**Appendix to**

**"Productivity spillovers through labor flows: The effect of the productivity gap, foreign-owned firms, and skill-relatedness"**

This file summarizes supplemental information for the article "Productivity spillovers through labor flows: the effect of the productivity gap, foreign-owned firms, and skill-relatedness" that has been submitted to the Review of Economics and Statistics.

## I. Detailed description of data management

We have access to the Hungarian administrative data integration database, which is an anonymized employer-employee linked panel dataset created by the matching of five administrative data sources, for the years 2003-2011, developed by the databank of HAS CERS. The database contains a 50% random sample of the population aged 15-74 living in Hungary in 2003 and the involved employees are traced over the period. The most important demographic features of employees (gender, age, place of residence in the year of entry), and information about their employment spells (months worked, ISCO-88 occupation code, monthly wage) as well as company characteristics (4-digit industry code according to the NACE'08 classification, number of employees, and specific rows of their balance sheets and financial statements including tangible assets, equity owned by private domestic, private foreign, and state owners, sales, pre-tax profits, material-type costs, personnel expenditures, wage bill) are known. All monetary variables are deflated by yearly industry-level producer price indices to calculate their real 2011 value.

The data is managed by the Databank of the Institute of Economics of the Hungarian Academy of Sciences and can be accessed for scientific research upon individual request. For more details consult http://adatbank.krtk.mta.hu/adatbazisok___allamigazgatasi_adatok. The necessary codes to go from the raw data to the results of the paper can be accessed at http://econ.core.hu/english/res/netecon_pub.html.

The raw data contains employee-employer links on a monthly basis. We defined the main employer for every worker and for every year as the workplace where the worker spent the highest number of months in the given year, and created yearly matrices of intercompany movements between these main employers. In particular, if an employee switches firm in the second half of year *t* or first half of year *t*+1, the receiving firm will be her employer in year *t*+1 and the sending firm will be her employer in year *t*.

However, our models assess the effect of labor mobility on firms' productivity on a yearly basis, which can lead to an endogenous connection between labor flows and productivity change (not discussed in the main text). The problem is illustrated in Figure I; productivity shocks (e.g. purchasing a machine) happening in the first half of year *t*+1 can affect the number of new hires in the first half of year *t*+1.

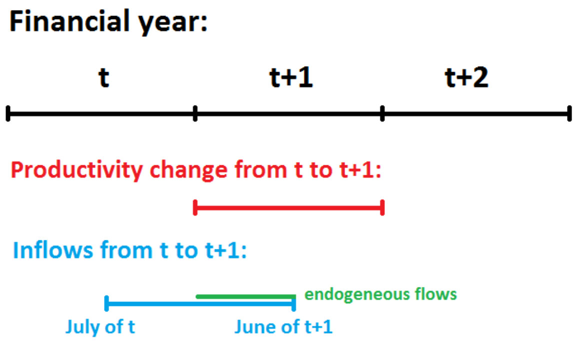


**Figure I. Periods of productivity change and labor mobility**

The potential of reversed causality shortly summarized above might distort our analysis. In order to exclude the possibility of such endogeneity, we conduct the analysis only for those new hires that were observed in year *t* or in January in year *t*+1 at the latest, and exclude all the cases of labor mobility that happened between February and June.

A certain time period has to pass for the new employee to exert a significant effect on firm productivity. With new employees working for a short period and not controlling for months worked at the receiving firm, we would underestimate the effect of new hires on yearly productivity growth. Therefore, in the productivity spillover analysis, only those workers were considered as new hires that stayed for at least 6 months with their new employer.

## II. Calculation of human capital

As described in the main text, the human capital of each worker is calculated for each year spent in the private sector. The gaps in private sector employment of at most 3 years are filled by linear interpolation. In case of gaps of at least 4 years, or when the worker only worked in the public sector before getting a job in the private sector, human capital is calculated by a wage regression on the subsample of public sector workers. In addition to the multi-dimensional fixed-effects approach, as a robustness check, we also estimated a pooled OLS regression with age, age-squared, gender and skill-levels of workers. Results are presented in Table I.

