Author: Márton Nagy
Course: Data Analysis 4 – Causal Analysis (Term Project)

Money matters?
Unraveling the impact of EU funding on local incomes in Hungary[1]

1. Introduction

The core idea behind the European Union (EU) providing funding for peripherical Member States is to boost economic convergence in the bloc. This would be beneficial in the long run for every EU member, as all would profit from a more flourishing common market. Thus, the extent to which EU funds translate to local development has been a much-researched question in the past decades[2]. However, most of the prior research focused on NUTS 2 or 3 regions which may not be the most informative spatial unit of analysis if we want to understand truly local effects. Thus, I propose an analysis of municipalities rather than higher-level regions to understand more local relationships.

My research question is the following: did EU funding contribute to increased average real incomes in Hungarian municipalities in the period between 2007 and 2022? My hypothesis is that it did, as the mechanisms explained in Section 3 make it plausible to think so. I believe this question is relevant for current Hungarian economic policy in two ways. First, as Hungary currently cannot access most of its EU funds, answering this question may help in assessing what might be the economic impacts on the municipal level of the lack of EU funds. Second, if it turns out that EU funds had little or no local income effects this might indicate a need for rethinking the allocation mechanisms for more effective funding in the future.

2. Data

The raw data behind the analysis comes from three sources. First, I have downloaded data scraped from the [official EU tendering website](#) of the Hungarian government[3] on individual funding decisions[4]. Second, I have downloaded the outcome and possible confounder variables from the National Spatial Development and Planning Information System's (TEIR) website[5]. Lastly, I have downloaded yearly Hungarian inflation data from the Central Statistical Office's (KSH) website[6].

The data wrangling process involved several steps. First, I had to aggregate individual funding decisions by year and municipality (after dropping observations where no funding amount or municipality was indicated). Importantly, a low number of funding decisions related to multiple municipalities. I have allocated these funds to individual municipalities by population share and flagged these observations. I also had to resolve some municipality naming issues (e.g. due to misspelling or due to having municipality parts instead of actual municipalities in the data). Then I dropped municipalities which have undergone administrative changes (e.g. split or merged), as these were not present in all the years. I have also dropped the capital from the analysis – the reason for this is that most of the EU funding registered to Budapest is given to government bodies and government-owned companies, the programs of which have a broader scope than to just develop the city. This way, I ended up with 3148 municipalities (out of the total 3155) observed over 16 years. Next, I merged the EU funding data with tables downloaded from TEIR[7]. Even though I had EU funding data from 2004 until 2024, my outcome and confounder variables were only available in the 2007-2022 period, so I limited my sample to this timeframe. At this point, I only had missing values in my *high prestige employment* confounder. As these affected only 27 municipalities for a small number of years, I

---

[1] Raw data files and source code is available at my [public GitHub repository](#).

[2] For the most notable articles, see: [Becker et al. (2010)](#), [Becker et al. (2013)](#), [Becker et al. (2018)](#).

[3] Orsós, M. (2025). *EU Love* [GitHub repository]. https://github.com/misrori/eu_love

[4] Note that an important shortcoming of this dataset is that it only shows funding *allocated* to a certain municipality in a certain year – but these funds may be paid out over several years. This fact implies using lags in FD models.

[5] TEIR (n.d.). *Országos Területfejlesztési és Területrendezési Információs Rendszer* [database]. https://www.oeny.hu/oeny/teir/#/

[6] KSH (2025). *Harmonizált fogyasztóiár-index az egyéni fogyasztás rendeltetés szerinti osztályozása (COICOP) alapján [előző év = 100,0%]*. https://www.ksh.hu/stadat_files/ara/hu/ara0003.html

[7] A detailed description of the variables used can be found in Section 3.

imputed them using simple linear extrapolation and added a flag for the imputation. Lastly, I calculated per capita values for EU funding and income and adjusted them to 2024 prices. I have also transformed employment related confounders to percentages. The summary statistics of these variables are presented in Table 1.

