# Agglomeration economies or selection? An analysis of Taiwanese science parks

**Syed Hasan**[1]

**Alessandra Faggian**[2]

**H. Allen Klaiber**[2]

**Ian Sheldon**[2]

[1]Department of Economics, Lahore University of Management Sciences, Lahore, Pakistan
[2]Department of Agricultural, Environmental and Development Economics, The Ohio State University, Columbus, OH, USA

The authors would like to thank participants at the 2013 annual meetings of the AAEA and NARSC, and participants in the IAMO symposium in Halle in 2014, for their valuable comments and suggestions on earlier versions of this paper. Hasan acknowledges the Institute of International Education for granting the doctoral scholarship which made this research possible.

**Corresponding Author:**

Ian Sheldon, Department of Agricultural, Environmental, and Development Economics, The Ohio State University, 2120 Fyffe Road, Columbus, OH 43210, USA.
Email: sheldon.1@osu.edu.

## Agglomeration economies or selection? An analysis of Taiwanese science parks


### Abstract

Agglomeration spillovers are a major driver of policy creating science parks across the world. However, agglomeration benefits may be offset by competition arising out of spatial proximity of firms. Analysis of Taiwanese firms' total factor productivity distribution shows that, depending on location choice, the impact of agglomeration and selection is heterogeneous across firm types. Spatial analysis is applied to evaluate the regional innovation policy of establishing science parks. A sectoral analysis of productivity distributions reveals that there is a positive relationship between technology-intensity of the production process and firm-productivity levels when firms are located in science parks.



**Keywords**: Productivity, agglomeration, selection, science parks

**Agglomeration economies or selection? An analysis of Taiwanese science parks**

## Introduction

Evaluating the effectiveness of local policies designating science parks requires an analysis of the productivity of firms located in those areas compared to firms located elsewhere. Given the large expenditures - often in terms of foregone tax revenues - associated with these policies, it is important to understand their effectiveness and the mechanisms underlying any possible productivity gains arising from them. To emphasize the level of public funds involved, it may be noted that in the case of Taiwan, estimated annual tax credits from a single policy instrument, *Statute for Upgrading Industries* (SUI), accounted for approximately one third of the total NT$100 billion tax revenue loss for the government which is close to 1 percent of country's GDP (Lien et al. 2007).

Estimation of firm-level productivity is recognized as a fundamental performance evaluation parameter in the heterogeneous firms' literature. For instance, recent models of trade and industry such as that presented in the seminal work of Melitz (2003) describe how the idiosyncratic productivity of heterogeneous firms determines their survival in domestic and foreign markets. However, Melitz does not address the issue of what actually determines firm productivity. Productivity is a key metric to evaluate firm success, so understanding how spatial policies designating science and technology parks might impact this productivity is of paramount importance.

As Parr (2002) points out *"the concept of agglomeration economies continues to represent an important aspect of locational analysis and regional economics*." (p.151). However, while the regional economics literature has focused largely on the effects of large cities (Rosenthal and

Strange, 2004), small cities (Gabe, 2004) and industrial clusters (Ciccone and Hall, 1996; Cooke, 2002) on productivity, a detailed study of the effect of science parks on productivity is still missing. Moreover, more recently, some contributions - both theoretical and empirical - point out that the increased productivity of firms located in large cities or industrial clusters might be related to self-selection rather than agglomeration economies (Combes et al. 2012; Baldwin and Okubo 2006).

There are three main issues with the current literature on the effect of science parks on productivity. First, site-specific methodologies, such as the case study of 72 US parks by Luger and Goldstein (1991), are often adopted, making generalization difficult and leaving little margin for meaningful policy recommendations. Second, the empirical evidence for firm profitability, survival rates etc., is generally mixed. Third, and most importantly, the variable of interest used for the evaluation of science parks might suffer from selectivity bias as pointed out by Phan, Siegel and Wright (2005). For example, an endogeneity problem may arise if the analysis uses the rate of firm survival as the dependent variable, given that science incubators are designed specifically to increase the life span of firms.

Our contribution addresses these issues by presenting a research methodology not restricted to a particular estimation model or specific park objective(s), but based instead on a robust theoretical foundation which provides a solid basis for generalization and policy evaluations of different contexts.

In order to do so, we extend Combes et al. (2012) using firm-level panel data from Taiwan for the period 2009 to 2011. Productivity distributions for firms in cities and science parks are simultaneously analyzed to identify the impact of agglomeration and selection effects. The analysis includes estimates for aggregate manufacturing and also specific industrial sectors defined on the basis of technology-intensity of the production process and measured through the percentage of employment in technology-oriented jobs (Hecker 2005).

The analysis generates the surprising result that only firms employing a higher proportion of technology-oriented workers display an increase in productivity when located in a science park. As the proportion of technology-oriented workers drops, so does the level of mean productivity. Controlling for self-selection bias, the differential effect of agglomeration economies and selectivity on firms' productivity is identified. Selection effects, although present, are much less in magnitude than agglomeration effects. Moreover, firms located in science parks also benefit from co-location with other firms in the same sector, which is consistent with the presence of Marshallian (localization) externalities. Interestingly, the opposite holds for Jacobian (diversification) externalities, which have a negative impact on the productivity of firms in science parks.

From a policy perspective, these results suggest that efficiency in the utilization of public incentives - offered via science parks - increases with the technological level of the industry and that, while science parks do help in correcting innovation market failures, they may turn out to be protection against market competition if extended to industries that are not technology-intensive.

## Regional Productivity and Spatial Innovation Policy

There seems to be a consensus on the positive relationship between productivity levels and regional density of employment, economic and industrial activity, so that enterprises located in large cities are - on average - more productive, see, for example, Fogarty and Garofalo (1988), and Tabuchi (1986). There are three main reasons often quoted in the literature for this relationship. The first is agglomeration economies: economies external to firms arising out of sharing and spillovers and ultimately causing increasing returns for the entire neighborhood. The second is competition-based selection: firm heterogeneity results in their varying placement across the productivity scale and as selection is tougher in large urban areas so only the most productive firms may survive or profitably operate there. The third is sorting: *ex ante*, more productive firms or talented individuals may choose to locate in larger cities.

With respect to agglomeration economies associated with urban regions, a detailed review of relevant studies and their findings is reported in Rosenthal and Strange (2004). A very significant contribution by the same authors is the estimate that productivity increases by 3-8 percent if city size is doubled. Marshallian externalities are generally attributed to agglomeration economies associated with firms located in large cities and industrial clusters with the theoretical underpinnings dating back more than a century to the influential work of Marshall (1890). The agglomeration economies literature explains productivity gains resulting from labor-market pooling, input sharing, and knowledge spillovers (Cainelli, Fracasso and Vittucci Marzetti 2014).

Apart from the agglomeration story, high level productivity observed in the case of large cities has recently been explained in terms of competitive selection associated with large markets. This explanation draws on Melitz (2003), who introduced product differentiation and international or interregional trade into the framework of industry dynamics due to Hopenhayn (1992). Within this framework, Melitz and Ottaviano (2008) incorporated variable price–cost markups and showed that larger markets attract more firms, which makes competition tougher. As firms cluster to gain agglomeration economies, the increased competition due to this clustering may reduce profits and thereby their willingness to locate in denser locations. Consequently from a process of “natural” selection, a remaining mass of higher than average productivity firms is observed in large cities or industrial clusters.

There is also a strand of literature indicating sorting of high productivity firms into large markets. This self-selection phenomenon raises serious endogeneity concerns when evaluating the impact of spatial clustering policies on firm productivity (Baldwin and Okubo 2006). Thus the sorting phenomenon is likely to confound much of the existing empirical literature on firm productivity along with the estimation of agglomeration benefits and congestion effects associated with clustering of firms.

The theoretical basis of this paper is the nested model of Combes et al. (2012) which distinguishes agglomeration from selection effects. Combes et al. extend Melitz and Ottaviano, by introducing agglomeration economies in the manner of Fujita and Ogawa (1982), and Lucas and Rossi-Hansberg (2002), and developed a model that includes both agglomeration and selection effects. Under monopolistic competition with free entry, profits decline as the number of competitors increases in one location. This results in reduced survival for less efficient firms. Combes et al., with two-digit industry-level data, structurally parameterize and use rightward shift, dispersion and left truncation of log-total factor productivity (TFP) distributions to indicate the strength of agglomeration and selection effects.

