Do Knowledge Spillovers through Worker Inflows Increase Establishments' Productivity? First Evidence from Germany

Bastian Stockinger[1]
IAB Nuremberg

Katja Wolf
IAB Nuremberg

June 2015

Preliminary version!
Do not quote!

Abstract

Several empirical studies find that worker inflows from more productive firms increase hiring firms' productivity. Supposedly, the reason is that workers' knowledge about superior technology and best practices spills over to hiring firms. We test whether this finding can be confirmed for Germany. Using a unique linked employer-employee data set and an establishment productivity proxy, we do not find evidence that worker inflows from more productive establishments increase productivity. In contrast, inflows from less productive establishments seem to have a positive productivity effect. This is probably because new hires from more productive establishments are negatively selected, and those from less productive establishments, positively selected. Thus, unlike previous studies, our analysis reveals the productivity potential in hiring the best workers from establishments down the productivity distribution. Instead of knowledge spillovers, our findings indicate a productivity-enhancing assortative matching of highly productive workers and firms.

Keywords: Knowledge Spillovers, Labor Mobility, Plant-Level Productivity
JEL Codes: D24, J61, J62, R23

[1] Corresponding author: Bastian Stockinger, Institute for Employment Research (IAB), Nuremberg, Germany. Phone: +49 (0)911 179 57 35. E-mail: bastian.stockinger@iab.de.

Contents

1. Introduction

Knowledge spillovers are one of several economic mechanisms typically referred to when explaining the sources of productivity, respectively, productivity growth or differences between firms, industries, or regions. Yet, their empirical identification is still a challenge, due to the multifaceted concept of knowledge and a lack of suitable data or identification strategies. By now, empirical research has come to a tacit consensus that knowledge spillovers can be identified by tracking worker flows between firms and analyzing the ensuing productivity of hiring firms. We review the literature leading to and taking this approach, critically assessing the proposed identification strategies and inferring the economic meaning of the results. From our assessment, we identify an approach that we regard as the most credible and convincing identification strategy to date. We refer to it as the spillover potential approach, as proposed by Stoyanov and Zubanov (2012, 2014) and, closely related, Serafinelli (2013). This approach builds on the idea that worker flows between firms or establishments of different productivity levels, notably form highly productive to less productive firms, are a channel through which productivity potential spills over.

We develop a data base on which we can probe this approach, a linked employer-employee data set from Germany. Our study is the first for Germany and at the same time, the first on a large economy. Our results do not suggest that worker inflows from more productive establishments increase hiring establishments' productivity. Instead, it is inflows from less productive establishments for whom we find a positive relation to productivity. Several measures to counter endogeneity bias suggest that our estimates may be an upper bound of the true causal effect of such hires. While in sharp contrast to the previous studies for other countries, our findings are less surprising when considering the recent evidence by Card et al. (2013) that West Germany has seen an increased sorting of well-paid workers into high-paying establishments over the last decades. This may also indicate a sorting of highly productive workers and firms. Reversely, it could be less productive workers who change jobs towards less productive firms – and these are the knowledge carriers suggested by previous studies. Our descriptive results indicate just that: Workers moving from more to less productive establishments are negatively selected from their sending establishments, while the opposite is true for movers in the other direction. We therefore find evidence of productivity gains from hiring the best workers of inferior establishments, but none of knowledge spillovers through hiring workers from superior establishments.

This paper proceeds as follows. The next two sections review theoretical considerations and previous empirical work on knowledge spillovers at the firm level. Section 4 presents our data and particularly how we identify our crucial variables, worker inflows and their characteristics. In Sections 5 and 6, the empirical model and descriptive statistics are presented. In Section 7 we discuss the econometric implementation of our model and estimation results. In Section 8, we draw preliminary conclusions.

2. Theoretical concepts and empirical approaches

A fundamental and empirically well-founded statement in the economics of knowledge, rooted in the work of Arrow (1962), Lucas (1988), and Romer (1986, 1990), is that the exchange of knowledge, which is a public good unless potential users are excluded from its use by law,[2] can exert positive external effects – knowledge spillovers. Being a key determinant of productivity and growth working through technological progress and learning, knowledge and its exchange have increasingly received further attention in the "knowledge production" literature (Griliches, 1979, 1998), which led the way also to empirical research at the micro (firm) level. Yet, these theoretical frameworks largely rely on the assumption of an external research sector, which increases the amount of public knowledge, and a production sector which may benefit from this public knowledge as an externality. To the best of our knowledge, there is no generally agreed-upon theoretical framework for knowledge spillovers at the micro level (i.e. between firms), but there are plenty of empirical approaches to detect such effects.

The main challenge for the empirical analysis of knowledge spillovers between firms is to find a concrete channel of knowledge exchange and its implementation. One empirically observable channel is the citation of patents, representing the exchange of innovative technological knowledge (Jaffe et al., 1993). Since the implementation of such knowledge is not trivial, few firms ever possess, apply for, or cite any patents, limiting the scope of empirical analysis based on patent citations. Moreover, empirical studies especially in the regional and urban economics literature suggest that even where close interconnections through patent citation exist (typically, in industrial clusters), the underlying mechanism seems to be the mobility of inventors and other workers within these clusters. That is, knowledge is being exchanged not in an abstract way (us-

[2] In contrast, the consumption of knowledge is always "non-rival," satisfying the second necessary condition of a public good.

ing patents as blueprints without consulting the inventor to interpret the information contained therein), but by personal contact between knowledge carriers (e.g., Breschi and Lissoni, 2009).

The fact that not all knowledge can be codified in patents, but that its exchange and implementation usually require personal interaction ("tacitness of knowledge"), has spurred a rich literature on localized knowledge spillovers, see e.g. Breschi and Lissoni (2001), Rosenthal and Strange (2004), Power and Lundmark (2004), and Abel et al. (2012). A classic proposition in this literature is that the regional stock of human capital fosters productivity and innovation in the regional economy. Yet, while a highly educated regional workforce and population might improve aggregate (regional) economic outcomes in various ways, it is unclear how it should affect productivity within the firms – that is, within the units of production, where knowledge supposedly comes into effect by enhancing productivity. Given the "tacitness" of knowledge, the most concrete and arguably most effective channel of knowledge spillovers is the mobility of workers, who carry knowledge from one firm to another. According to the studies of Almeida and Kogut (1999) and Song et al. (2003), it is the clustering of skilled workers, combined with a high degree of mobility, that accounts for the localization of knowledge spillovers in the semiconductor industry in Silicon Valley. In this context, Song et al. (2003) coined the expression "learning by hiring," suggesting that in knowledge-intensive labor markets, hiring may indeed serve the purpose of learning, rather than just replacing workers. Thus, knowledge spillovers are a strongly localized phenomenon because labor mobility is spatially concentrated.

Following the pioneer studies on Silicon Valley, a growing number of studies have considered worker mobility as a channel of knowledge spillovers, building on the idea that any (skilled) worker is a potential carrier of knowledge. A theoretical model including worker flows as the channel of spillovers has been developed by Dasgupta (2012), who seeks to explain knowledge diffusion processes through worker flows from multinational enterprises (MNEs) to host-country domestic firms. The basic proposition of this model and recent empirical studies is that there is potential for spillovers when workers move from "superior" firms, which should possess a great stock of knowledge and technological capacities, to "inferior" firms which benefit from the additional knowledge thus received. These empirical studies include, e.g., Stoyanov and Zubanov (2012, 2014) and Maliranta et al. (2009), who also find that firms do not fully compensate incoming workers (knowledge carriers) for their productivity effects, implying that worker inflows indeed are a channel of externalities. We review these and similar studies, which pave the way for own analysis, in more detail in the next section.

3. Review of empirical evidence

Previous empirical studies on mobility-induced knowledge spillovers argue that the occurrence and extent of spillovers depend on the characteristics of sending firms. A specific branch of literature focuses on inter-firm knowledge spillovers between multinational enterprises (MNEs) and domestic firms. The underlying assumption is that domestic firms receiving worker inflows from MNEs receive new knowledge on technology, marketing, et cetera, since MNEs are structurally more productive than non-MNEs (for a theoretical argument, see Helpman et al. (2004)). One of the first studies in this area is Görg and Strobl (2005) , who find that Ghanaian manufacturing firms whose chief executives have previously worked for MNEs achieve higher productivity levels than their domestic competitors. Balsvik (2011) finds evidence of spillovers from MNEs in the Norwegian manufacturing sector, as firms with high shares of workers with MNE experience achieve higher productivity levels. Similarly, Poole (2013) finds evidence of spillovers from worker flows between MNEs and domestic firms in Brazil, as identified by the wages of the receiving firms' incumbent workers.

The productivity gap between sending and receiving firms and its implications for knowledge spillovers have also been studied more generally (beyond the multinational-domestic context), also because focusing on characteristics of incoming workers' previous employers is less vulnerable to reverse causality than incoming workers' individual characteristics. Following this rationale, Stoyanov and Zubanov (2012) find that labor productivity and total factor productivity in Danish manufacturing firms are positively associated with the inflow of workers from more productive manufacturing firms, and the relationship gets stronger as the productivity gap between sending and hiring firms widens. Moreover, this positive association is statistically significant for worker inflows from more productive firms, but not inflows from less productive firms. Since inflows' individual ability is held constant, this means that otherwise equal workers from more productive firms have a positive productivity effect other workers do not have. The effect is small but robust (hiring an average amount of knowledge carriers, as compared to hiring none, corresponds to a productivity gain of 0.35 percent). Furthermore, the productivity gap of high-qualified incoming workers and managers is found to correlate much more strongly with receiving firms' productivity than that of less skilled inflows, suggesting that higher-skilled workers are more able to carry knowledge between firms. Taking several means to reduce endogeneity bias, Stoyanov and Zubanov (2012) thus identify the upper bound of a potentially causal effect of hiring employees from more productive firms on hiring firms' productivity. Closely related to Stoyanov and Zubanov's (2012) productivity gap approach, Serafinelli (2013) studies the impact of worker inflows from high-paying firms (a proxy for highly productive firms) on receiving (non-high-paying) firms' productivity, surviving a number of measures against reverse causality bias, and using local high-wage-firm downsizings as an instrument for the number of inflows from such firms. Similar to Stoyanov and Zubanov's (2012) results, it is found that inflows from non-high-paying firms do not have the same effect, even controlling for their (observed and unobserved) person-specific wage effect.

A number of related studies indicate qualitatively similar effects – hiring workers with particularly valuable experience is typically found to correlate (withstanding major sources of endogeneity bias) with firms' productivity, probably reflecting a positive externality to hiring firms. To mention just a selection, Møen (2005) finds that Norwegian manufacturers partly internalize knowledge spillovers from separating R&D workers by setting relatively steep tenure-earnings profiles for these workers. Kaiser et al. (2008) analyze Danish firms' patent applications, finding that the inflow of R&D workers is strongly related to the number of patent applications. Maliranta et al. (2009) come to similar conclusions concerning hiring firms' non-R&D activities, i.e. firms benefit from inflows' earlier R&D experience in terms of their non-R&D business. In sum, these studies substantiate the claim that firms can benefit from other firms' R&D activities by hiring workers previously employed there. Thus, the most important conclusion from the literature on knowledge spillovers through worker mobility between firms seems to be that worker inflows from highly productive, highly innovative, or in some other sense superior firms, can transfer part of their knowledge acquired there, and increase productivity in hiring firms.

