The rise of the 'emerging economies':
Towards functioning agricultural markets and trade relations?

Selection, agglomeration and firm productivity in Taiwan:
What impact on the high-tech sector?

Syed Hasan*, H. Allen Klaiber** and Ian Sheldon***

*AED Economics, Ohio State University, Columbus, OH 43210, US, hasan.47@osu.edu
**AED Economics, Ohio State University, Columbus, OH 43210, US, klaiber.16@osu.edu
*** AED Economics, Ohio State University, Columbus, OH 43210, US, sheldon.1@osu.edu

Abstract
Agglomeration spill-over benefits are a major driver of policy creating industrial and science parks across the world. However, agglomeration benefits can be offset by competition arising out of the spatial proximity of firms. In this paper we examine the impact of agglomeration and selection on Taiwanese firms' total factor productivity (TFP) distribution and show that agglomeration causes the log-TFP distribution to have a rightward mean-shift, but that this effect is heterogeneous across firm types. Firms located in science parks and classified in the high-tech sector, which includes biotechnology firms, have the highest log-TFP. Firms other than these, located in science parks and operating in the high-tech sector, have a productivity distribution lying between those located in densely populated and thinly populated regions. However, for low/medium-tech industries such as chemical manufacturing, the productivity distribution for firms located in science parks lags both that of large as well as small cities. Policy aimed at offering science park incentives should be industry-specific to generate positive productivity improvements, otherwise such incentives may simply be used as protection by inefficient firms.

Keywords: Agglomeration, Selection, Science Parks

JEL classification: D24 R12

1 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 to better understand the mechanisms underlying any productivity gains. Firm-level productivity is a key component of the heterogeneous firms' literature that draws on MELITZ's (2003) model of describing how the productivity of firms determines their survival in domestic and foreign markets. However, MELITZ and the subsequent literature does not say much about what determines firms' actual productivity, leaving it to a random draw once a firm has entered a market.

While productivity is a key metric of evaluating firm success, understanding how spatial policies designating science and technology parks affect productivity requires additional analysis. The regional economics literature has shown that firms located in large cities are often more productive than those located elsewhere (ROSENTHAL and STRANGE, 2004). For industrial clusters, a positive association between regional plant density and their productivity has been empirically confirmed giving support for policies encouraging firm clustering (CICCONE and HALL 1996).

However, recent theoretical developments in spatial economics, with heterogeneous firms indicates that, high productivity observed in large cities or industrial clusters may also result from competition-based selection (COMBES et al. 2012). Thus as competition increases, firms above a certain productivity threshold are likely to survive while firms below this threshold exit the market. In addition, BALDWIN and OKUBO (2006) using a MELITZ-type structure show that high productivity firms may self-select into large cities to avail themselves of the benefits of large markets.

Over the past three decades, Taiwan's government has undertaken several measures to encourage firms to spend more on innovative activity and to promote their technological capability. In this regard, the most significant policy measure concerns investment tax credits for R&D (LIEN et al. 2007). Figures show that the amount of R&D expenditure increased steadily from NT$94.828 billion in 1992 to NT$280.980 billion in 2005. Correspondingly, the amount of R&D tax credits has increased more than 10-fold from NT$1.529 billion in 1992 to NT$16.318 billion in 2005. However, in the wake of recent fiscal difficulties and revenue shortfalls the policy tool of R&D tax credits has been widely criticized as being beneficial only for a few large firms rather than for the remaining 97 percent of small and medium-sized firms in Taiwan. More importantly, the policies of R&D tax credits favouring specific industries or firms may result in tax base erosion and destroys the fairness of the taxation system. It is estimated that tax credits account for approximately one third of the NT$100 billion of total tax revenue loss for the Taiwanese government annually (LIEN et al.).

In the current paper we extend this line of research by using firm-level panel data for Taiwan spanning the years 2009-11 to simultaneously consider agglomeration and selection of firms in large cities and science parks. The analysis includes estimates for aggregate manufacturing and also specific industrial sectors segregated on the basis of technology-intensity of the production process. Importantly, we highlight firms in the high-tech sector, which includes those involved in biotechnology, and where high firm-level productivity may generate comparative advantage. The available empirical evidence points to the potential for biotechnology to contribute to agricultural productivity gains and food security in emerging

economies (CARTER et al. 2011). While Taiwan's biotech sector currently generates only NT$14 billion of exports, there is evidence that small, dedicated biotech companies that are R&D- intensive and operate primarily with venture capital, grants, initial public offerings and collaborative agreements can succeed in this sector (LAVOIE and SHELDON 2000). Through its promotion of innovative activities, the Industrial Development Bureau's long-term objective is for Taiwan to hold 3 percent of the world's biotechnology market.

In terms of spatial analysis, we divide counties in Taiwan into three exclusive categories based on population density. These regions include counties with above median population density (large), counties with below median population density (small) and counties housing science parks. Using this delineation, for each market we estimate firm's total factor productivity (TFP) while controlling for potential simultaneity and selectivity bias using the OLLEY and PAKES (1996) method. Figure 1 gives a visual indication of the correlation between employment density and productivity.

