TI 2016-021/VIII
Tinbergen Institute Discussion Paper

Place-based Policies, Firm Productivity and Displacement Effects: Evidence from Shenzhen, China

Hans Koster[a]
Fang Fang Cheng[b]
Michiel Gerritse[c]
Frank van Oort[d]

[a] *Faculty of Economics and Business Administration, VU University Amsterdam, Tinbergen Institute, the Netherlands, and London School of Economics, United Kingdom;*
[b] *Utrecht University, the Netherlands;*
[c] *University of Groningen, the Netherlands;*
[d] *Erasmus School of Economics, Erasmus University Rotterdam, Tinbergen Institute, the Netherlands, and Utrecht University, the Netherlands.*

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Place-based policies, firm productivity and displacement effects: Evidence from Shenzhen, China*

By HANS R.A. KOSTER,[a] FANG FANG CHENG,[b] MICHIEL GERRITSE[c] *and* FRANK G. VAN OORT[d]

This version: 31 March 2016

ABSTRACT — We analyse the economic impacts of place-based policies that aim to enhance economic development by stimulating growth and productivity of firms in designated areas. We use unique panel data from China with information on manufacturing firms' production factors, productivity and location, and we exploit temporal and spatial variation in place-based interventions due to the opening of science parks in the metropolitan area of Shenzhen. The identification strategy enables us to address the issues that *(i)* science parks are located in favourable locations and that *(ii)* high-productivity firms sort themselves in science parks. We find that productivity is approximately 15-25 per cent higher due to the policies. The results also show that local wages have increased in science parks. Weaker evidence suggests that displacement effects are sizeable.

JEL–code — H2, R3, R5
Keywords — place-based policies, transitional economies, science parks, productivity

I. Introduction

Many governments spend considerable amounts of money to stimulate employment growth, fight unemployment, and spur productivity. These investments are often not space-neutral but differ between regions, cities and even between neighbourhoods within cities. In developed countries, place-based policies tend to focus on distressed regions or neighbourhoods (Gobillon et al., 2012). In the European Union, for example, the Regional

---

* This work has benefited from a VENI research grant from the Netherlands Organisation for Scientific Research. We thank Maximilian von Ehrlich, Aart Gerritsen, Martijn Kobus, Jos van Ommeren as well as participants of seminars in Washington D.C., Rotterdam and Amsterdam.
[a] Corresponding author. Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, e-mail: h.koster@vu.nl. The author is also affiliated with the Spatial Economics Research Centre at the London School of Economics and the Tinbergen Institute.
[b] Urban and Regional Research Centre, Utrecht University, Heidelberglaan 2, 3508TC Utrecht, e-mail: f.cheng@uu.nl.
[c] Faculty of Economics and Business, University of Groningen. Nettelbosje 2, 9747AE Groningen, e-mail: michiel.gerritse@rug.nl.
[d] Erasmus School of Economics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062PA Rotterdam, e-mail: vanoort@ese.eur.nl. The author is also affiliated with the Urban and Regional Research Centre, Utrecht University, Heidelberglaan 2, 3508TC Utrecht, and the Tinbergen Institute.

Development Fund explicitly targets regions with high unemployment and a (nominal) income below 75 per cent of the EU average. Similarly, in the US, programmes such as federal urban Empowerment Zones (EZs) and Enterprise Communities are designed to use grants and hiring credits to benefit lagging neighbourhoods (see e.g., Busso et al., 2010).

The common rationale for both place-based and people-based (i.e., spatially blind) policies is to improve the prospects of poor and disadvantaged households (Barca et al., 2012; Neumark and Simpson, 2015). Place-based investments in lagging regions may be more effective in reaching deprived households than economy-wide investments (Garretsen et al., 2013; McCann and Ortega-Argilés, 2013). However, the focus on lagging regions may come at a (welfare) cost: the inefficiencies caused by place-based policies could be substantial. Glaeser (2008) provides several arguments against place-based policies. First, place-based policies that target deprived areas bring economic activity to the least productive places, thus lowering overall productivity. Second, productivity also falls if poor regional performance can be traced back to negative spillovers from local people or firms. Third, the distributional effects of place-based policies are unclear. For example, beneficiaries of the aid may be the richer people in the impacted area, thereby increasing inequalities within the region. Related to that point, the spatial extent of the effects of place-based investments may be unpredictable, so choosing a scale for a place-based policy can be problematic (Cheshire et al., 2014).

Whether place-based policies have large welfare costs depends on the responses of the people in the designated areas. Kline (2010) argues that place-based employment policies are most efficient when the demand and supply of labour are inelastic – in that case, the policy instrument produces little distortion of behaviour. Similarly, Busso et al. (2013) show that a larger heterogeneity in workers' preferences regarding commuting and residential locations leads to fewer job changes and capitalisation in wages instead of in land rents. Employment decisions and the degree to which workers and firms change location thus determine the welfare effects of the policies. As a result, the welfare costs or deadweight losses of such programmes can be approximated by interpreting the local economy's responses to the programme.

The empirical results on the effectiveness and the welfare costs of place-based policies are mixed (see for an overview Neumark and Simpson, 2015). Most studies find that positive effects are offset by substantial displacement effects within adjacent localities, but virtually all of the empirical studies on place-based policies examine programmes for *deprived* areas in *developed* countries. The welfare arguments may be different when applied to place-based policies for *leading* areas in *developing and transition* economies. Policies stimulate relatively productive firms and people and foster positive spillovers rather than negative spillovers. While almost unstudied, China, India, Brazil, South Africa, Russia, and many other transition economies extensively apply place-based policies and special economic zones to promote development (Rodriguez-Pose and Hardy, 2014). The available empirical evidence, therefore, is not representative of many of the place-based policies that have been implemented worldwide (Barca et al., 2012; Foray, 2015).

This paper investigates the economic impacts of substantial place-based investments in Shenzhen. In China, economic place-based policies were carried out primarily to promote foreign direct investment, technology transfers, and exports. Science parks are an actively used policy-instrument in this. Firms in science parks are expected to cooperate and interact more intensely by locating in close proximity to each other. In other words, place-based investment aim to foster agglomeration economies and innovation within science parks. In China, the number of science parks has increased dramatically in the last decade, which demonstrates the strong belief in the effectiveness of these policies (Cheng et al., 2014). Yet, with many economic, social and institutional characteristics differing from those in Western economies, the economic impact of science-parks in China and other developing countries are highly doubted and even called 'pipedreams' (Rodriguez-Pose & Hardy, 2014). However, systematic quantitative research supporting or questioning science park policies is lacking.

We use data on manufacturing industries in the Shenzhen metropolitan area to investigate the impact of science parks on firm productivity. We use a 10-year panel on approximately 10,000 firms. For each firm we have information on output, workforce size, capital stock, and wages. The data also provide information on a number of other firm characteristics and the location of the firms at the neighbourhood level. In Shenzhen's science parks, property rights to the land are guaranteed, and there is better access to fast internet. Moreover, different financial incentives are available for innovative firms. More information on the exact incentives will be given in the next Section.

Our analysis contributes to the literature in the following ways. First, it complements the knowledge on the effects of place-based strategies with a micro-data approach in a non-Western country.[1] By studying the Chinese context, we contribute to the discussion on the effectiveness of place-based development strategies in transition economies.

Second, we examine the productivity and spatial mobility effects caused by science park policies within a theoretical framework. This enables us to interpret our empirical results as 'sufficient statistics' (Chetty, 2009), allowing inferences on welfare using estimated elasticities. In our data, we cannot observe the different instruments of the programme, but we show that with minimal information, one may still obtain an estimate of the relative size of the productivity effects and the deadweight losses from the policy.

As a third contribution, the paper aims to identify the causal impacts of a place-based development programme. Often, governments single out specific locations for grants or to establish economic zones. Moreover, once a place-based policy is initiated (or even anticipated), productive firms and workers may sort into these areas because they benefit

[1] There is one notable exception; Wang (2013) uses data at the municipality level and finds that productivity, FDI and wages all have increased due to the policies that target Special Economic Zones in China.

from the policy or because the policy specifically applies to them. Ignoring the selection of a policy area and sorting processes in the evaluation of a place-based policy may lead to biased estimates of the true policy effects. To identify a causal effect of the science park policy, we need to control for unobserved locational traits and trends because science parks may have innate attractive geographic features and may evolve differently from non-targeted areas. To control for unobserved locational endowments, we rely on spatial differencing. This approach implies that we compare firms in science parks with firms in areas that are very close to science parks – areas that are very similar in geographical and functional characteristics. Hence, conditional on selecting areas very close to a science park boundary, the exact location of the boundary is considered to be random. We relax this assumption by including flexible functions of distance to the nearest science park boundary or geographic coordinates. To further address the issue that highly productive firms may sort themselves into science parks, for example because of entry requirements, we exploit the panel nature of our data and include firm fixed effects. Hence, we identify the effect of science parks by comparing productivity differences before and after the opening of a science park for a firm located in a science park vis-à-vis a firm that is located just outside a science park. A worry one might have is that the estimated parameter will be an overestimate of productivity effects if local displacement effects are important. We therefore also consider an alternative identification strategy in the sensitivity analysis, based on the observation that local industrial parks are often promoted to science parks later on, making them a feasible control group.

Our results show that area-based incentives have a substantial impact on firms' productivity in Shenzhen science parks. Even if we include firm fixed effects and use spatial differencing, firms' output is increased by 15-25 per cent due to science park policies. These large and economically meaningful effects are in line with the findings of Wang (2013) and contribute to the idea that place-based policies have more sizable effects in transition economies. We furthermore show that local wages have increased by approximately 10-15 per cent, which is smaller in magnitude than the productivity effects. Despite the large productivity effects, we also show that these policies seem to generate distortive effects. Workforce size increases in science parks, which suggests a displacement effect. Given the assumptions of our theoretical model, we estimate that the deadweight loss due to these displacement effects are sizeable (about 40 percent of the total effect on productivity). However, we note that these effects are statistically imprecise. We subject our results to an extensive sensitivity analysis, including an analysis based on another (cross-sectional) dataset.

The remainder of the paper is organised as follows. In Section II, we outline our theoretical framework and derive conditions under which we can approximate a deadweight loss of science park policies. Section III discusses the regional context and the data. In Section IV, we outline our empirical strategy. Section V presents the main results for productivity, wages, employment, and capital stock. Using these estimates and the structure of the theoretical model, we estimate the deadweight loss associated with these policies. In Section VI, we conduct an extensive sensitivity analysis. Section VII concludes.

II. Theoretical framework: place-based policies and deadweight losses

To structure our thoughts, we first discuss the potential welfare effects of place-based policies. The intuition follows Kline (2010), Kline and Moretti (2013) and Busso et al. (2013) and helps to interpret our empirical results as 'sufficient statistics' (Chetty, 2009) – allowing inferences on welfare using estimated elasticities. Our approach is more parsimonious than related structural models because we rely on spatial differencing to identify the policy effects. Comparing firms that are at very short distances from each other implies that a worker's residential decision does not vary much between the compared working locations. The design thus circumvents the issue that consumption amenities affect labour supply decisions and the issue that we do not observe such amenities. We assume that the place-based policies (i.e., setting up a science park) involve a bundle of instruments: they may directly affect productivity via an increased technology level (e.g., availability of broadband internet) and stimulate the use of capital and labour. In our data, we cannot observe the different instruments of the programme, but we show that with minimal information, one may still obtain an estimate of the relative size of the productivity effects and the deadweight losses from the policy.

We assume that there are two locations, $z = 1,2$. The science park policy only targets location 1. Firms in both locations employ labour and capital to produce a numéraire consumption good. Capital is hired from a 'large' competitive financial market where firms (re-)finance capital at a competitive rate. Workers and owners of land and capital consume the numéraire good and offer labour to firms in one of the two locations such that they can decide where to work. A worker's utility u_{iz} is equal to the consumption of the numéraire good and a worker-specific heterogeneity term:

$$u_{iz} = c_i + \xi_{iz} \,. \qquad (1)$$

The worker's heterogeneity term ξ_{iz} depends on the location z in which the worker is employed. This dependence of location allows for working location preferences that are separate from wages. As we compare firms at the boundaries of science parks, it is unlikely that ξ_{iz} captures differences in preferences for commuting and geographic attributes. Instead, differences in ξ_{iz} may be explained by a required mastery of English, administrative skills, levels of workplace stress, preferences for specific types of companies (international, high-tech) or prestige. For the purpose of the welfare analysis, we are agnostic about how exactly these preferences take form.

The production function of a firm j depends on labour ℓ_{jz}, capital k_{jz}, location-specific technology constant A_z and a firm-specific heterogeneity term ω_j. Firm productivity is thus a composite of an individual technology term and a location-specific productivity shifter. The total production is: $q_{jz} = q(A_z, \ell_{jz}, k_{jz}, \omega_j)$, so profit is given by:

$$(2) \qquad \max_{\ell,k} \pi_{jz} = q_{jz} - w_z(1-t_z)\ell_{jz} - r(1-s_z)k_{jz},$$

where we allow for any explicit or implicit subsidies to the employment of labour (t_z) or capital (s_z). The first-order condition of the profit-maximisation implies that $\mathrm{d}q_{jzd}/\mathrm{d}\ell_{jz} = w_z(1-t_z)$ so that the wage of a worker is determined by his marginal productivity, corrected for the subsidy.

We assume that the place-based policy can affect the technology in location 1 (A_1) and subsidises factor employment at costs $t_1 w_1 L_1$ and $s_1 r K_1$, where $L_1 = \int_{j \in z=1} \ell_j \mathrm{d}j$ and $K_1 = \int_{j \in z=1} k_j \mathrm{d}j$ are the cumulative factor uses for location 1. We assume that the individual firm productivity term ω_j and the location-specific technology term A_z are log-separable. This assumption implies that location-specific productivity shocks affect the log productivity of efficient and inefficient firms in the same way. Because there are no policies in location 2, it holds that $t_2 = s_2 = 0$.