Against the background of this literature, our main contribution is to present the first empirical evidence on productivity effects from worker flows between establishments of different productivity levels for Germany. The most closely related studies have all focussed on smaller economies, namely Norway (Balsik, 2011), Denmark (Stoyanov and Zubanov, 2012, 2014), and the Veneto region (Italy; Serafinelli, 2013). Our aim is to test the external validity of the approach taken in the three latter studies (the "spillover potential" approach, to be presented in detail in section 5) in the German context. Yet, we must not necessarily expect to confirm previous studies' findings for Germany. A recent study by Card et al. (2013) showed that the (West) German labor market has seen a substantial rise in wage inequality in the last decades, driven by increasing wage heterogeneity at the worker and establishment levels, and an increased tendency of high-wage workers to sort into high-wage establishments. Assuming that wages reflect productivity, this would mean that highly productive workers sort into highly productive establishments. Reversely, workers moving in the other direction may not be favorably selected in terms of their individual productivity, which seems to be relevant for their ability to transfer knowledge (recall Stoyanov and Zubanov's (2012) finding that higher-skilled worker inflows have stronger productivity effects). Stoyanov and Zubanov (2012, 2014) and Serafinelli (2013) take this potential "lemons bias" into account by controlling for workers' observed and unobserved individual productivity characteristics, and still find positive productivity effects from hiring downward-movers. Given the trend of increasingly assortative matching in Germany, however, it is questionable whether we can expect the same result for Germany. If wage sorting in Germany reflects productivity sorting, we could instead expect productivity gains from workers who move up, not down the establishment productivity distribution, reflecting efficiency gains from better matching rather than "top-down" knowledge spillovers along the establishment productivity distribution. To test these contradictory hypotheses, we exploit a rich and detailed source of employment and establishment data from Germany. The next sections present this data base and the empirical model we employ to test our hypotheses.

4. Data

We construct a linked employer-employee data set based on data provided by the Institute for Employment Research (IAB). Individual-level data are obtained from the Integrated Employment Biographies (IEB), establishment-level data from the Establishment History Panel (BHP) and the IAB Establishment Panel. The IEB contain precise information about individuals' labor market biographies. They are based on process-generated data from different administrative sources and contain daily information on every individual in Germany that is either in employment subject to social security, registered unemployed, or participating in measures of active labor market policy, excluding only civil servants and the self-employed. A detailed description of the IEB's construction is given in Vom Berge et al. (2013). The assignment of workers to establishments, as well as crucial variables such as begin and end dates of employment spells, are highly reliable as they are drawn from the official employment statistics of the Federal Employment Agency, which serves as the basis to compute contributions to the social security system. Since misreporting is subject to pecuniary penalties, the data are highly accurate and reliable. The IEB report the exact start and end dates for each employment spell, which allows us to calculate exactly the length of employment relationships. Furthermore, the information on unemployment spells in the IEB enables us to distinguish between worker inflows into establishments out of unemployment, and inflows out of other establishments.

We count an individual worker as an inflow in plant i if she was employed in another plant j before and both employment spells are at least seven days long. We ignore gaps between employment spells at the same establishment that are less than seven days long, i.e., we comprise consecutive spells with interruptions shorter than seven days into one employment relationship. The key criterion for the identification of inflows from other establishments is a change in the establishment identification number (ID). In this context three issues have to be discussed. First, a worker could be employed by two employers at the same time. For each point in time (i.e. each day), we assign each worker to a single employer, using the highest daily wage as the criterion of assignment. Second, as Hethey and Schmieder (2010) point out, establishment IDs appear and disappear not only in case of plant creation and closure, but also in case of spin-offs, acquisitions, restructurings, and changes of owner. In our context, this means that we must not consider flows between establishment IDs to be real labor flows if all or a substantial fraction of incoming workers come from the same establishment ID, as this might reflect a spin-off, restructuring, acquisition, or change of owner. For each establishment and year, we check for and remove clustered outflows from an establishment ID that, according to Hethey and Schmieder (2010), are probably incidents of an owner change, acquisition, or similar events. Third, we must ensure that establishments between which we observe worker flows are not part of the same firm. We use a Stata routine developed by Schäffler (2014) to estimate which establishments probably belong to the same firm, and disregard worker flows between such establishments. We cannot identify the firms that establishments belong to and use the firm as the level of productivity analysis, since we only have data on productivity-relevant variables for a sample of establishments. However, we understand the establishment as the natural unit of production, i.e. the spatially fixed unit in which a group of workers come together to produce a good or provide a service. Especially in the context of worker inflows, who supposedly transfer knowledge by personal contact to their new co-workers, we think that the establishment is the best suited level for the analysis of productivity effects.

Since the IEB contain no information on establishment-level variables like value added or capital, we draw these data from the IAB Establishment Panel, an unbalanced panel survey of German establishments, of which we use the waves 2003-2011 (see Fischer et al. (2009) for more information on the Establishment Panel). Due to the unique common establishment identifier in both data sets, we can merge individual employment biographies to each of these establishment observations. For details on the linking of employer and employee data, see e.g. Heining et al. (2013). In line with most of the previous literature, we only consider establishments in the manufacturing sector, which we define as the range of NACE[3] Rev. 1.1 (or, equivalently, ISIC[4] Rev. 3.1) divisions 15 through 41.[5] The interpretation of revenues (proxy for output) and intermediate inputs, and therefore value added, is more consistent when focusing on this sector; also, both of these variables are much better filled in manufacturing than in services or other sectors. A problem of the IAB Establishment Panel well recognized in the literature is that the entity referred to as the establishment may differ between the administrative records and the survey. To address this problem, we compare the total numbers of employees reported in the administrative register and the survey. Following Alda (2005), we therefore drop establishment observations for which the reported numbers deviate from each other by more than 40, 30, 20, or 10 percent, depending on the establishment size class (as reported in the administrative data).

5. Empirical model and estimation sample

5.1 The spillover potential approach

Essentially, the approach of our study is to relate establishments' productivity to worker inflows from "superior" firms, similar to Balsvik (2011), Stoyanov and Zubanov (2012, 2014), and Serafinelli (2013). In all of these studies, firm productivity is regressed on some measure of such

[3] Nomenclature Générale des Activités Economiques dans I'Union Européene.
[4] International Standard Industrial Classification.
[5] Within the period from which we draw data, the industry classification scheme has changed several times, notably, from the Classification of Industries 1993 (WZ93) to WZ03 in 2003 and from WZ03 to WZ08 in 2008. We deal with this problem by merging the industry code assigned by Eberle et al. (2011), who used intertemporal imputation of industry codes within establishments (establishments virtually never change industries) and a crosswalk between different classifications.

worker inflows, controlling for other productivity determinants. In some sense, Stoyanov and Zubanov's (2012, 2014) "productivity gap" approach is the most general, meaning that worker inflows from more productive firms in general (not from, e.g., MNEs in particular) should increase hiring firms' productivity. Serafinelli's (2013) approach is equally general, but uses the assumption that wages (respectively, wage fixed effects) are an adequate proxy of productivity. We will review all of these approaches in detail, to motivate the choice of our own approach and the measures we take to address potential biases.

The starting point in all of the cited studies is to distinguish worker inflows according to whether they possess a specific potential to exert spillovers on productivity. In Serafinelli (2013), firms are divided into high-wage firms (HWFs) and non-HFWs. They are classified into either category based on their fixed wage effect, as obtained from an AKM[6] regression of individual wages; the identification of firm fixed effects is based on worker mobility between firms (that is, individual-worker effects are explicitly removed from the firm fixed effect by following worker movements across firms and identifying also worker fixed effects). HWFs are the top third of the firm fixed wage effects distribution, Non-HWFs are the remaining two thirds. Only Non-HWFs are used for the analysis of worker inflow effects on productivity. Essentially, the approach can be summarized in the following estimation equation:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_H W_{it} H_{it} + controls_{it} + \varepsilon_{it} , \qquad (1)$$

with y = log value added, l = log labor, k = log capital, and i and t being the firm and time indices, respectively. The parameter of interest is β_H, the productivity coefficient of the number of hirings (not in logs) from high-wage firms. Yet, workers changing form high-paying to lower-paying firms might be negatively selected, giving rise to a potential "lemons bias." Serafinelli (2013) addresses this problem by weighting the number of worker inflows from high-wage firms by these workers' average individual ability (i.e. their individual fixed wage effect, also obtained from the AKM regression), represented here by W_{it}. His findings indicate that a high number of hirings from HWFs causes productivity gains, since a number of measures against endogeneity bias are employed, including an IV approach using regional HWF downsizings as an instrument.

In Stoyanov and Zubanov (2012), the supposed carriers of productivity spillovers are inflows from more productive firms. They are referred to as "spillover potentials" (SPs), all other inflows

[6] AKM refers to Abowd, Kramarz, and Margolis (1999), who proposed a procedure to disentangle worker and firm fixed effects on wages. The computation is further discussed in Abowd, Creecy, and Kramarz (2002).

are termed Non-SPs. We will use the same terminology in our own analysis. To identify the status of each worker flow (SP or Non-SP), Stoyanov and Zubanov (2012) compute the productivity gap as the TFP[7] difference between the sending and hiring firms for each hiring firm *i*, sending firm *j*, and year *t*. TFP has to be obtained by a first-stage production function estimation; value added and input data are obtained in their study from balance sheet data. To ensure that the productivity gap indicates the superiority/inferiority of a comparable sending firm for each worker inflow, firm productivity A_{it} (TFP) is normalized by the respective industry-year average. This productivity gap is averaged across all worker inflows to firm *i* in some year *t-s* as follows:

$$\overline{gap_{\iota,t-s-1}} = \frac{1}{H_{i,t-s}} \left(\sum_{j=1}^{J} A_{j,t-s-1} - A_{i,t-s-1} \right)$$

This is the mean of all (i.e. positive and negative) productivity gaps between inflows' sending establishments *j* and the hiring establishment *i*, in year *t-s-1*, assuming that there is exactly one inflow from each sending establishment (in practice, there can of course be several inflows from one establishment, which are then weighted according to their number). It is computed for year *t-s-1* since this is the last year that the newly hired workers potentially have spent completely at their sending establishments. The central measure of spillover potential embodied in worker inflows to establishment *i* at year *t* is

$$sh\_gap_{it} = \frac{H_{i,t-s}}{L_{it}} * \overline{gap_{\iota,t-s-1}}$$

where $\frac{H_{i,t-s}}{L_{it}} = S_{it}$ is the share of all worker inflows in year *t-s* in the hiring firm's total employment in year *t*. Stoyanov and Zubanov (2012, 2014) set *s=1*; we will consider both *s=1* and *s=2*.[8] Overall, thus, sh_gap_{it} captures both a quantitative and qualitative effect of worker inflows: the number of inflows as a fraction of total employment, and the positive or negative productivity differential with respect to their sending establishments. The joint effect of both dimensions should capture the impact that hiring workers from more (less) productive establishments has on

[7] Alternatively, value added per worker.
[8] In our setting, with s=1, inflows who arrived on Dec. 31 of year t-1 have the same weight as inflows who arrived on Jan. 1, t-1. With s=2, and ensuring that these inflows are still employed at year t, it is ensured that inflows have at least one year of tenure, and therefore might be a positive selection in terms of job matching. More fundamentally, these inflows have had more time to exert an influence on establishment productivity.

productivity. The main estimation equation derived in Stoyanov and Zubanov (2012) thus has the following structure:

$$A_{it} = \sum_{k=1}^{K} \alpha_k A_{i,t-k} + \vartheta sh\_gap_{it} + \gamma_1 ESTAB_{it} + \gamma_2 EMPL_{it} + \gamma_3 HIRE_{it} + \varepsilon_{it} \qquad (2)$$

where A_{it} is establishment i's productivity at year t, the $A_{i,t-k}$ are K lags thereof,[9] gap_{it} is as defined above, and $ESTAB_{it}$, $EMPL_{it}$, and $HIRE_{it}$ are control variables concerning the establishment, its incumbent workforce, and the newly hired workers, respectively. The variable sh_gap_{it} includes inflows from both more and less productive firms, but since the focus is on the former (SPs), it is also calculated separately for both groups:

$$sh\_gap\_SP_{it} = s_{it} * \overline{gap\_SP_{\iota,t-s-1}}\,,$$

$$\text{where} \quad s_{it} = \frac{H\_SP_{i,t-s}}{\tilde{L}_{it}}\,,$$

$$\text{respectively,} \quad sh\_gap\_Non\_SP_{it} = \frac{H\_Non\_SP_{i,t-s}}{\tilde{L}_{it}} * \overline{gap\_Non\_SP_{\iota,t-s-1}}\,,$$

($H\_SP_{i,t-s}$ and $H\_Non\_SP_{i,t-s}$ are the number of SPs/Non-SPs hired in period *t-s*), so equation (2) becomes

$$A_{it} = \sum_{k=1}^{K} \alpha_k A_{i,t-k} + \vartheta_1 sh\_gap\_SP_{it} + \vartheta_2 sh\_gap\_Non\_SP_{it} + \gamma_1 ESTAB_{it} + \gamma_2 EMPL_{it} + \gamma_3 HIRE_{it} + \varepsilon_{it} \qquad (2')$$

The main parameters of interest are the ϑs, the productivity coefficients of worker inflows from more/less productive establishments. The higher the share of inflows with a positive productivity gap in total employment, and the higher their average productivity gap itself, the higher is the potential for productivity spillovers through worker inflows. In contrast, no effect is expected for inflows from less productive firms (Non-SPs) or a high (absolute) amount of their mean gap, which is negative (but inverted into absolute values for a consistent interpretation of ϑ for both inflow groups).

