

---

# ANALÝZA INDIKÁTORŮ EVROPA 2020 REGRESNÍ ANALÝZOU

Dagmar Blatná, Lenka Hudrlíková\*

## Úvod

Evropská komise definovala novou desetiletou strategii, jež má následovat lisabonskou strategii. Nová strategie se nazývá Strategie pro inteligentní a udržitelný růst podporující začlenění (European Commission 2010b) a definuje směr, kterým se mají evropské státy ubírat a cíle, jež mají být dosaženy do roku 2020. Bylo stanoveno pět hlavních cílů:

- míra zaměstnanosti obyvatelstva ve věku 20–64 let by měla vzrůst nejméně na 75 %,
- každá země by měla investovat 3 % svého HDP do výzkumu a vývoje,
- v oblasti klimatu a energie by mělo být dosaženo tzv. cílů „20-20-20“, to znamená:
  - snížit emise skleníkových plynů oproti úrovním z roku 1990 nejméně o 20 % (pokud budou příznivé podmínky tak o 30 %),
  - zvýšit podíl obnovitelných zdrojů energie v konečné spotřebě energie na 20 %,
  - zvýšit energetickou účinnost na 20 %,
- snížit podíl dětí, které předčasně ukončují školní docházku, ze současných 15 % na 10 % a zvýšit nejméně na 40 % podíl obyvatel ve věku od 30 do 34 let s ukončeným terciárním vzděláním,
- snížit o 25 % počet Evropanů, kteří žijí pod hranicí chudoby daného státu, což vyvede z chudoby přes 20 milionů lidí.

Pro monitorování dosaženého stavu a plnění výše uvedených cílů Evropská komise vybrala z ukazatelů sledovaných Eurostatem soubor osmi indikátorů nazvaný EVROPA 2020:

- míra zaměstnanosti (v %),
- výdaje na výzkum a vývoj (% HDP),

---

\* Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky (blatna@vse.cz, lenka.hudrlikova@vse.cz).

- snižování emisí CO<sub>2</sub>,
- podíl obnovitelných zdrojů energie (v %),
- energetická účinnost (snížení spotřeby energie v mil. tun ropného ekvivalentu),
- předčasné ukončování školní docházky (v %),
- terciární vzdělání (v %),
- počet osob ohrožených chudobou či sociálním vyloučením.

Ukazatele souboru Evropa 2020 jsou individuální, vzájemně neprovázané, jelikož Strategie pro inteligentní a udržitelný růst podporující začlenění není ucelená koncepce vedoucí k jedné cílové hodnotě. Strategie pokrývá několik oblastí a stanovuje několik priorit, jež jsou nastavené odděleně a jsou konkretizované v sedmi nepropojených iniciativách, jako je např. „Program pro nové dovednosti“, „Mládež v pohybu“, „Boj proti chudobě“, „Inovace v Unii“, „Evropa náročná na zdroje“. Všechny dílčí cíle by ale měly vést k rozvoji celé společnosti ve všech oblastech života. Cílem článku je analyzovat ukazatele Evropa 2020 z pohledu vzájemných vztahů a vztahů s jinými ukazateli, jelikož pomocí řízení hodnot určitých ukazatelů by bylo možno zároveň dosáhnout cílových hodnot jiných ukazatelů. Při snaze o dosažení cíle nemusí být vždy přímo ovlivňován daný ukazatel, neboť pozitivní (resp. negativní) vliv na jeden ukazatel se může projevit příznivě i u jiných. Politici se tedy při svých rozhodnutích nemusí omezovat na hlavní monitorované jevy, ale správným nastavením politik lze cílovat i jiné ukazatele, pokud bude zřejmý vztah mezi ukazateli. Záměrem článku tedy bylo nalézt relevantní regresní model pro každý indikátor Evropa 2020. V původním datovém souboru bylo zahrnuto více než 40 ukazatelů z oblasti ekonomie, financí, vědecko-technického rozvoje, sociální oblasti, vzdělání apod., které by bylo možno využít jako vysvětlující proměnné změn indikátorů Evropa 2020. Tyto ukazatele byly vybrány záměrně na základě věcných vazeb ať už prokázaných nebo předpokládaných ekonomickou teorií. Mnohé z ukazatelů, které by z hlediska ekonomické teorie bylo vhodné aplikovat jako vysvětlující ukazatele, nebylo možno využít, často z důvodu, že nejsou pravidelně za jednotlivé země EU sledovány a publikovány, výběr byl do značné míry omezen také dostupností dat za všechny země EU. Článek není veden s cílem konfrontovat různé ekonomické teorie. Výsledky a závěry jsou odvozeny z kvantitativní analýzy dat zemí EU.

Všechna použitá data byla převzata z databáze Eurostatu, díky čemuž jsou definice jednotlivých ukazatelů v souladu s definicemi používanými Eurostatem a jsou srovnatelné. Analýza byla provedena na datech z roku 2010 (pouze v případě několika chybějících hodnot byly použité hodnoty roku 2009).

Nejčastěji používaným statistickým nástrojem k analýze závislostí je regresní analýza. V analýze dat indikátorů Evropa 2020 předpokládáme velkou variabilitu hodnot jednotlivých indikátorů i použitých vysvětlujících proměnných a existenci odlehklých pozorování, která mohou způsobit nesplnění podmínek použití klasické lineární regresní analýzy – normalitu, homoskedasticitu a nezávislost reziduí. V případech, kdy

jsou předpokládána (nebo dokonce i identifikována) odlehlá pozorování, jsou vhodným analytickým nástrojem k popisu závislostí metody robustní regrese, které jsou méně citlivé na porušení předpokladů klasického lineárního regresního modelu. Umožňují jednak odhalit odlehlá pozorování, ale i poskytnout dobré regresní odhady pro popis závislostí většiny dat analyzovaného souboru zemí EU. Z tohoto důvodu byla analýza indikátorů Evropa 2020 provedena klasickou lineární i robustní regresí. Výpočty byly provedeny s využitím statistického softwaru SAS 9.2 a S-Plus 6.

Článek je strukturován následovně. První část poskytuje bližší seznámení se strategií Evropa 2020. Následuje stručné představení použitých robustních metod regresní analýzy a diagnostických nástrojů. Výsledky analýzy jsou publikovány ve čtvrté části a poslední část poskytuje závěry a perspektivy pro další výzkum.

## 1. Literatura

Po publikování Strategie pro inteligentní a udržitelný růst podporující začlenění, vzniklo velké množství článků, analýz a reportů. Většina z nich je zaměřena na způsob výběru indikátorů (Martens 2010b), (Bongardt a Torres 2010), (Erixon 2010) či (Butković a Samardžija 2010). Relevance stanovených cílů je podrobena kritice (Colson) a jejich vzájemná propojenost je diskutována. Například (Nolan a Whelan 2011) ukazují, že růst zaměstnanosti nemusí mít nutně za následek nižší podíl lidí za hranicí chudoby. (Marx, Vandenbroucke a Verbist 2011) k testování podobné hypotézy využili regresní simulace. Realizovatelnost splnění cílů je zásadní otázka pro úspěšné splnění daných cílů. Touto otázkou se zabývají (Çolak a Ege 2011) či (Leschke, Theodoropoulou a Watt 2012). (Roth a Thum 2010) shledali, že cíle týkající se vzdělávání jsou velmi ambiciózní a téměř nesplnitelné v daném časovém horizontu. (Nolan a Whelan 2011) provedli analýzu, dle které cíl snížení chudoby není správně formulován. (Marlier a Natali 2010) poskytují komplexní pohled na sociální oblast v kontextu Evropa 2020. Další oblastí, v níž dochází ke kritice strategie a ukazatelů, je environmentální oblast, např. (Tišma a Čermak 2010) nebo (Papadaki 2012). (Stagnaro 2009) se zase zaměřil na evropský systém obchodování s emisními povolenkami, za účelem zhodnotit jeho efektivitu a následky, jež přináší i pro plnění environmentálních cílů strategie. Základní otázka je samozřejmě jak dosáhnout vytyčených environmentálních cílů (Klessmann 2009) či (Böhringer, Rutherford a Tol 2009). Jelikož strategie Evropa 2020 je následovníkem Lisabonské strategie, velká část článků se také zabývá porovnáním obou strategií (Martens 2010a) či přechodu z Lisabonu na Evropu 2020 (Hervás Soriano a Mulatero 2010), v rámci sociální oblasti např. (Natali 2010), (Stubbs a Zrinščak 2010), (Frazer, Marlier a Nicaise 2010), (Barbier 2011), (Lundvall a Lorenz 2012). Jak je zřejmé z uvedeného přehledu literatury, není této základní ekonomické strategii EU v českém odborném tisku věnována pozornost. Přes určitou kritiku již byla Strategie Evropa 2020 přijata a implementována na národních úrovních. Pozornost a úsilí by se nyní již mělo zaměřit na plnění vytyčených cílů.

