assess instrument validity, I also report the Hansen \( J \) test and the Difference Hansen test for overidentifying restrictions (Blundell and Bond, 1998). The Hansen \( J \)-statistic evaluates the joint validity of the moment conditions.\(^{15}\) The p-value reported indicates that the null hypothesis that the over-identifying restrictions are valid can not be rejected. In addition, in the case of GMM-SYS, the Difference-Hanson test examines if there is

---

\(^{14}\) If there is evidence of second-order serial correlation, instruments need to be dropped back more time periods to become valid instruments.

\(^{15}\) The Hanson \( J \) statistic is robust to heteroskedasticity or autocorrelation (Roodman 2006). It is the (robust) minimized value of the two-step GMM criterion function.
evidence that the additional moment conditions are valid. This test confirms the validity of the additional instruments. Overall, the estimation diagnostics indicate that estimation specifications are acceptable.

In Table 3, I present GMM estimation results for the sample of large firms. In this case the sample size is considerably reduced. In the original sample, about 30% of firms had more than 200 employees. In the final sample there are only 635 firms representing 22% of the total final sample. By dropping all multi-plant firms and firms that changed location, observations for large firms are disproportionally lost.

With the smaller sample size of large firms, there were signs of overfitting when using several lags of regressors as instruments in the difference equation. Thus, I have limited the instrument count. Specifically, I use only instruments at t-2 and leave out all instruments beyond t-2 (Roodman 2006). Market potential is only significant at the 10-percent level in the static GMM-SYS model of column (2). Here, the marginal effect of market potential is 0.131 and significantly lower than the corresponding result for small and medium sized firms. At the same time, the coefficient for capital is higher than for the small and medium sized firms while the coefficient for labour is lower as was observed with the pooled OLS estimates. This is as would be expected and reflects the higher capital share in large firms. In the GMM-FD model as well as in the dynamic specifications of the GMM-SYS estimations, results for the large firm sample are, however, largely inconclusive. This could be due to the reduced sample size. Also, dropping all multi-plant firms could bias the results. Running the estimations on the sample including all multi-plant firms, does indeed show that the results of column (2) are sensitive to their inclusion. While point estimates of the firm-level variables are similar, none is significant and nor is market potential significant. In the GMM-FD model and in the dynamic GMM-SYS estimation market potential remains non-significant.

In contrast re-estimation of the sample of small and medium sized firms without dropping multi-plant firms yields very similar results. In particular, concerning the market potential variable results remain largely unchanged both in magnitude and significance for small and medium sized firms.

As further robustness check I have also estimated dynamic models with the lagged dependent variable plus lags of all regressors. This does not result in qualitatively
different results. For small and medium sized firms, lagged value added, labour, capital, market potential and market potential squared are all again significant and of similar magnitude. Among the lagged independent variables only labour is significant. For large firms, the corresponding estimates are also similar in magnitude and significance with only the lagged capital utilization rate being significant among the lagged independent variables. Lagged market potential is always insignificant. Finally, I have run estimations with additional controls for localisation and urbanisation economies. Agglomeration economies have received a great deal of attention in the local and regional growth literature. In practice, however, effects of agglomeration economies are difficult to separate empirically from those of market potential. Nevertheless, in unreported estimations, I have added the log of employment in the same sector and same region and the log of total employment in the same region, in both cases outside the own firm, to control for localisation and urbanisation economies respectively.\footnote{All unreported estimates are available upon request.} In the preferred system GMM estimates, neither the proxy for localization economies nor the proxy for urbanization economies turns out significant. Estimates for capital, labour and market potential are almost the same. For example, for small and medium sized firms, the marginal effect of market potential is 0.294 in the static model and 0.168 for the contemporaneous effect and 0.350 for the long-run effect in the dynamic specification. For large firms, the proxies for localization and urbanization economies are also not significant. However, their inclusion in the static system-GMM regression makes the coefficients for market potential insignificant, again indicating that estimates for large firms are not robust. These results, however, reinforce the finding of a positive and significant effect of market potential for small and medium sized firms and suggest that market potential has a stronger impact on their productivity than locally-bounded agglomeration economies.

6. Alternative estimation approaches [WORK IN PROGRESS]

I further test the robustness of the findings to alternative estimation methods. Instead of adopting a one step approach as in the previous section where market potential enters directly into the production function, in ongoing work I test the robustness of the results to a two stage procedure where I’m first estimating TFP at the firm level and then test how market potential affects firms’ estimated TFP. In the first stage, I calculate TFP. In
In addition to OLS as benchmark, I follow Olley and Pakes (1996) who have developed a semi-parametric estimator where the simultaneity problem is solved by using the firm’s investment decision to proxy unobserved productivity shocks. In the second stage, I estimate the relationship between productivity and market potential and adopt an IV strategy to control for the potential simultaneity between TFP and market potential.