To ensure that the productivity coefficient of hiring SPs is not biased by their individual productivity characteristics, the vector $HIRE_{it}$ controls for individual fixed wage effects obtained from an AKM estimation, corresponding to W_{it} in (1), absorbing the effects of worker inflows' individual ability. If, as supposed, a positive and significant estimate of ϑ is found for SPs but not for

[9] How many lags are required is to be determined empirically.

Non-SPs while holding constant both groups' average individual ability, this would indicate that SPs' highly productive previous employers *j* have equipped them with productivity-enhancing knowledge that spills over to the hiring firms *i*.

Elaborating this approach further, Stoyanov and Zubanov (2013, 2014) propose to use a production function framework not only to obtain TFP, and measure the productivity gap, but also to estimate the effect of worker inflows on hiring firms' productivity – the equivalent of equation (2) (second stage). More precisely, the estimation approach is derived from a production function with labor as a heterogeneous input consisting of two groups: SPs and all other workers. The production function framework can be summarized as follows:

$$Y_{it} = A_{it} K_{it}^{\beta_K} L_{it}^{\beta_L}$$

is the production function in Cobb-Douglas form, where Y_{it} is the value added of firm *i* in year *t* (therefore, intermediate inputs are omitted). Labor in efficiency units is defined as

$$L_{it} = L_{it}^{rest} + \varphi_{it} L_{it}^{SP} = \left(L_{it}^{rest} + L_{it}^{SP}\right)(1 - s_{it} + s_{it}\varphi_{it}) = \tilde{L}_{it}[1 + s_{it}(\varphi_{it} - 1)] ,$$

with $\tilde{L}_{it}$ being the "nominal" total number of workers, $s_{it} = \frac{H\_SP_{i,t-s}}{\tilde{L}_{it}}$ being the share of SPs in total employment as above, and the productivity advantage of SPs over other workers being $\varphi \geq 1$. Inserting this in the production function yields

$$Y_{it} = A_{it} K_{it}^{\beta_K} \tilde{L}_{it}^{\beta_L} [1 + s_{it}(\varphi_{it} - 1)]^{\beta_L} ,$$

indicating that the labor productivity effect of hiring SPs is described by the factor

$$1 + s_{it}(\varphi_{it} - 1)$$

and their effect on total factor productivity is

$$[1 + s_{it}(\varphi_{it} - 1)]^{\beta_L}.$$

Using the (empirically true) statement that $s_{it}(\varphi_{it} - 1)$ is close to 0, we arrive at the following functional form in logs:

$$y_{it} = a_{it} + \beta_k k_{it} + \beta_l l_{it} + \beta_l (\varphi_{it} - 1) s_{it}$$

(Stoyanov and Zubanov, 2013, 2014). Since φ_{it} is an unknown model parameter, the TFP effect of hiring SPs ($[1 + s_{it}(\varphi_{it} - 1)]^{\beta_L}$) is identified empirically by the coefficient ϑ in the estimation equation

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \vartheta_1 \widehat{gap}_{it} \hat{s}_{it} + controls_{it} + e_{it} \qquad (3)$$

(log TFP is omitted now because it is contained in $\widehat{gap}_{it}$), which is the equivalent of estimation equation (2), where sh_gap_{it} is the equivalent of $\widehat{gap}_{it} \hat{s}_{it}$, and accordingly, ϑ_1 is equivalent in equations (2) and (3) (except that sh_gap_{it} in (2) includes both SPs and Non-SPs, while $\widehat{gap}_{it} \hat{s}_{it}$ in (3) includes only SPs). Thus, the effect of hiring SPs is identified directly within the estimation of a standard production function, and de facto treated as an additional factor input. According to this specification, again, the productivity effect of hiring SPs is linearly increasing in the size of the productivity gap times the share of SPs in total employment; the output gain due to hiring SPs is then $\hat{\vartheta} \widehat{gap}_{it} \hat{s}_{it}$.

Even though the estimation approaches (2) and (3) are structurally different, both arrive at a robustly similar result (Stoyanov and Zubanov 2012, 2014): A high share of SP inflows, weighted by the extent of their productivity gap, has a positive productivity effect, while no effect is found for Non-SPs. This result is generally in line with the findings of Serafinelli (2013). A crucial difference between Stoyanov and Zubanov's (2012, 2014) "productivity gap" approach and Serafinelli's (2013) "HWF" approach is that the former uses the magnitude of the productivity gap to weight the (relative) number of inflows (SPs vs. Non-SPs), while Serafinelli's (2013) identification is based on the mere number of HWF inflows, thus using less information.[10] In that sense, the productivity gap approach appears preferable to the HWF approach.

We implement both approaches (and different versions of them) to test whether Stoyanov and Zubanov's (2012, 2014) additional consideration of the productivity gap's magnitude improves the precision of the estimates. Applying a production function framework as in Stoyanov and Zubanov (2014), our estimation equation can be spelled out as follows:

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \alpha_k y_{i,t-k} + \beta_k k_{it} + \beta_l l_{it} + \tau_1 SP_{it} + \tau_2 Non\_SP_{it} + \gamma_1 ESTAB_{it} + \gamma_2 EMPL_{it} + \gamma'_3 HIRE\_SP_{it} + \gamma'_4 HIRE\_Non\_SP_{it} + \varepsilon_{it} , \qquad (4)$$

where lower-case letters indicate logs; y is log value added, of which we include K lags, the number of which is determined empirically; materials (intermediate inputs) are already accounted for by y; and $HIRE\_SP_{it}$ and $HIRE\_Non\_SP_{it}$ are the control variables as before, for SPs and Non-SPs, respectively. To test all measures proposed in the cited studies, we measure SP_{it} and Non_SP_{it} in four different ways:

[10] Recall that both approaches control for inflows' individual ability using AKM effects.

First, as in Serafinelli (2013), we consider the absolute number of SPs and Non-SPs, specifying that ϑ_1 and ϑ_1 are semi-elasticities. Second, we suppose that SP and Non-SP inflows' productivity effect increases by a constant coefficient not in their absolute number, but in a relative change therein, so we define SP_{it} and Non_SP_{it} as the log of the number of SPs/Non-SPs, estimating SP_{it} and Non_SP_{it} as elasticities. Third, approaching the specification of Stoyanov and Zubanov (2012, 2014), we compute the share of SPs and Non-SPs in establishments' total employment, assuming that it is the number of inflows in relation to the establishment size which matters for inflows' productivity effect. Fourth and finally, we adopt the productivity gap approach of Stoyanov and Zubanov (2012, 2014) by including not only the share of SPs/Non-SPs in total employment, but its interaction with the average productivity gap between their sending establishments and the receiving establishment, i.e. $sh\_gap\_SP_{it}$ and $sh\_gap\_Non\_SP_{it}$.

This final approach uses most information and is therefore preferred. However, we also present estimation results from the three other approaches to assess their explanatory power. Additionally, when adopting the productivity gap approach, we replicate all of Stoyanov and Zubanov's (2012, 2014) specifications, i.e. with (3) and without (2) consideration of capital. In the following, we discuss our implementation of the productivity gap approach on the basis of our data.

5.2 The productivity gap approach and its implementation using wage data

Unlike Stoyanov and Zubanov (2012, 2014), we do not have balance sheets or any other data on all sending establishments' output, sales, or inputs, limiting our possibilities of measuring the productivity gap between sending and hiring establishments ($\overline{gap_{i,t-s-1}}$). We therefore must construct a productivity proxy for all establishments between which we observe worker flows. To conduct an analysis comparable to the existing literature, similar to Serafinelli (2013), we proxy establishments' productivity using establishment fixed effects on their employees' wages. We obtain this fixed effect from OLS wage regressions of *all* regular full-time workers contained in the IEB's employment records, i.e. all full-time workers subject to social security contribution (excluding apprentices, interns etc.; some 20 million individuals). By using the complete IEB data set, we ensure that each establishment fixed effect reflects the establishment's wage level compared to all other establishments in the economy, such that the variable is correctly scaled with regard to the productivity gaps between any two establishments. We perform the regression separately for each of the years 1999-2008, thus identifying the establishment fixed effect not from variation in time, but variation across employees (every employee is contained only once per year, with his or her main job). More explicitly, we estimate

$$\begin{aligned} lnwi_{p,j,t-s-1} &= \beta_0 + \beta_1 male_p + \beta_2 age_{p,t-s-1} + \beta_3 German_{p,t-s-1} \\ &+ \sum\nolimits_{l=1}^{L} \beta_{4,l} occ\_stat_{l,p,j,t-s-1} + \sum\nolimits_{m=1}^{M} \beta_{5,m} qual_{m,p,t-s-1} \\ &+ \sum\nolimits_{n=1}^{N} \beta_{6,n} d\_occ2_{n,p,j,t-s-1} + \theta_{j,t-s-1} + \epsilon_{p,j,t-s-1} \end{aligned}$$

where $lnwi_{p,j,t-s-1}$ is the imputed log wage of worker *p* working at establishment *j* in year *t-s-1*; $male_p$ and $German_{p,t-s-1}$ are binary dummies taking on value one for male workers respectively German citizens; $age_{p,t-s-1}$ is the worker's age, $occ\_stat_{l,p,j,t-s-1}$ is a categorical variable indicating the occupational status of worker *p* in that particular job at plant *j* (e.g., blue-collar vs. white-collar, which can be related to different wage groups defined in collective agreements), $qual_{m,p,t-s-1}$ is a categorical variable of worker *p*'s qualification level at year *t-s-1*, and $d\_occ2_{n,p,j,t-s-1}$ is a two-digit occupation dummy. Wages, which are censored at the social security contribution limit (censoring concerns some 15 percent of employees), are imputed for censored observations adapting a method proposed by Gartner (2005).

(RESULTS OF WAGE REGRESSION HERE; TO BE DONE)

From this regression, we obtain the establishment fixed wage effect $\hat{\theta}_{j,t-s-1}$ for every sending establishment *j* (and every hiring establishment *i*). This parameter is the wage effect (premium) due to the establishment, net of most individual-level determinants. It is then regressed on a set of industry dummies at the three-digit level, again analogous to Stoyanov and Zubanov (2012), yielding a corrected establishment fixed effect $\widehat{\theta'}_{j,t-s-1}$. This correction accounts for systematic productivity differences, respectively differences due to the particular production processes, between industries, that we do not want the productivity gap variable to pick up. To some degree, this normalization should also improve the quality of $\hat{\theta}_{j,t-s-1}$ as a productivity proxy: A large part of $\hat{\theta}_{j,t-s-1}$ might be due to establishment-specific compensation policies but not the establishment's productivity, but $\widehat{\theta'}_{j,t-s-1}$ is cleared of potential wage-setting effects occurring at the industry level. Since wage bargaining in Germany typically takes place at the industry level, especially in the manufacturing sector, this correction should help us to obtain a slightly better productivity proxy. We thus compute the productivity gap between two establishments as the difference between these establishments in terms of $\hat{\theta}'_{j,t-s-1}$. Taking the average of all productivity gaps thus yields $\overline{gap_{\iota,t-s-1}}$.

Whether the establishment fixed effect is a good productivity proxy must be evaluated using more direct measures of productivity, where available. We have information on value added and capital from the IAB Establishment Panel for the receiving establishments, so we can assess the quality of $\hat{\theta}_{j,t-s-1}$ as a productivity proxy by looking at the correlations with several measures of productivity. Table 1 presents these correlations.[11] While far from a perfect fit, $\hat{\theta}_{j,t-s-1}$ is fairly correlated with log value added and log value added per worker. The correlation with TFP, obtained as the residual from a simple OLS regression of value added on capital and labor (all in logs), is rather low at .277. This may be due to an imprecise measurement of value added and capital (compared to balance sheet data), both of which we obtain from survey data. As capital is itself proxied using investment data and the perpetual inventory method, measurement error should be even more severe in the case of the TFP variable, which is indeed what we find.