Figure 1: Population density and TFP (county-level for Taiwan)
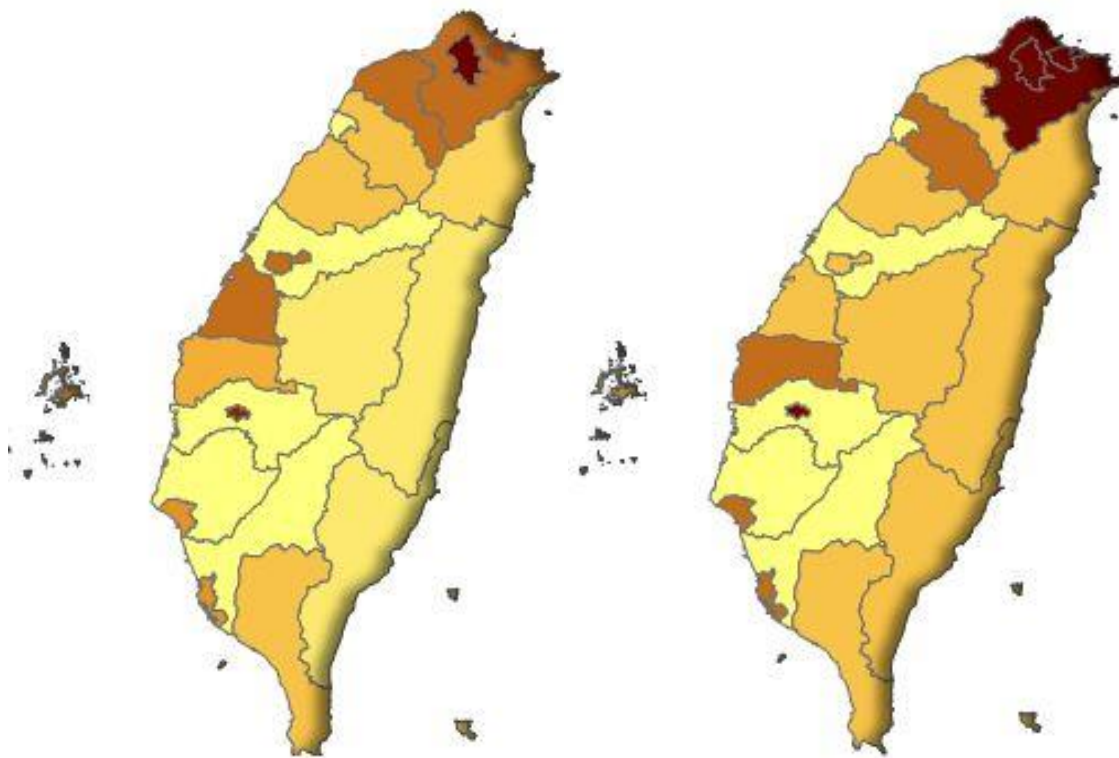
(a) County-market mean-TFP　　　　　　　　**(b) County-level population density**

Note: Map based on study data. The trend is increasing from light (yellow) to dark (brown).

Using estimates of firms' TFP, we conduct three analyses. First, we compare the regional productivity distributions of manufacturing and technical services firms in Taiwan and find that firms located in large cities have higher productivity levels compared to those located in science parks. Second we compare regional productivity distributions for narrowly defined industrial sectors (3-digit level NAICS) capturing the impact of benefits of locating in a science park. This analysis shows the surprising result that only firms using high-tech

production processes avail themselves of science park benefits and those using low-tech production have the lowest level of productivity. Finally we identify the impact of selection and agglomeration on firms' productivity based on their spatial location in science parks. We show that firms located in science parks only benefit from localization whereas those located in large cities benefit from both localization and urbanization. From a policy perspective, these results suggest that science parks do help in correcting innovation market failures but if extended to low-tech industries, they may turn out to be protection against market competition.

In this paper we identify that the efficiency in utilization of public incentives offered via science parks increases with the technology level of the industry. The research contribution of the paper is twofold. First, we add to the scarce literature that evaluates firms within science parks. Our main finding is that on average there is a positive relationship between the technology level of the industry and the total factor productivity of firms operating within that industry. Thus biotechnology firms, along with others operating in science parks, have the highest level of productivity. Second, we also supplement the more popular performance analysis method by comparing firms located in science parks with off-park firms. Here we find that at the aggregate level the TFP distribution of science park firms lags that of firms located in large cities but leads those located in small cities. The empirical evidence on the impact of science park intervention on innovative capability, survival rate, profitability and job creation is largely mixed and inconclusive (MONCK and PETERS 2009), providing little margin for policy recommendation. In contrast, the research methodology presented here is not restricted by any particular estimation model or specific park objectives. Instead it is based on a robust theoretical foundation and provides a minimum degree of homogeneity for similar evaluations.

2 LITERATURE REVIEW

Enterprises located in large cities are more productive. In the relevant literature there is a consensus on the positive relationship between productivity levels and regional density of labour, economic and industrial activity. Three main explanations have been presented for these observed phenomena. The first is agglomeration economies: economies external to firms arising out of sharing and spill-overs and ultimately causing increasing returns for the entire neighbourhood. 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 significant contribution by the same authors is the estimate that productivity increases by 3-8 percent if city size is doubled. External economies 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 literature explains productivity gains resulting from labour market pooling, input sharing, and knowledge spill-overs.

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 is based on the seminal work of MELITZ (2003), who introduced

product differentiation and international or interregional trade into the framework of industry dynamics of HOPENHAYN (1992). MELITZ and OTTAVIANO (2008) incorporated variable price–cost mark-ups into this framework and showed that larger markets attract more firms, which makes competition tougher. As firms cluster to gain agglomeration economies, the increased competition resulting from this clustering may reduce profits and thereby firm's willingness to locate in denser locations. As a result, low productivity firms may choose to avoid or exit densely populated regions such as cities or science parks. Consequently higher average productivity of firms and workers in larger cities may be the result of 'natural' selection of firms.