7. Discussion and conclusions [TO BE COMPLETED]

In this paper, I have looked directly at the effect of market potential on firm-level productivity. The market potential measure used takes into account the real transport network and its improvements. Such a transport based market potential reflects the geographical dimension of infrastructure and the services provided by infrastructure.

Different models and estimation techniques have been applied with the aim to reveal the real productivity effect resulting from improved market potential. The first results from the preferred system GMM estimations show that there are important productivity gains from transport infrastructure improvements through increased market potential for small and medium sized firms. Compared to estimates of the benefits from urban size and density, the results are in the upper range. Melo et al. (2009) provide a meta-analysis of estimates of urban agglomeration economies. For the period after 1990 they find an average elasticity estimate of 0.081. However, they also point out that studies that have used market potential have generally obtained higher estimates than those based on urban size or density. Puga (2010) in his review of the agglomeration economies literature positions the magnitude of productivity increases stemming from a doubling of city size between 3 and 8 percent. Combes et al. (2010) conclude that accounting for the endogeneity of location of workers and firms lowers the estimates of density to 0.02 and that of market potential to 0.034. The elasticity of market potential found in this paper for small and medium sized firms is markedly higher. First, small and medium sized firms may benefit disproportionally from improvements in market potential. As Forman et al. (2008) show, firms can substitute internal firm resources and external local resources. This makes the advantages of locations more important for smaller firms with less internal resources. Second, the results in this paper also suggests that ignoring the fact that infrastructure improvements extend markets may underestimate
the real effect from market potential as well as the benefits from agglomeration economies.

The paper contributes to the debate on whether there are wider economic benefits from transport infrastructure investments. Where transport infrastructure investment increases productivity, the direct user benefits estimated in traditional cost-benefit analysis will not capture total benefits. A full accounting of benefits must include gains from any additional economic activity it generates. To date, the estimation of indirect economic impacts of transport infrastructure investment is receiving increased attention but is still an unsolved issue.

A limitation of the ESEE survey data used in this paper is the lack of detailed zip-code location information of firms which meant that market potential had to be aggregated to the regional level. With detailed location information of firms, further research could estimate the spatial decay of market potential. A more refined analysis should also take into account traffic flows. As the results in this paper indicate, network improvements can be offset by rising congestion.

That said, the analysis presented in the paper provides a contribution to our understanding of productivity impacts of market potential and transport infrastructure investment in the manufacturing sector. This type of firm-level studies should be seen as a complement to aggregate productivity studies.

Acknowledgment: This work was supported by the Ministerio de Ciencia e Innovación [SEJ2006-08063] and a Ramón y Cajal research contract.
References:


Table 1. OLS estimation results:

<table>
<thead>
<tr>
<th></th>
<th>Small and medium sized firms</th>
<th>Large firms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Ln capital</td>
<td>0.265*</td>
<td>0.142*</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Capacity utilization rate</td>
<td>0.004*</td>
<td>0.003*</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>Ln labour</td>
<td>0.806*</td>
<td>0.706*</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Ln market potential</td>
<td>1.399*</td>
<td>2.810***</td>
</tr>
<tr>
<td></td>
<td>(0.480)</td>
<td>(1.624)</td>
</tr>
<tr>
<td>Ln market potential squared</td>
<td>-0.134**</td>
<td>-0.438**</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.214)</td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Observations</td>
<td>8872</td>
<td>8872</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.788</td>
<td>0.689</td>
</tr>
</tbody>
</table>

Note: Robust standard Errors in parenthesis. Significant coefficients are indicated by *, **, *** for significance at the 1%, 5% and 10% level, respectively. All estimations include time and industry dummies.
Table 2. GMM estimation results: small and medium sized firms

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: ln value added</th>
<th>GMM-FD</th>
<th>GMM-SYS</th>
<th>GMM-SYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln capital</td>
<td>0.147***</td>
<td>0.267*</td>
<td>0.125*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.046)</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>Capacity utilization rate</td>
<td>0.003**</td>
<td>0.004*</td>
<td>0.003**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>Ln labour</td>
<td>0.495*</td>
<td>0.873*</td>
<td>0.480*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.074)</td>
<td>(0.071)</td>
<td></td>
</tr>
<tr>
<td>Ln market potential</td>
<td>1.807</td>
<td>4.633*</td>
<td>2.666*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.566)</td>
<td>(1.548)</td>
<td>(0.935)</td>
<td></td>
</tr>
<tr>
<td>Ln market potential squared</td>
<td>-0.442</td>
<td>-0.543*</td>
<td>-0.320*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.372)</td>
<td>(0.195)</td>
<td>(0.116)</td>
<td></td>
</tr>
<tr>
<td>Ln value added (-1)</td>
<td></td>
<td></td>
<td>0.534*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.049)</td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td></td>
<td>t-2</td>
<td>t-2</td>
<td>t-3</td>
</tr>
<tr>
<td>m1 (p-value)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>m2 (p-value)</td>
<td>0.726</td>
<td>0.546</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Hansen test (p-value)</td>
<td>0.176</td>
<td>0.325</td>
<td>0.403</td>
<td></td>
</tr>
<tr>
<td>Difference Hansen test (p-value)</td>
<td></td>
<td></td>
<td>0.440</td>
<td>0.596</td>
</tr>
<tr>
<td>Observations</td>
<td>7240</td>
<td>8872</td>
<td>8104</td>
<td></td>
</tr>
</tbody>
</table>