In contrast to their approach, in this paper the focus is on a sample of specific sectors (computers and electronics, chemicals, and scientific and technical services) rather than on the aggregate manufacturing sector alone, so that sector-specific dynamics such as market conditions for supply of inputs, demand for output and the form of production functions can be controlled for in the estimation. Using Syverson's (2004) approach, we identify proxies for shift and truncation in the log TFP distributions across regions to estimate the impact of agglomeration and selection. The possibility of self-selection bias is controlled using the two-stage Heckman (1979) selection model.

The empirical analysis conducted in this paper depends on bias-free estimates of TFP distributions. The estimation method used is from Olley and Pakes (1996), their technique being robust to two econometric concerns: simultaneity and selectivity bias. However, the proxy variable for free inputs in Olley and Pakes' method is firm's investment. Often datasets report missing values regarding investments made by firms and thus a large number of observations have to be dropped in the estimation process. To avoid this, Levinsohn and Petrin's (2003) method is adopted, which uses intermediate inputs to proxy for productivity shocks. Given the limitations of the dataset available for this study, the return on capital is used as a proxy for investment while estimating TFP through Olley and Pakes' method.

Many definitions of science parks have been proposed, mostly by professional organizations, for example, the Association of University Research Parks (AURP) (1998) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) (2006), and by parks themselves as a way to define their activities. Common among these definitions is that a park is a type of public-private partnership that fosters knowledge flows often amongst park firms and universities and thereby contributes to regional economic growth and development. Empirical support for agglomeration effects in a park is provided by Jaffe (1989), Jaffe, Trajtenberg, and Henderson (1993), Audretsch (1998), and Rothaermel and Thursby (2005a, 2005b). A comprehensive overview of the research related to science park performance is given by Dabrowska (2011). Most studies focus on the impact of science park intervention on innovative capability, survival rates, profitability and job creation. The empirical evidence is largely mixed and inconclusive, providing little margin for policy recommendation (Monck, 2010).

Inspired by the success of California's Silicon Valley, the Taiwanese government embarked on a pursuit of upgrading its economy with technology and capital-intensive industries. In 1979, a statute was enacted for the establishment of an industrial park. The first park was established in December 1980 in Hsinchu city and it now stretches over both the city and county of Hsinchu. The park was a public project in its entirety as it was developed using public land and publicly funded infrastructure. The central government provides strong policy regulations along with preferential fiscal and other investment incentives. Similar science parks were subsequently established in central and southern Taiwan with the objective of providing a favorable environment with appropriate incentives to attract current technologies and skilled human resources.

## Model and Empirical Strategy

A model explaining the impact of external economies of scale and competitive selection on log TFP distribution as developed in Combes et al. (2012), and extended for regions by Arimoto, Nakajima and Okazaki (2009) is presented in this section. We further extend these models to accommodate sectoral impact on firm entry/ exit decision within a region.

### General Framework of the Model

Modifying and applying the model of Combes et al., the theoretical background of the productivity improvement effects in large cities and science parks in Taiwan is described.

***Preferences and Demand.*** First, the general framework of the model is introduced. A consumer's utility is given as follows,

$$U = q^o + \alpha \int_{i \in \Omega} q^i di - 1/2\gamma \int_{i \in \Omega} (q^i)^2 di - 1/2\eta (\int_{i \in \Omega} q^i di)^2 , \tag{1}$$

where $q^i$ denotes the consumption of variety $i$ from a set $\Omega$ of differentiated varieties of manufactured goods and $q^0$ is the *numeraire* good. Using the demand function for variety $i$ the utility maximization problem is solved subject to a budget constraint. Taking $P$ as the average price of varieties with positive consumption, the demand function for variety *i* can be written as follows:

$$q^i = \begin{Bmatrix} \frac{1}{\gamma+\eta\omega}(\alpha+\frac{\eta}{\gamma}\omega P)-\frac{1}{\gamma}p^i \text{ if } p^i \le \bar{h}, \\ \\ 0, \text{otherwise} \\ \end{Bmatrix}, \tag{2}$$

where $\bar{h}$ is the cutoff price at which the demand becomes zero.

***Production.*** The numeraire is produced under constant returns to scale using one unit of labor per unit of output which implies that the cost to firms of hiring one unit of labor is always unity. Differentiated products are produced under monopolistic competition. By incurring a sunk-entry cost *s* a firm manufactures a product using *h* units of labor per unit of output. The value of *h* differs across firms depending on their productivity and is randomly drawn, from a distribution with known probability density function *g* (*h*) and cumulative density function *G*(*h*) common to all regions. The total sales of a firm are $Q(h) = Cq(h)$, where $C$ is the mass of consumers. Firms maximize their profit as follows:

$$\pi(h) = \left[p(h)-h\right]Q(h). \tag{3}$$

In the monopolistically competitive industry with free entry firms enter until *ex-ante* profits can no longer offset the sunk-entry cost,

$$\frac{C}{4\gamma}\int_{o}^{\bar{h}}(\bar{h}-h)^2 dG(h)=s\,. \tag{4}$$

Using the optimal pricing rule the zero cut-off profit condition is derived as,

$$N \equiv \omega = \frac{2\gamma}{\eta}\frac{\alpha-\bar{h}}{\bar{h}-H} \tag{5}$$

where $N$ is the mass of surviving firms, which is equivalent to the number of varieties produced, and $H$ is the average cost of surviving firms.

***Agglomeration Effects.*** Combes et al. (2012) assume in their model that each worker supplies a single unit of labor. If the agglomeration effect is present, it is assumed that workers' productivity increases with the number of firms within a region. That is, effective labor supply by a single worker is $a(N), a' > 0, a'' < 0$ and $a'(0)=1$. On the other hand, if agglomeration of firms does not improve workers' productivity, for any value of $N$, $a(N)=1$. It is also assumed that if the agglomeration effect is present, it benefits workers across both the differentiated good and numeraire good sectors.

Given agglomeration effects, a firm of unit labor requirement $h$ hires labor such that $l(h)=Q(h)h/a(N)$.Taking logs of both sides, firm's log-productivity is obtained as $\phi=\ln(Q/l)=\ln[a(N)]-\ln(h)$. With $A\equiv\ln[a(N)]$ firm's log productivity is given as,

$$\phi = A-\ln(h). \tag{6}$$

Using the change of variable theorem the probability density function of firms' log-productivities is given as follows:

$$f(\phi)=\begin{Bmatrix} 0 & for\ \phi \le A-\ln\left(\bar{h}\right) \\ \dfrac{e^{A-\phi}g\left(e^{A-\phi}\right)}{G\left(\bar{h}\right)} & for\ \phi > A-\ln\left(\bar{h}\right) \end{Bmatrix}, \tag{7}$$

The numerator in (7) follows from use of (6) and the change of variables theorem, while the denominator $G\left(\bar{h}\right)$ takes care of the fact that firms with a unit labor requirement above $\bar{h}$ exit. Thus, the gain in productivity caused by external economies of scale due to the presence of *N* firms in the region as indicated by the term *A*, shifts the distribution of firms' log productivity to the right. This is referred to as the *agglomeration effect*.

***Selection Effect for Different Regions in Taiwan.*** In order to adopt the model for regional location of firms in any industry one more assumption is imposed. For any region, $r \in \{1\ldots R\}$, it is assumed that fixed sunk entry costs $s_r$, vary across regions based on the intensity of factor demands and provision of public policy incentives.