The total costs of the productivity improvements are $T_1 = T(A_1, Q_1)$, where $Q_1 = \int q_{j1} dj$ is the aggregate production in location 1. Note that a higher technology level A_1 requires a higher budget. The term Q_1 enters because given a level of productivity, higher production could imply a larger required budget: e.g., if firms enter or expand production, more broadband investment could be required to maintain the same level of technology. If technology inputs are non-rival, a higher level of production Q_1 implies that a higher budget is needed to achieve the same productivity level A_1.

The welfare function is the sum of individual utilities, firm profits, capital earnings and the net government budget. Consumers' wages equal their consumption of the numéraire. Thus, the welfare function is given by:

$$(3) \qquad \mathcal{W} = \int w_{iz} + \xi_{iz} \mathrm{d}i + \int \pi_{jz} \mathrm{d}j + r(K_1 + K_2) - \int t_1 w_1 \ell_{j1} \,\mathrm{d}j - \int s_1 r k_{j1} \,\mathrm{d}j - T(Q_1, A_1)\,.$$

The place-based policy causes direct productivity improvements (via the technology level) and changes in labour and capital levels. The costs of the policy are the implicit or explicit subsidies given to capital and labour and the costs of increasing the overall productivity level. Let us denote the application of the place-based policy by the variable p_1.

Using the welfare function, we can study the welfare effects of the science park policy. Differentiating with respect to p_1 and using the Envelope theorem on the profit function, the welfare effect is:

$$(4) \qquad \frac{\mathrm{d}\mathcal{W}}{\mathrm{d}p_1} = \frac{\partial Q_1}{\partial A_1}\frac{\mathrm{d}A_1}{\mathrm{d}p_1} - \left(\frac{\partial T_1}{\partial A_1} + \frac{\partial T_1}{\partial Q_1}\frac{\partial Q_1}{\partial A_1}\right)\frac{\mathrm{d}A_1}{\mathrm{d}p_1} - t_1 w_1 \frac{\mathrm{d}L_1}{\mathrm{d}t_1}\frac{\mathrm{d}t_1}{\mathrm{d}p_1} - s_1 r_1 \frac{\mathrm{d}K}{\mathrm{d}s_1}\frac{\mathrm{d}s_1}{\mathrm{d}p_1}.$$

The first term in (4) is the direct increase in production due to the science park policy through its increase in firm productivity, which accounts for the direct effect on production net of any increases in factor employment. In Section V.A, we estimate this effect using micro-data while controlling for changes in production factor usage. The second term in (4) is the composite effect on the required budget T_1. Increasing the technology level comes at the direct cost of increasing the budget. Indirectly, increased productivity might lead to higher production levels, which in turn could raise the budget; this second-order effect is captured in the second term $(\partial T_1/\partial Q_1)(\partial Q_1/\partial A_1)$.

The third term denotes the labour market effect of the policy: $t_1 w_1(\mathrm{d}L_1/\mathrm{d}t_1)(\mathrm{d}t_1/\mathrm{d}p_1)$. The welfare effects of changes in the workforce are considerably simplified by noting that a marginally affected worker is indifferent to working in the science park or not: possible wage differences are exactly reflected in his preference for locations ξ_{iz}.[2] Using this notion, the welfare loss from labour reallocation is the (implicit) subsidy multiplied by the workers that are displaced. Intuitively, if workers do not move, the policy is simply a transfer of which the net effect on welfare is zero in our welfare function. Only if workers move after the subsidy, there is a welfare loss: some workers preferred not to work in location 1 but are induced to do so by the labour subsidy. The degree to which science park policies subsidise labour is left implicit in the term $\mathrm{d}t_1/\mathrm{d}p_1$ because in our empirical application, we do not know the actual rate of subsidies. The elasticity of labour supply is important in determining the deadweight losses of the policy. To provide a supplementary empirical analysis of the labour supply responses, we investigate the wage responses to the policy in Section V.B. Note that although the wage responses are not present in the welfare effects, they are not ruled out by the theoretical model; changes in wages without changes in allocation are simply a welfare-neutral transfer.

The capital market is similarly affected by the place-based policy, as the fourth term in (4) illustrates: $s_1 r_1(\mathrm{d}K/\mathrm{d}s_1)(\mathrm{d}s_1/\mathrm{d}p_1)$. Similar to the labour-market-related policies, the welfare losses crucially depend on the capital mobility response. Intuitively, if capital does not move, the subsidy is simply a loss to the government offset by a gain to the capital owner. If capital moves, the subsidy compensates the higher productivity that capital could have achieved in another location.

We can obtain an indication of the relative size of the deadweight losses associated with the place-based policy that allows for a comparison of the direct benefits (in terms of changes in the technology level) with the deadweight losses of the factor employment responses. The sum of the welfare effects is the direct productivity effect, the deadweight losses due to factor employment changes, and the financing costs of the productivity effects (all of which are incorporated in equation (4)). We assume that the labour and capital-related policies change with the science park status, so $\mathrm{d}t_1/\mathrm{d}p_1$ and $\mathrm{d}s_1/\mathrm{d}p_1$ may be non-zero. The employment responses can then be written into elasticities: multiplying and dividing the welfare loss from labour reallocation by L_1 gives $-L_1 w_1 e_\ell$, where e_ℓ is the elasticity of labour supply with respect to the science park policies. Next, the term $L_1 w_1$ can be written as a cost share of

[2] The incentives to change job location can, of course, be restricted to wages by assuming zero heterogeneous preferences for location of work. This special case yields very stark theoretical welfare conclusions (see, for instance, Busso et al., 2013) but does not change our empirical welfare results.

production, $\alpha_\ell Q_1$. Rewriting the capital response in the same way and factoring the level of production Q_1, the net welfare effect ΔW can be written as:

$$\text{(5)} \qquad \Delta\mathcal{W} = Q_1(\Delta\mathcal{RW}) - \left(\frac{\partial T_1}{\partial A_1} + \frac{\partial T_1}{\partial Q_1}\frac{\partial Q_1}{\partial A_1}\right),$$

with:

$$\text{(6)} \qquad \Delta\mathcal{RW} = e_A \frac{\mathrm{d}A_1/\mathrm{d}p_1}{A_1} - \alpha_\ell e_\ell - \alpha_k e_k,$$

where we refer to $\Delta\mathcal{RW}$ as the gross relative welfare effect, e_A is the direct semi-elasticity of production to the science park policies (via productivity), and e_ℓ and e_k are the elasticities of labour and capital employment to the science park policies. By formulating the welfare effects as elasticities weighed by shares of production, the separate effect of each welfare gain and loss can be compared at a common scale – the output level. This approach does not allow us to derive absolute levels of welfare losses, but it does allow us to compare the direct productivity benefits with the loss due to labour (or capital) reallocation per unit of output. Thus, the welfare losses are interpreted relative to the productivity increases in percentages of output. The term $e_A\,(\mathrm{d}A_1/\mathrm{d}p_1)/A_1$ represents the percentage increase in output due to science park policies, keeping capital and labour constant – this term therefore represents the partial effect due to changes in the technology constant. This direct effect is estimated in Section V.A. The magnitude of the deadweight loss strongly depends on whether the place-based policy leads to changes in factor employment. Therefore, we estimate the effect of the science park on factor usage in addition to its effects on productivity in Section V.C. The labour and capital responses to the place-based policy are weighted with the cost shares α_ℓ and α_k. The cost shares can be recovered by estimating a Cobb-Douglas production function, which we estimate in Section V.A. Using the cost shares, the productivity gains and the factor movement deadweight losses are measured in terms of output, which allows us to compare the two quantities.

Intuitively, a high level of employment elasticity points to large deadweight losses, but especially if the cost share of labour is high: then, the explicit or implicit labour subsidies may also have been high. In Section V.C, we put the productivity effects and factor employment changes together to present an estimate of the relative size of the deadweight losses. Note that the last term in equation (5) represents the costs of public funds, which are unknown in our application, so we will only be able to measure the gross relative welfare effect.

This stylised model does not consider land use because data on land use are only available for a subset of firms in our dataset. In Section VI.E, we investigate whether this limitation impacts our main conclusions. For the welfare analysis, land use does not contribute much to the results, as science park land is in fixed supply.[3] There may also be agglomeration effects at play that drift across science park boundaries. In principle, this would lead to an

[3] If anything, in our data, firms in science parks pay somewhat higher land rents, which suggests a small role for land market instruments.

underestimate of the results because at the boundary the differences in productivity due to agglomeration economies are zero. However, in the sensitivity analysis we show that agglomeration economies drifting across the boundary are unlikely to be important in our setting. We therefore relegate the discussion of agglomeration effects in the theoretical model to Appendix A.

III. Regional context and data

A. Regional context

Shenzhen is considered one of the most important high-tech hubs in China. It was originally established as one of four special economic zones (SEZs) in 1979 to operate a socialist market economy because of its proximity to Hong Kong, which had been facing increasing pressure from rising labour costs and a tight land supply since the 1970s. Initially, the SEZs were geographically isolated but economically open areas where special and flexible economic policies were carried out primarily to promote foreign direct investment, technology transfers, and exports. The special treatment of these areas co-evolved with high economic and population growth rates. Figure 1 shows that the population of Shenzhen increased from a mere 310 thousand in 1980 to almost 11 million in 2010. A large proportion of the residents are migrants (more than 70 per cent).[4] Shenzhen received a substantial inflow of industrial activities from Hong Kong after the border was opened, most of which are low-tech and medium-tech manufacturing activities such as parts assembly (Enright et al., 2005). Compared with their foreign counterparts, Shenzhen's manufacturing firms have relatively thin profit margins that have been further eroded by rising land and labour costs in recent years (Gu and Chen, 2001; Shenzhen Planning Bureau, 2006; Linden et al., 2009).

Not long after the launch of the SEZs, the zones began to be emulated throughout China to foster economic growth and promote innovation. The concept was appealing because cities and regions could cordon off limited areas to offer special incentives to foreign and later also to domestic investors (Wu and Gaubatz, 2013, pp. 115). The next wave of zones was in the form of science parks, also called high-tech zones. In these areas, research institutions and firms are expected to cooperate and interact by locating in close proximity to each other. Some policies target more conventional manufacturing firms, while others promote high-tech enterprises and business services.

A strategy to develop science parks was also implemented in Shenzhen. The policy was designed to attract high-tech industries, stimulate innovation and foster entrepreneurship (Shenzhen Bureau of Trade and Industry, 2001; Ng, 2003). The first high-tech science park in China, the 'Shenzhen Science Park' (Shenzhen Keji Yuan, known as the Shenzhen High-tech

[4] Migrants are defined as residents that are not 'Hukou' holders. Hukou is a registration system of residents and can be granted to migrants when certain requirements are met. Hukou holders enjoy greater social security rights than non-Hukou holders.