There is a 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 effects from selection effects. For this, they extend MELITZ and OTTAVIANO (2008) by introducing agglomeration economies in the manner of FUJITA and OGAWA (1982) and LUCAS and ROSSI-HANSBERG (2002), and develop a model that includes both selection and agglomeration effects. Thus under monopolistic competition with free entry, profits decline as the number of competitors increase in one location. This results in reduced survival for low-productivity firms. COMBES et al. then structurally parameterize the strength of selection and agglomeration, and estimate the strength of these two effects using two-digit industry-level data.

In contrast to their approach, in this paper we focus on a sample of specific industrial sectors (computer and electronics, chemicals, and scientific and technical services) rather than the manufacturing sector alone, so that we can control for sector-specific factors such as market conditions for supply of inputs and demand for output, and the form of production functions. Further, using SYVERSON's (2004) approach we use proxies for shift, dispersion and truncation in the log-TFP distributions across the regions in order 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 a lot on the estimation of a bias-free TFP distribution. The estimation method used follows OLLEY and PAKES (1996), whose technique is robust to two econometric concerns: simultaneity and selectivity bias. However, the proxy variable for free inputs in their 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 often adopted, which uses intermediate inputs to proxy for productivity shocks. Given the limitations of the dataset available for this study, we use return on capital as a proxy for investment while estimating TFP through the OLLEY and PAKES method. However, for the biotechnology sector, where the dataset does not report any exits from the market, we have used LEVINSOHN and PETRIN's technique to estimate the log-TFP.

Many definitions of science parks have been proposed, mostly by professional organizations (e.g., AURP, 1998 and 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

4

partnership that fosters knowledge flows—often between park firms and universities and among park firms—and contributes to regional economic growth and development. Empirical support for the agglomeration effects in a park is provided by AUDRETSCH (1998), JAFFE (1989), JAFFE et al. (1993), and ROTHAERMEL and THURSBY (2005a, 2005b). However, there are also some disadvantages associated with being in a park. When a park attracts many firms which then have access to the same technologies, those firms may expect greater competition in the use of those technologies. On the human capital supply side, there is skilled and specialised labour available from universities involved in parks in the form of graduate students and consulting faculty, although there is more competition for that pool of human capital.

3 TOTAL FACTOR PRODUCTIVITY ESTIMATION

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

$$Y_{it} = A_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} \tag{1}$$

where for firm i at time t, Y_{it} is physical output, K_{it} and L_{it} are the inputs of capital and labour and A_{it} is the Hicks-neutral efficiency level of the firm. A_{it} is unobservable to the researcher. Equation (1) can be written in logarithmic form as:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \varepsilon_{it} \tag{2}$$

From (1) and (2) we can deduce that $\ln(A_{it}) = \beta_0 + \varepsilon_{it}$ where β_0 is the mean efficiency level across firms over time and ε_{it} is the deviation from the mean and can be further decomposed in an observable and unobservable component:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + v_{it} + u_{it} \tag{3}$$

In (3) firm-level productivity is $\omega_{it} = \beta_0 + v_{it}$ and u_{it} is the i.i.d. error term. The productivity level can be obtained from (3) by taking the exponential of the estimated ω_{it}.

3.1 Possible Sources of Bias in TFP estimation

The productivity estimate from (3) could suffer from simultaneity bias, competitive selection bias and multi-product bias, each of which is discussed in detail as follows:

3.1.1 Endogeneity or Simultaneity Bias

An OLS estimate of (3) can be unbiased only if the inputs to production are exogenous from the firm's productive efficiency. However, MARSCHAK and ANDREWS (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, for a two-input production function where labour is the only freely variable input and capital is quasi-fixed, that the capital coefficient will be biased downward if a positive correlation exists between labour and capital. Another relevant issue raised by OLLEY and PAKES (1996) relates to the entry and exit of firms which is 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 lead to biased results. If firms have some knowledge about their productivity level ω_{it} prior to their exit, this will generate correlation between ω_{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 ω_{it} without exiting. In sum, the selection bias or 'endogeneity of attrition'-problem will generate a negative correlation between ω_{it} and k_{it}, causing the capital coefficient to be biased downwards in a balanced sample.

3.1.2 Selection Bias
Firms' entry or exit into a market is determined to a great extent by their initial productivities as studied by FARIÑAS and RUANO (2005) for Spanish manufacturing firms. As the likelihood of a firm's survival is dependent on its productivity level ω_{it}, any knowledge about this prior to the decision to remain in the market or exit will generate correlation between ε_{it} and fixed capital (ACKERBERG et al. 2007). This correlation is due to the fact that firms with a higher capital supply will, ceteris paribus, be able to survive with lower ω_{it} relative to firms with a lower capital stock (VAN BEVEREN 2010).

3.1.3 Multiproduct Bias
Firms' decisions about the range of goods to produce are typically made at a more disaggregated level than is available in manufacturing data sets (BERNARD et al. 2009). If firms produce multiple products within the same industry and if these products differ in their production technology or in the demand they face, this will lead to biased TFP estimates, because the production function assumes identical production techniques across products manufactured by a single firm.

