**Online Appendix to Kampelmann and Rycx, "The impact of educational mismatch on firm productivity: evidence from linked panel data",** *Economics of Education Review*

1) <u>**Specification controlling for the birth cohort of workers**</u> **(used in section 5.2)**

In order to estimate the impact of educational mismatch on firm productivity when controlling for the birth cohort of workers, we reformulate equation (1) as follows:

$$\ln VA\_work_{jt} = \beta_0 + \beta_1 \ln VA\_work_{jt-1} + \beta_2\, REQ^*_{j,t} + \beta_3\, OVER^*_{j,t} + \beta_4\, UNDER^*_{j,t}$$
$$+ \mathbf{x_{j,t}}\, \boldsymbol{\beta_5} + \mathbf{z_{j,t}}\, \boldsymbol{\beta_6} + \gamma_t + u_{j,t} \qquad\qquad (2)$$

In equation (2):

a) $VA\_work_{j,t}$ is the productivity of firm $j$ at time $t$, measured by the average value added per worker.

b) $REQ^*_{j,t}$ is the average of the years of required education in firm $j$ at time $t$, controlling for cohort effects. It is equal to the employment weighted sum of the required years of education for all jobs occupied respectively by young (i.e. workers that are at most 35 years old) and older workers (i.e. those that are at least 36 years old) in firm $j$ at time $t$.

Mathematically: $REQ^*_{j,t} = \dfrac{1}{m_{young,j,t}}\sum\limits_{i=1}^{m_{j,t}} REQ\_young_{i,j,t} + \dfrac{1}{m_{older,j,t}}\sum\limits_{i=1}^{m_{j,t}} REQ\_older_{i,j,t}$

with:
- $m_{young,j,t}$ ($m_{older,j,t}$) the number of young (older) workers employed in firm $j$ at year $t$.
- $REQ\_young_{i,j,t}$ = [mode of education among young workers that are in the same occupation than worker $i$ (at the NACE 3-digit level in the entire economy) at time $t$] if worker $i$ is young, 0 otherwise.
- $REQ\_older_{i,j,t}$ = [mode of education among older workers that are in the same occupation than worker $i$ (at the NACE 3-digit level in the entire economy) at time $t$] if worker $i$ is older than 35 years, 0 otherwise.

c) $OVER^*_{j,t}$ is the average of the years of over-education in firm $j$ at time $t$, controlling for cohort effects. It corresponds to sum of the years of over-education among young and older workers in firm $j$ at time $t$, divided by the total number of workers employed in firm $j$ at time $t$.

Mathematically: $OVER^*_{j,t} = \dfrac{1}{m_{j,t}}\left(\sum\limits_{i=1}^{m_{j,t}} OVER\_young_{i,j,t} + \sum\limits_{i=1}^{m_{j,t}} OVER\_older_{i,j,t}\right)$

with:
- $m_{j,t}$ the total number of workers employed in firm $j$ at time $t$.
- $OVER\_young_{i,j,t}$ = $(Attained\_education_{i,j,t} - REQ\_young_{i,j,t})$ if $> 0$ & worker $i$ is young, 0 otherwise.
- $OVER\_older_{i,j,t}$ = $(Attained\_education_{i,j,t} - REQ\_older_{i,j,t})$ if $> 0$ & worker $i$ is older than 35 years, 0 otherwise.

- *Attained_education$_{i,j,t}$* the number of years corresponding to the highest level of education attained by worker $i$ employed in firm $j$ at time $t$.

d) *UNDER\*$_{j,t}$* is the average of the years of under-education in firm $j$ at time $t$, controlling for cohort effects. This variable is equal to the sum of the years of under-education among young and older workers in firm $j$ at time $t$, divided by the total number of workers employed in firm $j$ at time $t$.

Mathematically: $$UNDER*_{j,t} = \frac{1}{m_{j,t}}\left(\sum_{i=1}^{m_{j,t}} UNDER\_young_{i,j,t} + \sum_{i=1}^{m_{j,t}} UNDER\_older_{i,j,t}\right)$$

with:
- $m_{j,t}$ the total number of workers employed in firm $j$ at time $t$.
- *UNDER_young$_{i,j,t}$* = (*Attained_education$_{i,j,t}$* − *REQ_young$_{i,j,t}$*) if < 0 & worker $i$ is young, 0 otherwise.
- *UNDER_older$_{i,j,t}$* = (*Attained_education$_{i,j,t}$* − *REQ_older$_{i,j,t}$*) if < 0 & worker $i$ is older than 35 years, 0 otherwise.
- *Attained_education$_{i,j,t}$* the number of years corresponding to the highest level of education attained by worker $i$ employed in firm $j$ at time $t$.

e) $\mathbf{x_{j,t}}$, $\mathbf{z_{j,t}}$ and $\gamma_t$ are defined as in equation (1).

f) $u_{j,t}$ is the error term.