The free entry condition for any region *r* is given by Arimoto, Nakajima and Okazaki (2009) as follows:

$$\frac{C_r}{4\gamma}\int_o^{\bar{h}_r}(\bar{h}_r-h)^2 dG\left(h\right)=s_r\,. \tag{8}$$

We extend the regional model of Arimoto, Nakajima and Okazaki by adding an industry dimension to show that $\dfrac{\partial h_{rj}^d}{\partial s_{rj}}>0$, where subscript *j* refers to a particular industry. Therefore, for firms located in a region and operating in a particular industry, a unit increase in the entry cost raises the cut-off labor requirement. Hence, if the entry costs are lower either because of some policy incentive or due to less competition in factor demands, this lowers the cut-off labor requirement of surviving firms. Reversing the argument, the cut-off productivity level in a region goes up as more and more firms compete with each other for the available supply of factors of production such as land, labor etc. This higher cut-off level is observed as the left truncation of the log- total factor productivity distribution. This phenomenon is referred to as *competitive selection*. The proportion of firms in any sector $j$ that fail to survive product market competition in region $r$ is denoted by,

$$S_{rj} \equiv 1 - G\left(\bar{h}_{rj}\right). \tag{9}$$

Using the model outlined above, and also the implications noted in Arimoto, Nakajima and Okazaki , we test the following two hypotheses:

***Hypothesis 1*** **(*agglomeration*).** Agglomeration economies due to an increase in the number of firms in a region are likely to cause an increase in the mean log-total factor productivity distribution. This is observed as a comparative rightward shift in the log total factor productivity distribution for the firms located in that region.

***Hypothesis 2 (selection).*** The higher entry-cost in a region is likely to increase the cut-off unit labor requirement of surviving firms. This is observed through inter-regional comparison of the extent of left truncation of the log productivity distribution of firms

**TFP Estimation**

The econometric analysis conducted in this paper primarily hinges on estimation of TFP. The TFP distribution of firms located in any region is then predicted from the residual of their production functions. For this study firm-level TFP is calculated assuming that the technology for revenue generated is Cobb-Douglas in factors of production:

$$Y_{it} = A_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} , \tag{10}$$

where for firm *i* at time *t*, $Y_{it}$ is physical output, $K_{it}$ and $L_{it}$ are the inputs of capital and labor respectively. $A_{it}$ is the Hicks-neutral efficiency level of the firm and is unobservable to the researcher. Equation (10) can be written in logarithmic form as:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \varepsilon_{it} . \tag{11}$$

From (10) and (11) it can be deduced that $\ln\left(A_{it}\right) = \beta_0 + \varepsilon_{it}$ where $\beta_0$ is the mean efficiency level across firms over time and $\varepsilon_{it}$ is the deviation from the mean and can be further decomposed in an observable $v_{it}$ $v_{it}$ and unobservable component $u_{it}$ as :

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + v_{it} + u_{it} . \tag{12}$$

From (12), firm-level log-productivity is $\omega_{it} = \beta_0 + v_{it}$, and $u_{it}$ is the *i.i.d.* measurement error term.

**Possible Sources of Bias in TFP Estimation**

The productivity estimate from (12) may suffer from simultaneity bias, competitive selection bias and multi-product bias, each of which is discussed in detail as follows:

***Endogeneity.*** An OLS estimate of (12) is unbiased only if the inputs to production are exogenous from the firm's productive efficiency. However Marschak, Andrews and William (1944), long ago indicated that these inputs are not independently determined as firms themselves either observe or are able to predict their efficiency and hence determine the quantity of freely determined inputs accordingly. As the firm's productivity is not observed by the econometrician, its correlation with inputs causes simultaneity bias in the estimation (De Loecker 2007). The direction of the bias depends on the intensity of factor-use in the production process. Levinsohn and Petrin (2003) illustrate that for a two-input production function where labor is the only variable input and capital is quasi-fixed, the capital coefficient will be biased downward if a positive correlation exists between labor and capital.

***Selection Bias.*** Another issue raised by Olley and Pakes (1996) relates to the entry and exit of firms which was traditionally dealt with in TFP estimation by constructing a balanced panel, i.e., by omitting all firms that enter or exit over the sample period. However, several theoretical models such as that of Hopenhayn (1992) predict that the growth and exit of firms is motivated to a large extent by productivity differences at the firm level. Since low productivity firms have a stronger tendency to exit than their more productive counterparts, omitting all firms subject to entry or exit is likely to bias the results. This was also confirmed empirically by Fariñas and Ruano (2005) for Spanish manufacturing firms. If firms have some knowledge about their productivity level $\omega_{it}$ prior to their exit, this will generate correlation between $\omega_{it}$ and the fixed input capital (Ackerberg et al. 2007). This correlation has its origin in the fact that firms with a higher capital supply will, *ceteris paribus*, be able to withstand lower $\omega_{it}$ without exiting (Van Beveren 2010). In sum, the selection bias or “endogeneity of attrition”- problem will generate a negative correlation between $\omega_{it}$ and $k_{it}$, causing the capital coefficient to be biased downwards in a balanced sample.

## Econometric Analysis

### TFP Estimation Techniques

In this section, the techniques used in this paper for estimation of unbiased and consistent production function coefficients are described. As noted above, OLS estimates are likely to yield biased values of the coefficients. To avoid this, multiple methods are used to ensure robustness of the results. These include two stage least squares with instrumental variables (IV), the semi-parametric Olley and Pakes (1996) and the Levinsohn and Petrin (2003) techniques. Fixed effects estimation is not used as it depends on the strong assumption that productivity of firms is time-invariant. Also, as noted by Wooldridge (2009), the fixed effects estimator assumes strict exogeneity of the inputs which is not very likely and implies that inputs are not affected by the firm's knowledge of productivity.

***Instrumental Variables.*** One method to achieve consistency of coefficients in the production function is through use of instrumental variables for the endogenous independent variables, i.e., the freely alterable inputs in the production function. Ackerberg et al. (2007) explain Greene's (2004) requirements for the selection of valid instruments for the estimation of total factor of productivity. First, the instruments need to be correlated with the endogenous regressors (factor inputs). Second, the instruments should not enter the production function directly. Third, the instruments should not be correlated with the error term (and hence with productivity). Besides input prices, demand shifters are sometimes used as instruments in the literature. Keeping in mind data availability, recent values of county-level wages and population density are used as a measure of demand shifters to overcome the endogeneity of inputs problem.

***Olley and Pakes Estimation Methodology.*** Olley and Pakes were the first to introduce a semi-parametric estimation algorithm that takes both the selection and simultaneity problem directly into account. This estimator solves the simultaneity problem by using the firm's investment decision as proxy for unobserved productivity shocks.

Selection issues are addressed by incorporating an exit provision into the model. At the start of each period, each surviving firm decides whether to exit or to continue its operations. If it exits, it receives a particular sell-off value. If it continues, it chooses an appropriate level of variable inputs and investment. The firm is assumed to maximize the expected discounted value of net cash flows and investment and exit decisions will depend on the firm's perceptions about the distribution of future market structure.

Olley and Pakes' technique is based on three key assumptions. First, the only unobserved state variable is the firm's productivity which evolves as a first-order Markov process. Second, investment is monotonically related to productivity and hence during econometric analysis, non-negative values of the investment variable are required. This investment is shown as a function of capital and productivity, $i_{it} = i_t(k_{it}, \omega_{it})$. The monotonicity assumption allows its inversion as $\omega_{it} = h_t(k_{it}, \mathrm{i}_{it})$, so that productivity can be expressed in terms of capital and investment. Third, deflation on the basis of industry level prices implies that all firms face the same prices (Van Beveren 2010).

The Olley and Pakes technique proceeds in two stages. In the first-stage regression, using the relationship in (12) the free input variable(s) coefficients are derived. The second stage evaluates the temporal productivity level to compare it with the lower bound or the threshold. Using coefficients from the first stage and the survival probability and by applying a non-linear least squares method, the coefficient on the capital variable is estimated. Although Olley and

Pakes' technique is robust to simultaneity and selectivity problems, the empirical estimation using it may return unreliable results if either the investment variable has non-positive values or there are no firms exiting the market. An alternative is to use the Levinsohn and Petrin method which takes care of the simultaneity problem by using intermediate inputs as proxy for productivity instead of investment.

## Empirical Analysis

The main objective of this paper is to distinguish how agglomeration and competitive selection, affect productivity distribution of firms located in a science park compared to firms located elsewhere. To obtain reliable assessments of these effects, the first step is to generate statistically unbiased and consistent TFP estimates for firms.