4 ECONOMETRIC METHOD

4.1 TFP Estimation
The techniques used for estimation of unbiased and consistent production function coefficients are described as follows. As noted above, OLS estimates are likely to yield biased values of the coefficients, and as a consequence we use two-stage least squares with instrumental variables (IV), along with the semi-parametric OLLEY and PAKES (1996) and LEVINSOHN and PETRIN (2003) techniques. We did not use fixed effects estimation as it depends on the strong assumption that productivity 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.

4.1.1 Instrumental Variables
An alternative method to achieve consistency of coefficients in the production function is by using instrumental variables for the endogenous independent variables, i.e., the freely alterable inputs in the production function. Unlike the fixed effects estimator, IV methods do not rely on strict exogeneity of the inputs for consistent estimation (WOOLDRIDGE 2009).

GREENE (2004) has pointed out three requirements for achieving consistent estimates. First, instruments need to be correlated with the endogenous regressors (inputs). Secondly, the instruments cannot enter the production function directly. Finally, the instruments should not be correlated with the error term (and hence with productivity).

4.1.2 OLLEY and PAKES Method

OLLEY and PAKES (1996) were the first to introduce a semi-parametric estimation algorithm that takes both the selection and simultaneity problems directly into account. This estimator solves the simultaneity problem by using the firm's investment decision as a 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, given the information currently available. YASAR et al. (2008) have developed a routine in Stata to estimate the log of total factor productivity following the OLLEY and PAKES algorithm.

The OLLEY and PAKES method 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 needs to be monotonic with the productivity and hence during econometric analysis non-negative values of investment variable are required. Third, deflation on the basis of industry level prices implies that all the firms face the same prices. (VAN BEVEREN 2010).

The salient steps of the estimation process are as follows. 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 $\omega_{it} = h_t(k_{it}, i_{it})$, so that productivity can be expressed in terms of capital and investment.

The OLLEY and PAKES method proceeds in two stages. In the first-stage regression, using the above relationship in equation (2) 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 form the first stage and survival probability and by applying non-linear least squares method the coefficients on the capital variable is estimated. For details, see VAN BEVEREN (2010).

Although the 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 method is to use the LEVINSOHN and PETRIN (2003) method which takes care of the simultaneity problem by using intermediate inputs as a proxy for productivity instead of investment.

5 EMPIRICAL ANALYSIS

The main objective of this study is to establish how agglomeration and competitive selection affect regional productivity distribution. To get reliable estimates, the first and fundamental step is to arrive at unbiased TFP for the firm. Therefore much of the effort in this section is

focused on how to determine consistent TFP estimates. The estimates are computed keeping in consideration the practical issues pointed out in the previous section.

5.1 Data

To determine the selection and agglomeration effects on firms' TFP in Taiwan we use firm-level data disaggregated at the county-industry (3-digit NAICS) level from ISI Emerging Markets Information Services. The unbalanced panel data are for the years 2009 to 2011.The dataset has four main fields indicating the physical location, industry, operational status of the firm, and its listing and trading status on the stock market. The dataset also provides information about financial indicators relating to balance sheet and income statements such as non-current assets and sales revenues along with data on the profitability, liquidity and growth trend ratios. We extract information about each firm's total operating revenues, assets and number of employees to estimate the production function parameters.

For this paper, we supplement the dataset with county-level income and industry price data available at the website of National Statistics Office (DGBAS), Taiwan. These data are used to deflate the revenue figures and construct lagged instrumental variables to be used with the 2SLS/IV method.

5.1.1 Data Cleaning

We cleaned the raw data containing 4646 observations using several steps. First, we deflate the revenue figures by industry-level prices for the year 1996. Second, using box plots we examine the data for extreme outliers and remove the entities with top and bottom one percent TFPs to avoid their influence on the results. This results in a final dataset of 4627 observations. Table 1 gives the summary statistics of the mean and standard deviation of inputs and outputs used in the Cobb-Douglas production function.

Table1: Summary statistics-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
Labour	660	1,770	33,669

Table 2 shows the county-wise location of the selected industries. 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 the production. For NAICS 325 we have a total of 310 observations of which 71 are in the pharmaceutical sector (3254), and the rest are in basic chemical manufacturing (3251). It can be seen that computer and electronics firms (NAICS 334) constitute half of the total number of observations. From the total of 2150 observations, 389 are for semiconductor manufacturing (3344). Finally, we look at the scientific and technical industry (NAICS 541) where 20 observations are biotechnology firms. All industries have a presence in all three regions - science park counties, small cities and large cities.

Table 2: Summary statistics-II

State/County	NAICS 325	NAICS 334	NAICS 541
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

As the focus of the paper is on agglomeration and selection analysis, the geographical unit of estimation of each market is the county. This division is justified, due to the fact that for big cities the market effects are likely to spill over the entire county. 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 labour density statistics we classify Taipei County, New Taipei City, Keelung City and Chiayi City as the large cities. The counties designated Science Park counties are Hsinchu County, Tainan City, Yunlin City and Kaohsiung City.

5.2 Results

5.2.1 TFP estimates using OLS and 2SLS/I
The baseline TFP estimates are computed using OLS. The OLS estimate of (3) requires that $E(x_{it}\omega_{it}) = 0$. As for the firm it is possible to observe or anticipate its productivity and thus decide the level of the inputs. The more flexible is the nature of the input, the possibility of adjusting its level based on expected productivity becomes more likely. In this estimation it may be difficult for the firm to change its capital input but labour can be adjusted very easily in a short time. To test the reliability of OLS estimates we perform the Durbin Wu Hausman test of endogeneity (HAUSMAN 1978). The small p-value indicates that the estimates are not reliable (see Table 3).