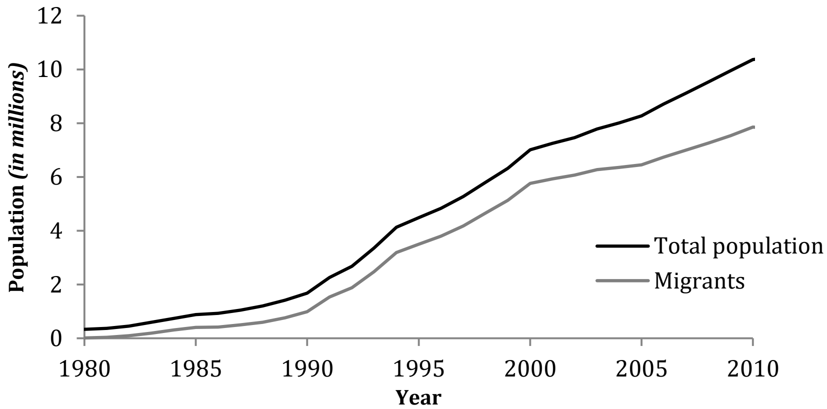


FIGURE 1 — POPULATION AND MIGRANTS IN SHENZHEN SINCE 1980

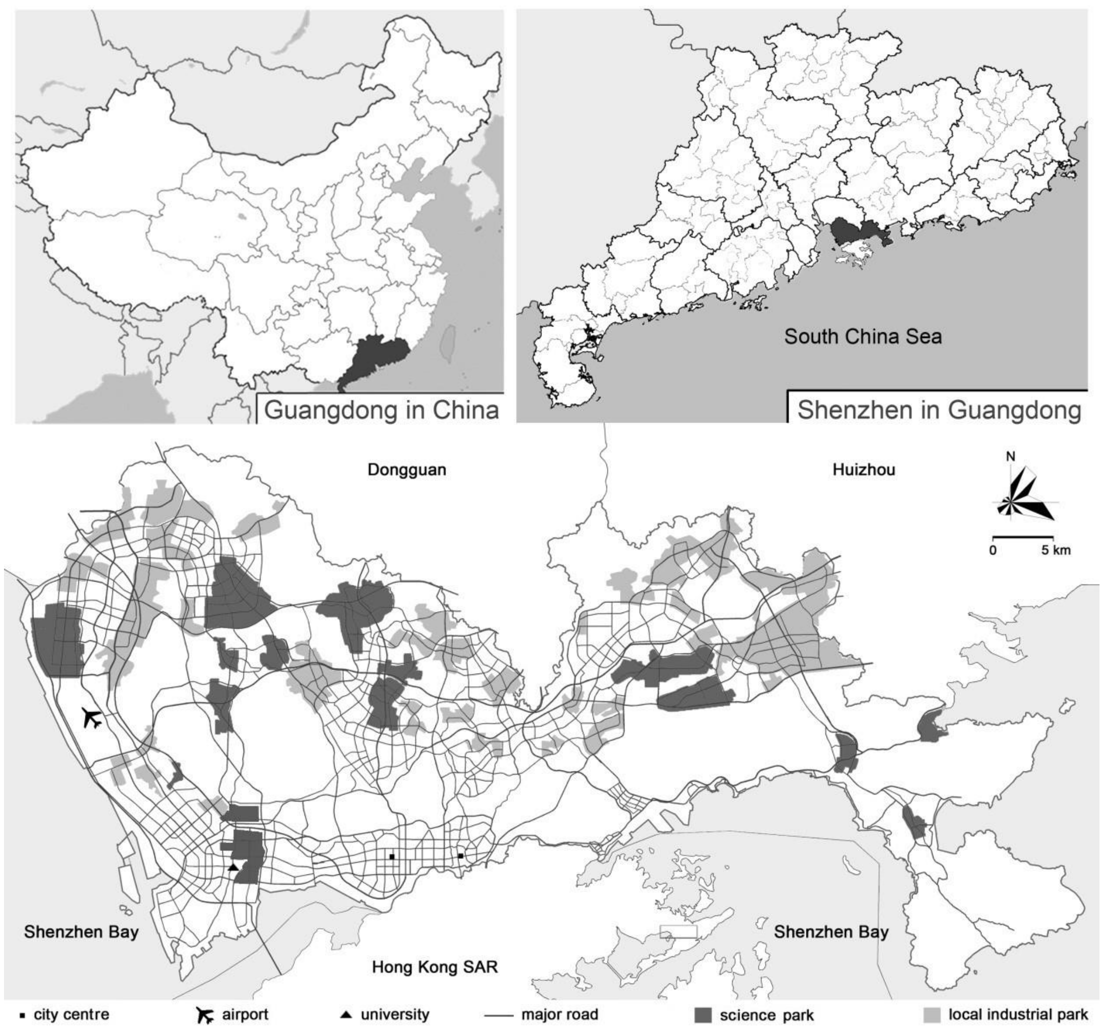


FIGURE 2 — GEOGRAPHIC DISTRIBUTION OF SCIENCE PARKS IN THE SHENZHEN AREA

Industrial park (SHIP) today) was established in 1985. It was selected and certified as one of the first six national-level science parks in China in 1996 (Ma and Chen, 2010). In addition, in 1998, the Shenzhen municipal government issued a policy document with regulations to further support the growth and development of high-tech industries. This document includes a package of preferential policies that apply exclusively to firms in science parks (Cheng et al., 2014). First, firms in these areas are offered a streamlined channel to government agencies that guarantees their ability to obtain more efficient services from the government. Second, science park agencies provide services in terms of financing, management, administration and marketing. Third, firms in science parks also have easier access to start-up funds and to a special fund that was exclusively established for science and technology firms in science parks (*Keji Sanxiang Jingfei*). This fund provides 10 million yuan each year to returned-from-overseas Chinese start-ups and 20 million yuan for R&D activities. Firms engaged in high-tech activities in science parks can receive R&D subsidies in amounts equivalent to 3-10 per cent of their total sales. Fourth, facilities such as banks, schools, restaurants, post offices, meeting places and supermarkets are offered in the science parks. Fifth, land policies are different inside science parks. Although all land in China is formally state-owned, firms in science parks typically enjoy exemptions from land use fees and other fees related to land leases. They are also exempt from property taxes for self-built or purchased properties for five years following start-up.[5]

The Shenzhen municipal government's policies have led to competition among the high-tech parks within the city (Cheng et al., 2014). By 2007, Shenzhen had fifteen science parks (see Figure 2). In addition, existing local industrial parks and other agglomerations may be upgraded in the future to science parks. The strategy is that by improving the economic conditions, developing market-based institutions and providing financial capital and subsidies to firms in science parks, the firms will be more likely to invest in physical and human capital, which will in turn lead to innovations and sustainable economic growth. Possibly as a result of these preferential policies, the Shenzhen region has seen a rapidly increasing number of high-tech firms since 2000.

B. Data and variables

The establishment-level data are acquired from the Chinese Industrial Statistics Database. This database is maintained by the Chinese National Statistical Bureau through compulsory registration and an annual firm survey collected by lower level statistical bureaus. The survey provides information on manufacturing firms. In total, there are over 3 million firm records over the period from 1998-2009. The database that we use in this paper only includes firms with annual turnover exceeding 5 million yuan (approximately 0.8 million USD). Only larger firms are included, but because the target population for the survey

[5] The extent to which these land use policies lead to changes in land consumption is investigated in Appendix B.

consists of companies that generate more than 90 per cent of China's total industrial output, any selection bias is expected to be small. In this study, we have data on firms located in the Shenzhen area for the years 1998, 1999, 2001, 2004, 2006 and 2007. After excluding unreliable observations (less than 5 per cent), we obtain 22,535 observations on 9,345 firms. Firms are not necessarily present in all years because they may have been too small in certain years, went out of business or relocated to areas outside Shenzhen.

The data provide information on firm location and the most important firm characteristics, such as fixed assets (capital), employment and taxes. Using the year of establishment, we calculate the firm's age and determine whether a firm is a start-up. We also have information on ownership status. Specifically, we know whether a firm is state-owned, privately owned, is owned by a firm from Hong Kong, Taiwan, or Macau (HTM), or is foreign-owned. Because we are uncertain whether state-owned firms have an incentive to maximise profits, we exclude them from the analysis (only 2 per cent of the observations). Of the total observations, 2.2 per cent relate to firms with multiple establishments. Because we are not sure whether these other establishments are located in other zones, we drop them from our dataset.

To control for the effect of spatial factors that may impact firm productivity, we use spatial differencing: we compare firms that are close to science park boundaries. To make the identification strategy more convincing, we collect additional data from the 2007 Land Use Survey conducted by the Shenzhen Planning Bureau. We calculate the distance from the centroid of each neighbourhood to the nearest employment centre, airport, seaport and the nearest highway ramp to control for accessibility. Additionally, we include a dummy variable that indicates whether a firm is located in the Special Economic Zone (SEZ), where preferential policies applied to firms in our period of analysis.[6] Zoning has also been taken into account. In Shenzhen, strict land use regulations prohibit firms from locating in restricted zones, such as ecological protection zones, water reservoirs, and areas near polluting production facilities and the nuclear power plant.

We have information on the location of firms at the neighbourhood level (*shequ* in Chinese). Neighbourhoods are generally quite small (316 hectares on average), so a science park usually consists of multiple neighbourhoods.[7] One problem is that the science park boundaries do not necessarily match the neighbourhood boundaries. However, the boundaries are sometimes very close to the neighbourhood boundaries, and sometimes no firms are located in the remainder of the neighbourhood no firms (e.g., because it is a restricted zone). We therefore let the science park boundaries overlap with neighbourhood boundaries when one of the latter observations applies. We then calculate the share of each

---

[6] Since the 1990s, preferential policies in the SEZ have gradually been suspended. The last and most important policy, that firms in the Shenzhen SEZ enjoy a tax rate of 15 per cent, compared with 33 per cent elsewhere, was suspended in 2008.

[7] These neighbourhoods are approximately twice the area of a census block in the United States (which is on average approximately 112 square hectares).

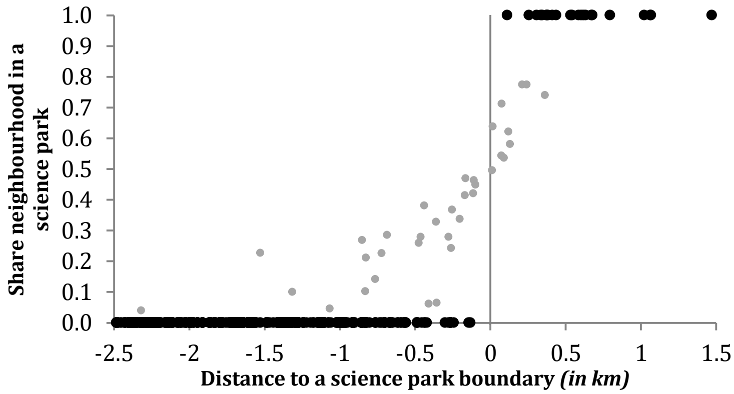
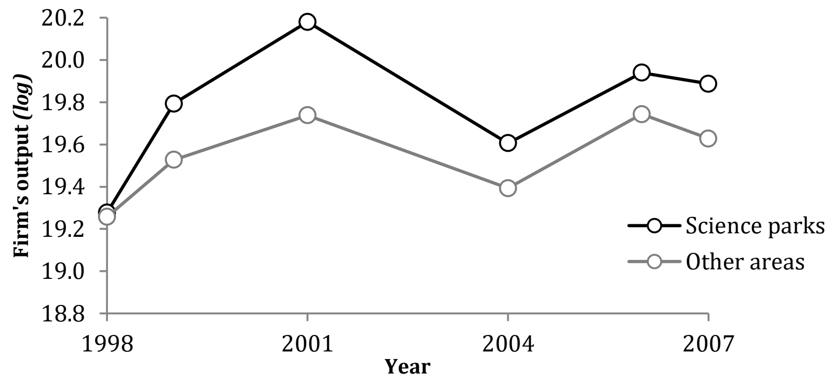


FIGURE 3 — AVERAGE DISTANCE OF THE NEIGHBOURHOOD TO THE SCIENCE PARK BOUNDARY

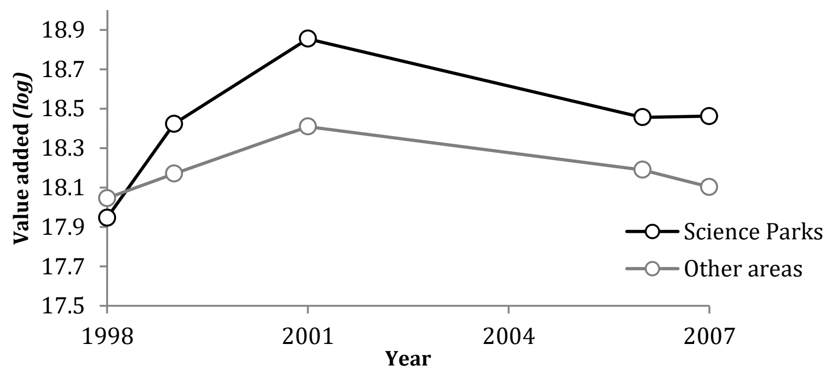
TABLE 1 — DESCRIPTIVE STATISTICS OF NATIONAL STATISTICS BUREAU SAMPLE

	Other areas				Science parks			
	μ	σ	min	max	μ	σ	min	max
Output *(in 10,000,000 ¥)*	137.777	1,178.534	0.037	75,736.800	297.104	2,254.364	0.163	75,736.800
Value added *(in 10,000,000 ¥)*	38.616	485.911	0.006	40,943.920	82.326	683.120	0.001	25,974.810
Wage per worker *(in 10,000 ¥)*	21.599	22.407	7.500	750.000	26.964	26.549	7.516	437.834
Science park *(in % of the neighbourhood)*	0.027	0.097	0.000	0.775	0.768	0.302	0.039	1.000
Distance to science park boundary *(in km)*	-3.954	3.152	-14.049	1.471	0.213	0.545	-2.321	1.471
Special economic zone	0.311				0.350			
Distance to employment centre *(in km)*	15.551	9.899	0.350	48.032	15.327	9.580	2.318	38.888
Distance to highway ramp *(in km)*	4.249	3.282	0.261	31.102	4.063	2.853	0.573	26.360
Distance to airport *(in km)*	23.369	14.011	0.849	77.648	20.638	12.992	4.796	70.108
Distance to seaport *(in km)*	19.779	8.963	0.700	39.886	20.031	8.508	8.491	33.761
Restricted zone	0.262				0.189			
Employees	397.451	1,120.086	2.000	82,067.010	567.525	2,480.965	2.000	82,067.010
Capital *(in 10,000 ¥)*	3.397	43.919	0.000	2,525.370	3.846	17.762	0.000	751.068
Start-up	0.010				0.013			
Age 0-5	0.425				0.445			
Age 6-10	0.319				0.312			
Age >10	0.256				0.243			
Privately owned firm	0.334				0.393			
HTM-owned firm	0.511				0.424			
Foreign-owned firm	0.155				0.183			
Tax rate	0.017	0.080	-1.000	1.000	0.019	0.084	-1.000	1.000

Notes. The effective number of observations for firms outside of science parks is 20,739 and for firms inside science parks is 4,279. For value added and wages we have 15,229 and 19,210 observations outside science parks, respectively, and 3,310 and 4,048 observations inside science parks, respectively. We exclude state-owned firms.



(A) LOG OF FIRMS OUTPUT



(B) LOG OF FIRMS VALUE ADDED

FIGURE 4 — OUTPUT AND VALUE ADDED IN SCIENCE PARKS AND OTHER AREAS

neighbourhood that is in a science park in the year prior to the year of observation. To calculate the distance to the science park boundary for each neighbourhood, we calculate for each point (using a 5 metre by 5 metre grid) within the neighbourhood the distance to the nearest science park boundary. Then, we take the average of the distances within the neighbourhood (where negative distances are locations outside of the science parks). Figure 3 plots the relationship between the average distance to the nearest science park boundary in a neighbourhood and the share of the neighbourhood that is part of a science park for 2007. Because we do not know the exact location of firms and we use micro data, measurement errors may occur, which can potentially lead to a downward bias of the estimated effect of science parks in a linear model. To avoid the problem of measurement error, we exclude observations in neighbourhoods that have shares that deviate from zero or one (approximately 1 per cent of the neighbourhoods, indicated by the grey dots in Figure 3). We also use another (cross-sectional) dataset in the robustness analysis for which we have the exact location of firms (Section VI.E).

In Table 1, we present descriptive statistics of the variables used in our analysis. The table shows that the average output is much higher for firms in science parks. However, this difference may be caused by some very large firms (such as Foxconn, a large high-tech multinational). The number of workers is indeed somewhat higher in science parks, but the amount of physical capital is not. Firm age and the number of start-ups are on average similar across science parks and other areas.

In Figure 4, we plot the average logarithm of firm output and the logarithm of the value added over the years for science parks and other areas. It is apparent that productivity in science parks is approximately 40 per cent higher compared with productivity in other areas (Figure 4A). Although we do not include any control variables, this figure provides some suggestive evidence that firms in science parks are more productive or that science parks attract productive firms. In Figure 4B, we plot value added over time. The difference between science parks and other areas is approximately 25 per cent. Note that we do not have data on value added for the year 2004.

For approximately 20 per cent of the firms, we observe a change in firm location. Only 20 percent of the moving firms originating from a regular area relocate into a science park. For about 75 percent of the relocations, the science park status does not change (so either stay inside or outside science parks). The remaining firms move from a science park into a regular area. The average moving distance is 8.9 kilometres and the median moving distance is 4.96 kilometres. Both of these observations suggest that the science park policy not just attracts firms from across the boundary.