## 2. Metodologie

Základní myšlenka robustní regrese spočívá v nalezení postupů, které modifikují klasickou metodu nejmenších čtverců (MNC) tak, aby byla redukována citlivost MNC na odlehlá pozorování, ale současně byly zachovány dobré vlastnosti odhadů získaných MNC. V případě nekontaminovaných dat jsou robustní regresní funkce blízké regresním funkcím získaným klasickou lineární regresí. Robustní regresní metody s vysokým bodem selhání jsou schopné identifikovat odlehlá pozorování (extrémní i vybočující odlehlá pozorování).<sup>1</sup>

Byla vyvinuta řada robustních metod řešících problémy s existencí odlehlých pozorování v regresi. V analýze byly využity dvě robustní metody s vysokým bodem selhání: LTS (the least trimmed squares) regrese a MM regrese. Odhady metodou LTS

(na rozdíl od MNC) získáme minimalizací součtu čtverců  $h$  reziduí  $\sum_{i=1}^h e_{(i)}^2$ , kde  $e_{(i)}^2$

je  $i$ -tá pořadová statistika čtverců reziduí (uspořádaných ve vzestupném pořadí).<sup>2</sup> LTS regrese je využita jako počáteční odhad parametrů v následující MM regresi, ale hlavně je vhodným analytickým nástrojem k odhalení odlehlých pozorování jak v jednoduchých, tak i ve vícenásobných závislostech.

MM odhady jsou definovány pomocí třístupňového postupu. V prvním stupni jsou vypočteny počáteční regresní odhady parametrů  $\hat{\beta}_j$ , které jsou konzistentní s vysokým bodem selhání, ale nemusí být vydatné. Ve druhém stupni je vypočten robustní odhad směrodatné odchylky chyb na základě reziduálních odchylek vypočtených z modelu s počátečními odhady  $\hat{\sigma}$ . Ve třetím stupni jsou iterativní procedurou vypočteny M-odhady regresních parametrů s vhodnou neklesající funkcí  $\psi$  řešením vztahu:

$$\sum_{i=1}^n \mathbf{x}_i \psi \left( \frac{y_i - \mathbf{x}_i^T \hat{\beta}}{\hat{\sigma}} \right) = 0,$$

kde  $\hat{\sigma}$  je robustní odhad směrodatné odchylky reziduí (vypočtený ve druhém kroku),  $\psi = \rho'$  je derivace vhodné ztrátové funkce  $\rho$ .<sup>3</sup> Podrobnější popis metod lze

1 V regresní analýze je vhodné rozlišovat mezi dvěma typy odlehlých pozorování. Pojmem extrémní odlehlá pozorování (*outliers*) se označují pozorování, u nichž se odlehlé hodnoty vyskytují u vysvětlované (závislé) proměnné  $y$ , zatímco odlehlé hodnoty u vysvětlujících proměnných (nezávislé proměnných) se nazývají vybočující pozorování (*leverage points*), přičemž tzv. dobrá vybočující pozorování (*good leverage points*) jsou vybočující pozorování, která nejsou současně extrémními odlehlými pozorováními. Jsou relativně vzdálená od většiny pozorování, ale leží blízko regresní funkce, kolem níž je soustředěna převážná část bodů. Dobrá vybočující pozorování mají omezený vliv na kvalitu regresních odhadů, někdy mohou dokonce i zlepšit jejich přesnost. Tzv. špatná vybočující pozorování (*bad leverage points*) jsou taková, která mají hodnoty vysvětlující proměnné vzdálené od většiny ostatních pozorování. Existence špatných vybočujících pozorování výrazně snižuje přesnost regresních odhadů, mnohdy i způsobuje mylnou informaci o směru závislosti.

2  $h$  je největší celé číslo mezi  $\left\lfloor \frac{n}{2} \right\rfloor + 1$  a  $\left\lfloor \frac{3n + p + 1}{4} \right\rfloor$ ,  $p$  je počet vysvětlujících proměnných, použitý software SAS používá defaultně nastavenou hodnotu  $\left\lfloor \frac{3n + p + 1}{4} \right\rfloor$ , která odpovídá bodu selhání kolem 25%.

3 Z důvodu srovnatelnosti výpočtů získaných použitým softwarem byla aplikována Tukeyho ztrátová funkce  $\rho$ , která je implementována v obou použitých softwarech SAS a S-Plus.

nalézt např. v (Ruppert a Carroll 1980), (Rousseeuw 2003), (Chen 2002), (Fox 2002), (Hubert, Rousseeuw a Van Aelst 2008), (Antoch a Vorlíčková 1992), (Yohai 1987) a v příručkách k statistickým programům (SAS), (UCLA) a (S-PLUS).

### 3. Identifikace odlehlých pozorování a výběr vhodného modelu

K identifikaci odlehlých pozorování v regresi byla vyvinuta celá řada numerických a grafických metod. Popis jednotlivých metod lze najít např. v (Rousseeuw 1984), (Rousseeuw a Van Zomeren 1990), (Rousseeuw 2003), (Fox 2002), (Olive 2002), (Swallow a Kianifard 1996), (Chen 2002) či (Hadi 1993). Z numerických nástrojů byly v analýze závislostí indikátorů Evropa 2020 využity následující: rezidua LTS regrese, standardizovaná rezidua, studentizovaná rezidua, robustní Mahalanobisova vzdálenost. Grafické metody identifikace, které jsou často využívány v průzkumové analýze dat, poskytují rychlou informaci o existenci odlehlých pozorování, mohou odlišit extrémní a vybočující pozorování i identifikovat vlivná pozorování, která mají významný vliv na regresní odhady (tzn. jsou současně extrémní i vybočující). Grafickými metodami lze rovněž posoudit, zda výsledný regresní model poskytuje normálně rozdělená rezidua. Z grafických nástrojů byly použity následující robustní diagnostické grafy: graf standardizovaných reziduí vs. robustních vzdáleností, graf standardizovaných reziduí u jednotlivých pozorování (zemí), normální Q-Q graf standardizovaných reziduí a graf jádrové hustoty reziduí.

Pro výběr vhodného regresního modelu byl v případě klasické lineární regrese použit index determinace, t-testy pro posouzení významnosti regresních parametrů a celkový F-test – podrobněji viz např. (Rao a Toutenburg 1995), (Hebák 2005), (Hušek 2007), (Hadi 1993) či (Víšek 1997). V případě robustní regrese byl vhodný model vybrán na základě robustních kritérií: robustní index determinace, robustní vzdálenost, robustní t-test, robustní F-test a Waldův test a robustních informačních kritérií: Akaikeho informační kritérium (AICR), Bayesovské informační kritérium (BICR) a robustní chyba predikce (RFPE). U klasických i robustních lineárních modelů byla samozřejmě také ověřována normalita reziduí.

### 4. Analýza, výsledky a diskuse

Jak je zřejmé z výše uvedeného přehledu literatury, práce zabývající se analýzou všech indikátorů, které byly Evropskou komisí doporučené k monitorování cílů této strategie a nalezením vztahů těchto indikátorů k ukazatelům sledovaným Eurostatem, dosud nebyla provedena nejen u nás, ale ani v některé zemi EU. Článek si neklade ambiciózní cíl provést komplexní analýzu vztahů indikátorů Evropa 2020, ale přispět k prozkoumání této rozsáhlé problematiky.

Pro každou vysvětlovanou proměnnou (jednotlivé indikátory Evropa 2020) byl použit soubor vysvětlujících proměnných vybraný z databáze Eurostatu jak na základě ekonomické teorie, tak i na základě intuice a dostupnosti dat. Byla vypočtena celá řada regresních modelů jak klasickou lineární regresí, tak i robustními metodami (LTS a MM), a to metodou postupného přidávání proměnných nebo metodou STEPWISE. Posouzení, který z vypočtených modelů je vyhovující, bylo provedeno jak z logického

hlediska, tak i hlediska výše uvedených statistických kritérií. Mnohdy výsledné regresní modely získané formálním statistickým postupem (metodou STEPWISE) nebyly interpretovatelné z věcného hlediska, na druhé straně, z hlediska ekonomické teorie předpokládané kombinace vysvětlujících proměnných, mnohdy neposkytly regresní modely vyhovující statistickým kritériím. Ukázalo se, že některé obecné teoretické předpoklady o vazbách ukazatelů nejsou přímo aplikovatelné, neboť soubor zemí EU je z hlediska indikátorů Evropa 2020 i vysvětlujících proměnných značně heterogenní. Zřejmě by bylo účelné provést komplexnější analýzu pro regiony i pro skupiny zemí rozříděné z hlediska ekonomické vyspělosti zemí (případně i z jiného hlediska).