### Data

To determine the agglomeration and selection effects on firms' TFP in Taiwan, firm data disaggregated at the county-industry (3-digit NAICS) level is used from the ISI Emerging Markets Information Services (EMIS). The data for the years 2009 to 2011 are in an unbalanced panel form. The dataset provides information about the physical location, industry, operational status of the firm, and its listing and trading status on the stock market. For each firm there is information available about financial indicators as shown in balance sheet and income statements, which include non-current assets, total employment and sales revenues. Indicators such as return on capital, profitability, liquidity and growth trend ratios are also part of the dataset.

The main dataset is supplemented with county-level income and industry price data available at the website of the National Statistics Office (DGBAS), Taiwan. These data are used for two purposes. The industry level price deflators are used to deflate the revenue figures. The county-level population and wages historical data are used as instrumental variables. The data up to a two-period lag are used to gauge local demand trends, while deep-lagged data are used to control selectivity bias of firms in high productivity regions.

The raw data containing 4662 observations are cleaned via a two-step process. In the first step, revenue figures are deflated using industry-level prices for the year 1996. Then using box plots, the data observations are examined for outliers and the entities with top and bottom one percent TFPs are removed in order to avoid their influence on the regression results. The remaining 4655 observations constitute the final dataset. Table 1 reports the summary statistics of the mean and standard deviation of inputs and output variables used in the Cobb-Douglas production function.

The county-wise location of select industries is shown in table 2. The dataset gives the 3-digit NAICS for all the firms, although it also provides 4-digit classification for a subset of these enterprises. This helps in detailed TFP analysis while segregating the firms in terms of the technology-intensity of their production. For NAICS 325 there is a total of 310 observations of which 71 are in the pharmaceutical sector (NAICS 3254), and the rest are in basic chemical manufacturing (NAICS 3251). It can be seen that computer and electronics firms (NAICS 334) constitute half of the total number of observations. Here, from the total of 2150 observations, 389 are in semiconductor manufacturing (3344). Finally, the scientific and technical industry (NAICS 541) is examined, in which case there are around 20 observations belonging to the biotechnology industry. All of these industries have a presence in all three regions, namely science park counties, small cities and the large cities.

For the spatial analysis, large cities/counties in Taiwan are divided into three exclusive regions based on population density and location of science parks. These regions are either with above median population density (hereinafter large cities), or with below median population density (hereinafter small cities) or those housing science parks. Using this delineation, we estimate firm's TFP for each market while controlling for potential simultaneity and selectivity bias using the Olley and Pakes (1996) method. In figure 1 a visual indication is given of the correlation between employment density and productivity. It can be observed that productivity levels are highest in large regions followed by regions that have science parks.

As the focus of this paper is on separating agglomeration and selection effects, the geographical unit of estimation is the county/city market. This choice is made to capture spillover effects of large markets. In the case of science parks, particularly the Hsinchu Science Park, ever increasing demand has forced a greater area in the county being designated as the science park. Based on population and labor density statistics, Taipei County, New Taipei City, Keelung City and Chiayi City are classified as the large cities. The regions designated Science Park counties are Hsinchu County, Tainan City, Yunlin City and Kaohsiung City.

## Results

The estimates for coefficients of input factors in the Cobb Douglas production function using different methods are reported in table 3.

### TFP Estimates Using OLS and Instrumental Variables

The baseline TFP estimates are computed using OLS. To obtain an unbiased OLS estimate of (12) it is required that input variables should not be correlated with the productivity term. However as the firm can either observe or anticipate its productivity it may adjust the level of flexible input(s).The OLS estimates for coefficients of labor and capital inputs are reported in column (1) of table 3. To test the reliability of the OLS estimates, the Durbin-Wu-Hausman test of endogeneity is performed. The small p-value indicates that the estimates are not reliable.

To avoid simultaneity bias we use 2SLS/IV with the return on capital and return on equity as instruments for the inputs in the production. The over-identification test indicates that the instruments are not correlated with the residual term. However, while the estimates as shown in column 2 of table 3 overcome the simultaneity bias, they still do not take care of the selectivity bias and hence the results are likely to be biased

**Olley-Pakes Method**

Keeping in mind the shortcomings of the techniques used above, TFP was predicted using the method proposed by Olley and Pakes (1996). The standard errors of all Olley-Pakes' estimation routines are bootstrapped using 200 replications to derive appropriate standard errors. From the results log-TFP distributions were drawn for each market. Here the return on capital is used as proxy for investments made by the firm along with control variables such as the number of employees to control for size. Using the TFP estimates from the Olley and Pakes method, the summary statistics for each of the regions are examined, as detailed in table 4. It is evident that at the aggregate level, large cities have the highest mean TFP value followed by that of science parks.

**Robustness Check for TFP Estimation**

As a robustness check, we obtain three additional productivity estimates namely: labor productivity, the TFP estimates from the Levinsohn and Petrin (2003) method and the Ackerberg et al. (2006) method. Ackerberg et al. have pointed out the likelihood of collinearity between inputs in both the Olley and Pakes, and the Levinsohn and Petrin approaches. These authors consider that labor may not be chosen independently, but is rather a function of capital and productivity. To avoid this eventuality, Wooldridge (2009) argues that coefficients for flexible and quasi-fixed inputs can be jointly estimated by GMM in a one-step procedure. As our dataset does not provide information on either labor skills or education levels, we use dynamic panel estimation to control for the unobserved worker heterogeneity in TFP estimation. These alternative productivity measures confirm that our results remain consistent. However, in the case of labor productivity, overestimation in the size and significance of agglomeration variables are observed.

**Agglomeration and Selection-Controlling Self Selection Bias**

The focus of this section of the paper is on firms falling under NAICS 334. This is the dominant industry in Taiwan and for which a sufficient number of observations are available for county markets in each of the three regions. In order to identify the impact of agglomeration, two variables namely localization and urbanization economies are used to capture the effect of specialization and diversification of economic activity respectively. Following Henderson, Kuncoro and Matthew *(*1995), localization *L* is quantified as the regional employment share (*E*) of the specific industry (defined at the three-digit NAICS level) in the manufacturing sector: Thus localization for industry *j*- in region *r*, at time *t* is given as, $L= [E_{jr}/E_r]/[E_j/E_{tot}]$. As in Mameli, Faggian and McCann (2014), urbanization is measured using the Herfindahl-Hirschman Index which is computed as $\sum_j s_{jrt}^2$ where $s_{jrt}$ is the employment share of two digit manufacturing industry *j* (except the industry under consideration) in region *r*, at time *t.*

Also, as noted earlier, agglomeration may entail diseconomies, for example, through higher land rents that trigger a selection process. As information on industrial land prices is not available, population density is used instead as a proxy, following a number of authors including Guimarães, Figueiredo and Woodward (2000). It can be argued that population density may in fact capture demand-side agglomeration economies, that is, firms locating near their potential markets. However, given the dominance of exporting firms in the industry under consideration, the relevant market for these firms is not the local Taiwanese market, hence it is reasonable to assume that population density does not capture market-size effects.

At this point the main econometrics-related concern that still needs to be addressed before conducting agglomeration analysis is the likelihood of self-selection of heterogeneous firms in markets with specific characteristics. While using the Olley and Pakes technique for estimation of production functions, the survival-based selection of firms is already taken care of. The focus is now on firm's self-selection as pointed out by Baldwin and Okubo (2006), where surviving firms sort into different regions depending on their productivity and regional characteristics, i.e., high productivity firms may concentrate in a region with a large market. To avoid confusion between these two separate phenomena, the former is referred to as competitive selection and the latter as self-selection. This self-selection, also referred to as sorting by high-productivity firms producing substitutable product varieties, is motivated by the potential of earning higher profits from a larger market. Hence, as a consequence of this, the higher productivity level observed in large markets may be attributed to the self-selection behavior.