Table 3: Production function coefficients

Model/Variables	OLS	2SLS/IV	OP
Capital	0.37^{***}	0.56^{***}	0.29^{**}
Labour	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	

Note: ** significant at 5% level, and *** significant at 1% level.

To avoid the simultaneity bias we also 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 overcome the simultaneity bias, they still do not take care of the selectivity bias and hence the results remain biased (see Table 3).

5.2.2 OLLEY and PAKES Method
Keeping in mind the shortcomings of the techniques used above, we finally estimate TFP using the method proposed by OLLEY and PAKES (1996). The standard errors of all OLLEY and PAKES estimation routines are bootstrapped using 200 replications. The TFP distributions were drawn for the cities with above and below the median population density and for the firms located in science parks. In this approach, we use the return on capital as proxy for investments made by the firm along with control variables such as the number of employees to control for size – see Table 3.

Using the TFP estimates form the OLLEY and PAKES method we examine the summary statistics for each of the regions as detailed in Table 4. It is evident that large cities have the highest mean value followed by that of the science park firms. The inter-quartile ratio (IQR) for each region indicates that big cities are the ones that benefit most from agglomeration economies showing the largest dispersion.

5.2.3 Agglomeration and Selection
In order to identify the impact of agglomeration we use two variables namely localization economies and urbanization economies. Following HENDERSON et al. (1995) we define localisation L as the regional employment share E of the specific industry (defined at the three-digit NAICS level) in the manufacturing sector: $L = [E_{jr}/E_r]/[E_j/E_{tot}]$ for industry j at time t in region r. Urbanisation U 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) at time t in region r.

In addition, as noted earlier, agglomeration may entail diseconomies, for example, through pollution or higher land rents. Since we do not have information on industrial land prices, we have used population density PD instead, following a number of authors including GUIMARÃES et al. (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, the relevant market for these firms in Taiwan is not the local market, hence there is the possibility that population density may not actually capture market-size effects.

Table 4: Region-wise log-TFP distribution statistics

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
std. dev.	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

Note: BM: Small City, SP: Science Park, AM: Large City

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 determining heterogeneous firms' location-choice decision, BALDWIN and OKUBO (2006) show that high-productivity firms self-select into large markets. This self-selection, also referred to as sorting by high-productivity firms producing substitutable product varieties, is motivated by the potential of higher profits from a large market. Hence, the average productivity level in large markets is expected to be higher than that in small markets. In using the OLLEY and PAKES (1996) technique for estimation of production functions we have already taken care of the survival-based selection of the firms, hence we focus now on BALDWIN and OKUBO's self-selection, where surviving firms sort into different regions depending on their productivity and regional characteristics, i.e., high (low) productivity firms concentrate in a region with a large (small) market.

We use the TFP estimates from the OLLEY and PAKES method to obtain distributional measures of regional market productivity median, IQR and 10th percentile values. By regressing these measures of regional productivity distribution on variables representing agglomeration economies and market competition we can establish their significance and direction. Finally, the effects of agglomeration economies and self-selection are numerically compared to identify which of the two contributes more to a region's productivity level.

5.2.4 Identification of Firm Self-Selection

To identify the process through which high-productivity firms sort into science parks and large cities we use a selection and an outcome equation. Considering firm's sorting in science parks the relevant selection equation is as follows:

$$z_{it}^* = \alpha_0 + \alpha C_{it} + \varepsilon_{it}, \text{ where } z_{it}^* \text{ is } z_{it} = 1 \tag{4}$$

z_{it} is the dummy variable of the select 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, firms' return on capital, and return on equity. The outcome equation is given as follows:

$$S_{prt} = \beta_0 + \beta_a A_{rt} + \beta_c X_{rt} + \upsilon_{prt} \text{ if } z_{it}^* > 0 \tag{5}$$

where S_{prt} is the pth percentile at time t of region r, A_{rt} are industry-specific agglomeration variables at time t for region r, X_{rt} are the region-time specific control variables and υ_{prt} is the error term. The expected sign of the agglomeration coefficient is positive.

To estimate the selection effect we use the Heckman two-step estimator for selection models (HECKMAN 1979). Such models are common in micro-econometric studies, in the estimation of wage equations or consumer expenditure. If the coefficient on the inverse Mill's ratio is statistically significant, there is selection bias. For this study we have the following instruments: log of population density in 1950, log of return on equity and capital. The validity of instruments is established using the SARGAN (1958) test of over-identification restrictions. The results confirm that the science-park or large city dummy is positive and significant in the analysis.

5.3 Agglomeration and Selection-Controlling Self Selection

After determining estimates for log TFP distributions for the regions the following regression analysis is conducted to find out the impact of agglomeration on right shift, dispersion and truncation of the distribution. As the issue relating to self-selection is still there we consider these as raw productivities. We now use the HECKMAN model to establish whether self-selection is positive and significant. The following variables are generated based on SYVERSON (2004):

  i.   Mean (median)-to check for relative shift
  ii.   IQR -for dispersion
  iii. 10th percentile - for truncation/cut-off

Using these as dependent variables we find their significance with respect to agglomeration variables (localisation and urbanisation) and selection (population density). We repeat these steps for big cities, small cities and science parks and for different industries at the 3-digit NAICS level.

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

Apart from using the HECKMAN (1979) model we also use dummy variables for each region in the instrumental variables regression to see if they are positive and significant and control the self-selection. If this type of endogeneity exists then instead of large concentration of firms in cities impacting the individual firm's productivity, it is the firm's self-selection which results in greater concentration of high productivity firms in any region. This possibility of reverse causality can be controlled through the use of instruments correlated with firm's productivity but not correlated with agglomeration.