Vzhledem k velkému počtu proměnných (7 indikátorů a celkově cca 40 vysvětlujících proměnných), bylo získáno mnoho vyhovujících lineárních regresních modelů popisujících závislost indikátorů na použitých ukazatelích z oblasti ekonomiky, financí, vzdělání, sociální ochrany, vědeckotechnického rozvoje a dalších oblastí. Není možné všechny modely prezentovat, proto u jednotlivých indikátorů je uveden nejlepší model vyhovující jak z hlediska výše uvedených statistických kritérií, tak i logických vazeb. Pro ilustraci jsou u vybraných modelů uvedeny i výsledky grafické identifikace odlehklých pozorování nebo graf pro posouzení normality reziduí. V případech, kde nebylo možno vybrat jednoznačně nejlepší regresní model, jsou uvedeny také alternativní regresní modely vyhovující jak z hlediska logického, tak i statistického.

## 5. Analýza indikátorů z hlediska stěžejních priorit a hlavních iniciativ

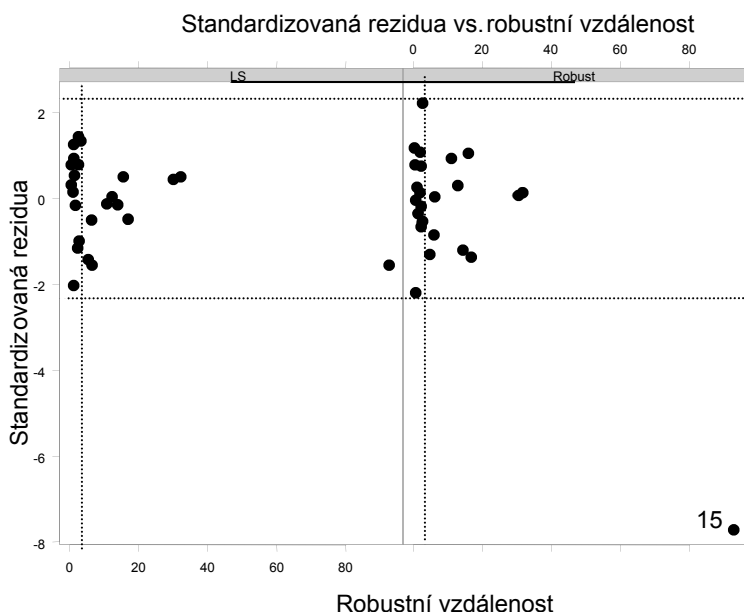
Strategie Evropa 2020 je rozdělena nejenom dle tří stěžejních priorit (inteligentní, udržitelný, inkluzivní růst), ale také je podpořena sedmi hlavními iniciativami.

Růst podporující začlenění má být dosažen díky iniciativě „Program pro nové dovednosti a pracovní místa“ – modernizace pracovních trhů a posílení postavení občanů rozvojem jejich dovedností v průběhu celého života za účelem zvýšení účasti na trhu práce a lepšího vyrovnání nabídky a poptávky na trhu práce, mimo jiné prostřednictvím mobility pracovních sil (European Commission 2010a). Nejdůležitějším ukazatelem v této oblasti je **míra zaměstnanosti** (ER, employment rate). Tento indikátor je jeden ze stěžejních v politickém rozhodování, a to samozřejmě nejenom v rámci Strategie Evropa 2020. ER má zásadní vliv na ekonomickou a sociální situaci země. Ukazatel je definován jako podíl zaměstnaných osob ve věku 20–64 na celkovém počtu obyvatel stejné věkové skupiny (Eurostat).

Regresní model pro závislou proměnnou ER, který splňuje výše uvedené požadavky na kvalitní model, lze vyjádřit pomocí vzorce:  $ER \sim GDPc + GGD + LTU$ . Na obrázku 1 v pravé části grafu je robustní diagnostika, v levé části grafu klasická. Body vně horizontálních čar zobrazují odlehlá extrémní pozorování, zatímco body napravo od vertikálních čar, značí vybočující pozorování. Robustní diagnostika odhalila jedno extrémní pozorování (15) Lucembursko a 11 vybočujících pozorování, klasická metoda neodhalila žádné extrémní odlehlá pozorování. Bod (15) Lucembursko je extrémní a zároveň vybočující, což znamená, že má významný vliv na odhady regresních parametrů v případě klasické lineární regrese. Lucembursko je obvykle označované za odlehlé pozorování v mnoha analýzách evropských dat, a to především díky specifikům této rozlohou malé země (např. dojíždění za prací z/do sousedního státu).

# **Obrázek 1**

## **Robustní diagnostický graf modelu $ER \sim GDPc + GGD + LTU$**



Pramen: Data EUROSTAT, vlastní zpracování.

Nezávislé proměnné jsou HDP na osobu (GDPc), hrubý veřejný dluh (GGD, general government gross debt) a dlouhodobá nezaměstnanost (LTU, long-term unemployment). Hrubý domácí produkt měřený na obyvatele (HDP) je měřítkem ekonomické výkonnosti země. Pro analýzu byl použit objemový index HDP na obyvatele vyjádřený v paritě kupní síly. Pokud je index za určitou zemi vyšší než 100, znamená to, že HDP na obyvatele této země je vyšší než průměr EU 27 a naopak. Údaje se uvádějí ve standardu kupní síly, tzn. společné měně, která stírá rozdíly v cenových hladinách mezi zeměmi a umožňuje tak srovnání HDP spíše mezi jednotlivými zeměmi než v čase. Samozřejmě vliv GDPc na míru zaměstnanosti je pozitivní (klasický párový koeficient korelace je 0,46; robustní 0,75). Přestože klasická produkční funkce předpokládá opačný kauzální vztah mezi HDP a ukazatelem práce, dle Okunova zákona může být vztah dvoucestný, tedy kauzální vztah vede jak od nezaměstnanosti k produkci, tak od produkce k nezaměstnanosti (Čadil et al. 2011). Proměnná GDPc tedy zůstala v modelu jako vysvětlující proměnná. Další proměnnou v modelu je ukazatel hrubého veřejného dluhu. Hrubý vládní dluh (GGD)<sup>4</sup> je vyjádřen jako konsolidovaný nevyrovnaný dluh (v nominální hodnotě ke konci roku) vyjádřený v procentech HDP. Vládní dluh může být financován buď půjčkou (např. pomocí vládních dluhopisů) či

4 Podrobná definice dle Eurostatu: [http://epp.eurostat.ec.europa.eu/portal/page/portal/product\\_details/dataset?p\\_product\\_code=TSDD410](http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?p_product_code=TSDD410).

zvýšením daní. Stát však v podstatě odčerpává finanční zdroje, které by jinak mohly použít soukromé subjekty na investice. Růst investic vede k podpoře konkurenceschopnosti, ekonomickému růstu a snížení nezaměstnanosti. To vše může mít negativní vliv na míru zaměstnanosti, jelikož pohyby v míře nezaměstnanosti jsou způsobeny hlavně změnami v odvětvích jako je stavebnictví, výroba, tzn. odvětvích závislých na zakázkách (velkých investicích). V souboru zemí EU nebyla prokázána významná závislost (klasický koeficient korelace je -0,16; robustní -0,30), ale zařazení GGD do vícenásobného regresního vztahu významně zvyšovalo index determinace. Poslední proměnnou je dlouhodobá nezaměstnanost (LTU, která udává procentuální podíl dlouhodobě nezaměstnaných (12 měsíců a déle) z celkového aktivního obyvatelstva. Vliv LTU na ER je negativní (klasický koeficient korelace je -0,63; robustní -0,90).

V tabulce 1 je uvedeno srovnání lineárních regresních modelů získaných klasickou a robustní regresí. Je zřejmé, že hodnoty regresních parametrů získaných klasickou a robustní regresí se liší.<sup>5</sup> Klasický lineární model není akceptovatelný jednak z hlediska nevýznamnosti regresních parametrů, ale i z hlediska nesplnění podmínky normality reziduí (viz obrázek 3 a obrázek 4). S ohledem na existenci vlivného pozorování a vybočujících pozorování, které vedou ke zkreslení regresních koeficientů i k nenormalitě reziduí lineárního LS modelu, jednoznačně preferujeme robustní model  $ER \sim GDPc + GGD + LTU$  uvedený v tabulce 1.