**Identification of Firm Self-Selection**

In order to estimate the selection effect, the Heckman (1979) two-step estimator for selection models is used here. Such models are common in micro-econometric studies, in the estimation of wage equations or consumer expenditure. The statistical significance of the coefficient of the inverse Mill's ratio indicates if there is any selection bias. To identify the process through which high-productivity firms sort into science parks and large cities, a selection and an outcome equation are used. Considering firm's sorting in science parks the relevant selection equation is as follows:

$$z_{it}^{*} = \alpha_0 + \alpha C_{it} + \varepsilon_{it}, \tag{13}$$

where $z_{it}^{*}$ is $z_{it} = 1$. $z_{it}$ is the dependent variable of the selection equation which is binary in nature and $C_{it}$ are the self-selection choice variables. The choice variables include lagged county level wages, lagged county population density, and firms' return on capital and return on equity. The outcome equation is given as follows:

$$S_{rt} = \beta_0 + \beta_a A_{rt} + \beta_c X_{rt} + \upsilon_{rt} \text{ if } z_{it}^{*} > 0, \tag{14}$$

where $S_{rt}$ is the region-time specific distributional measure, i.e., the mean (and the median as a robustness check) and the tenth percentile of the log-TFP distribution. $A_{rt}$ are industry-specific agglomeration variables for region $r$ at time $t$, $X_{rt}$ are the region-time specific control variables and $\upsilon_{rt}$ is the error term. The expected signs of localization and urbanization as defined here are positive and negative respectively, whenever such effects result in agglomeration economies. The following two dependent variables in (14) are generated following Syverson (2004):

1. Mean (median)-to capture the productivity gains due to agglomeration causing rightward shift of the log-TFP distribution

2. 10th percentile - to capture the selection effect as higher competition causes left truncation of the log-TFP distribution

Using these as dependent variables, their significance is evaluated with respect to the agglomeration variables (localization and urbanization) and selection (population density). These steps are repeated for big cities and science parks and for different industries at the 3-digit NAICS level. To account for the panel structure of the data, the data are transformed into a form suitable for fixed effects analysis, followed by application of the Heckman procedure.

The results reported in table 5 indicate that the magnitude of competitive selection is very small compared to agglomeration variables. Another interesting result is that firms located in science park benefit from specialization and firms located in large cities benefit form diversification of economic activity. Interestingly diversification and specialization have negative impact on firm productivity in case of science parks and large cities respectively.

**Robustness Check for Firm Self- Selection to Science Parks/Large Cities**

Apart from using the Heckman (1979) model we use regional dummy variables and instrumental variables to confirm and control for self-selection. For the latter approach the possibility of reverse causality can be controlled through the use of instrumental variables correlated with firm's productivity but not correlated with agglomeration economies.

For this study the following instruments are used: deep lag of population density (log of 1950 number), log of return on equity and capital are used as instrumental variables. The validity of the instruments is established using the Sargan (1958) test of over-identification restrictions. The results confirm that self-selection is present and may bias the estimates of agglomeration and competitive-selection.

## Discussion of Results

The results of this research can be divided into four parts. First, non-parametric comparisons are made of the log TFP distribution for aggregate manufacturing sector firms located in the three identified regions. This analysis helps in understanding the extent to which policy intervention is able to act as a productivity shock and disturb the equilibrium in which more productive firms are located in large cities. Consistent with the literature, for the overall manufacturing sector the highest mean log-TFPs are for firms located in large cities and the lowest mean log-TFP is for firms located in small cities. Additionally we find that the mean log-TFP for science park firms falls in between the two as shown in figure 2. This indicates that factors driving productivity-gain in large cities are not affected by policy incentives elsewhere, although establishment of science parks does lead to increased regional growth.

Second, for detailed analysis we disaggregate the manufacturing sector into firms classified under NAICS 325, 334 and 541. The TFP distributions for the selected industries are shown in figures 3, 4 and 5 respectively. The sector-based inter-regional comparison of manufacturing firms indicates that computer and electronics firms located in large cities are the ones with the highest level of mean log-TFP, followed by those in science parks and then by those located in small cities. However, as the technology-intensity of the production process is lowered, i.e., the chemicals industry, firms in science parks lose their comparative advantage and end up being those with the lowest mean log-TFPs. We extend this analysis to incorporate inter-regional comparison for service sector firms, specifically the scientific and technical services which employ a very high proportion of technology-oriented workers. The results show that firms located within science parks have higher mean log-TFP values in comparison to those located in large cities.

Third, intra-regional comparisons are made for manufacturing firms within science parks. Here sample comparisons are conducted for two sectors and we find that manufacturing firms in computer and electronics industry have the highest mean log-TFP whereas those in basic chemical manufacturing have much lower values. This suggests that within science parks, there is a direct relation between firm-level productivity and technology-intensity of the production process.

Fourth, the results are analyzed to establish the impact of agglomeration and competitive selection. The Heckman selection model confirms that firms do self-select in regions of high productivity and hence it is important to control for sorting behavior in the model. However, not all the productivity gains can be attributed to this phenomenon. Even after controlling for self-selection, it is found that the agglomeration variables namely, localization and urbanization, are statistically significant. The regression results indicate that selection due to competition is also significant. As shown in table 5, the coefficients for competitive selection are much smaller than those for the agglomeration variables. Moreover the mean of the regional log-TFP distribution for firms in science parks indicates that they do benefit from specialization of economic activity.

## Conclusion

Science parks have traditionally been established with the purpose of enhancing comparative advantage by supporting a regional innovation system through a place-based policy instrument (Cooke 2001). However, there are hardly any studies that evaluate the use of such policy measures via methods of regional economic analysis. This paper is an attempt to examine the impact of science parks on regional productivity levels and to establish the determinants of observed differences. As such this approach proposes uniform evaluation criteria for policy-driven industrial clusters.

In the case of Taiwan, firms located in science parks have productivity distributions proportional to the technology-intensity of their sector. Also, even after controlling for self-selection bias, firms located in science parks benefit from agglomeration economies arising out of specialization of economic activity. The results also confirm that self-selection in science parks by high productivity firms is empirically established. The elasticity of competition-based selection is much less in science parks compared to large cities. Thus science park incentives insulate firms from the competition they might face in open markets. This finding is also substantiated by the survival of low-technology chemicals manufacturing firms located in a science park with productivity distributions that lag even that of small cities.

The regional productivity distributions show that the relative intensity of economic and or industrial activity causes right shift and greater dispersion. Also firms in large cities face competition analogous to being in an open economy. Thus firms below a certain threshold level of productivity cannot survive there. All these findings confirm that the impact of industrial clusters such as science parks is not homogenous across firms and the resultant productivity shock is weak at the aggregate level.

The policy implication that arises from this study is based on robust theoretical foundations. An attempt has been made tease out the impact of policy on firm's productivity; the indicator of its heterogeneity. The interplay of selection and agglomeration with probability of sorting makes the analysis a daunting task. However, controlling for various observable and unobservable factors, the study suggests that incentives such as science parks do contribute in productivity improvements of the firms. The level to which benefits of science parks are taken advantage of by firms as depicted in their productivity depends on the underlying technology of their production process.

**Notes**

1. The dataset used in the study limits the analysis as it does not provide information about intermediate inputs or energy requirements of the firm. Also the dataset provides no details about the product mix of the firms. However, detailed examination of annual reports for firms operating in NAICS 334 show that technology, inputs and sales are broadly homogeneous.

## References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes. 2007. "Econometric Tools for Analyzing Market Outcomes". In *Handbook of Econometrics*, edited by J. Heckman and E. Leamer, 4171-4276. Amsterdam: North Holland.

Arimoto Y., K. Nakajima K, and T. Okazaki. 2009. "Agglomeration or Selection? The Case of the Japanese Silk-Reeling Industry, 1909-1916". RIETI Discussion Paper No. 10-E-003.

Association of University Research Parks (AURP). 1998. *Worldwide Research and Science Park Directory 1998*. New York: BPI Communications.

Audretsch, D. 1998. "Agglomeration and the Location of Innovative Activity". *Oxford Review of Economic Policy* 14:18-29.

Baldwin, R., and T. Okubo. 2006. "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting". *Journal of Economic Geography* 6: 323–346.