**Table I: Wage equations without and with employee fixed effects, separately on private and public sector employees**

| Method | Pooled OLS | | Employee FE | |
|---|---|---|---|---|
| Sample of employees | Private sector | Public sector | Private sector | Public sector |
| Age | 0.060*** | 0.039*** | 0.089 | 0.079 |
| | (0.001) | (0.003) | (416.32) | (105.41) |
| Age-squared | -0.001*** | -0.000*** | -0.001*** | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Male | 0.182*** | 0.093*** | - | - |
| | (0.004) | (0.009) | | |
| Low-skilled | 0.155*** | 0.139*** | - | - |
| | (0.007) | (0.033) | | |
| Mid-skilled | 0.009 | -0.008 | - | - |
| | (0.007) | (0.007) | | |
| Managers | 0.913*** | 1.128*** | 0.361*** | 0.624*** |
| | (0.012) | (0.035) | (0.01) | (0.021) |
| Professionals | 0.792*** | 0.790*** | 0.357*** | 0.524*** |
| | (0.019) | (0.032) | (0.01) | (0.016) |
| Technicians and assistants | 0.586*** | 0.536*** | 0.292*** | 0.349*** |
| | (0.015) | (0.014) | (0.011) | (0.013) |
| Office administrators | 0.475*** | 0.350*** | 0.241*** | 0.266*** |
| | (0.022) | (0.015) | (0.014) | (0.012) |
| Commercial workers | 0.387*** | 0.298*** | 0.241*** | 0.281*** |
| | (0.022) | (0.011) | (0.012) | (0.019) |
| Agriculture and forestry | 0.239*** | 0.121*** | 0.147*** | 0.130*** |
| | (0.018) | (0.012) | (0.007) | (0.009) |
| Blue-collars in industry and construction | 0.353*** | 0.267*** | 0.224*** | 0.226*** |
| | (0.014) | (0.009) | (0.01) | (0.008) |
| Assemblers and machine operators | 0.288*** | 0.279*** | 0.185*** | 0.213*** |
| | (0.023) | (0.026) | (0.01) | (0.010) |
| Army | 0.432*** | 0.844*** | 0.208*** | 0.629*** |
| | (0.080) | (0.028) | (0.031) | (0.067) |
| Firm FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Observations | 11,192,798 | 6,260,904 | 10,864,118 | 5,723,524 |
| R-squared | 0.687 | 0.759 | 0.843 | 0.849 |

Notes: Robust standard errors in parentheses. High-skilled: worked at least once in an occupation requiring tertiary education; Mid-skilled: worked at least once in an occupation requiring secondary education; Low-skilled: everybody else. The baseline occupation category is “Elementary occupations”. The baseline skill category is “High-skilled”. Employees present only in one year of the analysis do not have an individual FE, therefore they are excluded from Columns C and D. *** p<0.001, ** p<0.01, * p<0.5

Figure II shows the distribution of human capital calculated without and with employee fixed effects. Version 1 explains 69% of the variation of the log value of wage, whereas version 2 has an R-squared of 84%. The correlation between the two versions of human capital is 0.74. Since fixed effects can control for more individual-specific characteristics, version 2 is a better approximation of the worker's true human capital. Its closer-to-normal distribution also makes it more desirable for further analysis, therefore we continue with this measure.