**Tabulka 1**

**Robustní MM a klasické LS regresní modely  $ER \sim GDPc + GGD + LTU$**

Metoda odhadu	Parametry regresní funkce	Odhady regresních parametrů	Směrodatné odchylky regresních parametrů	t-test	$Pr(> t )$ (p-hodnota)	Waldův test	$Pr(>Chi)$ (p-hodnota)
MM	abs. člen	64,643	4,7174	13,7029	0,0000	322,78	0,0001
LS	abs. člen	72,3561	3,6580	19,7800	0,0000		
MM	GDPc	0,1277	0,0433	2,9499	0,0072	15,92	0,0001
LS	GDPc	0,0256	0,0221	1,1567	0,2593		
MM	GGD	-0,0754	0,0360	-2,0944	0,0490	8,87	0,0029
LS	GGD	-0,0248	0,0268	-0,9250	0,3646		
MMt	LTU	-0,7634	0,1115	-1,7176	0,0993	4,70	0,0302
LS	LTU	-1,1782	0,3967	-2,9697	0,0069		

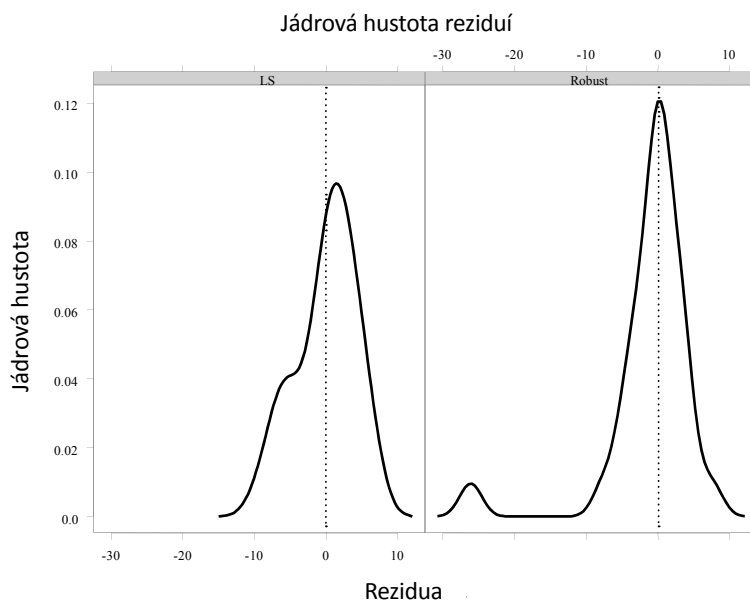
Pramen: Data EUROSTAT, vlastní výpočty.

<sup>5</sup> Hodnoty ve sloupci t-test jsou individuální t-testy jednotlivých regresních parametrů, Waldův test je užíván k testování významnosti regresních parametrů u robustní regrese. Symboly  $Pr(>|t|)$  a  $P(>Chi)$  vyjadřují minimální hladinu významnosti, na níž můžeme zamítnout hypotézu o nulové hodnotě parametru.



## Obrázek 2

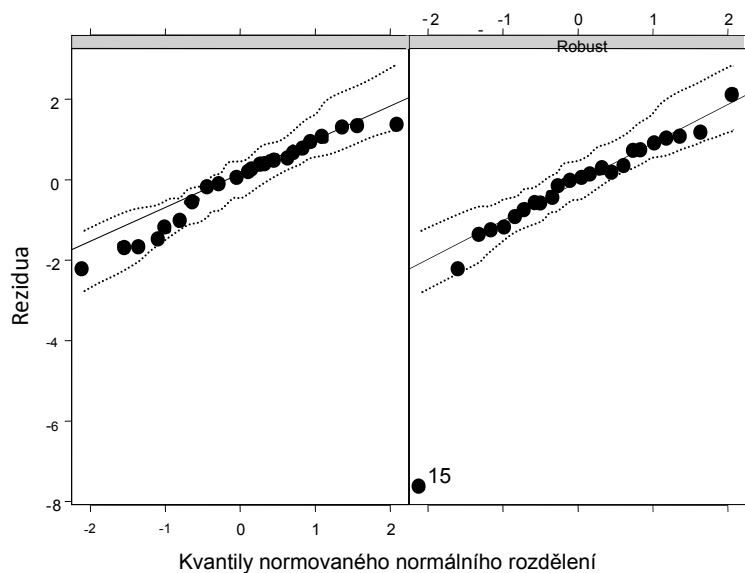
### Jádrová hustota reziduí klasického LS a robustního MM modelu $ER \sim GDPc + GGD + LTU$



Pramen: Data EUROSTAT, vlastní zpracování.

## Obrázek 3

### Normální Q-Q graf reziduí LS a MM modelu $ER \sim GDPc + GGD + LTU$



Pramen: Data EUROSTAT, vlastní zpracování.

Pro indikátor ER je možno jako vhodný model doporučit i další robustní model  $ER \sim LLL + ILCS$ , u něhož robustní regrese identifikovala jedno extrémní pozorování (12 Kypr). Vysvětlující proměnné jsou jiného charakteru. Ukazatelé vyjadřující participaci ve vzdělávání ve věkové skupině 25–64 let (LLL, Participation in education and training from 25 to 64 years) či individuální úroveň vzdělání v oblasti počítačů (ILCS, Individuals' level of computer skills) se zdají být snáze ovlivněny pomocí politických rozhodnutí a ustanovení. Tyto ukazatele nejsou tolik komplexní, a tudíž jejich cílová hodnota může být snáze řízena, a tedy i docílena.

Jak bylo řečeno v metodologické části, existuje mnoho kritérií pro volbu vhodného modelu. Nicméně v některých případech je téměř nemožné zvolit pouze jeden model jako jediný vhodný. Tabulka 2 uvádí oba doporučené robustní modely včetně charakteristik vhodnosti.

**Tabulka 2**

**Základní charakteristiky vhodnosti vybraných robustních modelů pro vysvětlovanou proměnnou ER**

Odlehlá pozorování	Robustní model	R <sup>2</sup>	AICR	BICR	deviance	RFPE
E: 15; V: 6,7,8, 9,13,15,22,24	64,623+0,128 GDPC–0,075 GGD–0,764 LTU	0,545	20,439	30,166	288,665	17.773
E: 12; V: 4,18,25, 26,27	54,707 + 0,508 LLL +0,641 ILCS	0,583	22,291	28,794	251,460	14.451

Pramen: Data EUROSTAT, vlastní výpočty. E značí extrémní pozorování, V značí vybočující pozorování.

Klíčovou roli ve strategii Evropa 2020 hraje vzdělání, specificky v oblasti inteligentního růstu, a to v iniciativě „Mládež v pohybu“, jež má za cíl zlepšení výsledků vzdělávacích systémů a zlepšit mezinárodní atraktivitu evropských vysokoškolských institucí. Dva cíle a relevantní ukazatele jsou tedy zaměřeny na vzdělávání: snížení míry nedokončení studia pod 10% a dosažení alespoň 40% podílu vysokoškolsky vzdělaného obyvatelstva věkové kategorii od 30 do 34 let. První indikátor je ukazatel **terciárního vzdělání** (TEA, tertiary educational attainment), jež je konstruován jako podíl populace ve věku 30–34 let, jež úspěšně dokončil terciární stupeň vzdělání (Eurostat). Výsledkem rozsáhlého testování modelů se závisle proměnou TEA byl zvolen model  $TEA \sim GEE + GDPc$  s dvěma nezávisle proměnnými: vládní výdaje na vzdělávání<sup>6</sup> jako podíl na HDP (GEE) a HDP na osobu (vyjádřeno v paritě kupní síly<sup>7</sup>).

6 Dle klasifikace COFOG.

7 Stejný ukazatel jako v ER modelu (GDPc).

Robustní LTS regresí byly u tohoto modelu identifikovány 4 vybočující pozorování (Bulharsko, Dánsko, Lucembursko a Rumunsko), žádné z nich ale není současně extrémní. Vzhledem k neexistenci vlivných pozorování, je klasický lineární regresní LS model stejný jako robustní model, proto v této situaci je klasický lineární regresní model plně vyhovující (klasické a robustní t-testy mají samozřejmě jiné hodnoty). Model je uveden v tabulce 3.