**Table 1: Summary statistics of raw variables (T:N = 16:3148)**

	Mean	SD	Min.	10%	25%	50%	75%	90%	Max.
Income p.c.	1,426,904	616,366	72,204	751,602	986,417	1,316,849	1,756,787	2,271,297	7,934,420
EU funding p.c.	185,721	1,115,127	0	0	0	10,834	118,393	413,146	1,24e8
Unemployment	3.9	3.6	0.0	0.6	1.4	2.9	5.3	8.6	30.7
Industry empl.[8]	7.0	5.8	0.0	3.2	4.8	6.6	8.6	10.9	100.0
Agric. empl.	1.1	1.4	0.0	0.0	0.2	0.6	1.3	2.5	26.2
High prest. empl.	11.9	7.0	0.0	4.7	7.4	10.7	14.9	20.6	100.0
Pop. density	69.4	114.6	0.7	15.2	24.5	40.8	69.3	126.9	2005.0

There are two important observations we can make from Table 1. First, most of our variables have rather a left-skewed distribution[9], so taking logs of all the variables may be a good idea. Thus, I have log-transformed all my variables (by adjusting zeros to a slightly positive number and adding a flag for these). Second, a significant share (almost 40%) of our country-year observations has zero EU funding. So, it might make sense to estimate effects not only on a continuous treatment variable, but also on a binary one. The use of weights, differencing and other modelling choices will be discussed in the next section.

3. Identification strategy

To aid thinking about causal relationships, Figure 1 presents a schematic DAG diagram on various variables that may be involved in the examined relationship. The variables were identified based on prior domain knowledge[10]. Note that even this chart is quite simplified as it only represents relationships with the causal and outcome variables.

Figure 1: DAG of the possible relationships between certain variables
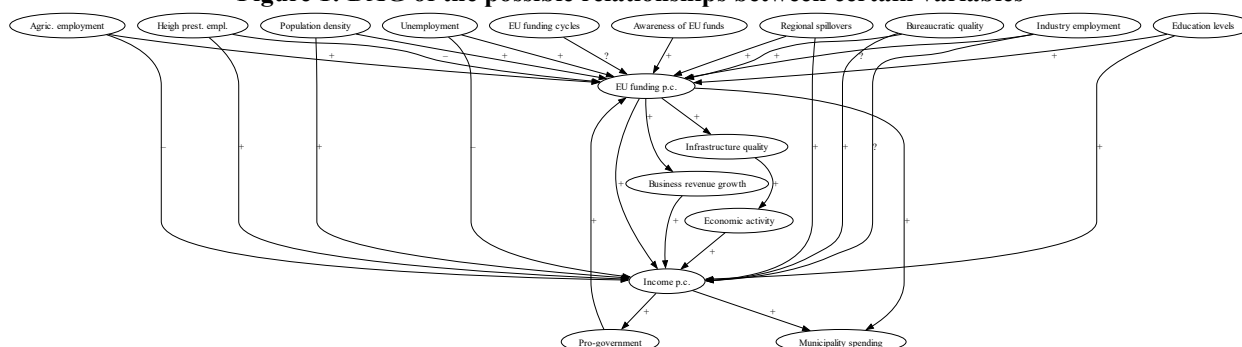
Note: symbols on edges represent hypothesized direction of relationship (+: positive; –: negative; ?: ambiguous)

First, let's examine the possible causal mechanisms. I believe there may be three major pathways through which EU funding might lead to increased incomes. First, local businesses revenue may grow because of EU funding (either because firms can make necessary developments to be more productive, or simply because they benefit from increased demand – e.g. because EU funds are used for construction, for which local firms are employed). This revenue growth can translate into wage growth as well. Second, EU funds may be used to enhance local infrastructure, which may attract more firms, thus increasing the demand for labor, driving wages up. Lastly, there may be a direct effect as well, given that some of the EU funds may be directly used to pay for certain welfare programs.

The DAG suggests that we should control for agricultural employment, industry employment, heigh prestige employment, unemployment, population density, education levels, regional spillovers, bureaucratic quality (common cause confounders), and pro-government

---

[8] As this variable was constructed by dividing the number of people working in industry with the work-age population, there ended up being some observations (6 in total) with over 100% values. I decided to cap these at 100%.
[9] Yearly histograms can be found in the Jupyter-notebook for both the raw and the log-transformed variables.
[10] As I have written my Bachelor's thesis on a related topic.

alignment (mechanism of reverse causality). However, we should not condition on business growth rates, infrastructure quality, economic activity (mechanism variables), municipality spending (collider variable), EU funding cycles, and awareness of EU funding (exogenous sources of variation). Unfortunately, some of the needed confounders are not available in the data. We do not have a measure for bureaucratic quality, political alignment (election results may be used, but these are only available every 4 years), and education levels (we only have non-yearly census data). Spatial spillovers could have been accounted for by using spatial autoregressive models, but this was out of the scope of this analysis. However, if we assume that these variables are mostly time-invariant, then using fixed-effects (FE) or first-differenced (FD) models should control for all time-invariant confounders. Also, if the reverse causality mechanism through political alignment is delayed, then including a lead term could account for that. Thus, the main models with which I aim to uncover a causal effect will be FE and FD models (the latter with lag and lead terms)[11],[12]. I will also add time fixed effects to all my models to account for aggregate trends, and I will experiment with adding municipality dummies to my FD models to account for geographically heterogenous trends. All models will be estimated weighted by population, as this is more informative given that incomes are realized on an individual level. As noted before, I will estimate models both with a continuous and a binary treatment variable.