IV. Econometric framework

A. Set-up

We assume that productivity q_{jzt} of firm j located in neighbourhood z in year t can be described by a Cobb-Douglas production function (see Moretti, 2004; Greenstone et al., 2010):

(7)
$$q_{jzt} = A_{zt} \prod_{M=1}^{M} x_{jtm}^{\alpha_m} \nu_{ijt},$$

where x_{jtm} is an input m, such as labour and capital, $m = 1, \dots, M$, α_m is a productivity parameter, A_{zt} denotes the location-specific technology level, and ν_{jzt} captures unobserved heterogeneity. It is assumed that $\log \nu_{jzt} = \eta_{jst} + \phi_t + \epsilon_{jzt}$, where η_{jst} denotes industrial sector s fixed effects (which may sometimes change over time for a firm), ϕ_t are time fixed effects and ϵ_{jzt} is an independently and identically distributed error term. We may assume that technology A_{zt} only depends on the share of a neighbourhood in a science park p_{zt} (e.g., because of institutional arrangements and technology spillovers between firms located there), so $A_{zt} = \mathrm{e}^{\beta p_{zt}}$. We note that p_{zt} is positive for firms in a science park in the year *after* the opening of the science park, which seems reasonable because we do not know the exact opening date of the science park. Equation (7) does not identify a causal effect of science parks *(i)* if the locations of science parks are not randomly distributed over space and *(ii)* if more productive firms sort themselves in science parks.

Most likely, the locations of science parks are not randomly chosen. For example, the selection may be based on the existing spatial distribution of high-tech industries. We address this issue by using spatial differencing. We compare science parks with neighbouring areas that are untreated but that are otherwise similar (Black, 1999; Bayer et al., 2007). Hence, we estimate a weighted regression where the weight is given by:

(8) $$w_{z\bar{t}} = \left(1 - \frac{d_{z\bar{t}}}{d_T}\right) 1_{d_{z\bar{t}}<d_T},$$

which implies that the weight is zero when a location is further than the threshold distance d_T kilometres from a science park boundary in $\bar{t} = 2007$. We include a set of boundary 'fixed effects' $\theta_{zb\bar{t}}$, which are calculated as the share of a neighbourhood that is closest to boundary b in $\bar{t}$ and zero otherwise.[8] Thus:

(9) $$\log q_{jzt} = \alpha \log x_{jt} + \beta p_{zt} + \theta_{zb\bar{t}} + \phi_t + \eta_{jst} + \epsilon_{jzt}.$$

We also control for observable neighbourhood variables denoted by n_{zt}. In Appendix B, we test for robustness of the results with respect to the choice of d_T and experiment with flexible spatial trends that may pick up unobserved factors. Note that spatial differencing implies that β will be an underestimate of the causal effect when the benefits of science park policies extend beyond the science park boundaries (e.g. fast internet). However, we have argued that most policies exclusively apply to firms inside science parks, so we do not consider this as a major problem.

However, the above equation does not control for the fact that firms may be more productive for unobserved reasons and sort themselves in science parks. It may even be that less productive firms are banned from science parks, while they are allowed to locate elsewhere. We therefore include firm fixed effects, which leads to:[9]

(10) $$\log q_{jzt} = \alpha \log x_{jt} + \beta p_{zt} + \zeta n_{zt} + \theta_{zb\bar{t}} + \phi_t + \eta_{jst} + \omega_j + \epsilon_{jzt},$$

where ω_j denotes a firm fixed effect. Hence, we assume that unobserved firm effects are log-separable from other effects. By including firm fixed effects, we exploit temporal variations in the assignment of science parks: we compare firms' productivity before and after the opening of a science park vis-à-vis the productivity of a firm located close to a science park. A second source of identifying variation is derived from firms that relocate. The above equation identifies a causal effect if unobserved firm productivity trends are uncorrelated with science parks.

However, one may argue that firms relocate in certain productivity stages. Furthermore, we may not properly control for all time-invariant location attributes that may be correlated

[8] Note that if we knew the exact location of firms rather than at the neighbourhood level, it would lead to the inclusion of dummies that equal one when the observation is closest to boundary b.

[9] It should be noted that by including firm fixed effects, we disregard the productivity effects for start-ups or firms that close down. Because we already control for firm age and whether a firm is a start-up in the previous specifications, this approach is unlikely to cause problems.

with the assignment of science parks. In the empirical section, we therefore also estimate a specification where we include firm-neighbourhood fixed effects ω_{jz}. Hence, we control for all time-invariant characteristics of a location and a firm. Because we identify the effect on staying firms, it is unlikely that the comparison of nearby companies along the science park boundary is confounded by selection of companies due to shocks to their future prospects to innovate and productivity. We will test this in more detail in Appendix B. Another worry one might have is that β will be an overestimate if firms just move across the boundary, as is often recorded in in the previous literature (see e.g. Einiö and Overman, 2012). We think this not a problem when we include firm-neighbourhood fixed effects because we do not identify the effect on moving firms but on firms that stay. Nevertheless, to fully rule out the possibility that the regression-discontinuity design leads to an overestimate, we pursue an alternative identification strategy in Section VI.B where control locations are not (necessarily) geographically close to treatment areas.

V. Regression results

A. Productivity effects

Table 2 reports the estimates based on Equations (7)-(10). We start with a simple regression based on Equation (7), where we regress the logarithm of output on whether a neighbourhood is part of a science park while controlling for workforce size, the capital stock and year fixed effects.[10] The coefficient suggests that firms are 26 per cent more productive in science parks.[11] This effect should not be interpreted as a causal effect of science park policies because the allocation of science parks may be non-random. In column (2), we include additional firm control variables and industrial sector fixed effects. The coefficient is somewhat lower and statistically insignificant.

When the location of science parks is not random and depends on geographic location and industrial conditions, this estimate is likely to be biased. For example, it might be that science parks are established at locations that have favourable geographical features, such as access to highways. We therefore use spatial differencing. This approach should address the problem of spatial unobservables, as we focus on areas close to science park boundaries, which should have a similar geography. On such a small spatial scale, the boundary of science parks is thought to be random. We then only include observations that are within 2.5

[10] To estimate the standard errors of the parameters of interest, one may cluster standard errors over space (at the science park area or neighbourhood level) or over time (at the firm level). Our pragmatic approach is that we cluster at the neighbourhood level because it leads to the largest standard errors and therefore to the most conservative conclusions. Nevertheless, it appears that the standard errors are very similar no matter whether we cluster at the science park, neighbourhood or firm level. We may also estimate a system of equations using seemingly unrelated regressions, including the regression equations for employment and capital. This approach would be more efficient and would therefore lead to lower standard errors.

[11] The marginal effect is calculated as $e^{\hat{\beta}} - 1$.

TABLE 2 — BASELINE REGRESSION RESULTS ON THE IMPACT OF SCIENCE PARKS ON PRODUCTIVITY
(Dependent variable: the logarithm of firms' yearly output)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.230*	0.174	0.219***	0.231***	0.239**	0.195**	0.165*
	(0.124)	(0.107)	(0.0778)	(0.0725)	(0.100)	(0.0881)	(0.0954)
Employees *(log)*	0.546***	0.614***	0.631***	0.635***	0.588***	0.587***	0.582***
	(0.0164)	(0.0152)	(0.0282)	(0.0288)	(0.0443)	(0.0438)	(0.0432)
Capital *(log)*	0.267***	0.225***	0.249***	0.247***	0.0949***	0.0926***	0.0886**
	(0.0113)	(0.00897)	(0.0158)	(0.0158)	(0.0329)	(0.0327)	(0.0351)
Start-up		-0.281***	-0.167	-0.160	-0.471***	-0.478***	-0.466***
		(0.0689)	(0.114)	(0.112)	(0.149)	(0.150)	(0.156)
Age 0-5		-0.0167	0.101*	0.105*	0.0522	0.0521	0.0460
		(0.0295)	(0.0540)	(0.0540)	(0.0720)	(0.0719)	(0.0757)
Age 6-10		0.0290	0.136***	0.133***	0.0737	0.0770	0.0799
		(0.0254)	(0.0408)	(0.0409)	(0.0556)	(0.0558)	(0.0583)
HTM-owned firm		-0.128***	-0.0689**	-0.0651**	0.190*	0.194*	0.208
		(0.0279)	(0.0335)	(0.0328)	(0.110)	(0.111)	(0.127)
Foreign-owned firm		0.172***	0.199***	0.204***	0.155	0.157	0.186
		(0.0372)	(0.0624)	(0.0624)	(0.101)	(0.101)	(0.113)
Special economic zone (SEZ)				0.180**	0.195	0.940**	
				(0.0861)	(0.347)	(0.473)	
Distance to employment centre *(log)*				-0.321**	0.0548	0.412**	
				(0.144)	(0.149)	(0.189)	
Distance to highway ramp *(log)*				-0.115***	-0.0813	-0.126	
				(0.0391)	(0.149)	(0.176)	
Distance to airport *(log)*				-0.0128	0.234	1.965	
				(0.150)	(0.308)	(1.709)	
Distance to seaport *(log)*				0.507*	-0.214	-1.217	
				(0.290)	(0.538)	(5.752)	
Restricted zone				0.0442	0.172	-0.00818	
				(0.0820)	(0.240)	(0.265)	
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311
R^2	0.549	0.597	0.641	0.643	0.942	0.943	0.947

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

kilometres of a science park boundary and include science park boundary fixed effects. In column (3), it can be observed that firms are 24 per cent more productive in science parks. The effect is almost identical if we include a range of neighbourhood control variables, such as distance to the nearest employment centre, distance to the international air- and seaports, and distance to the nearest highway ramp (column (4)). Because the coefficient hardly changes, this suggests that boundary fixed effects capture important locational endowments reasonably well.

In column (5) of Table 2 we include firm fixed effects to control for sorting. Because of the entry requirements, low-productivity firms may not be allowed to locate in science parks. However, by investigating the productivity of the same firm before and after the opening of a science park, we control for these entry requirements. The coefficient is now somewhat higher: science park policies seem to have increased firms' productivity by 27 per cent, which is very similar to the previous specifications. One may argue that neighbourhood variables may not capture all of the spatial variables that may be correlated with the boundaries of science parks. In column (6), we therefore add a flexible fifth-order polynomial function of geographic coordinates. The results indicate that the effect is slightly lower compared with the previous specification, but it still is statistically significant at the five per cent level. Another approach is to include firm-neighbourhood fixed effects to control for all time-invariant location attributes. Column (7) indicates that firms that do not relocate have seen an increase in productivity of 18 per cent after the opening of a science park.

The results undoubtedly suggest a strong and meaningful productivity effect of science parks, even if we control for firm selection. Our effects are in the same order of magnitude as those of Wang (2013), who finds even stronger productivity effects of special economic zones (up to 65 per cent).

The control variables have plausible signs. Production inputs, such as employment and capital, increase productivity. Start-ups are much less productive (15-61 per cent). Conditional on being a start-up, younger firms seem to be somewhat more productive. Furthermore, we find some evidence that foreign-owned firms have a higher output (17-23 per cent). The effect of distance to employment centres is unclear: when we include firm fixed effects, the sign switches. However, we note that if we would exclude the distance to the nearest employment centre, the effect of science parks is hardly affected.

B. Wages

We find a substantial productivity effect of the science park policy. The subsequent question then is whether these productivity increases have translated into higher wages. If, for instance, workers are very mobile and are elastically supplied, it should imply that wages have not increased due to science park policies. Conversely, when workers have strong idiosyncratic preferences and are inelastically supplied, we expect positive wage increases due to the opening of science parks. We therefore also run wage regressions. Wage is defined

TABLE 3 — REGRESSION RESULTS OF THE IMPACT OF SCIENCE PARKS ON FIRM WAGES
(Dependent variable: the logarithm of the average wage)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.202**	0.187**	0.159***	0.147***	0.112**	0.128***	0.112***
	(0.0878)	(0.0737)	(0.0486)	(0.0408)	(0.0479)	(0.0433)	(0.0427)
Firm variables (5)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,582)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,773)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	17,814	17,814	7,697	7,697	7,697	7,697	7,697
R^2	0.182	0.242	0.334	0.343	0.778	0.780	0.790

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.
*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

as the average wage per worker in a firm (implying that we divide the total wage bill by the number of employees). This measure may be slightly noisy if there are outliers (e.g., CEOs who earn a lot compared with the average worker). It is also not available for all firms (approximately 7.5 per cent of the observations are missing in this respect). Table 3 reports the regression results.

We again start with the naïve specification in Table 3, column (1). The coefficient seems to suggest that wages are 22 per cent higher in science parks. This effect is slightly smaller once we include firm control variables and industrial sector fixed effects (column (2)). In column (3), we use spatial differencing to improve on the identification of a causal effect of science park policies on wages. Again, we find a positive and meaningful effect of science parks on wages (17 per cent). This coefficient is very similar when we include neighbourhood variables in column (5). Once we control for firm sorting by including firm fixed effects in column (4), the effect of science park policies on wages is 11.9 per cent. This effect becomes statistically stronger when we include a flexible function of geographic coordinates (column (6)) or when we only focus on firms that have not relocated by including firm-neighbourhood fixed effects (column (7)). Hence, the results unambiguously suggest that wages have become higher due to science park policies, which implies that the policies have at least generated some benefits for local workers, if that the local costs of living have not increased. The wage effect, however, is somewhat smaller in magnitude than the productivity effect, which suggests that displacement effects might be present.

C. Displacement effects and deadweight losses

The previous analyses suggest a substantial productivity effect of science parks and higher wages for employees working in science parks. In a perfectly competitive market with zero profits and a somewhat elastic supply of labour, place-based policies may lead to wage differences but may also generate displacement effects. Table 4 reports the results of regressions where the dependent variable is either the workforce size or the capital stock of firms.