Table 1

	$\widehat{\theta_{Jt}}$	value added	log value added	VA per worker	log VA per worker	TFP
$\widehat{\theta_{Jt}}$	1.000					
value added	0.317	1.000				
log value added	0.630	0.560	1.000			
VA per worker	0.339	0.275	0.473	1.000		
log VA per worker	0.447	0.265	0.630	0.759	1.000	
TFP	0.277	0.099	0.355	0.718	0.941	1.000

Irrespective of the quality of our productivity proxy and like Serafinelli (2013), we face the problem that workers moving from high-paying to less high-paying firms could be negatively selected. To assess whether SPs in our sample are affected by such a "lemons bias," we use the information available to us concerning the potential selectivity of SPs (and Non-SPs) compared to similarly skilled co-workers at their sending establishments. We keep the estimate of $\epsilon_{p,j,t-s-1}$ from the above wage regression and normalize it by $\hat{\theta}_{j,t-s-1}$, the residual establishment fixed effect:

[11] By construction, we obtain two values of $\widehat{\theta_{Jt}}$ for each establishment observation, one from each of the wage regressions for the years t-s-1 (s=1, 2). The table only reports the values based on the former regression for s=1, but those from the s=2 regression are virtually identical.

$$\epsilon'_{p,j,t-s-1} = \frac{\epsilon_{p,j,t-s-1} - \overline{\epsilon_{j,t-s-1}}}{SD(\epsilon_{j,t-s-1})} = \frac{\epsilon_{p,j,t-s-1}}{SD(\epsilon_{j,t-s-1})}$$

The parameter $\epsilon'_{p,j,t-s-1}$ indicates each individual worker's relative pay position within his or her gender, nationality, qualification and occupation group, within the (sending) establishment, and controlling for age (as a proxy for experience). The establishment's mean individual residual $\overline{\epsilon_{J,t-s-1}}$ is equal to zero because the wage regression includes a constant. Thus, positive values of $\epsilon'_{p,j,t-s-1}$ indicate above-average pay, while negative values indicate the opposite. We can therefore determine for each worker inflow whether the worker is positively or negatively selected from his or her sending establishment. Concerning the correction for lemons bias in the final estimation, as yet, we can not (unlike the previous studies) use individual fixed effects independently of establishment fixed effects, in the spirit of AKM. This remains to be done. For the moment, we proxy worker inflows' individual ability by their education and experience.

5.3 Specification details and control variables

Some remarks are in order concerning the exact choice of data underlying our inflow variables. Since we consider the newly hired workers as knowledge carriers, we require them to satisfy several conditions. Most importantly, we exclude all inflows of unqualified workers, since we assume that they do not perform work tasks that involve the use of (considerable) technological knowledge. We thus require inflows to have a tertiary education or at least hold a vocational degree. Next, we choose to allow half a year (182 days) as the maximum gap between consecutive employment spells with different employers for a job move to be considered a relevant job-to-job transition. In case of a period of unemployment, it must not be longer than three months. We exclude all inflows employed as an apprentice or intern, as well as inflows of "marginal" employees.[12] However, we do include inflows out of (graduating from) vocational training who, when entering the establishment, become regular employees, since they have been employed for an average of three years in their training establishment and occupation. Finally, only incoming workers between the ages of 15 and 65, the official retirement age, are included.

Concerning the main production factors, labor input is given by total employment on June 30th of year t. It is defined as the full-time equivalent volume of work according to the BA's employ-

[12] Marginal employment is defined as employment not subject to social security contribution, with the monthly wage not exceeding (currently) 450 Euros, see Section 8, Subsec. 1, No. 1, of the German Social Code IV (SGB IV).

ment statistics.[13] To approximate the capital stock, we use the modified perpetual inventory method (PIM) by Müller (2008), deducing capital from net investment, which is surveyed in the Establishment Panel. The method uses investment data to infer the capital stock and industry-level depreciation rates for different categories of investment goods to account for depreciation. In our view, this method is adequate for the manufacturing sector, where the quality and depreciation of capital should be comparable within each of the different manufacturing industries. As emphasized by Ehrl (2013), whose procedure we also employ, the PIM must be further corrected for restructuring events such as insourcing, closure, sell-off, and spin-off of parts of the establishment.

The vector of control variables *ESTAB* includes categorical variables indicating whether the establishment is part of a larger enterprise, its legal form, the (self-reported) state of technical equipment, a dummy indicating young establishments (less that ten years old), the share of exports in total revenues, and a dummy indicating location in East Germany (which has still structurally lower productivity and wage levels than West Germany). *EMPL* is the vector of employment structure controls, containing the share of high-qualified employees (holding a university or university of applied sciences degree), the mean age, and the share of males among all employees. The vector *HIRE* controls for inflows' human capital, which might have a separate effect on productivity. It includes the share of high-qualified inflows in all inflows (those with a university or university of applied sciences degree), their mean age (proxy for work experience). These controls should account for workers' ability to carry knowledge between establishments (more highly qualified workers may be more able to do so), but not for the knowledge (superior or inferior knowledge, depending on the sending establishment's productivity) that they carry.[14] Since SPs less often increase their wage when moving to an establishment in our sample than do Non-SPs (see descriptive results), we include the share of SPs (Non-SPs) who increase their wage as a (rough) measure to rule out "lemons bias" and reverse causality bias (highly productive firms can pay higher wages and thus attract the workers they desire most) to some extent.

[13] Wo do not observe the exact hours of work, but only whether an employee works full-time, less than full-time but at least 18 hours, or less than 18 hours a week. Full-time equivalents are proxied by weighting "big" and "small" part-time with a factor of 0.6, respectively, 0.3.

[14] For establishment observations with zero inflows, obviously, we cannot compute the mean productivity gap, let alone the separate mean gaps for SPs and Non-SPs, or inflow control variables such as their share of high-qualified. In order not to lose these observations, the value of these variables is set to zero and we include a binary dummy taking the value one for these observations, which allows for them to have a systematically different constant effect (intercept).

Analogous to inflows, we can identify outflows in the same period (*t-2*), i.e. employees whose last employment period at the establishment ends in *t-1*, respectively *t-2*.[15] It could be necessary to control for outflows because the higher their number (relative to establishment size), the more likely an establishment is to replace them, raising the number of inflows. If establishments are not aware of, or not able to influence, the spillover potential to be gained from worker inflows (superior/inferior knowledge from sending establishments), including outflows as a control variable should not make a difference to the estimate of ϑ_1 and ϑ_2. The reason is that, if inflows are carriers of establishment-level knowledge (which the spillover potential approach assumes), any productivity effects of their hiring should be due to the knowledge transfer, regardless of why they have been hired. Furthermore, the parameters τ_1 and τ_2 estimate an interaction effect of the productivity gap and inflow intensity (i.e. these effects already account for the number of inflows as compared to establishment size). Finally, outflows do subtract knowledge from the establishment, which is to some extent replaced by inflows' knowledge, but the spillover potential we measure by sh_gap_{it} does not refer to this individual knowledge, but focuses on the part of knowledge transferrable between establishments. In fact, sh_gap_{it} is hardly correlated with outflow intensity (number of outflows divided by total number of employees; correlation coefficient below 0.1). Therefore, we do not to include outflows in our specification (neither have any of the preceding studies).

As the number and origin of inflows are subject to establishments' hiring decisions, it might be necessary to control for other channels of knowledge acquisition and accumulation. The most important of these channels are hiring graduates directly out of tertiary education[16] and apprentice training.[17] These are alternative (potentially competing) strategies aimed at accumulating productive knowledge. Again, if establishments are aware of (and able to adjust) the spillover potential of inflows in the face of their other knowledge-accumulating HR practices, we would need to control for graduate hiring and apprentice training. Again, however, correlations between these variables and sh_gap_{it} are below 0.1; therefore controlling for the shares of inflows from tertiary education and apprentices does not appear necessary from an empirical point of view. Finally,

[15] By definition, t-2 outflows must not return to the establishment within the year t-1, which keeps the timing analogous to inflows (t-2 inflows are present throughout t-1, while t-2 outflows remain absent).

[16] We cannot identify inflows out of tertiary education directly, but we can detect them quite plausibly as employees for whom there is no prior employment or unemployment (benefit receipt) information, who hold a university (or university of applied sciences) degree, and who are no older than 30 at the time of job entry.

[17] In Germany, vocational training mostly takes place on-the-job in firms.

our main regressions include year dummies, industry dummies at the two-digit level, and labor market region[18] fixed effects (where applicable).

6. Descriptive analysis

Our final sample contains 1,892 establishments and ranges over the years 2002 to 2010, i.e. we have up to nine (and at least four) observations per establishment. For somewhat more than half of all establishment observations, we observe a positive number of inflows; naturally, when we consider inflows from the previous year (*t-1*), the number of inflows is higher than with respect to year *t-2*.[19] Table 2 shows worker characteristics for stayers (hiring establishments' incumbent workers) and inflows separated into SPs and Non-SPs. Generally, inflows are younger than stayers. They are also much more often high-qualified (holding a university or university of applied sciences degree). Among the inflows, slightly less than 40 percent are SPs, meaning that we are more likely to observe inflows from less to more productive establishments than the other way around. A striking difference between SPs and Non-SPs is that the latter are much more likely to increase their wage when moving to the receiving establishment. We therefore control for the share of SPs/Non-SPs who increase their wage in our estimations.

Table 2

Worker characteristics	**Stayers**	**Inflows t-1**			**Inflows t-2**		
		All	**SP**	**Non-SP**	**All**	**SP**	**Non-SP**
Share high-qualified	0.098	0.228	0.296	0.184	0.215	0.279	0.176
Share male	0.788	0.797	0.781	0.807	0.802	0.785	0.812
Mean age	41.9	35.8	36.9	35.1	36.7	37.6	36.2
Share SPs		0.392			0.379		
Share intra-regional		0.569	0.532	0.592	0.596	0.566	0.615
Share wage increase		0.784	0.654	0.868	0.8	0.672	0.877
Share from same 2d ind.		0.165	0.162	0.168	0.157	0.155	0.158
Share from same 3d ind.		0.11	0.107	0.112	0.102	0.101	0.103
Share low-skilled (occ.)		0.378	0.303	0.427	0.379	0.312	0.419
Share mid-skilled (occ.)		0.464	0.49	0.447	0.471	0.493	0.458
Share high-skilled (occ.)		0.158	0.208	0.126	0.15	0.195	0.123
N (establishments)	11,797	6,815	5,140	4,906	6,348	4,675	4,531
N (individuals)	2,276,540	51,343	20,151	31,192	44,664	16,922	27,742

[18] Labor market regions are defined as in Kosfeld and Werner (2012), i.e. they are clusters of NUTS 3 regions (*Landkreise*, i..e. districs) strongly connected by commuting flows.
[19] This is because t-1 inflows can have arrived on Dec. 31st, year t-1 and they are only required to be employed at the hiring establishment on Jan. 1st, year t.

Inflows' wages are considered in more detail in Table 3. We see that, while SPs' wage levels (in their new job, i.e. at the receiving establishment) are not much behind Non-SPs', the difference is statistically significant at the 1% level. Wage changes, in contrast, are also economically significantly apart, with Non-SPs experiencing almost twice as large wage gains from changing establishments than SPs. The table also provides the results of a t-test on equality of means of the individual wage effect, which we obtained from the wage regression (4). Since this regression includes establishment fixed effects, the individual effect is the person's residual with respect to similarly qualified co-workers in his or her sending establishment. The differences in signs (negative for SPs, positive for Non-SPs) and the highly significant difference in magnitude between SPs' and Non-SPs' individual effects suggest that the former are negatively, and the latter positively selected from their sending establishments.