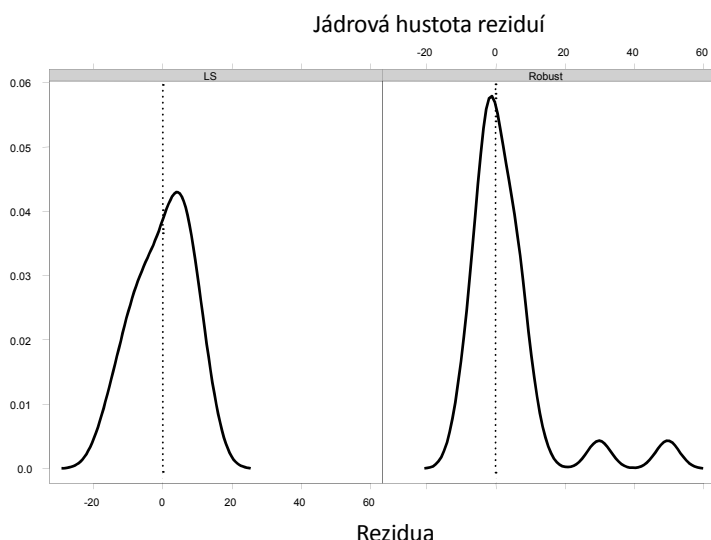
**Tabulka 3****LS regresní model TEA~ GEE + GDPc**

Parametry regresní funkce	Odhady regresních parametrů	Směrodatné odchytky regresních parametrů	t-test	$\Pr(> t )$ (p-hodnota)
abs. člen	-2,9209	7,4184	-0,3937	0,6973
GEE	5,0630	1,2331	4,1059	0,0004
GDPc	0,0888	0,0329	2,6998	0,0125

Pramen: Data EUROSTAT, vlastní výpočty.

V modelu TEA~ GEE + GDPc mají obě vysvětlující proměnné kladný vliv na výslednou hodnotu TEA. Intuitivně vyšší HDP na obyvatele znamená vyšší hodnotu ukazatele TEA. Stejně tak vyšší výdaje na vzdělávání mohou vést k vyšší hodnotě TEA. Pokud jsou vyšší výdaje na vzdělávání, může být nižší spoluúčast studentů a studenti nemusí z finančních důvodů opouštět vzdělávací systém. Poznotek z analýzy je v souladu s ekonomickou teorií.

Druhý potenciálně vhodný regresní model vyhovující všem výše uvedeným statistickým kritériím je model TEA~ GEE + NNI, kde místo HDP na obyvatele je vysvětlující proměnná čistý národní důchod (NNI, Net national income). NNI odpovídá hodnotě HDP opravené o primární rozdělení důchodů a spotřebu fixního kapitálu. V tomto případě bylo robustní regresí odhaleno sedm vybočujících pozorování (ČR, Německo, Estonsko, Irsko, Lucembursko, Malta a Rumunsko), z nich dvě jsou vlivná pozorování (7 Irsko a 15 Lucembursko), která měla za následek opačná znaménka dílčích regresních koeficientů u vysvětlující proměnné NNI v klasickém a robustním regresním modelu. Vzhledem k nevýznamnosti regresních parametrů klasického LS modelu a nenormalitě reziduí (viz obrázek 4), nelze klasický lineární model akceptovat, klasická regrese zde selhala. Přijetí klasického modelu by vedlo ke zcela nepříjemnému regresnímu modelu, robustní model TEA~ GEE + NNI je však akceptovatelný.

**Obrázek 4****Jádrová hustota reziduí klasického LS a robustního MM modelu TEA~ GEE + NNI**

Pramen: Data EUROSTAT, vlastní zpracování.

Pro indikátor TEA lze doporučit dva regresní modely vyhovující z hlediska použitých kritérií interpretovatelnosti (viz tabulku 4). Zvýšení hodnoty indikátoru TEA lze dosáhnout současným zvýšením výdajů na vzdělávání a zvýšením HDP na osobu, resp. NII.

**Tabulka 4****Výsledné regresní modely pro závisle proměnnou TEA**

Odlehlá pozorování	regresní modely		R <sup>2</sup>	AICR	BICR	deviance	RFPE
E: - V:2,4,8, 15,22	LS	-2,921+5,063 GEE +0,089 GDPc	0,540				
E:7,15; V:3,5,6,7, 15,16,17,21	MM	-104,940+4,116 GEE +1,385 NNI	0,532	22,29	29,58	1050,6	14,26

Pramen: Data EUROSTAT, vlastní výpočty. E značí extrémní pozorování, V značí vybočující pozorování, tučně označená jsou vlivná pozorování.

Druhý ukazatel týkající se vzdělávání je **míra nedokončení studia** (ELE, Early leavers from education and training). Ukazatel je definován jako podíl populace ve věku 18–24 let, která dosáhla nejvýše nižšího stupně sekundárního vzdělání, tomu odpovídají stupně ISCED 0,1,2,3, a která nebyla dále ve vzdělávacím systému během

4 týdnů předcházejících průzkumu (Eurostat). S cílem najít vhodný model bylo opět analyzováno více než 40 proměnných obsažených v databázi Eurostatu. Všechny vyhovující modely zahrnovaly proměnnou – počet osob s vyšším sekundárním nebo terciárním vzděláním (PUSE, Persons with upper secondary or tertiary education attainment). Proměnné ELE a PUSE jsou velmi silně korelovány, jelikož v podstatě je jeden doplňkem druhého. Potenciální modely neobsahující proměnnou PUSE měly velmi nízký index determinace a další ukazatele vhodné jako vysvětlující proměnné v databázi nejsou, případně nejsou k dispozici údaje za všechny země EU. Z těchto důvodů není pro indikátor ELE prezentován jako vyhovující žádný model.

Boj proti chudobě a sociálnímu vyloučení je předmětem iniciativy “Evropská platforma pro boj proti chudobě” (European Commission 2010c). Hlavní ukazatel v této iniciativě je **počet osob ohrožených chudobou či sociálním vyloučením** (PRPSE, People at risk of poverty or social exclusion). Ukazatel je tvořen součtem hodnot tří dílčích ukazatelů: počet osob ohrožených chudobou, počet osob materiálně deprivovaných a počet osob žijících v domácnostech, vyznačujících se velmi nízkým zapojením do pracovního procesu (Eurostat). Pomocí procedury STEPWISE byl konstruován regresní model  $PRPSE \sim ELE + RULC + GGD + PUSE$  ( $R-sq = 0.53$ ). Jeden z prediktorů je ukazatel předčasné ukončené školní docházky (ELE), což je vlastně sám o sobě jeden z ukazatelů Evropa 2020. Další prediktory jsou: růst skutečných jednotkových nákladů práce (Real unit labour cost – growth (RULC)), veřejný dluh vyjádřený jako podíl HDP (GGD) a počet osob s vyšším sekundárním nebo terciárním vzděláním (PUSE). V tomto modelu robustní LTS regrese odhalila 4 extrémní a 7 vybočujících pozorování, z toho 4 země byly identifikovány jako vlivná pozorování – (5) Německo, (9) Španělsko, (11) Itálie, (22) Rumunsko. Klasickou diagnostikou nebyla odlehlá pozorování zjištěna. Z grafické identifikace bylo zřejmé, že LS model neposkytuje normálně rozdělená rezidua. Kromě toho, znaménko dílčího regresního koeficientu u PUSE nemělo logicky správný směr. Z uvedených výhod je zřejmé, že použití klasického LS modelu bez odpovídající identifikace odlehlých pozorování a testování normality by mohlo vést k přijetí neakceptovatelných modelů. Robustní model se stejnými vysvětlujícími proměnnými je rovněž nepoužitelný jak z hlediska interpretace, indexu determinace ( $R-sq.=0,063$ ), tak i ze statistického hlediska nevýznamnosti parametrů a nenormality reziduí. Pro indikátor PRPSE tedy nelze doporučit žádný z testovaných LS ani MM modelů s uvažovaným souborem vysvětlujících proměnných.

Iniciativa „Inovace v Unii“ má vést ke zlepšení rámcových podmínek a přístupu k financování výzkumu a inovací, čímž by se posílil inovativní řetězec a zvýšil objem investic v celé Evropské unii. Takzvaný inteligentní růst je zaměřen na vzdělávání, výzkum a inovace a digitální společnost, tudíž na indikátor – **Hrubé domácí výdaje na výzkum a vývoj jako podíl na HDP** (GERD, Gross expenditure on R&D). Pro podíl výdajů na vědu a výzkum na HDP bylo nalezeno několik vhodných modelů (viz tabulku 5). V případech, kdy nebyly identifikovány vlivné body (modely 2, 5 a 6) jsou

klasický LS a robustní model velmi blízké a lze použít klasické lineární regresní LS modely, v případě identifikovaných vlivných pozorování pak robustní model.