In Panel A of Table 4 we test the impact of place-based policies on workforce size. The naïve specification in column (1) suggests that firms in science parks do not have a larger workforce than firms outside of those areas. This result also holds if we include firm control variables and industry fixed effects (column (2)), if we use spatial differencing (column (3)), and if we include neighbourhood variables (column (4)). However, when we control for firm sorting by including firm fixed effects in column (5), the results show that the coefficient becomes positive and statistically significant. This result suggests that firms are 16 per cent larger due to science park policies. When we include a flexible function of geographic coordinates in column (6), the point estimate is similar but not statistically significant at conventional levels. This result also holds if we include firm-neighbourhood fixed effects in column (7). Hence, although the estimates are imprecise, the policies seem to have generated some displacement effects with respect to labour.

In Panel B of Table 4, we investigate whether the increase in workforce size is accompanied by increases in the capital stock. This inquiry is interesting, as displacement effects may refer to both changes in the labour force and in the capital stock (see equation (5)). Column (1) provides the naïve regressions without any locational or firm controls. The coefficient suggests that firms operating in science parks use more capital. This result also holds if we include firm control variables and if we focus on observations close to the science park boundaries, including science park fixed effects. However, this effect is unlikely to be a causal effect of science park policies because high-tech firms that have relatively high shares of capital use are more likely to end up in science parks because of entry requirements. Indeed, once we include firm fixed effects, the coefficient becomes statistically insignificant and essentially equal to zero (columns (4)-(7)).

Using the framework developed in Section II, we can estimate the productivity effect net of changes in factor use, which we will refer to as the 'gross relative welfare effect', and the deadweight losses associated with science park policies. To estimate the net welfare effect, we would need information on the costs of the programme, which is not public. It should be noted that the results presented in Table 5 depend on a model with two generic regions, and capital and labour as the only production factors, so the results should be interpreted with

TABLE 4 — REGRESSION RESULTS OF THE IMPACT OF SCIENCE PARKS ON WORKFORCE SIZE AND THE CAPITAL STOCK

(Dependent variable: the logarithm of firms' employment)

Panel A: Workforce size	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	-0.0579	-0.0313	0.0507	0.0728	0.149**	0.131	0.110
	(0.0886)	(0.0933)	(0.0585)	(0.0593)	(0.0745)	(0.0822)	(0.0889)
Firm variables (5)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311
R^2	0.012	0.165	0.206	0.212	0.936	0.936	0.941
Panel B: Capital stock	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.465***	0.401***	0.467***	0.485***	0.0121	-0.0305	-0.0538
	(0.155)	(0.116)	(0.144)	(0.142)	(0.109)	(0.107)	(0.0926)
Firm variables (5)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311
R^2	0.053	0.202	0.235	0.239	0.911	0.911	0.918

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

caution. Given that it is not possible to observe individual instruments of the science park with our data (such as actual explicit or implicit factor employment subsidies or the cost of funding), we therefore assume that the causal effects of science parks on firms' productivity and employment choices are the result of one policy bundle, which is captured by the variable p_z. We then use equation (6) to estimate the gross welfare effect and deadweight losses:

$$\text{(11)} \quad \Delta\hat{\mathcal{RW}} = e_A \frac{\mathrm{d}A_{zt}/\mathrm{d}p_{zt}}{A_{zt}} - \alpha_\ell e_\ell - \alpha_k e_k = \left(\mathrm{e}^{\hat{\beta}_A} - 1\right) - \hat{\alpha}_\ell\left(\mathrm{e}^{\hat{\beta}_\ell} - 1\right) - \hat{\alpha}_k\left(\mathrm{e}^{\hat{\beta}_k} - 1\right),$$

where the circumflexes refer to estimated values. The welfare effect comprises three terms: the direct productivity increase $(\mathrm{e}^{\hat{\beta}_A} - 1)$; and the two deadweight losses on the markets for labour and capital. They are of comparable magnitude here, because they are all scaled to relative change to output value. Table 5 reports the results when we estimate equation (11). The column numbers refer to their respective specifications in Tables 2 and 4. To calculate the standard errors, we use seemingly unrelated estimation (SUE) and the delta method because we have nonlinear combinations of coefficients.

Column (1) in Table 5 relates to the naïve regression. The results suggest a positive gross welfare effect of 14 per cent, but this effect is far from statistically significant, partly because of a seemingly substantial positive displacement effect of capital. In columns (2), (3) and (4), the gross relative welfare effect is also statistically insignificant. However, these results are mainly due to a strong and seemingly large displacement effect related to capital. The results suggest a total deadweight loss of approximately 45 to almost 60 per cent of the total productivity effect $(\hat{\alpha}_\ell(\mathrm{e}^{\hat{\beta}_\ell} - 1)/(\mathrm{e}^{\hat{\beta}_A} - 1))$. However, these displacement effects are unlikely to be causal effects of science park policies because high-tech, capital-intensive firms sort themselves into science parks.

When we move to our preferred estimates based on spatial differencing and firm fixed effects (columns (5)-(7)), we find a gross relative welfare effect of 10.9 per cent, although this effect is not statistically significantly different from zero at conventional levels (p-value = 0.211) (see column (5)). The results in column (5) of Table 5 further suggest that the deadweight loss related to labour is 12.7 percent, or about 50 percent of the productivity effect. Note again that because the estimated effects are functions of different coefficients, the confidence intervals are quite wide. This also holds for the results in columns (6) and (7). The gross relative welfare effect appears to be positive and economically meaningful, but the deadweight loss related to labour becomes smaller once we include firm-neighbourhood fixed effects in column (7). Nevertheless, it is still about 40 percent of the productivity effect $(\hat{\alpha}_\ell(\mathrm{e}^{\hat{\beta}_\ell} - 1)/(\mathrm{e}^{\hat{\beta}_A} - 1))$

Thus, it seems that the gross relative welfare effect is about 5-16 percent, albeit imprecise. The science park policy elicits workers who would otherwise not work in the science park. Meanwhile, the responses to the use of capital do not seem to imply deadweight losses. First, the capital employment is not statistically significantly sensitive to science park policies in the most reliable estimates (Table 5, columns (4)-(7)), and second, its cost share is relatively low, with a Cobb Douglas weight of 0.08. Note that the assumption of a fixed capital financing rate in the theoretical model should not affect the result: with inelastic capital responses, the deadweight losses due to capital market distortions are small.

TABLE 5 — ESTIMATES OF THE GROSS RELATIVE WELFARE EFFECT AND THE DEADWEIGHT LOSS

	(1) SUE	(2) SUE	(3) WSUE	(4) WSUE	(5) WSUE	(6) WSUE	(7) WSUE
Gross relative welfare effect, $\Delta\hat{\mathcal{R}}\mathcal{W}$	0.131	0.0976	0.0636	0.0580	0.109	0.157*	0.116
	(0.166)	(0.162)	(0.101)	(0.0990)	(0.0872)	(0.0899)	(0.0933)
Deadweight loss – labour	-0.0307	-0.0188	0.0328	0.0479	0.127***	0.0879**	0.0677
	(0.0459)	(0.0556)	(0.0388)	(0.0404)	(0.0378)	(0.0381)	(0.0414)
Deadweight loss – capital	0.158**	0.111***	0.148***	0.154***	0.00355	-0.00417	-0.00464
	(0.0653)	(0.0383)	(0.0551)	(0.0552)	(0.00826)	(0.00711)	(0.00541)
Deadweight loss – total	0.127	0.0921	0.181**	0.202**	0.131***	0.0838**	0.0631
	(0.0900)	(0.0764)	(0.0824)	(0.0871)	(0.0427)	(0.0408)	(0.0419)
Firm variables (5)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311

Notes: WSUE stands for Weighted Seemingly Unrelated Estimation. The column numbers refer to the coefficient estimates from Tables 2, 4 and 5. Standard errors are calculated using the delta method and clustered at the neighbourhood level. The standard errors are in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

The high deadweight loss estimate of labour relative to the direct policy benefits differs from related studies on U.S. place-based policies. Busso et al. (2013) find deadweight losses of up to roughly fifty per cent of our estimates (comparing their figures that are not corrected for the marginal costs of public funds). In part, the variation could be explained by methodological differences. We compare firms around a border, whereas Busso et al. compare 'runner-up' tracts that actually experienced the policy. The obstacles to changing jobs may therefore be lower in our sample. The role of amenities and housing contributes relatively little to the deadweight loss estimates of Busso et al., so it is unlikely that they explain the difference from our analyses. An obvious and important explanation for the differences is the institutional context of the policies. The Chinese labour market has different institutions and cultures, and production is relatively labour-intense, which potentially provides extra weight to labour market policies (Hsing, 2010). The bargaining position of Chinese manufacturing workers appears to be relatively weak, and workers may be easily substitutable due to a large supply. Moreover, the rationale of the two place-based policies is different: the Chinese science park policies target relatively well-performing areas, whereas the US empowerment zones intend to stimulate economically lagging regions.

Clearly, workers' motives to move to economically leading areas where the policy leads to (relatively high-skilled) job demand are different from the motives to move into economically lagging areas (Cazes and Verick, 2013). The fact that the results point towards a deadweight loss in labour may seem surprising given that high-tech, capital-intensive firms in particular are attracted to science parks. However, high-tech industries in China are much more labour-intensive than high-tech industries in many Western countries.

VI. Sensitivity analysis

A. Introduction

We have argued that the estimated productivity and wage effects may be interpreted as a causal effect of place-based policies. In this subsection, we report what we consider as the most important robustness checks. We first consider an alternative identification strategy using local industrial parks as control areas instead of locations close to science park boundaries. We then proceed with the use of an alternative measure of productivity. Third, we examine whether the presence of agglomeration economies, proxied by firm density, may provide an explanation for the productivity effect. If agglomeration economies are an important driver of the productivity effect, there may be substantially different welfare implications. Finally, we use another (cross-sectional) dataset to corroborate our results and investigate whether the omission of information on land use matters for the results.

We relegate another set of sensitivity analyses to Appendix B. There, we investigate whether the inclusion of firm-specific linear trends affect the main results. We also we investigate whether the effect of science park policies is different for domestic and foreign-owned firms and we check whether our results can be explained by differences in tax regimes. We continue with testing and relaxing the assumptions made in the spatial differencing estimation strategy. Eventually, we test robustness of the results by deliberately ignoring measurement error by also including observations in neighbourhoods that are partly in science parks.

B. Alternative identification strategy: local industrial parks vs. science parks

The approach to identify a causal effect of science park policies is to combine spatial differencing with temporal differencing. We may also use another source of identifying variation based om classification of different dedicated areas. There are two categories: local industrial parks and science parks. Only in the latter attractive institutional arrangements described in Section II are offered. However, local industrial parks are often upgraded to science parks later on and can therefore be considered as a feasible control group. A similar approach has been used by Wang (2013) and Busso et al. (2013) who also use 'runner-up' locations as a counterfactual. If the upgrading of local parks into science parks is random over time, or is at least not correlated with ϵ_{jzt}, β measures a causal impact of science parks on productivity. We then estimate a weighted regression where the weight is equal to the share of the neighbourhood in a local industrial park or science park in 2007.

TABLE 6 — SENSITIVITY ANALYSIS: LOCAL INDUSTRIAL PARKS VS. SCIENCE PARKS
(Dependent variable: the logarithm of firms' yearly output)

	(1) WLS	(2) WLS	(3) WLS	(4) WLS
Science park	0.197***	0.153***	0.160	0.155
	(0.0448)	(0.0400)	(0.105)	(0.110)
Firm variables (7)	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	Yes	Yes	Yes
Year FE (6)	Yes	Yes	Yes	Yes
Industry FE (33)	Yes	Yes	Yes	Yes
Firm FE (3,856)	No	Yes	Yes	Yes
Firm-neighbourhood FE (4,001)	No	No	No	Yes
Industrial park FE (50)	Yes	Yes	Yes	Yes
Number of observations	8,427	8,427	8,427	8,427
R^2	0.648	0.651	0.940	0.942

Notes: We estimate weighted regressions where the weight is equal to the share of the neighbourhood in a local industrial park and science park in 2007. Standard errors are clustered at the neighbourhood level and in parentheses.
*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

Table 6 reports the results for productivity. In column (1) we include science and local industrial park fixed effects, and control for firm characteristics, industry, and year fixed effects. Firms in science parks seem to be 21 percent more productive. In column (2) we control for neighbourhood attributes, leading to a slightly lower but similar and highly statistically significant coefficient (16.5 percent). In column (3) we control for unobserved firm heterogeneity by including firm fixed effects. The coefficient of science parks is very similar to the previous specification, but it is somewhat imprecisely estimated and only statistically insignificant at the 13 percent level. This is not too surprising as the effective (i.e. weighted) number of observations is only 5,217. Also when we include firm-neighbourhood fixed effects, the coefficient remains very similar, but again, somewhat imprecise. In general, the point estimates are almost identical to the ones obtained by spatial differencing in Table 2.

C. Value added

As an alternative to the total value of production, we can use an alternative measure of productivity. We then take the total value added as the measure of firms' output. We note that we do not have observations for the year 2004, so the number of observations is lower than in the baseline regressions. Table 7 reports the results.

TABLE 7 — SENSITIVITY ANALYSIS: THE IMPACT OF SCIENCE PARKS ON VALUE ADDED

(Dependent variable: the logarithm of firms' value added)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.322*	0.257*	0.187*	0.171*	0.241*	0.252*	0.204
	(0.168)	(0.145)	(0.104)	(0.0938)	(0.136)	(0.136)	(0.175)
Firm variables (7)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (5)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,194)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,338)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	14,015	14,015	6,062	6,062	6,062	6,062	6,062
R^2	0.535	0.569	0.599	0.606	0.903	0.904	0.908

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

* Significant at the 0.10 level

Column (1) is the naïve regression of locating in science parks while controlling for workforce size and capital usage, as well as year fixed effects. The results indicate that the value added is 38 per cent higher for firms in science parks. When we control for other firm variables and industry fixed effects, the effect is 29 per cent (column (2)). We then use spatial differencing without and with neighbourhood variables (columns (3) and (4), respectively). The effects of science park policies are 20.5 and 18.5 per cent, respectively. In column (5), where we control for firm fixed effects, the effect becomes somewhat stronger (27 per cent). The effect is similar once we include a flexible function of geographic coordinates (column (6)). Column (7) in Table 7 includes firm-neighbourhood fixed effects, so we only identify the effect of science parks based on science park openings. The point estimate is then very similar to previous specifications, but it is imprecise and not statistically significantly different from zero at conventional significance levels (p-value $= 0.246$). In any case, because we have fewer observations, the results are less precise than the baseline results. However, the point estimates are very similar to the regression results reported in Table 2.