Table 3

	Inflows t-1				Inflows t-2			
	All	SPs	Non-SPs	p	All	SPs	Non-SPs	p
Wage	89.28	87.70	90.17	0.000	90.01	87.80	91.21	0.000
Wage change	22.64	13.96	27.45	0.000	23.98	15.31	28.60	0.000
Ind. effect ($\epsilon'_{p,j,t-s-1}$)	0.0508	-0.0239	0.1004	0.000	0.0381	-0.0450	0.0895	0.000

Note that wages refer to the receiving establishment, wage changes refer to the difference between sending and receiving establishment, and the individual effect refers to the sending establishment

Table 4

(Only t-1 inflows)	All firms	Zero hiring	Pos. hiring	hire SPs	hire Non-SPs
Log value added	14.924	13.753	15.779	15.922	16.205
Log VA per worker	10.995	10.804	11.134	11.131	11.222
Log capital	14.220	12.659	15.361	15.560	15.868
Log labor	3.928	2.949	4.644	4.790	4.982
Labor (heads)	192.976	39.324	305.301	361.239	391.693
East dummy	0.510	0.614	0.434	0.437	0.368
N	11,797	4,982	6,815	5,140	4,906

Turning to the establishment level, Table 4 displays characteristics of establishments, separated by whether they had any hiring, zero hiring, hiring of SPs, or hiring of Non-SPs in period t-1 (statistics based on t-2 inflows omitted, but very similar). Clearly and unsurprisingly, establishments with a positive number of hires are larger and have higher value added and capital levels than non-hiring establishments. Among those which hire any workers, those hiring SPs are smaller and have less value added and capital than those hiring at least one Non-SP worker. This was to be expected: By definition, hiring SPs means hiring from more productive establishments; thus, the larger and more productive an establishment, the less likely it is for a given worker inflow to be an SP. As we can also see, we have a disproportionately large share of establishment observations in East Germany (this is due to the sample design of the Establishment Panel; in the entire population, the share of East establishments would be in the order of 20 percent). Yet, among the hiring establishments (and those hiring Non-SPs in particular), East German establishments are underrepresented in the sample.

Table 5

A: Sending and hiring establishments' FEs ($\widehat{\theta}'_{j,t-s-1}$; if >0 inflows; t-1 inflows)					
	Obs	Mean	Std. Dev.	Min	Max
Receiving estab.	6815	0.111	0.311	-0.923	1.061
Mean of sending estab.	6815	0.104	0.280	-1.911	1.282
B: Summary statistics for hiring establishments (weighted by number of inflows; t-1 inflows)					
			Inflows' mean productivity gap		
	Obs	Estab. FE	All inflows	SPs	Non-SPs
All establishments	6815	0.279	-0.077	0.199	-0.245
Small (<50)	1888	-0.083	0.085	0.210	-0.131
Large (>=50)	4927	0.303	-0.088	0.198	-0.252

Turning to the hiring of SPs and Non-SPs, let us consider the core explanatory variables of our model in more detail: the establishments' fixed effect (our productivity proxy) and the mean gap, in terms of this variable, between sending and hiring establishments. Panel A of Table 5 presents summary statistics of the establishment fixed effect, both for the receiving establishment and for the mean of inflows' sending establishments. On average, the fixed effect of receiving establishments is slightly higher, in line with the above finding that inflows of Non-SPs are more common than inflows of SPs. Panel B explores the establishment fixed effect in more detail, with respect to the difference between small (<50 employees) and large (>=50 employees) hiring establishments, and regarding SPs and Non-SPs. Naturally, larger establishments have a higher fixed effect, reflecting their higher productivity and wage levels. When we consider the mean productivity gap, on the other hand, larger establishments fare worse than smaller ones, since there are more relatively unproductive establishments to hire from. When separating the mean gaps for SPs and Non-SPs, accordingly, the mean positive (SP) gap is smaller for larger establishments, and the mean negative (Non-SP) gap is larger in absolute terms.[20] These findings are qualitatively in line with Stoyanov and Zubanov's (2012).

7. Econometric analysis

7.1 Estimation technique

To analyze spillover effects, we estimate establishment-level production functions. We can refer to a great body of literature dealing with the econometric issues involved. In a very comprehensive paper, Eberhardt and Helmers (2010) (hf. EH) review the most important problems encountered by econometricians using "fat" panel data (large N, short T) at the firm or establishment level. We strongly refer to their paper for its comprehensiveness and emphasis on the imperfections of the data typically used (availability and quality of output and capital data, need for proxies, etc.).

Essentially, unobserved total factor productivity (TFP) is composed of firms' mean efficiency, period-specific effects, firm-specific effects, and an idiosyncratic component, and since the latter is observed by the firm but not the econometrician, there can be unobserved factors influencing firms' input choices, implying that failing to control for these factors renders OLS and fixed-effects estimates inconsistent. More explicitly, the main problem arises from the possibility of the firm to observe its idiosyncratic TFP effect *before* choosing its levels of capital and labor; the idiosyncratic effect thus is an omitted variable that needs to be controlled for. Otherwise, it is being transmitted to the observed inputs (capital and labor), i.e. the production factors' coefficients take up the idiosyncratic effect and are thus biased upward. In contrast, a downward bias can result from imprecise measurement of inputs (attenuation bias). Contained in the unobserved idiosyncratic TFP effect is also the problem of simultaneity or reverse causality, i.e. the simultaneous or reversed determination of factor inputs with respect to the realized output. In our context, this means that if we find a positive correlation between establishments' productivity and their hiring of certain workers, this might mean that the worker inflows increase productivity due to the superior knowledge from their previous employer, or that highly productive establishments attract these workers because they foresee their positive productivity effect (and because they may be able to receive a positive externality from their hiring).

[20] Note that in the econometric analysis, the mean gap – respectively, the share*gap variable for Non-SPs – is inverted from negative to positive values, so coefficients can be interpreted analogously for SPs and Non-SPs.

EH discuss three approaches to combat these endogeneity biases. The first approach, instrumenting factor inputs using factor prices, can be ignored in the case of our study since the core explanatory variables in our model, the share and productivity gap of worker inflows, are not input factors in a strict sense. For the same reason, we do not address estimation issues arising from the assumption of perfect competition, and the proposed solutions, as are discussed in Van Beveren (2007, 2012). Neither do we emphasize the problem of selection bias arising from the survival (attrition) of highly (un-)productive firms.[21] Instead, we focus on the problem of endogeneity (reverse causality) bias arising from establishments' anticipation of their productivity level and their according choice of inputs. The two main approaches to minimize this bias are, first, control function approaches trying to model the idiosyncratic TFP shock explicitly, and second, dynamic panel data (DPD) approaches making use of internal instruments in panel data sets. The first class of estimators has been developed by Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), Ackerberg et al. (2006) (ACF), and Wooldridge (2009) (WOP); the second class is rooted in the work of Arellano and Bond (1991) (AB) and Blundell and Bond (1998, 2000) (BB).

To construct the control function for the idiosyncratic TFP shock observed by the firm but not the researcher, OP, LP, ACF, and WOP need to assume that this shock is the only unobservable entering the investment (respectively, intermediate inputs) function. This "scalar unobservable assumption" (EH) cannot be tested. More specifically, to identify the labor coefficient, which should be more important, given our core explanatory variables, than identifying the capital coefficient, the structural estimators assume a discrete sequence of establishments' decisions about the particular factor inputs. Again, this assumption cannot be tested empirically (EH, p. 24). At best, the assumption could be plausible in some particular production processes (industries), but we do not expect it to hold across the entire manufacturing sector (let alone other sectors). An advantage of the OP approach in particular is that the selectivity of surviving establishments is controlled for, which is achieved by including a dummy variable indicating whether an establishment exited the market in a given year (cf. Yasar et al., 2008). A particular problem of the OP estimator is that, due to the monotonicity condition regarding the investment-productivity relationship, it cannot be implemented for establishment observations which report zero investment (Van Beveren, 2007, 2012), which is the case for a substantial number of our observations (21 percent).

[21] Yet, in an attempt to implement the Olley and Pakes (1996) estimator, this problem is addressed by design of the estimator.

Using the longitudinal dimension of panel data, the DPD estimators control for time-invariant unobserved establishment heterogeneity. In the context of production function estimation, this means that if the unobserved productivity shock biasing input decisions were time-constant for every establishment, any within-estimator (such as fixed effects) would remove this bias entirely, which of course is a very restrictive and implausible assumption. The DPD estimators indicated above, by using internal IVs, take an additional step to combat this endogeneity bias. Furthermore, unlike the "structural" estimators (OP, LP, etc.), the DPD estimators allow one to test all crucial assumptions made about the data-generating process (DGP). It could thus be argued that, overall, the DPD estimators are a more conservative choice than any of the "structural" (control function) estimators. On the other hand, due to using only within-establishment variation in a fat panel, one may fail to identify effects with any precision using these estimators. Aiming to maximize the robustness of our findings, we employ both classes of estimators.

A final limitation we face, as already pointed out by Stoyanov and Zubanov (2012), is that we cannot control for unobserved hiring preferences regarding the origin of newly hired workers, to the degree that these are not directly related to the unobserved productivity shock we are trying to absorb using structural estimators. Neither can we be sure that such preferences are time-invariant and we get rid of them by using the DPD estimators.

7.2 Results

7.2.1 Main results

Table 6 presents the results from an OLS estimation of equation (4) gradually introducing additional control variables. According to a simple test, two lags of log value added are required to account for serial correlation. Focusing on our explanatory variable of interest, we see that the share*gap of SPs, i.e. the "quantity-times-quality" of skilled worker inflows from more productive establishments, is not significantly related to the hiring establishments' value added. In contrast, for Non-SPs, there is a relatively large and significant positive relationship. This pattern of results is robust to stepwise introduction of additional control variables, and it holds for both *t-1* and *t-2* inflows. The set of controls variables for worker inflows (*HIRE*; not displayed) yields the expected signs: For instance, a high share of high-qualified workers among the SPs/Non-SPs and a high share of wage increases among them are positively related to hiring establishments' value added.

Table 6

OLS regression of log value added, inflows from previous year (t-1)												
L.Log value added	0.588	***	0.585	***	0.577	***	0.556	***	0.551	***	0.544	***
L2.Log value added	0.219	***	0.223	***	0.219	***	0.226	***	0.220	***	0.214	***
Log capital stock	0.012	***	0.012	***	0.012	***	0.011	***	0.012	***	0.011	***
Log labour	0.208	***	0.204	***	0.217	***	0.228	***	0.228	***	0.236	***
Share SPs * mean gap SPs	0.259		0.308		0.154		0.150		0.134		0.086	
Share Non-SPs * mean gap Non-SPs	1.517	**	1.719	**	1.901	**	1.745	**	1.654	**	1.591	**
Year dummies			X		X		X		X		X	
Industry dummies (2d.)					X		X		X		X	
Region dummies (LMR)							X		X		X	
Establishment controls									X		X	
Empl. structure controls											X	
Observations	7298		7298		7298		7298		7298		7298	
R-squared	0.952		0.953		0.953		0.955		0.955		0.955	
OLS regression of log value added, inflows from year before previous (t-2)												
L.Log value added	0.590	***	0.588	***	0.579	***	0.558	***	0.552	***	0.546	***
L2.Log value added	0.219	***	0.223	***	0.219	***	0.225	***	0.219	***	0.213	***
Log capital stock	0.012	***	0.012	***	0.012	***	0.011	***	0.012	***	0.012	***
Log labour	0.215	***	0.211	***	0.223	***	0.235	***	0.235	***	0.242	***
Share * mean gap SPs	-0.724		-0.752		-0.918		-0.904		-0.993		-1.052	*
Share Non-SPs * mean gap Non-SPs	1.369	*	1.632	**	1.771	**	1.589	**	1.531	*	1.478	*
Year dummies			X		X		X		X		X	
Industry dummies (2d.)					X		X		X		X	
Region dummies (LMR)							X		X		X	
Establishment controls									X		X	
Empl. structure controls											X	
Observations	7298		7298		7298		7298		7298		7298	
R-squared	0.952		0.953		0.953		0.955		0.955		0.955	

Dependent variable is log value added. Standard errors clustered at establishment level. HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