**Tabulka 5**

**Robustní MM a klasické LS regresní modely pro závisle proměnnou GERD**

Odlehlá pozorování		regresní modely	index det.	AICR	BICR	Devian.	RFPE
E: 13 V: 4,6,13,19, 21,25,26	MM	-1,936 + 0,018 CPL + 0,037 SRE + 0,022 IRUI	0,624	29,952	31,103	4,029	15,19
E: -; V: 22	LS	-2,860 + 0,162 GESP + 0,025 IRUI	0,797				
E: 4; V: 4,10,25, 26,27	MM	-1,772 + 0,070 LLL + 0,155 GESP	0,637	22,389	28,763	3,900	13,64
E: 12,25; V: 5,12,15, 18,25	MM	-7,042 + 0,111 ER + 0,013 LPH	0,538	22,53	29,958	6,418	24,25
E: -; V: 2,13,22, 25,26	LS	-1,517 + 0,045 HBA + 0,096 SRE	0,688				
E: -; V: 2,4,13,19, 22,25,26	LS	-1,960 + 0,018 CPL + 0,024 HBA + 0,029 SRE	0,785				

Pramen: Data EUROSTAT, vlastní výpočty. E značí extrémní pozorování, V značí vybočující pozorování, tučně označená jsou vlivná pozorování.

První model je tvořen třemi vysvětlujícími proměnnými: komparativní cenová hladina (CPL, Comparative price level), podíl obnovitelných zdrojů<sup>8</sup> (SRE, Share of renewables) a osoby pravidelně užívající internet (IRUI, Individuals regularly using the Internet).<sup>9</sup> V druhém modelu je ukazatel IRUI spolu s ukazatelem veřejných výdajů na sociální oblast<sup>10</sup> (GESP, general government expenditure on social protection as a percentage of GDP). Třetí model obsahuje GESP avšak spolu s ukazatelem vyjadřujícím participaci ve vzdělávání ve věkové skupině 25–64 let (LLL, indicator measuring participation in education and training from 25 to 64 years). Čtvrtý model zahrnuje dvě vysvětlující proměnné: míru zaměstnanosti (ER) a produktivitu práce za hodinu (LPH, Labor productivity per hour). Pátý model je založen na ukazatelích vyjadřujících počet domácností se širokopásmovým připojením (HBA, Households with broadband access) a podíl obnovitelných zdrojů (SRE). Poslední uvedený model je kombinací proměnných CPL, HBA a SRE. V případě vysvětlované proměnné GERD je těžké určit, který z modelů je vhodnější než jiný, jelikož jich více splňuje statistické požadavky

<sup>8</sup> Ukazatel je SRE je jeden z ukazatelů Evropa 2020, a tudíž jeho bližší popis je v příslušné sekci.

<sup>9</sup> Pravidelně znamená nejméně jednou týdně v průměru během posledních tří měsíců před šetřením.

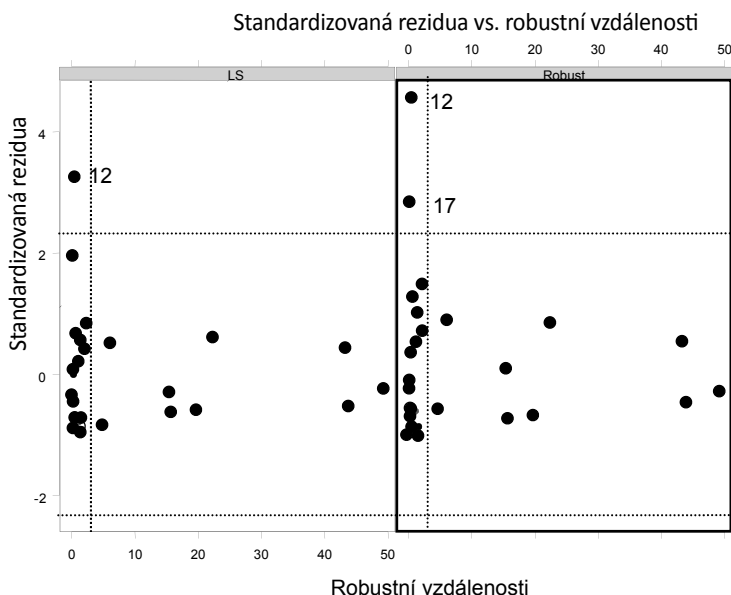
<sup>10</sup> Dle klasifikace COFOG.

a jsou dobře interpretovatelné i z věcného hlediska. Zde záleží na politické a hospodářské strategii jednotlivých zemí, který model použít v případě potřeby splnění cíle navýšení GERD.

Část strategie zabývající se udržitelným růstem je řízena iniciativami “Klima, Energie a Mobilita” a „Evropa méně náročná na zdroje“. Cílem je podpora oddělování hospodářského růstu od využívání zdrojů, a to pomocí snížení úrovně uhlíku, většího využívání obnovitelných zdrojů energie, modernizace dopravního odvětví a podpory energetické účinnosti. Indikátorem v této oblasti je **Snižování emisí CO<sub>2</sub>** (GGE, Greenhouse gas emissions), které je monitorováno od Kyotského protokolu. Ukazatel je definován jako vývoj uměle vytvořené emise šesti plynů, váženo jejich globálním oteplovacím potenciálem (Eurostat). Pro ukazatel GGE byl zvolen model s dvěma prediktory **GGE ~ GDP.gr + HICP**. V modelu byla identifikována dvě extrémní pozorování (12 Kypr a 17 Malta), ale žádné z nich není zároveň vybočující pozorování, klasická diagnostika identifikovala jako odlehlé extrémní pozorování pouze 12 Kypr (viz obrázek 6).

#### Obrázek 6

##### Diagnostický graf modelu **GGE ~ GDP.gr + HICP**



Pramen: Data EUROSTAT, vlastní zpracování.

Z důvodu existence odlehlých pozorování a nenormality reziduí u klasického modelu, preferujeme robustní model uvedený v tabulce 6. První prediktor je růst HDP (GDP.gr, GDP growth) a druhý je harmonizovaný index spotřebitelských cen (HICP, Harmonized Indices of Consumer Prices) konstruovaný s cílem zachytit změny cen

spotřebitelských zboží a služeb. Jedná se o mezinárodně srovnatelný ukazatel míry inflace. Regresní koeficienty u růstu HDP (GDPgr) i u indexu spotřebitelských cen (HICP) mají záporné znaménko, což ukazuje na negativní vliv obou vysvětlujících proměnných na ukazatel GGE. Výsledný regresní model ukazuje, že zřejmě v evropských zemích existuje tzv. trade off mezi ekonomickým růstem a environmentálně udržitelným rozvojem.

**Tabulka 6**

**Charakteristiky vhodnosti modelu GGE~ GDPgr + HICP**

Odlehlá pozorování	regresní modely		index det.	AICR	BICR	RFPE
E:12,17; V:2,6,7,8,13, 14,16,22	MM	375,556 -6,701 GDPgr -2,413 HICP	0,580	22,065	28,986	14,237
	LS	414,580 -7,435 GDPgr -2,697 HICP	0,546			

Pramen: Data EUROSTAT, vlastní výpočty. E značí extrémní pozorování, V značí vybočující pozorování.

V rámci udržitelného růstu byla představena iniciativa „Evropa méně náročná na zdroje“ (European Commission 2011) s hlavním ukazatelem **podílu obnovitelných zdrojů na finální energetické spotřebě** (SRE, Share of renewables in gross final energy consumption)<sup>11</sup>. Jedná se o energii spotřebovávanou na území jednotlivých států, jejíž původ je vodní, geotermální, větrné, solární, nebo energii z biomasy a biologického rozpadu odpadu (Eurostat). Všechna kritéria pro výběr vhodného modelu v této oblasti splňuje regresní model **SRE ~ ED+ EGRS**, u něhož bylo identifikováno osm vybočujících a pět extrémních pozorování, z toho tři (13 Lotyšsko, 14 Litva 26 Švédsko) jsou vlivná pozorování. Skandinávské země jsou v této oblasti odlehlá pozorování zjevně z důvodu odlišných přírodních podmínek od zbytku Evropy, což vede k jiné struktuře zdrojů energie. Vzhledem k identifikovaným vlivným pozorováním a faktu, že rezidua u klasického modelu nemají normální rozdělení, jednoznačně preferujeme robustní model (uvedený v tabulce 7) s vysvětlujícími proměnnými energetická závislost (ED, Energy dependence) a elektřina generovaná z obnovitelných zdrojů (EGRS, Electricity generated from renewable sources). Z vypočteného vztahu je zřejmé, že podíl obnovitelných zdrojů na finální spotřebě je možné zvýšit snížením energetické závislosti, která ukazuje, jak ekonomika země závisí na importu při pokrytí svých energetických potřeb a současným zvýšením podílu elektřiny generované z obnovitelných zdrojů na národní spotřebě elektřiny.