## 2) <u>Specification according to workers' age (used in section 5.3)</u>

To test the impact on firm productivity of mean years of over- and under-education respectively among young and older workers (while controlling for cohort effects), we reformulate equation (2) as follows:

$$\ln VA\_work_{j,t} = \beta_0 + \beta_1 \ln VA\_work_{j,t-1} + \beta_2 REQ*_{j,t} + \beta_3 YOUNG\text{-}OVER*_{j,t}$$
$$+ \beta_4 OLDER\text{-}OVER*_{j,t} + \beta_5 YOUNG\text{-}UNDER*_{j,t}$$
$$+ \beta_6 OLDER\text{-}UNDER*_{jt} + \mathbf{x_{jt}}\,\boldsymbol{\beta_7} + \mathbf{z_{jt}}\,\boldsymbol{\beta_8} + \gamma_t + p_{j,t} \qquad (3)$$

In equation (3):

a) *VA_work$_{j,t}$* is the productivity of firm $j$ at time $t$, measured by the average value added per worker.

b) *REQ\*$_{j,t}$* is the average of the years of required education in firm $j$ at time $t$, controlling for cohort effects (see equation (2)).

c) *YOUNG-OVER\*$_{j,t}$* is the average of the years of over-education among young workers in firm $j$ at time $t$, controlling for cohort effects. It corresponds to the sum of the years of over-education among young workers (i.e. workers that are at most 35 years old) in firm $j$ at time $t$, divided by the total number of workers employed in firm $j$ at time $t$.

2

Mathematically: $YOUNG-OVER*_{j,t} = \dfrac{1}{m_{j,t}} \sum\limits_{i=1}^{m_{j,t}} OVER\_young_{i,j,t}$

with:

- $m_{j,t}$ the total number of workers employed in firm $j$ at time $t$.
- $OVER\_young_{i,j,t} = (Attained\_education_{i,j,t} - REQ\_young_{i,j,t})$ if $> 0$ & worker $i$ is young, 0 otherwise.

d) *OLDER-OVER\*$_{j,t}$* is the average of the years of over-education among workers older than 36 years in firm $j$ at time $t$, controlling for cohort effects. It is equal to the sum of the years of over-education among older workers in firm $j$ at time $t$, divided by the total number of workers employed in firm $j$ at time $t$.

Mathematically: $OLDER-OVER*_{j,t} = \dfrac{1}{m_{j,t}} \sum\limits_{i=1}^{m_{j,t}} OVER\_older_{i,j,t}$

with:

- $m_{j,t}$ the total number of workers employed in firm $j$ at time $t$.
- $OVER\_older_{i,j,t} = (Attained\_education_{i,j,t} - REQ\_older_{i,j,t})$ if $> 0$ & worker $i$ is at least 36 years old, 0 otherwise.

e) *YOUNG-UNDER\*$_{j,t}$* is the average of the years of under-education among young workers in firm $j$ at time $t$, controlling for cohort effects. It is equal to the sum of the years of under-education among young workers in firm $j$ at time $t$, divided by the total number of workers employed in that firm $j$ at time $t$.

Mathematically: $YOUNG-UNDER*_{j,t} = \dfrac{1}{m_{j,t}} \sum\limits_{i=1}^{m_{j,t}} UNDER\_young_{i,j,t}$

with:

- $m_{j,t}$ the total number of workers employed in firm $j$ at time $t$.
- $UNDER\_young_{i,j,t} = (Attained\_education_{i,j,t} - REQ\_young_{i,j,t})$ if $< 0$ & worker $i$ is young, 0 otherwise.

f) *OLDER-UNDER\*$_{j,t}$* is the average of the years of under-education among older workers in firm $j$ at time $t$, controlling for cohort effects. To obtain this measure, we sum the years of under-education among older workers in firm $j$ at time $t$ and divide by the total number of workers employed in firm $j$ at time $t$.

Mathematically: $OLDER-UNDER*_{j,t} = \dfrac{1}{m_{j,t}} \sum\limits_{i=1}^{m_{j,t}} UNDER_{i\_older,j,t}$

with:

- $m_{j,t}$ the total number of workers employed in firm $j$ at time $t$.
- $UNDER\_older_{i,j,t} = (Attained\_education_{i,j,t} - REQ\_older_{i,j,t})$ if $< 0$ & worker $i$ is at least 36 years old, 0 otherwise.

g) $\mathbf{x_{j,t}}$, $\mathbf{z_{j,t}}$ and $\gamma_t$ are defined as in equation (1).

h) $p_{j,t}$ is the error term.