Table 7

OLS results, number of inflows	Inflows from previous year (t-1)						Inflows from year before previous (t-2)					
L.Log value added	0.557	***	0.551	***	0.545	***	0.558	***	0.553	***	0.546	***
L2.Log value added	0.226	***	0.220	***	0.214	***	0.225	***	0.219	***	0.213	***
Log capital stock	0.011	***	0.012	***	0.011	***	0.011	***	0.012	***	0.012	***
Log labour	0.225	***	0.224	***	0.232	***	0.238	***	0.237	***	0.245	***
Number of SPs	-0.000		0.000		0.000		-0.002	*	-0.001		-0.001	
Number of Non-SPs	0.000		0.000		0.000		-0.001		-0.001		-0.001	
Establishment controls			X		X				X		X	
Empl. structure controls					X						X	
Observations	7298		7298		7298		7298		7298		7298	
R-squared	0.955		0.955		0.955		0.955		0.955		0.955	
OLS results, log inflows	**Inflows from previous year (t-1)**						**Inflows from year before previous (t-2)**					
L.Log value added	0.556	***	0.551	***	0.544	***	0.558	***	0.552	***	0.546	***
L2.Log value added	0.226	***	0.220	***	0.214	***	0.225	***	0.219	***	0.213	***
Log capital stock	0.011	***	0.012	***	0.011	***	0.011	***	0.012	***	0.012	***
Log labour	0.225	***	0.224	***	0.232	***	0.236	***	0.235	***	0.242	***
Log number SPs	-0.005		-0.001		0.001		-0.025	**	-0.020	**	-0.019	*
Log number Non-SPs	0.004		0.003		0.005		0.013		0.013		0.014	
Establishment controls			X		X				X		X	
Empl. structure controls					X						X	
Observations	7298		7298		7298		7298		7298		7298	
R-squared	0.955		0.955		0.955		0.955		0.955		0.955	
OLS results, share of inflows	**Inflows from previous year (t-1)**						**Inflows from year before previous (t-2)**					
L.Log value added	0.556	***	0.550	***	0.544	***	0.558	***	0.552	***	0.546	***
L2.Log value added	0.226	***	0.220	***	0.214	***	0.225	***	0.219	***	0.213	***
Log capital stock	0.011	***	0.012	***	0.011	***	0.011	***	0.012	***	0.012	***
Log labour	0.225	***	0.225	***	0.233	***	0.235	***	0.234	***	0.242	***
Share SPs	-0.010		-0.006		-0.015		0.005		0.007		-0.001	
Share Non-SPs	0.058	*	0.050	*	0.042		0.078	**	0.073	**	0.069	**
Establishment controls			X		X				X		X	
Empl. structure controls					X						X	
Observations	7298		7298		7298		7298		7298		7298	
R-squared	0.955		0.955		0.955		0.955		0.955		0.955	

Dependent variable is log value added. Standard errors clustered at establishment level. Year, 2-digit industry and labor market region (LMR) dummies included.

*HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01*

When we consider simpler measures of SP/Non-SP inflows, we do not get a similarly robust pattern of results. If we just include the number of each kind of inflows (upper third of Table 7), analogous to Serafinelli (2013), we do not see any significant relationship, except a marginally negative coefficient for SPs from period *t-2* in one specification – namely the questionable specification excluding important establishment and employment structure controls. In contrast, we detect a negative relationship between hiring SPs and log value added when including the number of SPs/Non-SPs in logs (only for *t-2* inflows). This effect also slightly decreases as we include additional establishment and employment controls. However, in line with the other estimation results, these findings indicate that SPs are less favorably related to productivity than are Non-SPs. In the bottom third of the table, we regress log value added on the share of SPs/Non-SPs in total employment, yielding similar results as in our preferred specification using the share*gap variable. Yet, the share*gap specification, which uses more information than the simpler specifications, reveals a more robust pattern of results, so we choose this as our preferred specification.

Table 8

OLS results, VA per worker, inflows t-1						
L.log value added per worker	0.548	***	0.540	***	0.534	***
L2. log value added per worker	0.254	***	0.245	***	0.238	***
Share SPs * mean gap SPs	-0.671		-0.400		-0.390	
Share Non-SPs * mean gap Non-SPs	0.577		0.804		0.856	
Establishment controls			X		X	
Empl. structure controls					X	
Observations	7298		7298		7298	
R-squared	0.701		0.703		0.705	
OLS results, VA per worker, inflows t-2						
L. log value added per worker	0.550	***	0.541	***	0.535	***
L2. log value added per worker	0.254	***	0.246	***	0.238	***
Share SPs * mean gap SPs	-1.390	**	-1.083	**	-1.057	**
Share Non-SPs * mean gap Non-SPs	0.475		0.850		0.928	
Establishment controls			X		X	
Empl. structure controls					X	
Observations	7298		7298		7298	
R-squared	0.700		0.702		0.704	

Dependent variable is log value added per worker. Standard errors clustered at establishment level. Year, 2-digit industry and labor market region (LMR) dummies included. HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

Having resolved the issue of what explanatory variable to use (i.e. the difference between Serafinelli's (2013) count measure and Stoyanov and Zubanov's (2012, 2014) share*gap measure), we may address the question in which estimation framework the relationship between productivity and worker inflows should be investigated – that is, we estimate the alternative specification (2) rather than (4), both of which use the share*gap variable on the right-hand side. In Table 8, we regress log value added per worker on two lags of itself, our share*gap variable for SPs and Non-SPs, and the same controls as above. While the pattern of signs is consistent across *t-1* and *t-2* inflows, only in the latter case do we find a significant relation with productivity – hiring SPs, or hiring them from much more productive establishments, is inversely related to labor productivity. However, as argued by Stoyanov and Zubanov (2012, 2014), since (Non-)SPs' overall productivity effect is the crucial outcome formulated in our hypothesis, we should rather investigate their correlation with total factor productivity (TFP), which we have done implicitly already in the above estimations. In Table 9, we address this relationship explicitly, after obtaining TFP as the residual from a simple Cobb-Douglas regression of log value added on log capital and log labor. These results confirm that, while we still estimate largely negative signs for the quantity and quality of SPs, the only robustly significant finding (for both *t-1* and *t-2* inflows, controlling for establishment covariates) is that hiring Non-SPs correlates positively with hiring establishments' productivity.

While the results so far consistently point in one direction, our estimates may suffer from endogeneity bias. To address this issue, we apply Levinsohn and Petrin's (2003) (LP) estimator, proxying unobserved productivity shocks, which may bias hiring decisions, by intermediate inputs. We implement the LP estimator using the *levpet* command in Stata, developed by Petrin et al. (2004). There are two versions of this estimator, one based on including revenues and intermediate inputs separately and one directly using value added, i.e. the difference of the two. We present results from using both versions (Table 10). The results confirm our previous findings. In its revenue version, the LP estimator generates similar but smaller parameters than the OLS production function (equation 4) and the OLS regression of TFP (equation 2), suggesting that the OLS estimates are biased upwards, as expected if reverse causality is a problem. Aware of the shortcomings of this estimator, we take the results as corroborating our interpretation that hiring Non-SPs, and hiring them from much less productive establishments, could be causally related to higher productivity, while hiring workers from highly productive, high-paying establishments does not seem to affect hiring establishments' productivity.

Table 9

OLS results, TFP, inflows t-1						
L.TFP	0.543	***	0.540	***	0.535	***
L2.TFP	0.240	***	0.236	***	0.230	***
Share SPs * mean gap SPs	-0.055		0.156		0.178	
Share Non-SPs * mean gap Non-SPs	1.085		1.257	*	1.322	*
Establishment controls			X		X	
Empl. structure controls					X	
Observations	7298		7298		7298	
R-squared	0.642		0.644		0.645	
OLS results, TFP, inflows t-2						
L.TFP	0.543	***	0.539	***	0.534	***
L2.TFP	0.241	***	0.237	***	0.232	***
Share SPs * mean gap SPs	-0.504		-0.247		-0.208	
Share Non-SPs * mean gap Non-SPs	1.193		1.482	**	1.573	**
Establishment controls			X		X	
Empl. structure controls					X	
Observations	7298		7298		7298	
R-squared	0.642		0.644		0.645	

Dependent variable is log TFP, obtained from regression of log value added on log capital and log labor. Standard errors clustered at establishment level. Year, 2-digit industry and labor market region (LMR) dummies included. HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

Table 10

LP results	Inflows t-1		Inflows t-2	
Revenue version				
Log intermediate inputs	0.654	***	0.664	***
Log capital stock	0.011		0.002	
Log labour	0.339	***	0.338	***
Share SPs * mean gap SPs	0.253		-0.409	
Share Non-SPs * mean gap Non-SPs	1.285	***	1.119	**
Observations	11797		11797	
Value added version				
Log capital stock	0.016	***	0.016	***
Log labour	0.711	***	0.710	***
Share SPs * mean gap SPs	-0.297		-0.809	
Share Non-SPs * mean gap Non-SPs	2.409	***	2.681	***
Observations	11797		11797	

Dependent variable is log revenues (revenue version) respectively log value added (value added version. Standard errors obtained by bootstrapping with 50 replications. Year, 2-digit industry and labor market region (LMR) dummies included. ESTAB, EMPL and HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

An estimation issue not yet addressed is unobserved heterogeneity of establishments that could bias the productivity parameters of inputs and SP/Non-SP inflows, arising from any time-constant unobserved establishment productivity determinants. These could be controlled for by any estimator that uses only within-establishment variation in the data. Since, however, there are likely also time-variant unobserved productivity determinants at the establishment level, we use the longitudinal information in our data also to instrument first differences by lagged levels of the endogenous explanatory variables (in particular, the share*gap variables). We refer to this as the AB (Arellano and Bond, 1991) estimator. To use additional information and improve the efficiency of the estimation, we also use Blundell and Bond's (1998, 2000) estimator (BB).

Table 11 displays the results. While both estimators pass the Sargan test of instrument exogeneity at conventional confidence levels – the potentially more efficient BB estimator somewhat less convincingly –, we fail to identify a significant coefficient for any of our explanatory variables of interest. As in the LP results, even the capital coefficient, which is not measured very precisely but proxied using perpetual inventory, is rendered insignificant by the instrumentation. This suggests that there is not enough within-variation, or too much noise, in the data to identify productivity effects. We therefore proceed using the LP estimator as a means of alleviating endogeneity bias.

Table 11

AB/BB results	**Inflows t-1**				**Inflows t-2**			
	AB		**BB**		**AB**		**BB**	
L.Log value added	0.200	***	0.390	***	0.212	***	0.419	***
L2.Log value added	0.043		0.102	***	0.049		0.123	***
Log capital stock	0.003		0.010		0.003		0.013	*
Log labour	0.412	***	0.533	***	0.421	***	0.533	***
Share * mean gap SPs	-0.020		-0.360		-0.305		-0.694	
Share * mean gap Non-SPs	-0.286		-0.097		0.506		0.278	
Observations	5384		7298		5384		7298	
Sargan p-value	0.370		0.156		0.430		0.146	

Dependent variable is log value added. Standard errors clustered at establishment level. Year dummies included. ESTAB, EMPL and HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

7.2.2 Disaggregation of inflows by skill group and industry origin

To gain further insight on the relationship between SP/Non-SP inflows and productivity, we divide them into several different groups. First, assuming that higher-skilled workers may be more effective knowledge carriers (as suggested by the findings of Stoyanov and Zubanov, 2012), we distinguish three skills groups based on the workers' occupation in the hiring establishment. Appendix Table A 1 lists the occupational groups falling into the three categories (low-, mid-, and high-skilled). Second, we separate inflows by whether they come from the same or another 2-digit industry (as shown in the descriptive section, even with this broad industry definition, inflows from within the same industry are relatively rare (some 15 percent)). In line with a host of findings in the literature, one might expect that inflows from within the industry can transfer more knowledge about technology and best practices from their previous employer, enhancing their productivity effect at the hiring establishment.