Cainelli, G., Fracasso, A., and Vittucci Marzetti G. 2014. "Spatial Agglomeration and Productivity in Italy: A Panel Smooth Transition Regression Approach". *Papers in Regional Science*, DOI: 10.1111/pirs.12103

Ciccone, A., and R. Hall. 1996. "Productivity and the Density of Economic Activity". *American Economic Review* 86: 54-70.

Combes, P., G. Duranton, L. Gobillon, D. Puga, and S. Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection". *Econometrica* 80: 2543-2594.

Cooke, P. 2001. "From Technopoles to Regional Innovation Systems: The Evolution of Localized Technology Development Policy". *Canadian Journal of Regional Science* 24: 21-40.

Cooke, P. 2002. "Biotechnology Clusters as Regional, Sectoral Innovation Systems". *International Regional Science Review* 25: 8-37.

Dąbrowska, J. 2011. "Measuring the Success of Science Parks: Performance Monitoring and Evaluation". Paper prepared for XXVIII IASP Conference on Science and Technology Parks. Copenhagen.

De Loecker, J. 2007. "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity". NBER Working Paper No. 13155.

Emerging Markets Information Service (EMIS). New York: Internet Securities accessed via Ohio State University Library.

Fariñas, J., and S. Ruano. 2005. "Firm Productivity, Heterogeneity, Sunk Costs and Market Selection". *International Journal of Industrial Organization* 23: 505–534.

Fogarty, M.S., and G.A. Garofalo. 1978. "Urban Spatial Structure and Productivity Growth in the Manufacturing Sector of Cities". *Journal of Urban Economics* 23: 60-70.

Fujita, M., and H. Ogawa. 1982. "Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations". *Regional Science and Urban Economics* 12: 161–196.

Gabe T. 2004. "Establishment Growth in Small Cities and Towns". *International Regional Science Review* 27: 164-186.

Greene, W. 2004. "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model". *Journal of Econometrics* 126: 269-303.

Guimarães, P., and O. Figueiredo, and D. Woodward D. 2000. "Agglomeration and the Location of Foreign Direct Investment in Portugal". *Journal of Urban Economics* 47: 115-135.

Hausman, J. 1978. “Specification Tests in Econometrics”. *Econometrica* 46: 1251-1271.
Hecker, D.E. 2005. “High-Technology Employment: A NAICS-Based Update”. *Monthly Labor Review* 128: 57–72.
Heckman, J.J. 1979. “Sample Selection Bias as a Specification Error”. *Econometrica* 47: 153–161.
Henderson, J., A. Kuncoro, and M. Matthew. 1995. “Industrial Development in Cities”. *Journal of Political Economy* 103: 1067-1090.
Hopenhayn, H. 1992. “Entry, Exit, and Firm Dynamics in Long Run Equilibrium”. *Econometrica* 60:1127-1150.
Jaffe, A. 1989. “Real Effects of Academic Research”. *American Economic Review* 79: 957-970.
Jaffe, A., M. Trajtenberg, and R. Henderson. 1993. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations”. *Quarterly Journal of Economics* 108: 577-598.
Levinsohn, J., and A. Petrin. 2003. “Estimating Production Functions Using Inputs to Control for Unobservables”. *Review of Economic Studies* 70: 317-342.
Lien, W.J., J.C. Wang, S.W. Wang, and F.H. Tsai. 2007. “The Economic Impact of Taiwan’s Investment Tax Credits and its Direction of Adjustment”. Taipei: Chung-Hua Institution for Economic Research.
Lucas Jr., R.E., and E. Rossi-Hansberg. 2002. “On the Internal Structure of Cities”. *Econometrica* 70: 1445–1476.
Luger, M. I., and Goldstein, H. A. 1991. *Technology in the Garden*,” Chapel Hill: UNC Press
Mameli, F., Faggian, A., and McCann, P. 2014. "Estimation of Local Employment Growth: Do Sectoral Aggregation and Industry Definition Matter?" *Regional Studies* 48: 1813-1828
Marschak, J., J. Andrews, and H. William. 1944. “Random Simultaneous Equations and the Theory of Production”. *Econometrica* 12: 143–205.
Marshall, A. 1890. *Principles of Economics*. London: Macmillan.
Melitz, M. 2003. “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”. *Econometrica*, 71: 1695-1725.
Melitz, M., and G. Ottaviano. 2008. “Market Size, Trade and Productivity”. *Review of Economic Studies* 75: 295–316.
Monck, C. 2010. “Performance Monitoring and Evaluation”. UK Science Park Association (UKSPA) Conference Proceedings. Saffron Walden: UKSPA.
National Statistics. 2010. Taiwan: Directorate-General of Budget, Accounting and Statistics Executive Yuan (DGBAS).
Olley, G., and A. Pakes. 1996. “The Dynamics of Productivity in the Telecommunication Equipment Industry”. *Econometrica* 64: 1263–1297.
Parr, J.B. 2002. “Missing Elements in the Analysis of Agglomeration Economies”. *International Regional Science Review*, 25: 151-168.
Phan, P., D. Siegel, and M. Wright. 2005. “Science Parks and Incubators: Observations, Synthesis and Future Research”. *Journal of Business Venturing* 20: 165-182.
Rosenthal, S., and W. Strange. 2004. “Evidence on the Nature and Sources of Agglomeration Economies”. In *Handbook of Regional and Urban Economics* edited by V. Henderson and J.F. Thisse, 2119-2171. Amsterdam: North Holland.
Rothaermel, F., and M. Thursby. 2005a. “Incubator Firm Failure or Graduation? The Role of University Linkages”. *Research Policy* 34: 1076-1090.

_____________________________2005b. “University-Incubator Firm Knowledge Flows: Assessing their Impact on Incubator Firm Performance”. *Research Policy* 34: 302-320.

Sargan, J.D. 1958. “The Estimation of Economic Relationships Using Instrumental Variables”. *Econometrica* 26: 393-415.

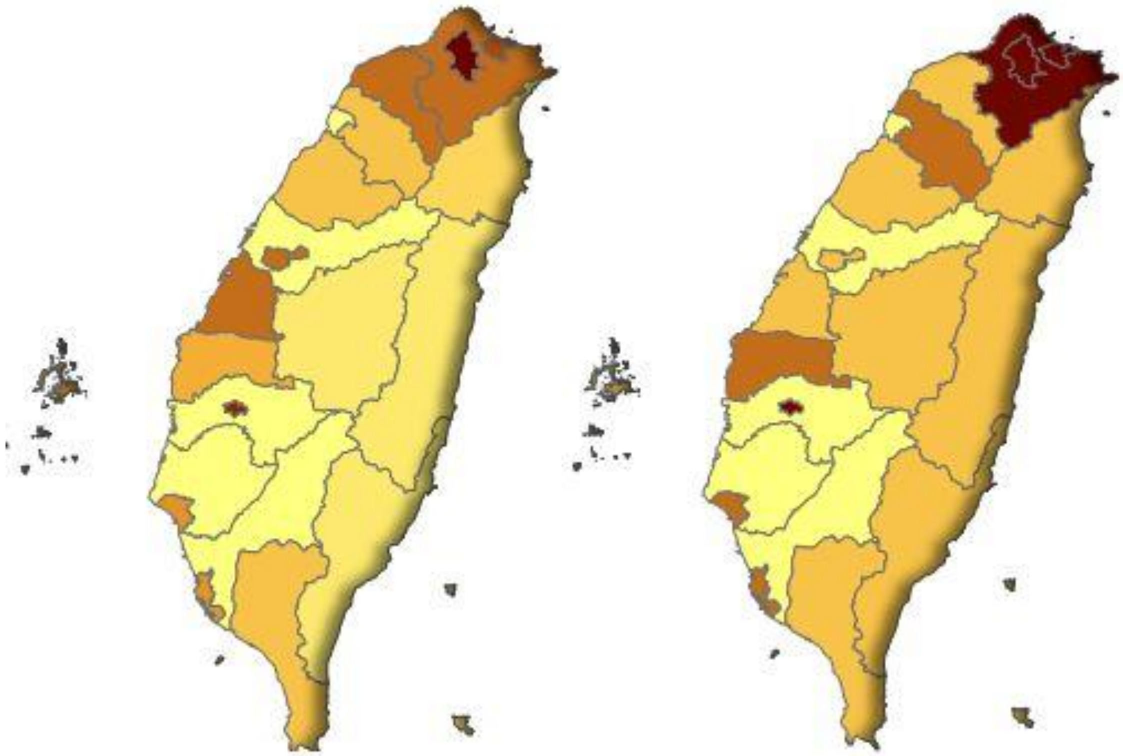
Syverson, C. 2004. “Market Structure and Productivity: A Concrete Example”. *Journal of Political Economy* 112: 1181–1222.