D. Agglomeration economies

A part of the productivity effect may be caused by the presence of agglomeration economies. It has been widely confirmed that productivity advantages through industrial concentration

TABLE 8 — SENSITIVITY ANALYSIS: AGGLOMERATION ECONOMIES

(Dependent variable: the logarithm of firms' output)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.206*	0.152	0.204***	0.206***	0.241**	0.188**	0.171*
	(0.108)	(0.0932)	(0.0696)	(0.0695)	(0.102)	(0.0918)	(0.0988)
Agglomeration $\rho = 2.5$ *(log)*	0.0822***	0.0859***	0.0906***	0.0993***	-0.00419	0.0269	-0.0185
	(0.0260)	(0.0231)	(0.0337)	(0.0332)	(0.0696)	(0.0743)	(0.0989)
Firm variables (7)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311
R^2	0.551	0.600	0.642	0.643	0.942	0.943	0.947

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{\iota}} = (1 - d_{z\bar{\iota}}/d_T)1_{d_{z\bar{\iota}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

stimulate interactions and foster economic growth, which we refer to as agglomeration economies (Glaeser et al., 1992; Glaeser, 2008; Greenstone et al., 2010). Because of a concentration of high-tech firms in science parks, firms may be more productive due to input and output sharing, labour market pooling and knowledge spillovers, among other things. If these nonmarket interactions are important, area-based incentives might lead to an increase in social welfare, and job creation strategies may be (more) efficient.

However, it is well understood that the intensity of interactions between firms decay over space continuously, which implies that spatial differencing would lead to an *underestimate* of the productivity effects of clustering. To test whether agglomeration economies are important, we calculate the spatially weighted density of firms, following Lucas and Rossi–Hansberg (2002) and Koster et al. (2014). The weighted density of all firms $\mathcal{A}_{zt}$ for a certain location z in year t is given by:

$$\mathcal{A}_{zt} = \rho \sum_{\tilde{z}} \mathrm{e}^{-\rho d_{z\tilde{z}}} f_{\tilde{z}t} \tag{12}$$

where $d_{z\tilde{z}}$ is the distance between the centroid of neighbourhood z and the centroid of another neighbourhood $\tilde{z}$, ρ is a decay parameter and $f_{\tilde{z}t}$ is the count of firms in a certain neighbourhood $\tilde{z}$. We assume that $\rho = 2.5$, which implies that after 500 metres, the spatial

weight is approximately one-third. To avoid collinearity, we do not control for distance to the nearest employment centre in the different models. The results are reported in Table 8.
We find that the coefficients of science parks are only slightly lower when we include the agglomeration variable. Column (1) suggests a positive agglomeration effect: doubling the number of firms increases productivity by 5.7 per cent. The coefficient related to science parks is very similar to the baseline specification. When we use spatial differencing in column (3) and include neighbourhood control variables in column (4), the results are almost unaffected. In columns (5), (6) and (7) in Table 8 we include firm fixed effects, which implies that we identify the effects of agglomeration over time and by staying firms. The coefficients related to science parks are essentially the same compared to the baseline specifications. Agglomeration no longer has an effect on productivity, but it could also be that there is insufficient identifying variation to estimate this effect, given the relatively large standard errors. We have also investigated whether these results hold for different values of the decay parameter ρ, and if we only include high-tech industries in $\mathcal{A}_{zt}$, but the coefficient related to science parks is approximately the same and is hardly affected by the inclusion of the agglomeration variables.

E. SBTI data

The quality of Chinese national data has been criticised, as they may reflect politicised aggregations of data submitted by local and other statistical bureaus (Au and Henderson, 2006). While this is usually not the case for local micro data, it is worthwhile to investigate whether our results hold when another dataset is used.

Therefore, we alternatively use another cross-sectional, establishment-level dataset that is derived from the 2007 Shenzhen Industrial Enterprise Survey. This dataset is collected and maintained by the Shenzhen Bureau of Trade and Industry (SBTI) through a compulsory annual firm survey. It provides firm-specific information, including exact firm location (rather than at the neighbourhood level), employment and annual turnover. The SBTI dataset does not cover all manufacturing firms. Instead, only firms with annual turnover exceeding 5 million CNY (approximately 0.8 million USD) are surveyed. After excluding unreliable observations (less than 5 per cent), we have 8,837 firms in our dataset that generate 97 per cent of the industrial output in Shenzhen (Shenzhen Statistics Bureau 2008). We do not have detailed information on capital, but we include the use of electricity and water as proxies. Because we know the exact location of firms, we use the original boundaries of science parks rather than using data aggregated at the neighbourhood level. We also control for the share of highly educated employees and the share of employees that engage in research and development activities. Table B2 in Appendix B reports the descriptive statistics, which look similar to our baseline sample.

Table 12 reports the regression results. In column (1), we only include the science park dummy and firms' workforce size and use of electricity and water. The productivity effect is

TABLE 9 — SENSITIVITY ANALYSIS: SBTI DATA
(Dependent variable: the logarithm of firms' yearly output)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.167*	0.0648*	0.103**	0.111***	0.166***	0.135***	0.123***
	(0.0794)	(0.0329)	(0.0353)	(0.0330)	(0.0339)	(0.0307)	(0.0302)
Employees *(log)*	0.508***	0.628***	0.670***	0.667***	0.661***	0.678***	0.676***
	(0.0178)	(0.0247)	(0.0355)	(0.0352)	(0.0101)	(0.00996)	(0.0102)
Electricity usage *(log)*	0.0148***	0.0181***	0.0164***	0.0170***	0.0169***	0.0183***	0.0181***
	(0.00341)	(0.00454)	(0.00335)	(0.00352)	(0.00236)	(0.00318)	(0.00328)
Water usage *(log)*	0.0395***	0.0432***	0.0579***	0.0605***	0.0574***	0.0594***	0.0591***
	(0.0109)	(0.00563)	(0.00827)	(0.00775)	(0.00320)	(0.00317)	(0.00309)
Share college graduates		1.474***	1.531***	1.445***	1.297***	1.367***	1.350***
		(0.0934)	(0.0969)	(0.0887)	(0.0355)	(0.0594)	(0.0556)
Share R&D employees		-0.0313	-0.107	-0.153	-0.212**	-0.423***	-0.430***
		(0.427)	(0.553)	(0.556)	(0.0915)	(0.0814)	(0.0816)
Special economic zone (SEZ)				0.245***			
				(0.0602)			
Distance to employment centre *(log)*				-0.0329	0.0447	0.0394	0.156***
				(0.0290)	(0.0462)	(0.0332)	(0.0358)
Distance to highway ramp *(log)*				-0.0707***	-0.119***	-0.156**	-0.186**
				(0.0221)	(0.0345)	(0.0595)	(0.0661)
Distance to airport *(log)*				-0.0221	0.462*	0.745**	0.758*
				(0.0347)	(0.245)	(0.331)	(0.386)
Distance to seaport *(log)*				0.0271	0.408*	0.542*	2.841***
				(0.101)	(0.240)	(0.311)	(0.819)
Restricted zone				-0.0771	0.00333	-0.0617	-0.0979*
				(0.0611)	(0.0320)	(0.0468)	(0.0521)
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	No	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood FE (191)	No	No	No	No	Yes	Yes	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	8,873	8,837	4,184	4,184	4,184	2,748	2,748
R^2	0.383	0.480	0.533	0.536	0.558	0.574	0.579

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In columns (6) and (7), $d_T = 1.25$. In column (7), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

18.2 per cent. If we include additional firm characteristics and industrial sector fixed effects, we find that firms in science parks are 6.7 per cent more productive. When we employ spatial differencing, we again find a positive effect of science parks of 10.8 per cent (column (3)). This effect is hardly affected once we include the neighbourhood control variables (column (4)). In column (5), we include neighbourhood fixed effects, which implies that we compare firm productivity between science parks and areas outside of science parks but *within* the neighbourhood. Note that the effects of neighbourhood characteristics are also identified within neighbourhoods.[12] The effect of science parks is somewhat stronger: science parks seem to increase productivity by 18.1 per cent. Column (6), Table 9, improves on this result by reducing the threshold distance by a half to 1.25. The coefficient is very similar to the previous specification. In column (7), we also control for a flexible function of geographic coordinates. We still find a positive and meaningful effect of science parks on firms' productivity of 13.1 per cent. These effects are very much in line with the baseline results presented in Table 2, which increases our confidence in the results. We might also estimate regressions for employment and capital, but because we cannot include firm fixed effects to control for sorting effects, we think that those results will not be very informative.

We interpret the productivity effect as the causal effect of science parks on technology, which is captured by A_i in our theoretical model. We then control for labour and capital and assume that the consumption of land is fixed. However, when land is much cheaper in science parks, firms may substitute capital or labour for land, and (part of) the productivity effect may be explained by changes in the consumption of land. We investigate this issue further in Appendix B by exploiting additional information on rents and the use of land that is (only) available in the SBTI data. The results show that this issue does not seem to be a problem.

VII. Conclusions

In this paper, we analyse the economic impact of place-based governmental investments in science parks in the Chinese city of Shenzhen. Virtually all of the empirical studies on place-based policies traditionally examine programmes for *deprived* areas in *developed* economies. However, the welfare arguments may be different when applied to place-based policies in *leading* areas of *developing and transition* economies. We argue that especially in China, institutional circumstances, pronounced demographic and economic transitions, and a substantial rural-urban migration, enabled the application of placed-based policies in the form of science parks and special economic zones in cities on scales that are unprecedented in Western economies (Wu and Gaubatz, 2013). These policies might stimulate relatively productive firms and people and foster positive spillovers rather than reinforcing negative spillovers, as is often observed in Western countries.

We note that place-based policies can have large welfare costs depending on the responses of the people and places that the policies are applied to. As a result, the welfare costs or deadweight losses of such programmes can be approximated by interpreting local economy's responses to such a programme. Because the theoretical underpinnings as well as the empirical results on the effectiveness and welfare costs of place-based policies are mixed, we contribute to the place-based policy discussion in three ways. First, we introduce a theoretical model in which welfare gains and losses, productivity, wages and employment are

[12] Because the Special Economic Zone (SEZ) boundaries overlap with neighbourhood boundaries, we cannot estimate the coefficient once we include neighbourhood fixed effects.

simultaneously introduced and measured in relation to each other. This stylised model informs us of the magnitude of the displacement effects and potential deadweight losses of the localised policies. Second, we empirically test the effectiveness of place-based development strategies on firm-level productivity by focusing on science park development in Shenzhen, China while controlling for observed and unobserved heterogeneity, sorting and selection. This approach enables us to be reasonably persuasive with respect to the measurement of a causal effect of the policies. Third, we contribute to the potentially important discussion on differences in context between developed and transition countries. While almost unstudied, China, India, Brazil, South Africa, Russia, and many other transition countries extensively use place-based policies, science parks and special economic zones to promote development. The empirical evidence to date is therefore arguably not representative of many of the place-based policies that are in place worldwide.

Our results show that area-based incentives have a substantial impact on firms' productivity in Shenzhen's science parks. Even if we include firm fixed effects and use spatial differencing, firms' output has increased by 15-25 per cent due to science park policies. These large and economically meaningful effects are in line with the findings of Wang (2013) and contribute to the idea that place-based policies have much more profound effects in developing countries and transition economies. We subject our results to an extensive sensitivity analysis, including an analysis based on another identification strategy and a (cross-sectional) dataset. We also test the impact of science parks on wages and employment. We find positive causal effects of the policies on wages. We also find weak evidence that firms have hired more people due to the science park policies. Using a stylised theoretical model, we estimate that these displacement effects may imply a deadweight loss of up to 40 per cent of the total productivity effect. This result suggests that place-based policies may have substantial distortive effects on local economies, but we note that the estimated effects are statistically imprecise.

These outcomes are important for determining the effectiveness of place-based policy strategies in developing countries. On the one hand, the welfare and productivity effects in science parks are impressive and are remarkably large. On the other hand, in recent years, following the Torch Programme, numerous special zones and science parks (whichever they are named) have proliferated throughout China. In addition to the nationally designated areas, there are also provincial-, county- and city-sponsored development zones – such as in Shenzhen. Wu and Gaubatz (2013) observe that science parks attract not only foreign direct investment, they also stimulate domestic investments. Our analyses nevertheless show that science parks may also lead to displacement effects on labour that are on average larger than those observed as a result of place-based policies in developed countries. Place-based policies may therefore help the development of the designated areas while hampering productive development nearby.

References

Au, C. C., and Henderson, J. V. (2006). Are Chinese Cities Too Small? *Review of Economic Studies*, *73*(3), 549–576.

Barca, F., McCann, P., and Rodríguez-Pose, A. (2012). The Case for Regional Development Intervention: Place-based versus Place-neutral Approaches. *Journal of Regional Science*, *52*(1), 134–152.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, *115*(4), 588–638.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, *114*(2), 577–599.

Busso, M., Gregory, J., and Kline, P. (2013). Assessing the Incidence and Efficiency of a Prominent Place Based Policy. *American Economic Review*, *103*(2), 897–947.

Cazes, S., and Verick, S. (2013). *The Labour Market of Emerging Economies: Has Growth Translated into More and Better Jobs?* London: Palgrave MacMillan.

Cheng, F. F., Van Oort, F. G., Geertman, S., and Hooimeijer, P. (2014). Science Park Planning and the Co-location of High-tech Small- and Medium Sized Firms in China's Shenzhen. *Urban Studies*, *51*(5), 1073–1089.

Cheshire, P. C., Nathan, M., and Overman, H. G. (2014). *Urban Economics and Urban Policy: Challenging Conventional Policy Wisdom*. Cheltenham: Edward Elgar Publishing.

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics*, *1*(1), 451–488.