6 DISCUSSION OF RESULTS

The results can be divided into three parts. The first relates to the non-parametric comparisons of the log TFP distribution for manufacturing firms located in the three identified regions. For the overall manufacturing sector the firms located in the cities with above median population density have the highest mean TFPs. Also the TFP distribution here is more dispersed and has the highest minimum value indicating greater within region competition. The firms located in science parks depict similar characteristics (see Figure 2).

Figure 2: Kernel density plots for the three regions
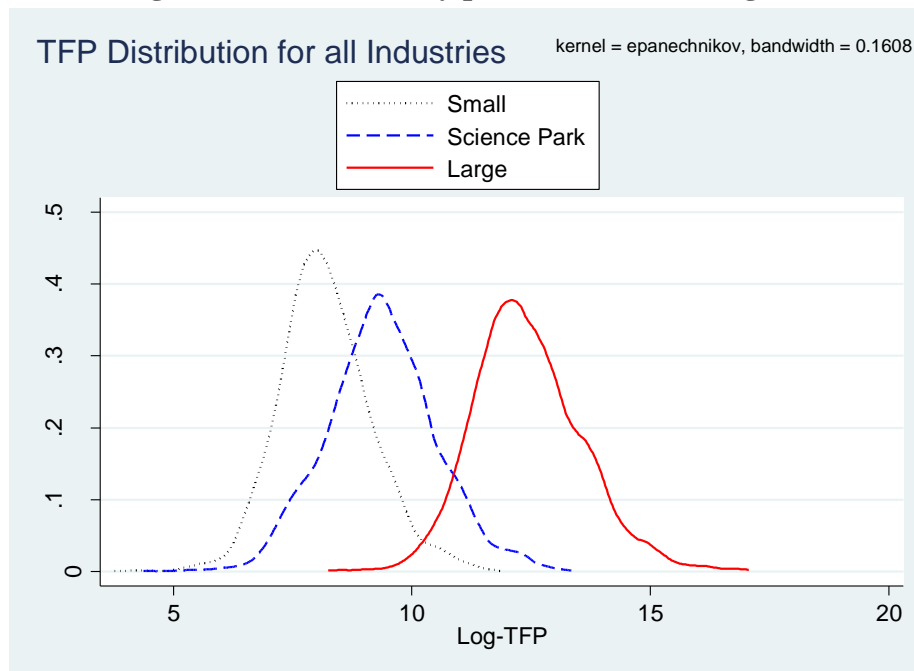
As the analysis becomes more industry-specific for NAICS 334, 325 and 541, we find that firms in science parks show varying trends in TFP distribution. The TFP distributions for the selected sectors are shown in figures 3, 4 and 5 respectively. The chemicals industry shows that firms located in science parks have the lowest productivity level whereas the scientific and technical services industry including the biotechnology sector has the highest productivity firms located in science parks.

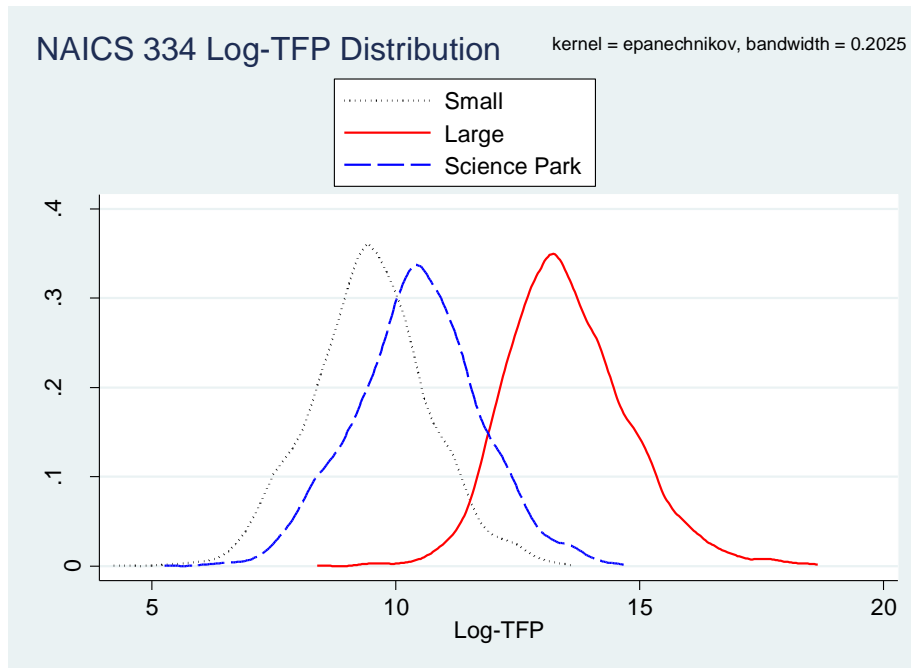**Figure 3: Kernel density plots for computer and electronics firms for three regions**