Tabuchi, T. 1986. "Urban Agglomeration, Capital Augmenting Technology, and Labor Market Equilibrium". *Journal of Urban Economics* 20: 211-228.

United Nations Educational, Scientific and Cultural Organization (UNESCO). 1994 <http://www.unesco.org/pao/s-parks/what.htm>.

Van Beveren, I. 2007. “Total Factor Productivity Estimation: A Practical Review”. LICOS Discussion Paper No. 182/2007.

Wooldridge, J.M. 2009. “On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Un-observables”. *Economics Letters* 104: 112–114.



(a) County-market Mean TFP (b) County–level Population Density

**Figure 1**. Population density and TFP (County-level for Taiwan). Map based on data used in the study. The trend is increasing from light (yellow) to dark (brown)

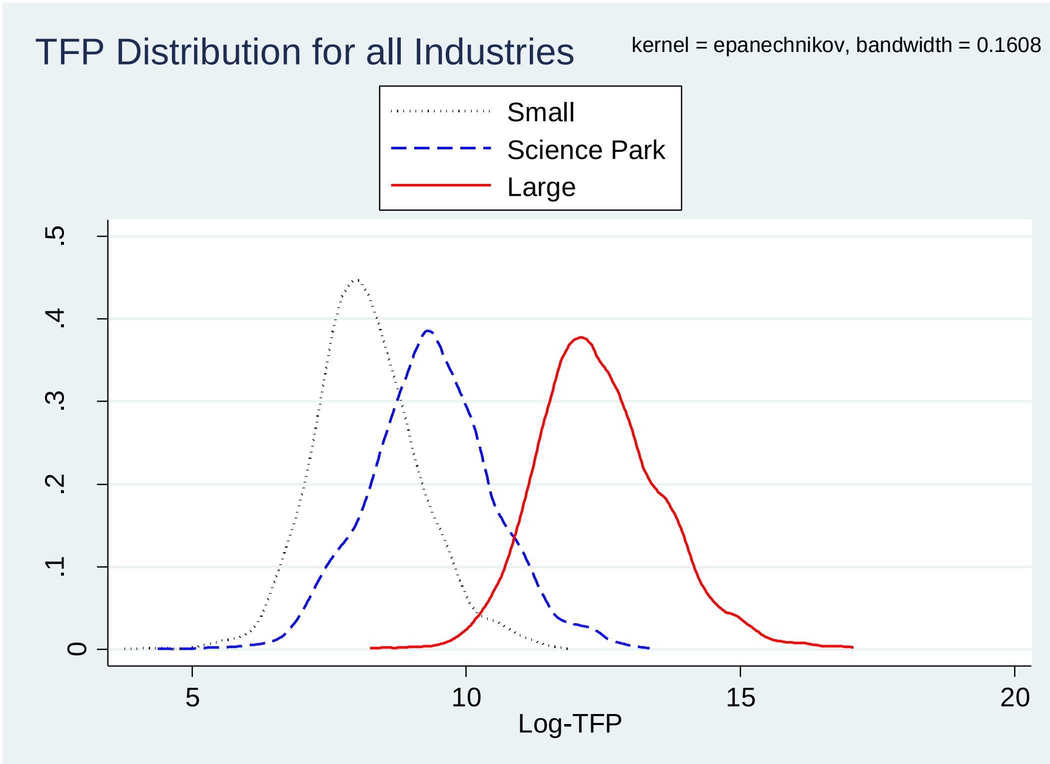



**Figure 2.** Kernel density plots for the three regions

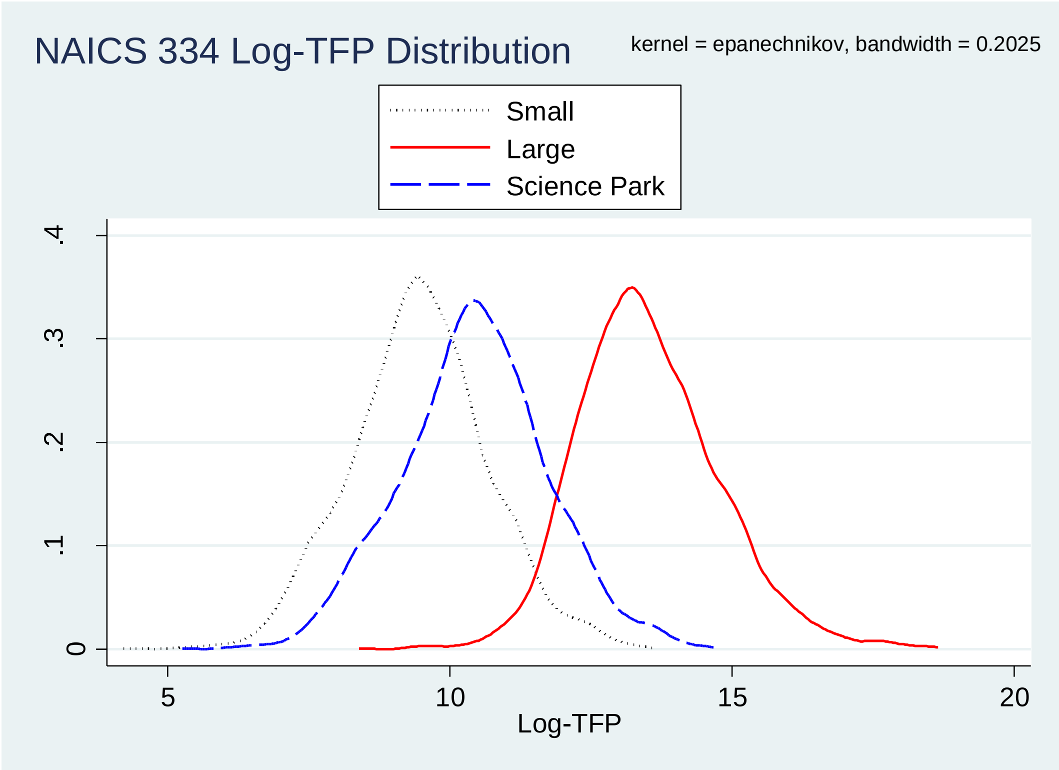


**Figure 3.** Kernel density plots for Computer and Electronics Firms for the three regions

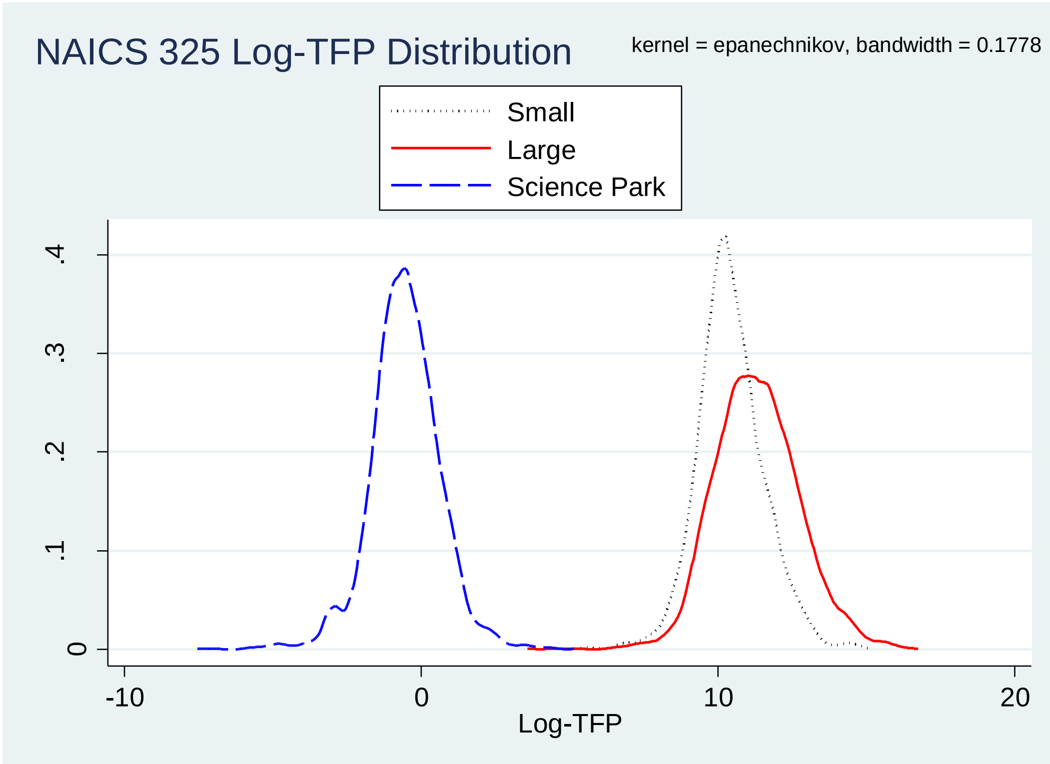



**Figure 4.** Kernel density plots for Chemical Manufacturing Firms for the three regions

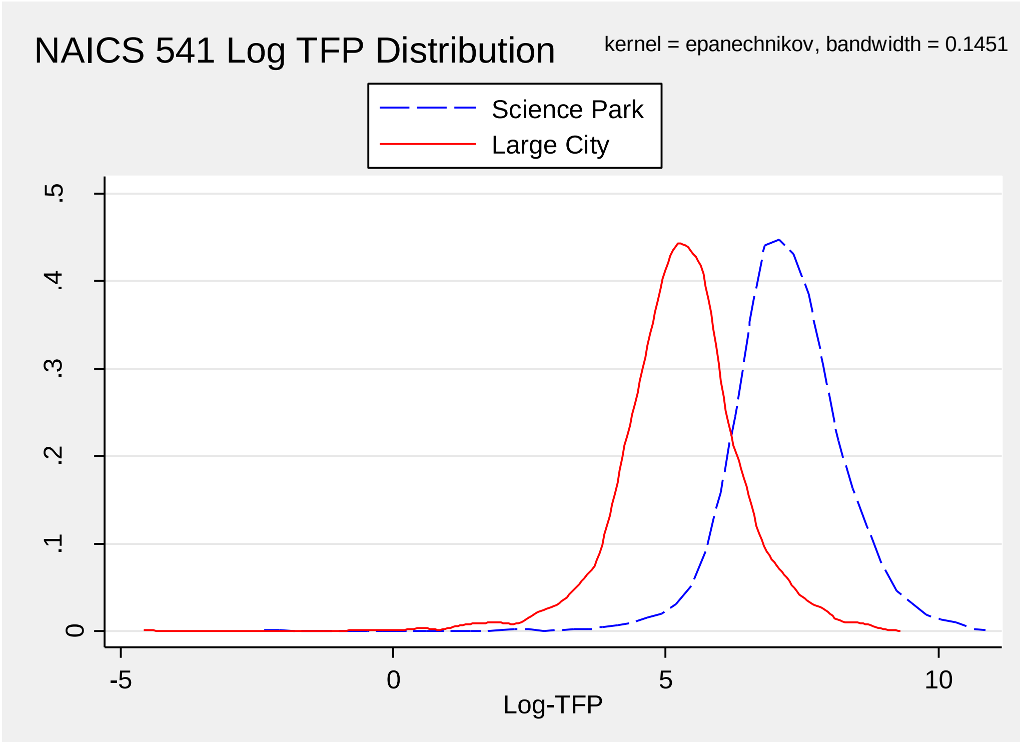



**Figure 5.** Kernel density plots for Scientific and Technical Services Firms

**Table 1.** Summary Stats-I

| Variable | Mean | Std. Dev. | Max |
|---|---|---|---|
| Capital | 9,031,678 | 41,300,000 | 1,520,000,000 |
| Revenue | 9,724,772 | 44,300,000 | 845,000,000 |
| Labor | 660 | 1,770 | 33,669 |

**Table 2.** Summary Stats-II

| | NAICS | | |
|---|---|---|---|
| | **325** | **334** | **541** |
| State/County | | | |
| Changhua County | 6 | 14 | 2 |
| Chiayi City | 3 | 3 | 1 |
| Hsinchu County | 20 | 474 | 11 |
| Hualien County | 0 | 0 | 1 |
| Kaohsiung City | 29 | 82 | 9 |
| Keelung City | 0 | 6 | 1 |
| Miaoli County | 12 | 40 | 1 |
| Nantou County | 6 | 4 | 1 |
| New Taipei City | 23 | 600 | 22 |
| Penghu County | 0 | 0 | 1 |
| Pingtung County | 0 | 6 | 1 |
| Taichung City | 18 | 78 | 6 |
| Tainan City | 28 | 68 | 2 |
| Taipei | 138 | 430 | 55 |