Table 12

LP revenue version	**Inflows t-1**		**Inflows t-2**	
l_intermed	0.654	***	0.665	***
Log capital stock	0.012		0.001	
Log labour	0.342	***	0.340	***
Share*mean gap SPs, low-skilled	0.090		-0.018	
Share*mean gap SPs, mid-skilled	0.084		-0.631	
Share*mean gap SPs, high-skilled	1.315		-0.601	
Share*mean gap Non-SPs, low-skilled	1.973	**	1.146	
Share*mean gap Non-SPs, mid-skilled	0.791		0.969	**
Share*mean gap Non-SPs, high-skilled	5.614	**	6.448	**
Observations	11797		11797	
LP VA version	**Inflows t-1**		**Inflows t-2**	
Log capital stock	0.016	***	0.016	***
Log labour	0.715	***	0.712	***
Share*mean gap SPs, low-skilled	-0.308		-0.385	
Share*mean gap SPs, mid-skilled	-0.958		-1.362	
Share*mean gap SPs, high-skilled	1.824		-0.187	
Share*mean gap Non-SPs, low-skilled	3.751	***	2.478	
Share*mean gap Non-SPs, mid-skilled	1.593	*	2.718	***
Share*mean gap Non-SPs, high-skilled	8.807		11.783	***
Observations	11797		11797	

Dependent variable is log revenues (revenue version) respectively log value added (value added version. Standard errors obtained by bootstrapping with 50 replications. Year, 2-digit industry and labor market region (LMR) dummies included. ESTAB, EMPL and HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

Table 12 displays the results for skill groups. Again, we estimate both the revenue-based and value-added-based specifications of the LP estimator. By and large, the results confirm the general picture: When skill groups are considered, it is especially higher-skilled Non-SPs whose hiring is positively related to productivity (*t-2* inflows). In the case of *t-1* inflows, the picture is somewhat mixed, but it is still exclusively Non-SP skill groups for whom we find positive coefficients, and considering that *t-1* inflows could be a rather "noisy" group in terms of their tenure in the hiring establishment (which could, technically, be as short as two days), *t-2* inflows should provide a clearer picture. This finding indeed suggests that higher-skilled workers are better knowledge carriers. Concerning industrial origin (Table 13), a robustly significant finding is that Non-SPs from different industries are positively related to productivity. Since industry switchers are the majority among either group of inflows, this was to be expected empirically. Thus, we do not find evidence of industrial proximity driving knowledge spillovers.

Table 13

LP revenue version	**Inflows t-1**		**Inflows t-2**	
Log intermediate inputs	0.654	***	0.663	***
Log capital stock	0.012		0.003	
Log labour	0.341	***	0.340	***
Share*mean gap SPs, same ind.	-0.521		0.760	
Share*mean gap SPs, diff. ind.	0.358		-0.442	
Share*mean gap Non-SPs, same ind.	1.885		0.005	
Share*mean gap Non-SPs, diff. ind.	1.240	***	1.461	***
Observations	11797		11797	
LP VA version	**Inflows t-1**		**Inflows t-2**	
Log capital stock	0.016	***	0.016	***
Log labour	0.714	***	0.713	***
Share*mean gap SPs, same ind.	-0.112		1.704	
Share*mean gap SPs, diff. ind.	-0.256		-0.899	
Share*mean gap Non-SPs, same ind.	3.569	**	0.198	
Share*mean gap Non-SPs, diff. ind.	2.272	*	3.349	***
Observations	11797		11797	

Dependent variable is log revenues (revenue version) respectively log value added (value added version. Standard errors obtained by bootstrapping with 50 replications. Year, 2-digit industry and labor market region (LMR) dummies included. ESTAB, EMPL and HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

7.2.3 Robustness checks

In the Appendix, we test the robustness of our results by splitting the sample. First, we split establishments according to their size in terms of total employment (heads); as in Stoyanov and Zubanov (2012), the dividing line is 50 employees (Table A 2). While this splitting renders most share*gap coefficients insignificant (probably due to the limited sample size), we find that the positive coefficient of Non-SP hiring is higher for larger establishments. Again, while reversed, this finding is in line with Stoyanov and Zubanov (2012) in the sense that they find greater productivity coefficients for SP hires in larger firms, which could be due to better resource endowments which allow those firms to allocate (and train) new hires more effectively. Another robustness check addresses the regionally biased sample design of the IAB Establishment Panel and performs the estimation separately for West and East German establishments (Table A 3). We find that the overall results tend to be driven by the (undersampled) West German establishments, while the opposite negative/positive effects for SPs and Non-SPs are generally confirmed.

Another robustness check we consider is whether the global financial and economic crisis of 2008/09, which strongly affect sales and hiring behavior particularly in the German manufacturing sector, somehow biases our results. Therefore, we consider two splits of the sample period: First, we distinguish between the pre-crisis years 2002-2007 and the during/post-crisis years 2008-2010; second, we single out the core crisis period (2008/09) and perform the estimation separately for the year 2002-2007 and 2010, and 2008/09, respectively (Table A 4). Again, sample size limitations inhibit the identification of significant coefficients, yet again, the only significant coefficients we find conform to the overall pattern: Non-SPs (*t-1*; in the 2002-2007 and 2010 sample) have a positive, SPs (*t-2*; 2002-2007 sample) a negative productivity coefficient.

One could still be worried that our main result, the positive productivity coefficient of Non-SP hires, is due to the fact that Non-SPs are movers into better-paid jobs, regarding Non-SPs' far greater wage increases connected to their job moves. So far, we have controlled for this simply by including the share of SPs and Non-SPs who increase their wage upon the job move. This is a coarse but conservative measure, considering that our wage data are top-coded and therefore it is much easier for us to identify wage increases in a binary manner than to compute the exact wage change. Still, in another robustness check, we include the mean absolute change in inflows' (uncorrected) wages (Table A 5). Another robustness check in the same table uses the individual wage effect of SPs and Non-SPs in their sending establishments ($\epsilon'_{p,j,t-s-1}$; i.e. their relative wage position) as a control for their individual ability, in this case referring to the average productivity level of similarly skilled co-workers in their sending establishment. The results indicate that omitting the amount of wage increases does bias the share*gap coefficient upward, since it is now estimated lower than in our basic specification. However, the sign and significance pattern for SPs and Non-SPs does not change, supporting our baseline finding that SPs tend to decrease, and Non-SPs tend to increase, hiring establishments' productivity.

7.3 Summary of results

To sum up, we find that hiring SPs does not seem to be related to establishment-level productivity gains. In fact, in the specifications where we identify a significant effect for them, it tends to be negative rather than positive. In contrast, and contrary to the findings of previous studies, we find that hiring Non-SPs is associated with significant productivity gains. Our specification accounts for the relative number of Non-SPs in terms of total establishment size and, in particular, their (inverse) quality in terms of their mean productivity gap, meaning that a Non-SP hire is more highly weighted, the larger his or her productivity gap (i.e. the lower his or her sending establishment's productivity relative to the hiring establishment). This finding is surprising given our hypothesis derived from previous empirical studies, but less surprising considering our descriptive finding that Non-SPs are positively (and SPs negatively) selected from their sending establishments.

Our estimates suggest that the share*gap coefficient ranges around 1.1.[22] Considering that the mean establishment in the sample hires some 2.8 percent of its workforce, that some 60 percent of new hires are Non-SPs, and that the mean productivity gap of Non-SPs is (-)0.245, the mean share*gap variable for Non-SPs is about (-)0.004. Holding constant their mean productivity gap, this would mean that hiring an average number of Non-SPs (as compared to hiring none) is associated with a productivity (value added) increase of the factor 1.1*0.004 = 0.0044, or 0.44 percent. This number is close to Stoyanov and Zubanov's (2012) estimate for the mean hiring of (productivity-gap-weighted) SPs (0.35 percent). Since our means at ruling out reverse causality bias are certainly imperfect, since we cannot rule out biases from unobserved hiring preferences (establishments might prefer Non-SPs for unclear reasons), and since we get much lower and insignificant estimates using internal IVs in the DPD estimators, we may regard this result as an upper bound of the true causal effect.

[22] LP revenue version, t-2 inflows.

8. Preliminary conclusions

In this paper, we attempt to identify knowledge spillovers from worker inflows into German manufacturing establishments, applying an approach similar to Stoyanov and Zubanov (2012, 2014) and Serafinelli (2013). This approach uses the productivity and wage levels of newly hired workers' previous employers to identify the productivity effects of hiring. We separate inflows into "spillover potentials" (SPs) and Non-SPs, i.e. inflows from more and less productive establishments, using establishments' fixed wage effect (obtained from a regression using the entire population of regular full-time employees in Germany) as proxy of productivity.

Unlike our predecessors, we do not find any positive productivity effect of hiring SPs. In contrast, we find robustly significant positive effects of hiring Non-SPs, despite controlling for individual skills and the wage increases that SPs and Non-SPs experience when changing jobs, as well as their relative wage position in the sending establishment. A potential explanation for our main finding lies in the fact that Non-SPs are positively, and SPs negatively selected from their sending establishments, as indicated by their individual wage effect (relative wage position).

The research question asked in the title of this paper, "do knowledge spillovers through worker inflows increase German establishments' productivity," could thus be answered as follows: Yes, but these productivity effects are due to workers ascending, not descending, in the establishment productivity distribution. In this important respect – the diffusion of knowledge and skills, a key element of productivity growth –, Germany seems to differ systematically from other (smaller) economies. Whether we are willing to refer to the found productivity effects as knowledge spillovers, i.e. external effects, certainly needs a more profound discussion. Since our results do not indicate that hiring establishments benefit from inflows' previous employers but from the workers' individual capacities, a more sound interpretation might be that highly productive workers are better matched in highly productive firms, and are only able to unfold their full productivity potential there. Our findings might also reflect sorting of highly productive workers into (already) productive establishments, where they might increase productivity even further – potentially increasing the dispersion of establishment productivity at the upper end of the distribution. Such a sorting pattern could be expected on the grounds of Card et al.'s (2013) finding that high-earning workers increasingly sort into high-paying establishments in Germany.

The main, albeit preliminary conclusion we may draw from previous studies and our analysis, is that hiring workers from more productive firms can increase hiring firms' productivity ("learning by hiring"), but this effect does not seem to prevail in the German manufacturing sector, by far the largest labor market investigated to date. We suggest that hiring the top-performers from less productive firms could be equally – or even more – effective, depending on the labor market investigated. Further research should thus explore the explanatory power of assortative matching as an alternative explanation for observed productivity effects, as well as the structural and institutional differences between labor markets that might determine the mobility pattern and the productivity effects of (increasingly) heterogeneous workers moving between (increasingly) heterogeneous firms.

References

Abel, Jaison R.; Ishita Dey and Todd M. Gabe. 2012. "Productivity and the Density of Human Capital." *Journal of Regional Science*, 52(4), 562-86.

Abowd, John M.; Creecy. Robert H. and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer- Employee Data." *Unpublished mimeo, Cornell University. Accessed at http://ideas.repec.org/p/cen/tpaper/2002-06.html.*

Abowd, J., F. Kramarz, D. Margolis. 1999. "High-Wage Workers and High-Wage Firms." *Econometrica*, 67(2), 251-333.

Ackerberg, Daniel; Kevin Caves and Garth Frazer. 2006. "Structural Identification of Production Functions." *MPRA Paper*, 38349.

Almeida, Paul and Bruce Kogut. 1999. "Localization of Knowledge and the Mobility of Engineers in Regional Networks." *Management Science*, 45(7), 905-17.

Arellano, Manuel and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies*, 58(2), 277-97.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention," H. M. Groves, *The Rate and Direction of Inventive Activity: Economic and Social Factors*. UMI, 609 - 26.

Balsvik, Ragnhild. 2011. "Is Labor Mobility a Channel for Spillovers from Multinationals? Evidence from Norwegian Manufacturing." *Review of Economics and Statistics*, 93(1), 285-97.

Blossfeld, Hans-Peter. 1987. "Labor-Market Entry and the Sexual Segregation of Careers in the Federal Republic of Germany." *American Journal of Sociology*, 93(1), 89-118.

Blossfeld, H.P. 1983. "Höherqualifizierung Und Verdrängung - Konsequenzen Der Bildungsexpansion in Den Siebziger Jahren." *VASMA-Projekt, Arbeitspapier*, 28.

Blundell, Richard and Stephen Bond. 2000. "Gmm Estimation with Persistent Panel Data: An

Application to Production Functions." *Econometric Reviews*, 19(3), 321-40.

____. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*, 87(1), 115-43.

Breschi, Stefano and Francesco Lissoni. 2001. "Knowledge Spillovers and Local Innovation Systems: A Critical Survey." *Industrial and Corporate Change*, 10(4), 975-1005.

____. 2009. "Mobility of Skilled Workers and Co-Invention Networks: An Anatomy of Localized Knowledge Flows." *Journal of Economic Geography*, 9(4), 439-68.

Card, David; Heining, Jörg; Kline, Patrick. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics*, 128(3), 967-1015.