11 Více informací o ukazatelích, jejich definicích a konstrukci lze nalézt v Renewable Energy Directive 2009/28/EC, the Energy Statistics Regulation 1099/2008 a v DG ENERGY transparency platform [http://ec.europa.eu/energy/renewables/index\\_en.htm](http://ec.europa.eu/energy/renewables/index_en.htm).



**Tabulka 7**

**Robustní MM a klasický LS regresní model pro závisle proměnnou SRE**

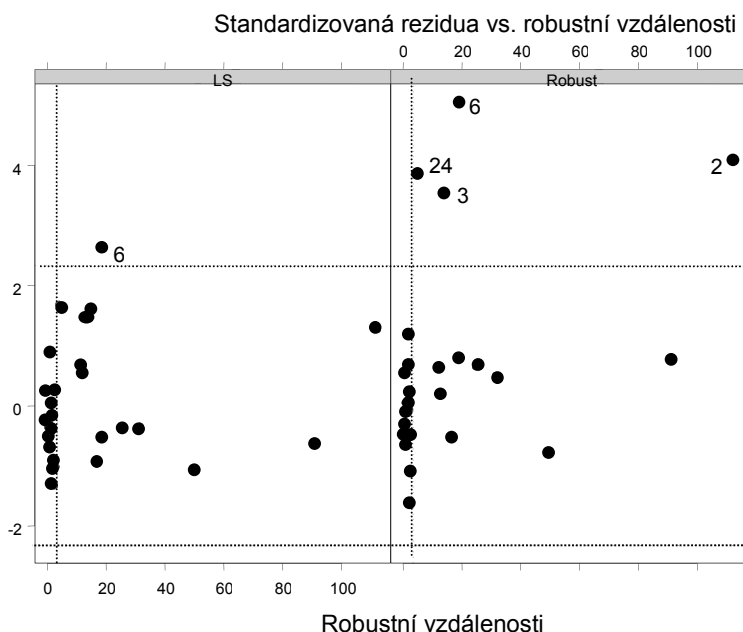
Odlehlá pozorování		regresní modely	index det.	AICR	BICR	RFPE
E:6,13,14,25,26; V:4,13,14, 17,19,21,22,23,26	MM	5,276 -0,046 ED + 0,437 EGRS	0,521	22,650	30,715	20,594
	LS	7,974 - 0,081 ED +0,529 EGRS	0,738			

Pramen: Data EUROSTAT, vlastní výpočty. E značí extrémní pozorování, V značí vybočující pozorování, tučně označená jsou vlivná pozorování.

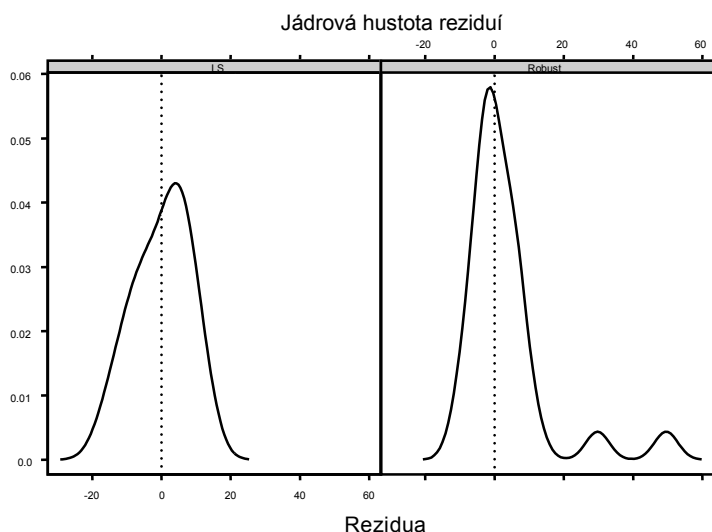
**Energetická intenzita ekonomiky** (EI, Energy intensity of the economy) je podíl mezi hrubou domácí spotřebou energie a HDP (Eurostat). Ukazatel tedy vyjadřuje energetickou spotřebu ekonomiky a její celkovou energetickou účinnost. Hrubá domácí spotřeba energie je počítána jako součet pěti druhů energie dle původu: uhlí, elektrina, ropa, zemní plyn a obnovitelné zdroje. Jako nejvhodnější byl vybrán regresní model **EI ~ CPL + RULC + HICP**. Robustní diagnostika odhalila čtyři extrémní pozorování ((2) Bulharsko, (3) Česká republika, (6) Estonsko a (24) Slovensko) a 11 vybočujících pozorování. Všechna čtyři extrémní pozorování jsou současně vybočující, tudíž jsou vlivnými body. Klasická diagnostika však dokázala odhalit jen jeden vlivný bod, (6) Estonsko (viz obrázek 7).

**Obrázek 7**

**Diagnostický graf pro model EI ~ CPL + RULC + HICP**



Pramen: Data EUROSTAT, vlastní zpracování.

**Obrázek 8****Jádrová hustota EI ~ CPL + RULC + HICP**

Pramen: Data EUROSTAT, vlastní zpracování.

Tabulka 8 uvádí jednoznačně preferovaný výsledný robustní regresní model, LS model neposkytl normálně rozdělená rezidua (viz obrázek 8).

**Tabulka 8****Robustní regresní model EI ~ CPL + RULC + HICP**

Parametry regresní funkce	Odhady regresních parametrů	Směrodatné odchylky regresních parametrů	t-test	Pr(> t ) (p-hodnota)	Waldův test	Pr(>Chi) (p-hodnota)
abs. člen	-765,817	298,1462	-2,5686	0,0172	1,51	0,2196
CPL	-1,4316	0,7859	-1,8216	0,0816	10,81	0,0010
RULC	12,1216	4,9892	2,4296	0,0233	10,01	0,0016
HICP	10,0790	2,1048	4,7887	0,0001	6,45	0,111

Pramen: Data EUROSTAT, vlastní výpočty.

Z provedené analýzy je zřejmé, že dosáhnout změnu energetické spotřeby ekonomiky a její celkové energetické účinnosti lze dosáhnout ovlivňováním komparativní cenové hladiny (CPL) země, skutečných nákladů práce (RULC) a míry inflace (vyjádřené harmonizovaným indexem spotřebitelských cen (HICP)).

## Závěr

Společná hospodářská politika, jak daných cílů v oblasti Strategie Evropa 2020 dosáhnout, nebyla EU formulována, jednotlivé stěžejní priority a je podporující hlavní iniciativy, nejsou vzájemně provázány. Cílem článku bylo analyzovat indikátory Evropa 2020 z pohledu vztahů s jinými makroekonomickými ukazateli, které jsou k dispozici za všechny sledované země EU a bylo je možno využít jako vysvětlující proměnné při konstrukci relevantních regresních modelech pro indikátory souboru Evropa 2020. Taková analýza dosud nebyla v rámci EU provedena.

Z podstaty ukazatelů, jež zachycují výkonnost evropských zemí, lze předpokládat, že hodnoty některých ukazatelů lze ze statistického pohledu označit za odlehlá pozorování a případně je následně identifikovat i jako vlivná pozorování, proto byla k analýze použita vedle klasické lineární regrese i robustní regrese s vysokým bodem selhání MM regrese.

Důkladná analýza ukazatelů Evropa 2020 vedla k nalezení vhodných regresních modelů pro téměř všechny indikátory, ve dvou situacích vhodný model uveden nebyl z důvodu nenalezení vhodných vysvětlujících proměnných mezi publikovanými ukazateli v databázi Eurostatu. Předpoklad existence vlivných odlehlých pozorování v analýze dat zemí EU se plně potvrdil, proto u většiny indikátorů byl jako nejvhodnější model doporučen robustní model. U některých identifikátorů nebylo možno jednoznačně určit jediný nejvhodnější regresní model a jsou uvedeny alternativní modely. Výběr, který z uvedených vypočtených modelů by mohl být k ovlivnění výše analyzovaného indikátoru použit, pak závisí na tom, jak má daná země nastavenou hospodářskou politiku a které ukazatele (resp. jejich kombinace) jsou politickými a ekonomickými zásahy snadněji ovlivnitelné (snížení, resp. jejich zvýšení).

Provedená analýza rovněž ukázala, že některé obecné teoretické předpoklady o vazbách indikátorů Evropa 2020 a vysvětlujících proměnných nejsou pro soubor zemí EU obecně platné, zřejmě z důvodu, že země EU tvoří soubor značně heterogenní a bylo by účelné provést komplexnější statistickou analýzu pro regiony i pro skupiny zemí rozříděné z hlediska ekonomické vyspělosti zemí (případně i z jiného hlediska, např. regionálního).