Before moving on to discussing results, I present some more detailed descriptions of the variables I do have in my dataset. Income, the outcome variable, is measured by the yearly total income subject to personal income tax (in 2024 price HUF), divided by population. EU funding, the causal variable, is an aggregate measure of individual funding decisions (as described in Section 1), in 2024 price HUF, also divided by population. Unemployment rate is the number of people unemployed for more than 180 days, divided by the work-age population, in percentages. Industry employment rate is the number of people working in industrial positions, divided by the work-age population, in percentages (agricultural employment is perfectly analogous). High prestige employment rate has been downloaded as a percentage value by default. Lastly, population density is measured in people/km^2.

4. Model estimates

The main results of the most appropriate regression models are presented in Table 2. I present models with either a continuous or a binary treatment variable. For the FD models, I present cumulative effect estimates[13], as these are more informative given that the allocated funding is actually paid out over multiple years.

First, let's concentrate on the models with the continuous treatment variable. Let's note the difference in the estimates of the FD and FE models. The reason behind the difference may be the fact that I could not control for the reverse effects in the FE model (even though the lead terms turned out to be significant in the FD models and do impact significantly the estimates[14]) and municipality specific trends (which also impact the results substantially). Another reason may be that real effect of EU funding takes more time than 3 years to kick in – this longer run effect may be captured better by the FE model, as it cannot really be reliably estimated by FD models given the relatively short time period available for analysis. All in all, I believe that the

---

[11] I have also run a simple diff-in-diff model, with a binary treatment denoting whether the municipality received high or low levels of funding in total in the period between 2007 and 2022. Even though the PTA seemed to hold, the model was quite uninformative given that it could not take into account the timing of the funding. See the details in the submitted Jupyter-notebook.

[12] Of course, many more models (simple OLS models, models without controls, and less complex FD models) have been estimated during the analysis, the detailed results of which may be found in the submitted Jupyter-notebook. Note that the simple cross-sectional OLS models for different years showed a significant relationship between the causal and the outcome variable, so there is indeed a rationale for the analysis.

[13] The individual lagged coefficients generally show a relationship decreasing in magnitude over time.

[14] The lead term coefficients were positive in all models, meaning that municipalities tend to receive funding after their income levels have already increased.

FE model gives a more biased estimate than the FD model containing lead terms and district dummies (this is also the case for the binary treatment set-up) – but note that with the FD model, I can only uncover rather short-term effects.

Second, the results suggest that it is not the magnitude of EU funding, but rather the fact whether a municipality received EU money that explains variation in income per capita. Thus, no meaningful relationship is shown between the EU funding per capita and income per capita. However, all estimated models show a significant pattern between receiving funding and income per capita – so using a binary treatment variable may be a meaningful and more informative choice.

Considering the above, I believe that my best model is Model (3) with binary treatment – the FD model containing both lags, leads, and municipality-specific trends. The cumulative coefficient of this model suggests that, controlling for changes in the given confounders, within 3 years of a change from no EU funding to receiving EU funding, real income per capita tends to change by 1.03%[15] on average relative to the municipality-specific trend.

Table 2: Main regression results

	(1) – FD Δln(income p.c.)	(2) – FD Δln(income p.c.)	(3) – FD Δln(income p.c.)	(4) – FE ln(income p.c.)
	continuous treatment models			
Δln(EU funding p.c.) cumul.	-0.00005 (0.00067)	0.00126 (0.00099)	-0.00129 (0.00109)	
ln(EU funding p.c.)				0.00115** (0.00052)
R^2	0.759	0.756	0.782	0.950
	binary treatment models			
Δ(funding dummy) cumul.	0.01633*** (0.00232)	0.03092*** (0.00293)	0.01031*** (0.00362)	
funding dummy				0.03470*** (0.00267)
R^2	0.759	0.756	0.782	0.949
No. of lags	3	3	3	–
No. of leads	0	2	2	–
Muni. dummies	no	no	yes	–
T:N	12:3148	10:3148	10:3148	16:3148

Notes: Clustered standard errors in parentheses. The R^2 for the FE models is the within R^2. All models include confounders as well as year fixed effects. Confounders include all confounders explained in Section 3, as well as appropriate flags created during data cleaning and transformation (thus the continuous treatment models also include the no funding dummy*, as it was created because of the log transformation, and which is one if the* funding dummy *is zero). Confounders are added in differences (with lag terms as well) to the FD models, but in levels to the FE model. All models are weighted by average population over 2007-2022.*

5. Discussion

As mentioned earlier, some important confounders were unfortunately not available in my dataset. As a result, the main effect estimate can only be considered a causal relationship if we assume that all these omitted confounders are mostly time invariant, or they only effect level differences, not changes – which is rather a though assumption to make, especially for regional spillovers. However, according to the DAG in Figure 1, all the identified, but omitted confounders are related to both the causal and the outcome variable in the same direction. If we accept this, then the estimated effects are positively biased, meaning that the actual effect should be closer to zero (or it could even have an opposite sign).