Dasgupta, Kunal. 2012. "Learning and Knowledge Diffusion in a Global Economy." *Journal of International Economics*, 87(2), 323-36.

Eberhardt, Markus and Christian Helmers. 2010. "Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators." *Oxford University, Department of Economics Discussion Paper Series*, 513.

Eberle, Johanna; Peter Jacobebbinghaus; Johannes Ludsteck and Julia Witter. 2011. " Generation of Time-Consistent Industry Codes in the Face of Classification Changes * Simple Heuristic Based on the Establishment History Panel (Bhp)." *FDZ-Methodenreport*, 05/2011 (en).

Ehrl, Philipp. 2013. "Agglomeration Economies with Consistent Productivity Estimates." *Regional Science and Urban Economics*, 43(5), 751-63.

Fischer, Gabriele; Florian Janik; Dana Müller and Alexandra Schmucker. 2009. "The Iab Establishment Panel – Things Users Should Know." *Schmollers Jahrbuch*, 129(1), 133-48.

Gartner, Hermann. 2005. "The Imputation of Wages above the Contribution Limit with the German Iab Employment Sample." *FDZ Methodenreport*, 2.

Görg, Holger and Eric Strobl. 2005. "Spillovers from Foreign Firms through Worker Mobility: An Empirical Investigation*." *Scandinavian Journal of Economics*, 107(4), 693-709.

Griliches, Zvi. 1998. "Issues in Assessing the Contribution of Research and Development to Productivity Growth," Z. Griliches, *R&D and Productivity: The Econometric Evidence.* University of Chicago Press, 17-45.

____. 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." *The Bell Journal of Economics*, 10(1), 92-116.

Heining, Jörg; Theresa Scholz and Stefan Seth. 2013. "Linked-Employer-Employee Data from the Iab: Liab Cross-Sectional Model 2 1993-2010 (Liab Qm2 9310)." *FDZ Datenreport*, 2.

Helpman, Elhanan; Melitz, Marc J.; Yeaple, Stephen R. 2004. "Export Versus Fdi with Heterogeneous Firms." *American Economic Review*, 94(1), 300–16.

Hethey, Tanja and Johannes F. Schmieder. 2010. "Using Worker Flows in the Analysis of Establishment Turnover – Evidence from German Administrative Data." *FDZ Methodenreport*, 6.

Jaffe, Adam B.; Manuel Trajtenberg and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics*, 108(3), 577-98.

Kaiser, Ulrich; Hans Christian Kongsted and Thomas Rønde. 2008. "Labor Mobility and Patenting Activity." *University of Copenhagen Centre for Applied Microeconometrics Working paper*, 7.

Kosfeld, Reinhold and Alexander Werner. 2012. "Deutsche Arbeitsmarktregionen – Neuabgrenzung Nach Den Kreisgebietsreformen 2007–2011." *Raumforschung und Raumordnung*, 70(1), 49-64.

Levinsohn, James and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies*, 70(2), 317-41.

Lucas, Robert E. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1), 3-42.

Maliranta, Mika; Pierre Mohnen and Petri Rouvinen. 2009. "Is Inter-Firm Labor Mobility a Channel of Knowledge Spillovers? Evidence from a Linked Employer–Employee Panel." *Industrial and Corporate Change*, 18(6), 1161-91.

Møen, Jarle. 2005. "Is Mobility of Technical Personnel a Source of R&D Spillovers?" *Journal of Labor Economics*, 23(1), 81-114.

Mueller, Steffen. 2008. "Capital Stock Approximation Using Firm Level Panel Data, a Modified Perpetual Inventory Approach." *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 228(4), 357-71.

Olley, G. Steven and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6), 1263-97.

Petrin, Amil; Poi, Brian P.; Levinsohn, James. 2004. "Production Function Estimation in Stata Using Inputs to Control for Unobservables." *Stata Journal*, 4, 113-23.

Poole, Jennifer P. 2013. "Knowledge Transfers from Multinational to Domestic Firms: Evidence from Worker Mobility." *Review of Economics and Statistics*, 95(2), 393-406.

Power, Dominic and Mats Lundmark. 2004. "Working through Knowledge Pools: Labour Market Dynamics, the Transference of Knowledge and Ideas, and Industrial Clusters." *Urban Studies*, 41(5-6), 1025-44.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5), S71-S102.

____. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, 94(5), 1002-37.

Rosenthal, Stuart S. and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economices," J. V. T. Henderson, J.F., *Handbook of Regional and Urban Economics.* Elsevier, 2119-71.

Schäffler, Johannes. 2014. "Reloc Linkage: A New Method for Linking Firm-Level Data with the Establishment-Level Data of the Iab." *FDZ-Methodenreport*, 05, 26 p.

Serafinelli, Michel. 2013. "Good Firms, Worker Flows and Productivity " *MPRA Paper*, 47508.

Song, Jaeyong; Paul Almeida and Geraldine Wu. 2003. "Learning-by-Hiring: When Is Mobility More Likely to Facilitate Interfirm Knowledge Transfer?" *Management Science*, 49(4), 351-65.

Stoyanov, Andrey and Nikolay Zubanov. 2014. "The Distribution of the Gains from Spillovers through Worker Mobility between Workers and Firms." *European Economic Review*, 70, 17-35.

____. 2013. "Money on the Table? Firms' and Workers' Gains from Productivity Spillovers through Worker Mobility." *IZA Discussion Paper*, 7702.

____. 2012. "Productivity Spillovers across Firms through Worker Mobility." *American Economic Journal: Applied Economics*, 4(2), 168-98.

Van Beveren, Ilke. 2012. "Total Factor Productivity Estimation: A Practical Review." *Journal of Economic Surveys*, 26(1), 98-128.

____. 2007. "Total Factor Productivity Estimation: A Practical Review." *LICOS Discussion Papers*, 18207.

Vom Berge, Philipp; Marion König and Stefan Seth. 2013. "Sample of Integrated Labour Market Biographies (Siab) 1975-2010." *FDZ-Datenreport*, 1 (en).

Wooldridge, Jeffrey M. 2009. "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables." *Economics Letters*, 104(3), 112-14.

Yasar, Mahmut; Rafal Raciborski and Brian Poi. 2008. "Production Function Estimation in Stata Using the Olley and Pakes Method." *Stata Journal*, 8(2), 221.

Appendix

Table A 1

Skill group	Blossfeld (1983, 1987) occupation classes	Freq. t-1/t-2
High-skilled	Engineers, professions, managers	0.378 / 0.379
Mid-skilled	Skilled manual occupations, technicians, skilled services, semiprofessions, skilled commercial and administrative occupations	0.464 / 0.471
Low-skilled	Agricultural occupations, unskilled manual occupations, unskilled services, unskilled commercial and administrative occupations, occupations n.e.c.[23]	0.158 / 0.150

Skill groups based on the classification of Blossfeld (1983, 1987). Categorization uses the Classification of Occupations 1988 (KldB88) of the Federal Employment Agency.

Table A 2

OLS results, establ. size	**Inflows t-1**				**Inflows t-2**			
	Small establ. (<50 empl.)		**Large establ. (>=50 empl.)**		**Small establ. (<50 empl.)**		**Large establ. (>=50 empl.)**	
L.Log value added	0.493	***	0.562	***	0.495	***	0.563	***
L2.Log value added	0.219	***	0.216	***	0.218	***	0.213	***
Log capital stock	0.011	***	0.019	**	0.011	***	0.020	***
Log labour	0.307	***	0.202	***	0.303	***	0.211	***
Share * mean gap SPs	0.542		-0.122		-1.700	*	-3.430	
Share * mean gap Non-SPs	0.146		2.639		0.440		2.655	
Observations	3788		3510		3788		3510	
R-squared	0.856		0.924		0.855		0.924	

Dependent variable is log value added. Standard errors clustered at establishment level. Year, 2-digit industry and labor market region (LMR) dummies included. HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

Table A 3

OLS results, West/East	**Inflows t-1**				**Inflows t-2**			
	West		**East**		**West**		**East**	
L.Log value added	0.596	***	0.498	***	0.596	***	0.498	***
L2.Log value added	0.176	***	0.236	***	0.176	***	0.236	***
Log capital stock	0.014	**	0.011	***	0.014	**	0.011	***
Log labour	0.222	***	0.273	***	0.222	***	0.273	***
Share * mean gap SPs	0.384		-1.269		0.384		-1.269	
Share * mean gap Non-SPs	1.554		0.804		1.554		0.804	
Observations	3527		3771		3527		3771	
R-squared	0.963		0.936		0.963		0.936	

Dependent variable is log value added. Standard errors clustered at establishment level. Year, 2-digit industry and labor market region (LMR) dummies included. HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

[23] Occupations not elsewhere classified are exclusively outside the production, service, and administration sectors, and thus contain mainly household helpers and the like.

Table A 4

OLS results, Crisis years								
A: Inflows t-1	**2002-2007**		**2008-2010**		**w/o 2008/09**		**only 2008/09**	
L.Log value added	0.528	***	0.546	***	0.538	***	0.545	***
L2.Log value added	0.218	***	0.239	***	0.214	***	0.255	***
Log capital stock	0.010	***	0.014	**	0.010	***	0.015	**
Log labour	0.249	***	0.202	***	0.245	***	0.181	***
Share * mean gap SPs	0.203		-0.095		0.293		-0.636	
Share * mean gap Non-SPs	1.569		1.581		1.678	*	1.663	
Observations	4335		2963		5241		2057	
R-squared	0.957		0.956		0.956		0.958	
B: Inflows t-2	**2002-2007**		**2008-2010**		**w/o 2008/09**		**only 2008/09**	
L.Log value added	0.529	***	0.550	***	0.539	***	0.550	***
L2.Log value added	0.217	***	0.237	***	0.213	***	0.255	***
Log capital stock	0.010	***	0.014	**	0.010	***	0.014	**
Log labour	0.255	***	0.210	***	0.251	***	0.189	***
Share * mean gap SPs	-1.558	*	-0.195		-1.211		-0.508	
Share * mean gap Non-SPs	1.523		1.471		1.439		1.605	
Observations	4335		2963		5241		2057	
R-squared	0.956		0.956		0.956		0.957	

Dependent variable is log value added. Standard errors clustered at establishment level. Year, 2-digit industry and labor market region (LMR) dummies included. HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.

Table A 5

LP, control for mean wage increase/mean individual wage effect in sending establishment												
LP/Rev results, t-1	**Inflows t-1**						**Inflows t-2**					
Log labour	0.339	***	0.341	***	0.340	***	0.338	***	0.340	***	0.336	***
Share * mean gap SPs	0.253		0.344		0.410		-0.409		-0.339		-0.042	
Share * mean gap Non-SPs	1.285	***	1.219	***	2.013	***	1.119	**	0.896	*	1.571	***
Log capital stock	0.011		0.011		0.015		0.002		0.000		0.006	
Log intermediate inputs	0.654	***	0.653	***	0.652	***	0.664	***	0.665	***	0.665	***
Mean wage incr. SPs			0.000	**					0.000			
Mean wage incr. Non-SPs			0.000	**					0.001	***		
Mean indiv. effect SPs					-0.002						-0.001	
Mean indiv. effect Non-SPs					-0.012	**					-0.006	
Observations	11797		11315		10079		11797		11332		10213	
LP/VA results, t-1												
Log labour	0.711	***	0.713	***	0.713	***	0.710	***	0.712	***	0.710	***
Share * mean gap SPs	-0.297		-0.146		-0.429		-0.809		-0.762		0.010	
Share * mean gap Non-SPs	2.409	**	2.282	***	4.171	***	2.681	***	2.355	***	3.437	***
Log capital stock	0.016	***	0.016	***	0.015	***	0.016	***	0.017	***	0.020	***
Mean wage incr. SPs			0.001	**					0.000			
Mean wage incr. Non-SPs			0.002	***					0.002	***		
Mean indiv. effect SPs					-0.006						-0.002	
Mean indiv. effect Non-SPs					-0.021						-0.007	
Observations	11797		11315		10079		11797		11332		10213	

Dependent variable is log value added. Standard errors obtained by bootstrap with 50 replications. Year, 2-digit industry and labor market region (LMR) dummies included. ESTAB; EMPL and modified HIRE control variables included. *p<0.1, **p<0.05, ***p<0.01.