| Taitung County | 0 | 0 | 0 |
| Taoyuan County | 23 | 339 | 10 |
| Yilan County | 4 | 3 | 0 |
| Yunlin County | 0 | 3 | 1 |
| Total | **310** | **2,150** | **125** |

**Table 3.** Production Function Coefficients

| Model/Variables | OLS | IV/2SLS | OP |
|---|---|---|---|
| | **(1)** | **(2)** | **(3)** |
| **Capital** | 0.37*** | 0.56*** | 0.29 ** |
| **Labor** | 0.56*** | 0.21 *** | 0.47** |
| **R-squared** | 0.62 | 0.57 | |
| **Sargan Test(p-value)** | - | 0.77 | |
| **DWH (p-value)** | 0.0002 | 0.265 | |

Notes: * Significant at 10% level, **significant at 5% level, and *** significant at 1% level

**Table 4.** Region-wise Log-TFP Distribution Stats
(AM: Large City, BM: Small City, SP: Science Park)

| Stats | BM | SP | AM |
|---|---|---|---|
| N | 840 | 1427 | 2388 |
| mean | 4.106923 | 8.32283 | 11.76685 |
| sum | 13409.1 | 27174.04 | 38418.78 |
| max | 8.708421 | 12.10286 | 17.08633 |
| min | -2.43337 | 1.005013 | 4.605112 |
| range | 11.14179 | 11.09784 | 12.48122 |
| sd | 1.043982 | 1.096338 | 1.116861 |
| variance | 1.089898 | 1.201957 | 1.247379 |
| skewness | 0.098394 | 0.024502 | 0.384866 |
| kurtosis | 4.706174 | 4.342106 | 4.483479 |
| p50 | 4.063416 | 8.291471 | 11.66475 |
| p5 | 2.549411 | 6.604828 | 10.17849 |
| p10 | 2.932698 | 6.96137 | 10.49661 |
| p25 | 3.461043 | 7.634048 | 11.01647 |
| p50 | 4.063416 | 8.291471 | 11.66475 |
| p75 | 4.691591 | 8.983652 | 12.43476 |
| p90 | 5.439116 | 9.73185 | 13.19089 |
| p95 | 5.90151 | 10.1574 | 13.67761 |
| p99 | 6.822562 | 11.15639 | 14.81686 |
| IQR | 1.230548 | 1.349604 | 1.418283 |

**Table 5.** Agglomeration and Selection in Science Park

| | MED | | 10-TILE |
|---|---|---|---|
| | **LOC** | **URB** | **Selection** |
| **SP** | 0.161*** | 0.571*** | 6.08exp(-06)*** |
| | (0.077) | (0.106) | (9.61exp(-07)) |
| **AM** | -0.71*** | -1.251*** | 0.0002*** |
| | (0.268) | (0.351) | (0.00001) |

Notes: * significant at 10% level, **significant at 5% level, and
*** significant at 1% level, clustered standard errors in parenthesis