5.1. Robustness of results

Though not in Table 2, I have also estimated FD models without lags or leads, and without

---

[15] As the model is weighted by population, this may be interpreted as the average expected real income change for any randomly chosen individual, given that they live in a municipality included in the sample (remember that we have excluded the capital and those municipalities which have undergone administrative changes). Also, note that this is only an ATET estimate at best, given that municipalities do self-select into treatment when deciding whether to apply for funds.

adding confounders. The effect estimates range from 0.1% to 3.09%[16]. Both adding lag and lead terms significantly increased the coefficients, but adding municipality dummies had a negative impact. Adding confounders also had an impact on all model specifications[17]. In addition, the FE estimate differs significantly from the best FD estimate (as discussed more in detail in Section 4). Thus, my estimate is not quite robust across different identification strategies. However, I do believe that the chosen model captures the most endogenous variation in the causal variable, thus it is the closest to a true causal estimate. Regarding the validity of the results, the internal validity of the estimates is questionable due to many omitted factors (as discussed briefly above). External validity may be okay (given that we accept the results' internal validity) for other timeframes in the near future, but I would not expect the results to generalize to many years in advance or to different parts of the EU.

5.2. Heterogeneity of effects

The estimates discussed above only show an average relationship across all municipalities. However, one might argue that the effects may be heterogeneous based on different characteristics of the municipalities. One such factor that may influence the effects could be the size of the municipalities. To test this idea, I re-estimated Model (3) with binary treatment on subsamples of small (less than 5000 inhabitants) and large (more than 5000 inhabitants) municipalities[18]. These models show that within 3 years of a change from no EU funding to receiving EU funding, real income per capita increases by 0.99% for small municipalities, but by 3.18% for large ones, relative to their municipality-specific trends[19]. I have also examined the possibility of heterogeneity for the continuous treatment model. However, the coefficients turned out to be still insignificant for both subsamples – though the coefficient was now positive for small, but negative for large municipalities.

6. Conclusion

Although the analysis turned out to be rather inconclusive mainly due to the lack of important confounders in my dataset, we can still identify some important insights. Most importantly, my main model showed that the 3-year cumulative effect of changing from no EU funding to having EU funding on real income per capita is at maximum 1.03% for the whole sample. Thus, we have some evidence suggesting that indeed, EU funding contributes to local real income growth (though we cannot rule out that this effect may be smaller or even negative in reality).

Second, we must note that the magnitude of this possible effect is rather miniscule (especially so for small municipalities). This strongly suggests that EU funding does not effectively translate into local income growth. The reason behind this may be two-fold. On the one hand, it is possible that the real effect takes even more than 3 years to show up (as discussed in Section 4). On the other hand, municipalities simply might not have the necessary absorptive capacities to translate funding into income growth[20]. If the latter is the case, that may have some relevant policy indications – e.g., local and national policymakers should take action to strengthen the absorptive capacities of municipalities so that EU funding translates more strongly into local income growth. Also, as the possible effects of EU funding seem to be heterogeneous across different municipality sizes, policymakers should identify why larger municipalities are better at translating EU funding into income growth. Then, these best practices could be used to strengthen the link between EU funding and income growth for smaller municipalities as well.

Lastly, regarding Hungary's current inability to utilize EU funds due to rule of law concerns, we can conclude that the estimates suggest that sustained lack of access to EU funds may hinder local income growth in the coming years.

---

[16] Contemporaneous effect estimate only with controls, vs. Model (2) in Table 2.

[17] However, for the chosen FD model this change was rather negligble (only 0.01 percentage point), meaning that this model indeed captures most of the endogenous variation.

[18] The detailed results of these models can be found in the submitted Jupyter-notebook.

[19] Both of these results are significant at 1%, thus there is strong evidence for the heterogeneity of effects.

[20] Prior literature mentioned at the beginning of this paper also suggests that absorptive capacity is indeed a key factor in translating EU funding to positive economic outcomes.