Einiö, E., and Overman, H. G. (2012). The Effects of Spatially Targeted Enterprise Initiatives: Evidence from UK LEGI. *Mimeo, London School of Economics*.

Enright, M., Scott, E., and Chang, K. M. (2005). *Regional Powerhouse: The Greater Pearl River Delta and the Rise of China*. Singapore: Wiley.

Foray, D. (2014). *Smart Specialisation. Opportunities and Challenges for Regional Innovation Policy.* London: Routledge.

Garretsen, H., McCann, P., Martin, R., and Tyler, P. (2013). The Future of Regional Policy. *Cambridge Journal of Regions, Economy and Society*, *6*, 179–186.

Glaeser, E. L. (2008). *Cities, Agglomeration, and Spatial Equilibrium*. Oxford: Oxford University Press.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, *100*(6), 1126. doi:10.1086/261856

Gobillon, L., Magnac, T., and Selod, H. (2012). Do Unemployed Workers Benefit from Enterprise Zones? The French Experience. *Journal of Public Economics*, *96*(9-10), 881–892.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, *118*(3), 536–598.

Gu, C., and Chen, G. (2001). Expanding the Industrial Horizon in Shenzhen: Globalization, Industrial Restructuring and Transformation. *Economic Geography (in Chinese)*, *21*(3), 261–265.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Hsing, Y. T. (2010). *The Great Urban Transformation: Politics of Land and Property in China.* Oxford: Oxford University Press.

Imbens, G. W., and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, *142*(2), 615–635.

Kline, P. (2010). Place-Based Policies, Heterogeneity, and Agglomeration. *American Economic Review*, *100*(2), 383–387.

Kline, P., and Moretti, E. (2013). Place-Based Policies with Unemployment. *American Economic Review*, *103*(3), 238–243.

Koster, H. R. A., Van Ommeren, J. N., and Rietveld, P. (2014). Agglomeration Economies and Productivity: A Structural Estimation Approach using Commercial Rents. *Economica*, *81*(321), 63–85.

Linden, G., Kraemer, K. L., and Dedrick, J. (2009). Who Captures Value in a Global Innovation Network? *Communications of the ACM*, *52*(3), 140.
Lucas, R., and Rossi–Hansberg, E. (2002). On the Internal Structure of Cities. *Econometrica*, *70*(4), 1445–1476.
Ma, F., and Chen, H. (2010). China's High-tech Park Development Starts from Shenzhen. *Economic Geography (in Chinese)*, *30*(10), 1585–1588.
McCann, P., and Ortega-Argilés, R. (2013). Smart Specialization, Regional Growth and Applications to European Union Cohesion Policy. *Regional Studies*, 1–12.
Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, *94*(3), 656–690.
Neumark, D., and Simpson, H. (2015). Place-based Policies. In G. Duranton and J. V. Henderson (Eds.), *Handbook of Regional and Urban Economics 5*. Amsterdam: Elsevier.
Ng, M. K. (2003). Shenzhen. *Cities*, *20*(6), 429–441.
Rodriguez-Pose, A., and Hardy, D. (2014). *Technology and industrial parks in emerging economies. Panacea or pipedream?* Berlin: Springer.
Shenzhen Bureau of Trade and Industry. (2001). *Guidance on Transforming Traditional Industries with High-tech or Advanced Applicable Technologies*. Shenzhen: Shenzhen Bureau of Trade and Industry.
Shenzhen Planning Bureau. (2006). *Transformation and Evolution*. Shenzhen: Shenzhen Bureau of Trade and Industry.
Wang, J. (2013). The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities. *Journal of Development Economics*, *101*(1), 133–147.
Wu, W., and Gaubatz, P. (2013). *The Chinese City*. London: Routledge.

Appendix A. Place-based policies and agglomeration externalities

One objection to the use of factor employment changes in welfare measures is that there may be positive externalities, especially agglomeration externalities. These externalities potentially bias the welfare conclusions. To observe the argument, suppose that the multiplicative technology A_1 is taken as given by producers, but it in fact depends on regional employment: $Q_1 = A_1(L_1, T_1) \cdot f(K_1, L_1)$. The first-order derivative with respect to welfare is then the sum of the displacement effect and the externality:

$$(13) \qquad -t_1 w_1 \frac{\mathrm{d}\ell_1}{\mathrm{d}t_1} + \sum\nolimits_z \frac{q_z}{A_z}\frac{\partial A_z}{\partial \ell_z}\frac{\mathrm{d}\ell_z}{\mathrm{d}t_1}.$$

The latter term is the agglomeration effect due to labour reallocation in all regions z. For the moment, we will assume that the net effect can be positive if the policy concentrates workers in agglomerated areas – intuitively, moving one worker from a small region to a large, more productive region increases the net agglomeration benefit. The rate at which an average worker moving to region 1 changes aggregate productivity is his gain in region 1 minus the agglomeration benefits lost in an average origin location (z^{-1}):

$$(14) \qquad \psi_\ell = \frac{q_1}{A_1}\frac{\partial A_1}{\partial \ell_1} + \sum\nolimits_{z^{-1}} \frac{q_z}{A_z}\frac{\partial A_z}{\partial \ell_z}\frac{\mathrm{d}\ell_z/\mathrm{d}t_1}{\mathrm{d}\ell_1/\mathrm{d}t_1}$$

Suppose that the cost share of labour is α_ℓ (in our empirical specification, this is the Cobb-Douglas parameter for labour). Multiplying and dividing by t_1 and L_1 and using the definition gives an expression of the welfare change as:

$$(15) \qquad -w_1 \ell_1 \frac{\mathrm{d}\ell_1}{\mathrm{d}t_1}\frac{t_1}{\ell_1}\left(1 - \frac{\psi_\ell}{t_1 w_1}\right),$$

where ψ_ℓ is the average agglomeration benefit per moving worker, and $t_1 w_1$ is the per worker subsidy cost. High agglomeration benefits relative to the subsidy can diminish the welfare loss or even turn it into a gain. Thus, with agglomeration externalities, an estimate of the welfare costs could be overstated, or the welfare effect could even be interpreted as a loss when there is a net positive effect.

In our methodology, the agglomeration effects do not influence our estimates of displacement. The reason is that we compare firms that are close to science park borders that have no substantial barriers. Therefore, any agglomeration effects due to firm density are comparable between just inside and just outside the science park border. Hence, they should impact the treatment and control groups alike. To check this in the data, we investigate agglomeration externalities by incorporating access measures to the other firms in the area – they do not matter for our estimates of the science park effect. See Section VI.D for more details.

Appendix B. Additional sensitivity analyses

A. Firm-specific trends

In the current analysis we analyse the effect of science park policies using spatial and temporal differencing. More specifically, we compare the productivity changes between a firm that is in a science park with a firm that is just across the border of a science park. It is therefore unlikely that the comparison of nearby companies along the science park boundary is confounded by selection of companies due to shocks to their future prospects to innovate and productivity. For example, a firm may come up with a more potential portfolio of innovative projects than a rival company, so that the latter firm is not allowed to locate in a science park. Because we compare stayers, we think that unobserved shocks to firm's productivity are unlikely to play a major role. Nevertheless, we may include firm-specific linear trends, so that we identify the effect of science parks based on non-linear changes in productivity.

In columns (1) and (2) of Table B1 we replicate the baseline specifications in columns (5) and (7) in Table 2, so we use spatial differencing, but we extend the model by including for each firm a linear trend. It is shown that the coefficients are quite imprecise, which is not too surprising because the firm-specific trends soak up most of the relevant identifying variation. Nevertheless, the point estimates are similar to the baseline specifications, in particular once we include firm-neighbourhood fixed effects. The latter implies that we identify the effect based on stayers. In columns (3) and (4) we use the alternative identification strategy with local industrial locations as control locations. Again, the results are imprecise, but the point estimates are very similar to the previous specifications, which is reassuring. Hence, although including firm-specific trends leads to large standard errors, we do not find strong evidence that our results can be explained by firm-specific unobserved shocks.

TABLE B1 — SENSITIVITY ANALYSIS: FIRM-SPECIFIC TRENDS
(Dependent variable: the tax rate)

	Spatial differencing		*Local industrial parks*	
	(1)	(2)	(3)	(4)
	WLS	WLS	WLS	WLS
Science park	0.0863	0.123	0.0950	0.141
	(0.126)	(0.197)	(0.153)	(0.181)
Firm variables (7)	Yes	Yes	Yes	Yes
Year fixed effects (6)	Yes	Yes	Yes	Yes
Industry FE (33)	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Firm-specific year trends	Yes	Yes	Yes	Yes
Firm-neighbourhood FE	No	Yes	No	Yes
Science park boundary FE (15)	Yes	Yes	Yes	Yes
Number of observations	8,311	8,311	8,427	8,427
R^2	0.959	0.960	0.983	0.984

Notes: The weights for the weighted least squares (WLS) specifications in columns (1) and (2) are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. The weight is equal to the share of the neighbourhood in a local industrial park and science park in 2007 in columns (3) and (4). Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

B. Firm heterogeneity and ownership

One may wonder whether the productivity effect we find applies to all firms or for example only to large multinational enterprises that will probably mainly produce for international markets. We then interact the science park dummy with the ownership status. The results reported Table B2 are essentially a replication of Table 2. Columns (1) and (2) seem to suggest that indeed the productivity effect mostly applies to foreign-owned firms. Also columns (3) and (4) seem to suggest that the effect is the most pronounced for foreign-owned firms, while there is no effect for firms owned by enterprises based in Hong Kong, Taiwan, or Macau (HTM) However, if we move to the more believable specifications with firm fixed effects in columns (5)-(7), it is shown the effects between the different types of firms are similar and not statistically significantly different from each other (the p-value = 0.689 in column (5) and 0.615 and 0.919 in columns (6) and (7) respectively). However, in column (7) the results become statistically imprecise with coefficients being statistically significant only around the 20 percent level. Nevertheless, we do not find much robust evidence that the science park effect only benefits one type of firm.

TABLE B2 — SENSITIVITY ANALYSIS: OWNERSHIP AND THE PRODUCTIVITY EFFECT
(Dependent variable: the logarithm of firms' output)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.240**	0.161*	0.229***	0.238***	0.337**	0.302**	0.202
× Domestic firm	(0.0997)	(0.0937)	(0.0693)	(0.0598)	(0.160)	(0.138)	(0.146)
Science park	-0.0166	0.0619	0.0968	0.119	0.195	0.150	0.154
× HTM-owned firm	(0.0998)	(0.0905)	(0.0755)	(0.0765)	(0.122)	(0.114)	(0.126)
Science park	0.639***	0.404**	0.433***	0.446***	0.232**	0.183*	0.142
× Foreign-owned firm	(0.190)	(0.179)	(0.149)	(0.146)	(0.0965)	(0.0987)	(0.113)
Firm variables (7)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year fixed effects (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311
R^2	0.551	0.598	0.643	0.645	0.942	0.943	0.947

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

C. Tax rate

Firms may be more productive in science parks solely because of lower taxes and tax exemptions. The tax rate is officially lower in science parks, and some firms receive tax holidays for a certain number of years. To investigate whether there are structural differences in tax rates between firms in science parks and other areas, we repeat the analysis, but now the dependent variable is the tax rate. The coefficient in column (1), Table B3, suggests that the tax rate is not statistically significantly different for firms in science parks. When controlling for other firm variables and industrial sector fixed effects, the tax rate is still not statistically significantly different for firms inside and outside of science parks. In column (3), where we use spatial differencing, we find weak evidence that taxes are lower in science parks. When we include neighbourhood control variables in column (4), the results suggest that tax rates are approximately 0.7 percentage points lower in science parks. When we include firm fixed effects, the effect of science parks on tax rates is statistically insignificant, although the point estimate suggests a sizeable effect on taxes of approximately 3 percentage points. The point estimate is similar once we include a flexible function of geographic coordinates and firm-neighbourhood fixed effects in columns (6) and (7).

TABLE B3 — SENSITIVITY ANALYSIS: TAX RATES
(Dependent variable: the tax rate)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	-0.000603	-0.00208	-0.00514*	-0.00753**	-0.0312	-0.0372	-0.0437
	(0.00254)	(0.00231)	(0.00303)	(0.00327)	(0.0224)	(0.0253)	(0.0289)
Firm variables (7)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year fixed effects (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311
R^2	0.067	0.084	0.101	0.103	0.555	0.556	0.566

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

TABLE B4 — SENSITIVITY ANALYSIS: INCLUDING THE TAX RATE IN THE PRODUCTIVITY REGRESSIONS
(Dependent variable: the logarithm of firms' output)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.229*	0.170	0.211***	0.220***	0.214**	0.165*	0.130
	(0.125)	(0.107)	(0.0769)	(0.0720)	(0.101)	(0.0858)	(0.0906)
Tax rate	-1.563***	-1.614***	-1.453***	-1.454***	-0.806***	-0.812***	-0.783***
	(0.143)	(0.136)	(0.215)	(0.216)	(0.245)	(0.245)	(0.256)
Firm variables (5)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	19,215	19,215	8,311	8,311	8,311	8,311	8,311
R^2	0.557	0.606	0.648	0.650	0.943	0.944	0.948

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

We can also directly control for tax rate in the productivity regressions, although we are aware that the tax rate is likely to be endogenous because, for example, larger firms may be more effective in negotiating tax exemptions. Table B4 reports the results for the impact of science parks on firms' output while controlling for the tax rate. The table shows that the point estimates are generally slightly lower but are very similar to the baseline specifications reported in Table 2. It can be observed that the point estimates are very similar to our baseline specifications, although with somewhat higher standard errors. Hence, the fact that firms in science parks pay lower taxes can only explain a small part of the productivity effect related to science park policies.

D. Spatial differencing: sensitivity

We also investigate the robustness regarding the assumptions when we use spatial differencing. The results are summarised in Table B5 and are compared with the baseline specifications that are reported in columns (5) and (7) of Table 2. We first investigate whether the results are robust to assumptions on the weighting function. In column (1), we use a tricube weighting function instead of a linear weighting function, which is given by:

$$w_{z\bar{t}} = \left(1 - \left(\frac{d_{z\bar{t}}}{d_T}\right)^3\right)^3 1_{d_{z\bar{t}}<d_T}. \tag{16}$$

The results indicate that the effect of science park policies is very similar. This result also holds for the specification using firm-neighbourhood fixed effects (column (5)).