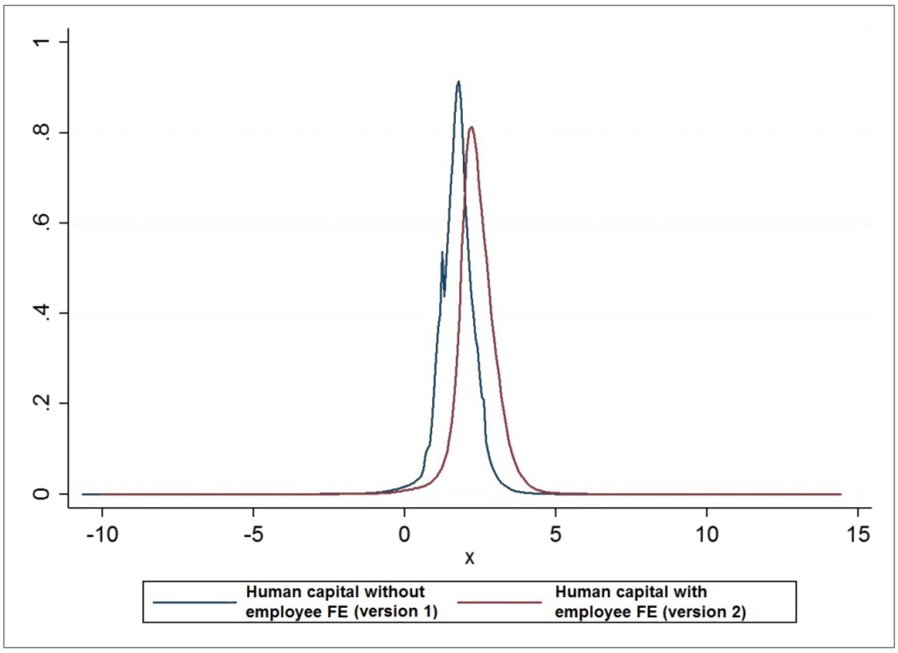


**Figure II. Density plots of human capital without employee FE (version 1) and with employee FE (version 2)**

Figures III and IV show the distributions of human capital with employee fixed effects by gender and skill level. Looking at the curves, we can infer that there is no significant difference between the value of the work-related abilities of men and women, although the variation is higher in the case of women. There is a clear difference between the distributions of human capital by skill level, particularly to the advantage of highly skilled workers. These descriptive findings confirm our decision to use human capital calculated with worker fixed effects.

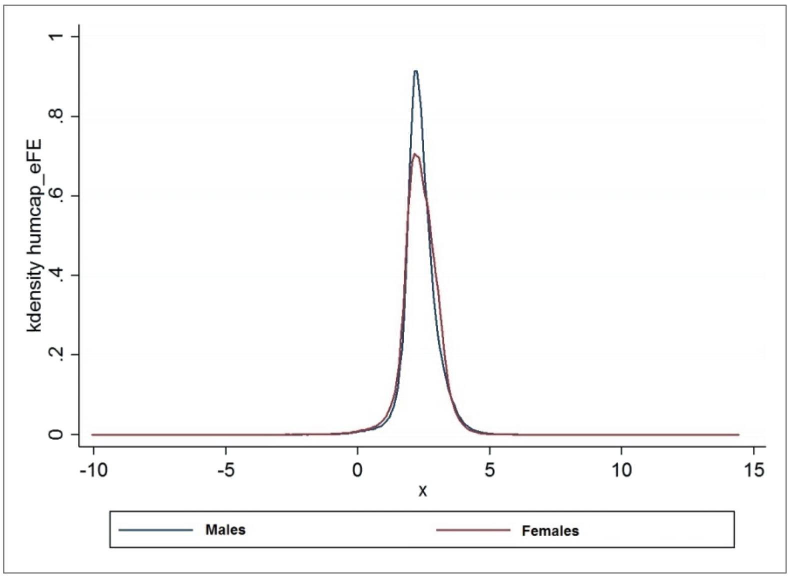


**Figure III. Distribution of human capital with employee FE by gender**

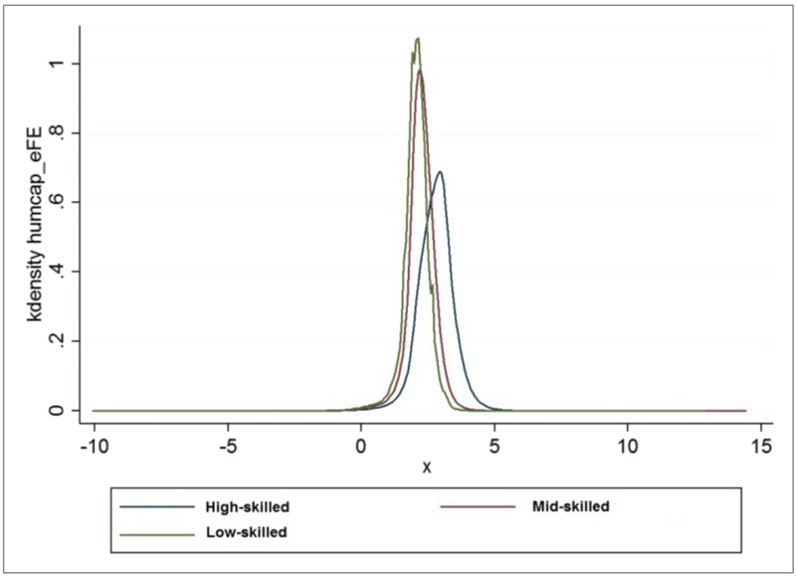


**Figure IV. Distribution of human capital with employee FE by skill levels**

High-skilled: worked at least once in an occupation requiring tertiary education; Mid-skilled: worked at least once in an occupation requiring secondary education; Low-skilled: everybody else.