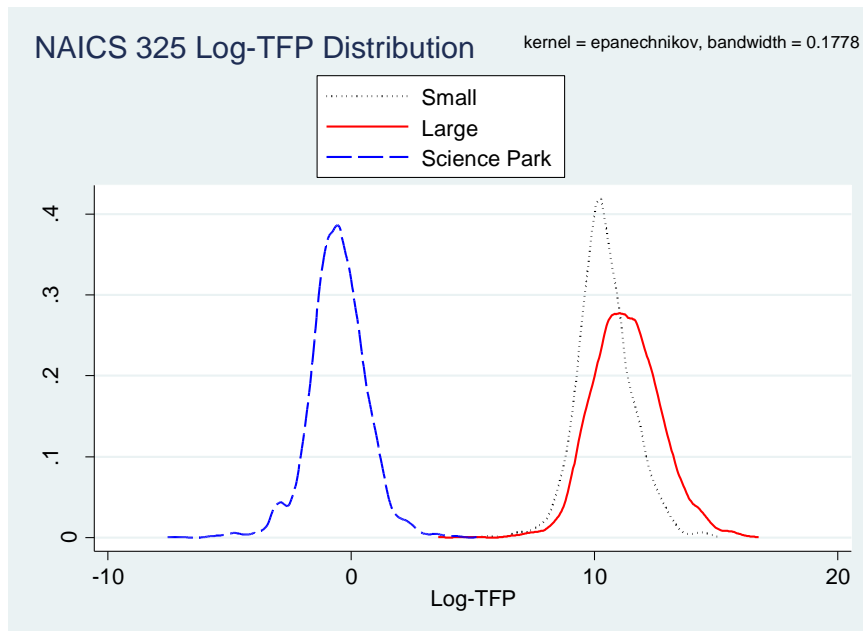**Figure 4: Kernel density plots for chemical manufacturing firms for three regions**
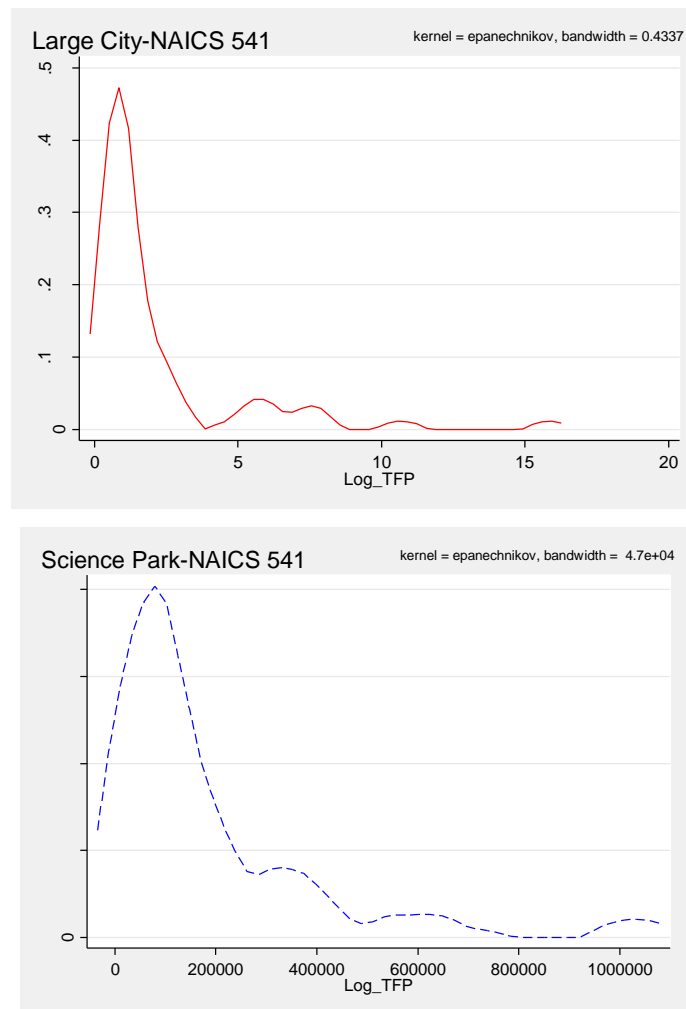
Figure 5: Log –TFP Distribution for NAICS 541

Note: Biotechnology sector classified here

Analysing the results for the impact of agglomeration and selection, the HECKMAN (1979) selection model confirms that firms do self-select in regions of high productivity. However, even after controlling for self-selection, we find that the agglomeration variables for localization and urbanization are statistically significant. The regression results indicate that selection due to competition is also significant. However the coefficients for selection are much smaller than those of the agglomeration variables (see Table 5).

Table 5: Agglomeration and selection in science parks

IQR			MED			10-TILE		
L	U	PD	L	U	PD	L	U	PD
1.89***	-3.47***	-.07**	0.48***	-0.35***	0.31***	-.70***	1.84***	.04***
(.10)	(0.16)	(.02)	(0.01)	(.06)	(.00)	(.06)	(.09)	(.01)

Note: **significant at 5% level, and ***significant at 1% level. Standard errors are noted in parenthesis.

7 CONCLUSION

Firms located in science parks have higher productivities even after controlling for self-selection bias. Moreover, firms located in science parks benefit from both types of agglomeration economies namely specialization and diversification. 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 than the elasticity of agglomeration variables. Thus science park incentives insulate firms from the competition they might face in open markets. This fact is even demonstrated in the case of low-tech firms which have a productivity distribution that lags 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. Also industrial clusters such as science parks are not always sufficient to provide positive productivity shocks to incumbent firms.

The policy implication that arises from this study is based on robust theoretical foundations. We have tried to tease out the impact of policy on firm's productivity; the indicator of its heterogeneity, and a key part of the literature on firms and trade stimulated by the work of MELITZ (2003). 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 to productivity improvements of firms, including those in the biotechnology sector. The extent to which benefits of science parks are taken advantage of by firms as depicted in their productivity depends on the underlying technology of production.