In columns (2) and (3), we test whether our results are robust to changing the boundary threshold. As suggested by Imbens and Lemieux (2008), we present the results for threshold distances that are twice and half the size of the originally chosen threshold distance. In column (2), we set $d_T = 1.25$ (half of the original threshold distance). The point estimate is very similar. The corresponding specification using firm-neighbourhood fixed effects in column (6) also confirms that the findings are hardly influenced by the choice of boundary threshold. However, due to the lower number of observations, the standard error is somewhat higher, so the point estimate is not statistically significantly different from zero in the latter specification. In column (3), we set the threshold distance to twice the size of the original threshold, so $d_T = 5$. The coefficient in column (3) is very similar to the baseline specification in Table 2. Column (7) also suggests a very comparable effect, but it is somewhat imprecise.

When using spatial differencing, one may also include a flexible trend of the assignment variable (Hahn et al., 2001). In our case, this would imply that we should include a spatial trend. We therefore include a flexible function of the average distance to a science park boundary. We estimate this function using a fifth-order polynomial. Column (4), Table B5,

TABLE B5 — SENSITIVITY ANALYSIS: SPATIAL DIFFERENCING

(Dependent variable: the logarithm of firms' output)

	(1) WLS	(2) WLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
	Tricube weighting	$d_T = 1.25$	$d_T = 5.00$	*Distance to science park*	*Tricube weighting*	$d_T = 1.25$	$d_T = 5.00$
Science park	0.241**	0.199**	0.256**	0.166*	0.166*	0.145	0.164
	(0.100)	(0.0970)	(0.0994)	(0.0902)	(0.0972)	(0.107)	(0.101)
Firm variables (5)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Distance to science park boundary $\Omega(\cdot)$	No	No	No	Yes	No	No	Yes
Year FE (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	Yes	Yes	Yes
Science park boundary FE (15)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	8,311	3,509	13,154	8,311	8,311	3,509	13,154
R^2	0.943	0.946	0.939	0.942	0.947	0.949	0.945

Notes: Standard errors are clustered at the neighbourhood level and in parentheses.
*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

shows that the effect is positive, albeit somewhat lower, and statistically significantly different from zero at the 10 per cent level.[13] Hence, different assumptions regarding identification strategy do not affect our main conclusions.

E. Measurement error

A further robustness analysis focuses on the presence of measurement error in the variable of interest. In the previous analyses, we excluded firms in neighbourhoods that have a share that deviates from one or zero to avoid measurement error. In this subsection, we include all observations.[14] It may be that the excluded neighbourhoods are a non-random subset of the

[13] Note that because distance to a science park boundary is time-invariant, it will drop when we include firm-neighbourhood fixed effects, so we do not report a similar specification with firm-neighbourhood fixed effects because the results are the same as those reported in column (7) of Table 2.

[14] In a special case, we may say something about the bias of the estimated parameter β due to the measurement error in p_{zt}. Let us assume that $\log q_{jzt} = \beta p_{zt} + \epsilon_{jzt}$, and let us also assume that firms are uniformly distributed within neighbourhoods and that $p_{zt} = p^*_{zt} + \mu_{zt}$, where p^*_{zt} is an (unobserved) dummy indicating whether a firm is in a science park. Then, $\text{plim}(\beta) = \hat{\beta}\left(1 - \sigma_\mu/\sigma_p\right)$. Because p^*_{zt} is either one or zero, it may be shown that $\sigma_z = (1/n)\sum_{j=1}^{n} p_{zt}(1 - p_{zt})$, which implies that when p_{zt} is always zero or one, $\sigma_\mu = 0$. One may estimate an errors-in-variables regression with reliability equal to $\left(1 - \sigma_\mu/\sigma_p\right)$. However, we also include other variables in the regression analysis

TABLE B6 — SENSITIVITY ANALYSIS: MEASUREMENT ERROR
(Dependent variable: the logarithm of firms' output)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.156	0.111	0.183**	0.181***	0.163**	0.134**	0.103
	(0.117)	(0.101)	(0.0737)	(0.0669)	(0.0687)	(0.0645)	(0.0736)
Firm variables (7)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	No
Year fixed effects (6)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE (3,707)	No	No	No	No	Yes	Yes	Yes
Firm-neighbourhood FE (3,910)	No	No	No	No	No	No	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	22,116	22,116	11,212	11,212	11,212	11,212	11,212
R^2	0.557	0.603	0.644	0.646	0.942	0.942	0.948

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In column (6), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

population, as they are the neighbourhoods that are close to science park boundaries. An advantage of including all observations is that, due to the higher number of observations, the effects may be more precisely estimated. However, due to the measurement error, the effects are likely to be downward biased. Table B6 reports the results.

Columns (1) and (2) show that firms in science parks are not necessarily more productive. However, the estimated effects are quite imprecise. Column (3) focuses on neighbourhoods that are close to science park boundaries. In this case, science parks seem to increase productivity by 20 per cent. This effect is almost identical if we include neighbourhood control variables. In column (5), we include firm fixed effects. The productivity effect is 18 per cent, while it is slightly lower if we control flexibly for geographic coordinates in column (6). Column (7) includes firm-neighbourhood fixed effects. The effect is slightly lower and not statistically significant. This result is not very surprising if we take into account that the effect will be downward biased due to the measurement error. In any case, the results largely confirm the estimates that we found earlier.

(such as distance to the nearest employment centre), which may be correlated with the variable of interest and may also be measured with error. Hence, we cannot easily solve the problem by estimating an errors-in-variables regression.

TABLE B7 — DESCRIPTIVE STATISTICS OF SBTI SAMPLE

	Other areas				Science parks			
	μ	σ	min	max	μ	σ	min	max
Output *(in 10,000,000 ¥)*	115.624	2,221.319	5.010	187,469.400	267.946	2,011.515	5.010	41,918.490
Science park	0.000	0.000	0.000	0.000	1.000	0.000	1.000	1.000
Distance to science park (*in km)*	-4.434	3.322	-14.728	-0.005	0.428	0.323	0.007	1.891
Special economic zone (SEZ)	0.295	0.456	0.000	1.000	0.253	0.435	0.000	1.000
Dist. to employment centre (*in km)*	7.855	4.877	0.218	31.390	8.182	4.711	0.274	24.709
Dist. to highway ramp (*in km)*	4.836	3.340	0.093	37.304	5.004	3.793	0.052	30.623
Dist. to airport (*in km)*	27.725	15.520	1.097	91.726	23.836	15.900	5.066	85.045
Dist. to seaport (*in km)*	25.373	10.261	1.230	48.329	25.788	8.880	10.810	40.555
Restricted zone	0.059	0.237	0.000	1.000	0.000	0.000	0.000	0.000
Employees	302.671	2,458.576	1.000	199,908.000	493.163	2,604.361	1.000	77,472.000
Electricity usage *(in 10,000,000 ¥)*	127.524	463.554	0.000	9,609.000	163.356	496.233	0.000	7,729.575
Water usage *(in 10,000,000 ¥)*	3.153	14.998	0.000	810.324	4.623	21.867	0.000	529.873
Share college graduates	0.203	0.234	0.000	1.000	0.236	0.261	0.000	1.000
Share R&D employees	0.036	0.100	0.000	1.000	0.058	0.136	0.000	1.000

Notes. The total number of observations is 8,873. The number of observations in science parks is 1,524 and in local industrial parks is 2,302.

F. SBTI data: other results

Table B7 reports descriptive statistics for the alternative dataset based on the 2007 Shenzhen Industrial Enterprise Survey. The results show that mean output as well as the variance in output are of the same order of magnitude as in the other dataset. Additionally, the average distances to science park boundaries are similar, which suggests that there is no geographic bias in the data selection. The share of college graduates is approximately 20 per cent and is slightly higher in science parks. Additionally, the share of R&D employees is higher in science parks (5.8 per cent versus 3.6 per cent).

We interpret the productivity effect as the causal effect of science parks on technology, which is captured by A_i in our theoretical model. We then control for labour and capital and assume that the consumption of land is fixed. However, when land is much cheaper in the science parks, firms may substitute capital or labour for land, and (part of) the productivity effect may be explained by changes in the consumption of land. To investigate this issue further, we exploit additional information on rents and the use of land that is (only) available in the SBTI data.

Table B8 reports the results for the rent regressions, where the log of total rents paid for factory space is the dependent variable. This information is available for approximately 75 per cent of the observations. In the naïve regression reported in column (1), it seems that rents are 16.8 per cent higher in science parks, although the effect is only marginally significant. This result is slightly lower if we control for neighbourhood characteristics (column (2)). When we focus on areas close to science parks and include science park fixed

TABLE B8 — REGRESSION RESULTS ON THE IMPACT OF SCIENCE PARKS ON FACTORY RENTS
(Dependent variable: the logarithm of firms' total rents paid)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.156*	0.116***	0.0856	0.00270	-0.0310	-0.0817	0.0569
	(0.0852)	(0.0337)	(0.0709)	(0.115)	(0.132)	(0.121)	(0.0612)
Firm control variables (5)	No	No	No	No	No	Yes	Yes
Neighbourhood variables (6)	No	Yes	Yes	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	Yes	Yes
Industry FE (33)	No	No	No	No	No	Yes	Yes
Neighbourhood FE (191)	No	No	No	Yes	Yes	Yes	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	6,707	6,707	3,289	3,289	2,124	2,124	2,111
R^2	0.003	0.051	0.039	0.125	0.122	0.127	0.495

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In columns (5), (6) and (7), $d_T = 1.25$. In columns (6) and (7), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

effects, the effect becomes statistically insignificant (column (3)). In column (4), where we include neighbourhood fixed effects, the rent effect becomes statistically insignificant. This result also holds if we reduce the boundary threshold to 1.25 km and when we control for a flexible function of geographic coordinates in columns (5) and (6), respectively. However, one might argue that we do not include any characteristics of the building (e.g., size or quality), which is advisable in these types of 'hedonic' regressions. To proxy for building characteristics, we therefore include firm characteristics in column (7). For example, it might be expected that firms with a larger workforce and a larger capital stock also need more space. We are aware that these variables are potentially endogenous (as the price of land also may impact workforce size and the size of the capital stock). The coefficient seems to suggest that, if anything, land is more expensive in science parks, rather than cheaper, which makes it more believable that our results are not explained by changes in the consumption of land.

To test this more directly, we use information on land consumption. For approximately 17 per cent of the observations, we have information on the consumption of land. We then interact the science park variable with a dummy variable that indicates whether land information is present. Furthermore, we control for the log of land consumption and include a dummy variable that indicates whether land use is missing. If the science park effect is much lower for firms for which we have information on the consumption of land, it may indicate that part of the effect is not explained by changes in technology but by changes in land consumption. Table B9 reports the results.

TABLE B9 — REGRESSION RESULTS ON THE IMPACT OF SCIENCE PARKS ON PRODUCTIVITY, INCLUDING LAND CONSUMPTION
(Dependent variable: the logarithm of firms' output)

	(1) OLS	(2) OLS	(3) WLS	(4) WLS	(5) WLS	(6) WLS	(7) WLS
Science park	0.139**	0.0432	0.0854**	0.0973**	0.143***	0.125***	0.117***
	(0.0613)	(0.0342)	(0.0378)	(0.0352)	(0.0222)	(0.0245)	(0.0252)
Science park × (1 – Land missing)	0.0332	0.0353	0.0582	0.0376	0.0272	-0.00252	0.00266
	(0.121)	(0.116)	(0.0907)	(0.0793)	(0.0422)	(0.0283)	(0.0306)
Land *(log)*	0.175***	0.150***	0.157***	0.163***	0.173***	0.180***	0.182***
	(0.0324)	(0.0282)	(0.0301)	(0.0295)	(0.0187)	(0.0153)	(0.0139)
Land missing	1.041***	0.902***	1.059***	1.089***	1.188***	1.249***	1.271***
	(0.333)	(0.276)	(0.283)	(0.278)	(0.167)	(0.137)	(0.128)
Firm control variables (5)	No	Yes	No	No	No	Yes	Yes
Neighbourhood variables (6)	No	No	No	Yes	Yes	Yes	Yes
Geographic coordinates $\Upsilon(\cdot)$	No	No	No	No	No	No	Yes
Industry FE (33)	No	Yes	Yes	Yes	Yes	Yes	Yes
Neighbourhood FE (191)	No	No	No	No	Yes	Yes	Yes
Science park boundary FE (15)	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	8,873	8,837	4,184	4,184	4,184	2,748	2,748
R^2	0.419	0.500	0.550	0.554	0.576	0.590	0.596

Notes: The weights for the weighted least squares (WLS) specifications are given by $w_{z\bar{t}} = (1 - d_{z\bar{t}}/d_T)1_{d_{z\bar{t}}<d_T}$, where $d_T = 2.5$. In columns (5), (6) and (7), $d_T = 1.25$. In columns (6) and (7), we include a flexible function $\Upsilon(\cdot)$ of geographic coordinates, which is approximated by a fifth-order polynomial function of geographic coordinates, including interactions. Standard errors are clustered at the neighbourhood level and in parentheses.

*** Significant at the 0.01 level
** Significant at the 0.05 level
* Significant at the 0.10 level

Column (1) is a naïve regression without locational control variables, but with controls for workforce size and the use of electricity and water. Science parks seem to impact productivity positively. In column (2), we also control for industrial sector fixed effects and other firm characteristics, which leads to a statistically insignificant effect of locating in a science park. It should be noted that firms for which we do not have information on land consumption for some reason seem to be more productive. In columns (3)-(7), we again use spatial differencing. In all of the specifications, the coefficient related to science parks is positive and statistically significant at the five per cent level. Additionally, in all of the specifications, the interaction effect of science parks and a dummy that indicates whether we have information on land consumption is statistically insignificant. The interaction effect is close to zero, which suggests that firms for which we have information on land use encounter a similar productivity effect; thus, controlling for land consumption does not seem to influence our main conclusions.