### III. Visualization of the skill-relatedness network

In Figure V, the skill-relatedness network of four-digit industries is plotted using a spring algorithm, which brings related industries close to each other. It is visible from the network that there is a correlation between the official NACE classification and skill-relatedness, because industries in similar sectors tend to group together. However, one can observe a much more complex structure of industry relations of technological proximity than one can deduce from industry classification (Neffke, Henning, & Boschma, 2011).

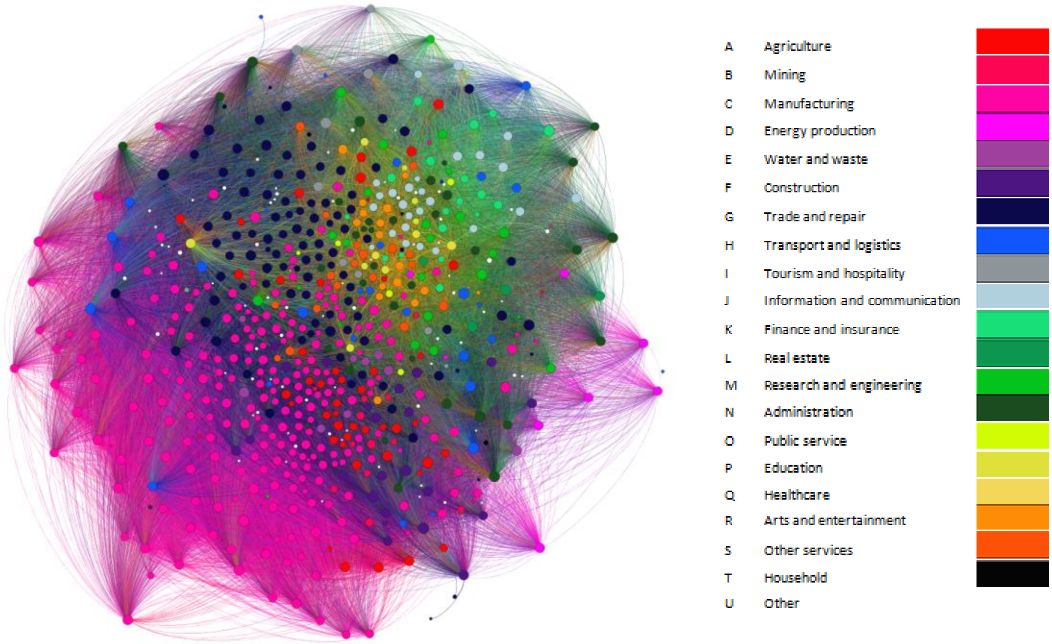


**Figure V: Skill-relatedness network in Hungary, 2003-2012**

Notes: Nodes are industries defined by 4-digit NACE codes and color-codes refer to sectors of 1-digit NACE codes. We included edges with weights $\overline{R_{ij}}$ >0. The natural logarithm of employment is used to depict the size of the industry, which is reflected by the size of the nodes. The position of the nodes is determined by the Force Atlas 2 algorithm in Gephi.

### IV. Robustness check with alternative skill-relatedness matrices

In order to demonstrate the robustness of the skill-relatedness effect on productivity spillovers, we present the results of two additional estimations. In the first estimation, we only look at the movements of managers, high-skilled employees, and middle-skilled high-wage employees when analyzing productivity spillovers, and also construct the skill-relatedness matrix from the above flows.

Results in Table II imply that flows from related industries and from the same industry outperform flows from unrelated industries (Column A), even after controlling for the productivity gap and its interactions (Column B). However, only the productivity gap and its interactions with the share of related flows remain significant when company ownership variables are introduced (Column C).

An additional robustness check utilizes the skill-relatedness matrix calculated from Swedish labor flow data. This last check is very important to demonstrate that our main finding still holds when the relatedness of industries is identified from exogenous data sources. The Swedish skill-relatedness matrices were calculated for the years 2003-2010 similarly to the Hungarian ones. For this period, there were 32,301 industry pairs (out of 258,840 possible combinations) where both the Hungarian and the Swedish data indicated mobility. The correlation coefficient of the two skill-relatedness measures was 0.35 for these cases.

Results reported in Table III suggest that skill-related movements to the company as well as the interaction of the productivity gap and skill-relatedness increase productivity.