Note: Significant coefficients are indicated by *, **, *** for significance at the 1%, 5% and 10% level, respectively. All estimations include time and industry dummies. Estimations are carried out with two-step robust standard errors, and with the finite-sample correction to the two-step covariance matrix derived by Windmeijer (2005).
Table 3. GMM estimation results: large firms

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: ln value added</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GMM-FD</td>
</tr>
<tr>
<td>Ln capital</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>(0.251)</td>
</tr>
<tr>
<td>Capacity utilization rate</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>Ln labour</td>
<td>0.292</td>
</tr>
<tr>
<td></td>
<td>(0.283)</td>
</tr>
<tr>
<td>Ln market potential</td>
<td>1.861</td>
</tr>
<tr>
<td></td>
<td>(5.383)</td>
</tr>
<tr>
<td>Ln market potential squared</td>
<td>-0.120</td>
</tr>
<tr>
<td></td>
<td>(0.813)</td>
</tr>
<tr>
<td>Ln value added (-1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td>t-2</td>
</tr>
<tr>
<td>m1 (p-value)</td>
<td>0.001</td>
</tr>
<tr>
<td>m2 (p-value)</td>
<td>0.709</td>
</tr>
<tr>
<td>Hansen test (p-value)</td>
<td>0.528</td>
</tr>
<tr>
<td>Difference Hansen test (p-value)</td>
<td>0.598</td>
</tr>
<tr>
<td>Observations</td>
<td>1991</td>
</tr>
</tbody>
</table>

Note: Significant coefficients are indicated by *, **, *** for significance at the 1%, 5% and 10% level, respectively. All estimations include time and industry dummies. Estimations are carried out with two-step robust standard errors, and with the finite-sample correction to the two-step covariance matrix derived by Windmeijer (2005).
### Table A1: Sample: number of firms by year

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>1399</td>
</tr>
<tr>
<td>1991</td>
<td>1155</td>
</tr>
<tr>
<td>1992</td>
<td>1037</td>
</tr>
<tr>
<td>1993</td>
<td>891</td>
</tr>
<tr>
<td>1994</td>
<td>1056</td>
</tr>
<tr>
<td>1995</td>
<td>928</td>
</tr>
<tr>
<td>1996</td>
<td>851</td>
</tr>
<tr>
<td>1997</td>
<td>773</td>
</tr>
<tr>
<td>1998</td>
<td>988</td>
</tr>
<tr>
<td>1999</td>
<td>881</td>
</tr>
<tr>
<td>2000</td>
<td>1031</td>
</tr>
<tr>
<td>2001</td>
<td>921</td>
</tr>
<tr>
<td>2002</td>
<td>928</td>
</tr>
<tr>
<td>2003</td>
<td>697</td>
</tr>
<tr>
<td>2004</td>
<td>693</td>
</tr>
<tr>
<td>2005</td>
<td>1153</td>
</tr>
<tr>
<td>Total</td>
<td>15382</td>
</tr>
</tbody>
</table>

### Table A2. Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min.</th>
<th>Max.</th>
<th>Average annual growth rate (A log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value added (in thousands)</td>
<td>7202.7</td>
<td>19123.7</td>
<td>0.15</td>
<td>765048.1</td>
<td>0.037</td>
</tr>
<tr>
<td>Labour (effective total hours worked, in thousands)</td>
<td>255.4</td>
<td>677.1</td>
<td>3</td>
<td>16519</td>
<td>-0.005</td>
</tr>
<tr>
<td>Capital stock (in thousands)</td>
<td>6006.6</td>
<td>26401.8</td>
<td>0.69</td>
<td>1242320</td>
<td>0.074</td>
</tr>
<tr>
<td>Capacity utilization (%)</td>
<td>81.5</td>
<td>15.3</td>
<td>5</td>
<td>100</td>
<td>-0.006</td>
</tr>
<tr>
<td>Market potential index(^a)</td>
<td>53.8</td>
<td>15.41</td>
<td>25.4</td>
<td>92.5</td>
<td>0.014</td>
</tr>
</tbody>
</table>

*Notes: \(^a\) based on Eq.(1).*