REFERENCES

ACKERBERG, D., L. BENKARD, S. BERRY, and A. PAKES (2007). Econometric tools for analysing market outcomes, in: HECKMAN, J and E. LEAMER (eds.): *Handbook of Econometrics*, Amsterdam, North Holland, pp. 4171-4276.

ASSOCIATION OF UNIVERSITY RESEARCH PARKS (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(2): 18-29.

BALDWIN, R., T. OKUBO (2006). Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting, *Journal of Economic Geography* 6(3): 323–346.

BERNARD, A.B., S.J. REDDING and P.K. SCHOTT (2009). Products and productivity, *Scandinavian Journal of Economics* 111(4): 681-709

CARTER, C.A., G. MOSCHINI, and I. SHELDON (2011). *Genetically Modified Food and Global Welfare*, Bingley, Emerald Group Publishing.

CICCONE, A., R., HALL (1996). Productivity and the density of economic activity, *American Economic Review* 86(1): 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(6): 2543-2594.

DE LOECKER, J. (2007). Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalisation on Productivity, *NBER Working Paper No: 13155*, Cambridge, MA.

EMERGING MARKETS INFORMATION SERVICE (EMIS). New York: Internet Securities, < http://www.securities.com/products/emis.html>.

FARIÑAS J., S. RUANO (2005). Firm productivity, heterogeneity, sunk costs and market selection, *International Journal of Industrial Organization* 23(7-8): 505–534.

FUJITA M., H. OGAWA (1982). Multiple equilibria and structural transition of non-mono-centric urban configurations, *Regional Science and Urban Economics* 12(2):161–196.

GREENE, W. (2004). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model, *Journal of Econometrics* 126(2): 269-303.

GUIMARÃES, P., O. FIGUEIREDO and D. WOODWARD (2000). Agglomeration and the location of foreign direct investment in Portugal, *Journal of Urban Economics* 47(1): 115-135.

HAUSMAN, J. (1978). Specification tests in econometrics, *Econometrica* 46(6): 125l- 1271.

HECKMAN, J.J. (1979). Sample selection bias as a specification error, *Econometrica* 47(1): 153–161.

HENDERSON, J., A. KUNCORO and M. MATTHEW (1995). Industrial development in cities, *Journal of Political Economy* 103(5): 1067-1090.

HOPENHAYN, H. (1992). Entry, exit, and firm dynamics in long run equilibrium, *Econometrica* 60(5): 1127.1150.

JAFFE, A. (1989). Real effects of academic research, *American Economic Review* 79(5): 957-970.

JAFFE, A., M. TRAJTENBERG and R. HENDERSON (1993). Geographic localization of knowledge spill-overs as evidenced by patent citations, *Quarterly Journal of Economics* 108(3): 577-598.

LAVOIE, B., I. SHELDON (2000). The source of comparative advantage in biotechnology: A real options approach, *Agribusiness – An International Journal* 16(1): 56-67.

LEVINSOHN, J., A. PETRIN (2003). Estimating production functions using inputs to control for un-observables, *Review of Economic Studies* 70(2): 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, R.E., E. ROSSI-HANSBERG (2002). On the internal structure of cities, *Econometrica* 70(4): 1445–1476.

MARSCHAK J., J. ANDREWS and H. WILLIAM (1944). Random simultaneous equations and the theory of production, *Econometrica* 12(3-4): 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(6): 1695-1725.

MELITZ, M., G. OTTAVIANO (2008). Market size, trade and productivity, *Review of Economic Studies* 75(1): 295–316.

MONCK, C., K. PETERS (2009). Science Parks as an Instrument of Regional Competitiveness: Measuring Success and Impact, *IASP Annual Conference Proceedings*, Malaga, IASP, < http://www.sqw.co.uk/files/8113/8712/8061/57.pdf>

NATIONAL STATISTICS (2010). Directorate-General of Budget, Accounting and Statistics Executive Yuan (DGBAS), Taiwan.

OLLEY, G., A. PAKES (1996). The dynamics of productivity in the telecommunication equipment industry, *Econometrica* 64(6): 1263–1297.

ROSENTHAL, S., W. STRANGE (2004). Evidence on the nature and sources of agglomeration economies, in: HENDERSON, V. and J.F. THISSE (eds.): *Handbook of Regional and Urban Economics*, Amsterdam, North Holland, pp. 2119–2171.

ROTHAERMEL, F., M. THURSBY (2005a). Incubator firm failure or graduation? The role of university linkages, *Research Policy* 34(7): 1076-1090.

ROTHAERMEL, F., M. THURSBY (2005b). University-incubator firm knowledge flows: Assessing their impact on incubator firm performance, *Research Policy* 34(3):302-320.

SARGAN, J.D. (1958). The estimation of economic relationships using instrumental variables, *Econometrica* 26(3): 393-415.

SYVERSON, C. (2004). Market structure and productivity: A concrete example, *Journal of Political Economy* 112(6): 1181–1222.

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 Series 182/2007*, Leuven, Catholic University of Leuven.

WOOLDRIDGE, J.M. (2009). On estimating firm-level production functions using proxy variables to control for un-observables, *Economics Letters* 104(3): 112–114.

YASAR, M., R. RACIBORSKI and B. POI (2008). Production function estimation in Stata using the Olley and Pakes Method, *Stata Journal* 8(2): 221–231.