**Table II. Skill-relatedness and productivity spillovers; only high-skilled sample of movers**

| | Column A | Column B | Column C |
|---|---|---|---|
| Lagged productivity | 0.614*** | 0.646*** | 0.645*** |
| | (0.006) | (0.006) | (0.006) |
| Human capital | 0.146*** | 0.133*** | 0.133*** |
| | (0.014) | (0.014) | (0.014) |
| Lagged human capital | -0.002 | -0.003 | -0.004 |
| | (0.012) | (0.012) | (0.012) |
| Share of SR2 inflows | 0.005 | -0.005 | -0.029 |
| | (0.030) | (0.030) | (0.031) |
| Share of SR3 inflows | 0.078* | 0.062* | 0.037 |
| | (0.031) | (0.030) | (0.031) |
| Share of SR4 inflows | 0.085** | 0.077** | 0.055 |
| | (0.029) | (0.028) | (0.029) |
| Share of same industry inflows | 0.111*** | 0.093*** | 0.073* |
| | (0.029) | (0.028) | (0.029) |
| Productivity gap | | 0.063*** | 0.059*** |
| | | (0.012) | (0.011) |
| PG of SR2 inflows | | 0.032 | 0.035 |
| | | (0.025) | (0.025) |
| PG of SR3 inflows | | 0.076* | 0.079** |
| | | (0.030) | (0.029) |
| PG of SR4 inflows | | 0.079** | 0.083** |
| | | (0.028) | (0.028) |
| PG of same industry inflows | | 0.211*** | 0.214*** |
| | | (0.031) | (0.031) |
| From private domestic | | | 0.035* |
| | | | (0.016) |
| From private foreign | | | 0.043* |
| | | | (0.020) |
| Observations | 54,791 | 54,791 | 54,791 |
| R-squared | 0.581 | 0.585 | 0.585 |

Notes: Industry-region-year FE models. Firm ID-clustered robust standard errors in parentheses. SR1 [-1;-0.5] is used as baseline skill-relatedness; further categories are SR2: [-0.5; 0], SR3: [0; 0.5]; SR4: [0.5;1]. Additional controls are the characteristics of the receiving firm (total assets, ownership, size), general inflow-outflow measures (share of outflows, fluctuation, share of workers without a job in the previous year).*** p<0.001, ** p<0.01, * p<0.05

**Table III. Skill-relatedness and productivity spillovers; Swedish skill-relatedness matrix**

| | Column A | Column B | Column C |
|---|---|---|---|
| Lagged productivity | 0.666*** | 0.678*** | 0.678*** |
| | (0.008) | (0.008) | (0.008) |
| Human capital | 0.158*** | 0.149*** | 0.148*** |
| | (0.019) | (0.019) | (0.018) |
| Lagged human capital | 0.003 | 0.006 | 0.007 |
| | (0.017) | (0.017) | (0.017) |
| Share of SR2 inflows | 0.013 | 0.008 | 0.010 |
| | (0.010) | (0.009) | (0.012) |
| Share of SR3 inflows | 0.003 | 0.015 | 0.017 |
| | (0.018) | (0.019) | (0.021) |
| Share of SR4 inflows | 0.042* | 0.045* | 0.047* |
| | (0.019) | (0.019) | (0.021) |
| Share of same industry inflows | 0.053 | 0.104 | 0.106 |
| | (0.079) | (0.073) | (0.073) |
| Productivity gap | | 0.017 | 0.017 |
| | | (0.009) | (0.009) |
| PG of SR2 inflows | | 0.001 | 0.001 |
| | | (0.011) | (0.011) |
| PG of SR3 inflows | | 0.034* | 0.034* |
| | | (0.016) | (0.015) |
| PG of SR4 inflows | | 0.017 | 0.018 |
| | | (0.020) | (0.020) |
| PG of same industry inflows | | 0.302*** | 0.302*** |
| | | (0.076) | (0.076) |
| From private domestic | | | -0.005 |
| | | | (0.011) |
| From private foreign | | | 0.005 |
| | | | (0.014) |
| Observations | 31,549 | 31,549 | 31,549 |
| R-squared | 0.631 | 0.632 | 0.632 |

Notes: Industry-region-year FE models. Firm ID-clustered robust standard errors in parentheses. SR1 [-1;-0.5] is used as baseline skill-relatedness; further categories are SR2: [-0.5; 0], SR3: [0; 0.5]; SR4: [0.5;1]. Additional controls are the characteristics of the receiving firm (total assets, ownership, size), general inflow-outflow measures (share of outflows, fluctuation, share of workers without a job in the previous year). *** p<0.001, ** p<0.01, * p<0.05