Článek nemá ambici posuzovat, zda je reálné dosáhnout cílových hodnot ukazatelů, touto problematikou u některých indikátorů se zabývaly články uvedené v části Literatura. V analýze nepoužíváme vzdálenost jednotlivých zemí od cílových hodnot. Nutno připomenout, že evropské země jsou značně heterogenní a nelze vyvodit jednotnou univerzální politiku pro všechny země.

Článek je především analyticky orientován a neposkytuje návod pro hospodářskou politiku, avšak mohl by sloužit jako jeden z podkladů k její formulaci, jelikož poskytuje informace o dalších ukazatelích, které mají vliv na indikátory strategie Evropa 2020 a současně by mohl přispět pro revizi stanovených cílů a cest k jejich dosažení v rámci jednotlivých iniciativ, neboť na reálných datech zemí EU kvantifikuje regresní závislosti indikátorů Evropa 2020 na ukazatelích (obsažených v databázi Eurostatu) z různých oblastí ekonomie, financí, vědecko-technického rozvoje, vzdělání, sociální oblasti apod. Z výše uvedených důvodů izolovanosti jednotlivých iniciativ, učinit nějaký souhrnný sjednocující závěr z analýzy nebylo možné. V některých oblastech provedená analýza také upozornila, že některé obecné teoretické předpoklady o vazbách ukazatelů nejsou pro soubor zemí EU obecně platné.

**Seznam zemí**

1. Belgie	10. Francie	19. Rakousko
2. Bulharsko	11. Itálie	20. Polsko
3. Česká republika	12. Kypr	21. Portugalsko
4. Dánsko	13. Lotyšsko	22. Rumunsko
5. Německo	14. Litva	23. Slovinsko
6. Estonsko	15. Lucembursko	24. Slovensko
7. Irsko	16. Maďarsko	25. Finsko
8. Řecko	17. Malta	26. Švédsko
9. Španělsko	18. Nizozemsko	27. Velká Británie

**Seznam ukazatelů**

ER	Míra zaměstnanosti, věková skupina 20–64.
GERD	Hrubé domácí výdaje na vědu a výzkum v % HDP.
GGE	Snížení emisí CO <sub>2</sub> , základ = 1990.
PRPSE	Počet osob ohrožených chudobou či sociálním vyloučením.
SRE	Podíl obnovitelných zdrojů na konečné energetické spotřebě.
TEA	Terciární vzdělání, věková skupina 30–34.
EI	Energetická účinnost v mil. tun ropného ekvivalentu.
ELE	Předčasné ukončování školní docházky.
CPL	Komparativní cenové hladina.
ED	Energetická závislost.
EGRS	Elektrina generovaná z obnovitelných zdrojů energie.
GDP.gr	Meziroční růst hrubého domácího produktu.
GDPc	HDP na obyvatele v PPS.
GEE	Veřejné výdaje na vzdělávání v % HDP.
GESP	Veřejné výdaje na sociální ochranu v % HDP.
GGD	Hrubý veřejný dluh.
HBA	Počet domácností s širokopásmovým připojením.
HICP	Harmonizovaný index spotřebitelských cen.
ILCS	Individuální úroveň vzdělání v oblasti počítačů.
IRUI	Osoby pravidelně užívající internet.
LLL	Participace ve vzdělávání, věková skupina 25–64.
LPH	Produktivita práce na hodinu.
LPP	Produktivita práce na zaměstnanou osobu.
LTU	Dlouhodobá nezaměstnanost.
NNI	Čistý národní důchod.
PUSE	Osoby s vyšším sekundárním nebo terciárním vzděláním, věková skupina 25–64.
RULC	Růst skutečných nákladů práce.

## Literatura

- ANTOCH, J.; VORLÍČKOVÁ, D. *Vybrané metody statistické analýzy dat*. Praha : Academia, 1992. ISBN 80-200-0204-9.
- BARBIER, J. C. *Changes in political discourse from the Lisbon strategy to Europe 2020: tracing the fate of social policy*. ETUI, 2011.
- BÖHRINGER, Ch.; RUTHERFORD, T. F.; TOL, R. S. J. The EU 20/20/2020 targets: An overview of the EMF22 assessment. *Energy Economics*. 2009, vol. 31, s. 268–273.
- BONGARDT, A.; TORRES, F. Europe 2020–A Promising Strategy? *Intereconomics*. 2010, vol. 3, s. 136.
- ÇOLAK, M. S.; EGE, A. An Assessment of EU 2020 Strategy: Too Far to Reach? *Socil Indicators Research*. 2011, s. 1–22.
- COLSON, P. The Lisbon Strategy, and now Europe 2020: Are these Strategies Relevant?
- ERIXON, F. The Europe 2020 Strategy: time for Europe to think again. *European View*. 2010, vol. 9, no. 1, s. 29–37.
- EUROPEAN COMMISSION. An Agenda for new skills and jobs: A European contribution towards full employment. COM (2010) 682. 2010a.
- EUROPEAN COMMISSION. Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth. Communication from the Commission: Publications Office. 2010b.
- EUROPEAN COMMISSION. The European Platform against Poverty and Social Exclusion: A European framework for social and territorial cohesion. Brussels : European Commission, 2010c.
- FOX, J. *Bootstrapping regression models. An R and S-PLUS Companion to Applied Regression: A Web Appendix to the Book*. Thousand Oaks, CA : Sage, 2002. <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-bootstrapping.pdf>.
- FRAZER, H.; MARLIER, E.; NICAISE, I. *A social inclusion roadmap for Europe 2020*. Garant, 2010.
- HADI, A.; SIMONOFF, S. Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*. 1993, vol. 88, no. 424, s. 1264–1272.
- HEBÁK, P.; HUSTOPECKÝ, J.; MALÁ, I. *Vícerozměrné statistické metody (2)*. Praha : Informatorium, 2006. 240 s. ISBN 80-7333-036-9.
- HUŠEK, R. 2007. *Ekonometrická analýza*. Praha : Ekopress, 1999. ISBN 80-86119-19-X.
- CHEN, C. Robust regression and outlier detection with the ROBUSTREG procedure. Presented at Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference 2002. SUGI Paper 265-27. SAS Institute, Inc., Cary, NC.
- KLESSMANN, C. The evolution of flexibility mechanisms for achieving European renewable energy targets 2020 – ex-ante evaluation of the principle mechanisms. *Energy Policy*. 2009, vol. 37, no. 11, s. 4966–4979.
- LESCHKE, J.; THEODOROPOULOU, S.; WATT, A. How do economic governance reforms and austerity measures affect inclusive growth as formulated in the Europe 2020 Strategy? *A triumph of failed ideas: European models of capitalism in the crisis*. ETUI, 2012. 243 s.
- LUNDVALL, B.-Å.; LORENZ, E. From the Lisbon Strategy to Europe 2020. In MOREL, N.; PALIER, B.; PALME, J. (eds.). *Towards a social investment welfare state? Ideas, policies and challenges*. Chapter 13. Bristol : Policy Press, 2011, s. 333–351.
- MARLIER, E.; NATALI, D. (eds). *Europe 2020: Towards a More Social EU?* Peter Lang, 2010.
- MARTENS, W. Europe 2020 – A promising strategy? *Intereconomics*. 2010a, vol. 45, no. 3, s. 136–170.
- MARTENS, W. Europe 2020 and Beyond. *European View*. 2010b, vol. 9, no. 1, s. 1–3.
- MARX, I.; VANDENBROUCKE, P.; VERBIST, G. *Can higher employment levels bring lower poverty in the EU? Regression based simulations of the Europe 2020 target*. Discussion Paper series. Forschungsinstitut zur Zukunft der Arbeit, 2011.
- NATALI, D. The Lisbon Strategy, Europe 2020 and the crisis in between. *Europe 2020: Towards a More Social EU?* Peter Lang, 2010, s. 69–93.
- NOLAN, B.; WHELAN, Ch. T. The EU 2020 poverty target. UCD Geary institute discussion paper series. Dublin : Geary Institute, 2011.

- OLIVE, D. J. Applications of robust distances for regression. *Technometrics*. 2002, vol. 44, no. 1, s. 64–71.
- PAPADAKI GIOTI, O. European Environmental Policy and The Strategy 'Europe 2020'. *Regional Science Inquiry*. 2012, vol. 4, no. 1, s. 151–158.
- RAO, C.; TOUTENBURG, H. Linear models: least squares and alternatives. In *Springer series in statistics*. New York : Springer Verlag, 1995.
- ROTH, F.; THUM, A.-E. The Key Role of Education in the Europe 2020 Strategy. CEPS Working Document. 2010, 338.
- ROUSSEEUW, P. J. Least median of squares regression. *Journal of the American Statistical Association*. 1984, vol. 79, no. 388, s. 871–880.
- ROUSSEEUW, P. J.; LEROY, A. M. *Robust Regression and Outlier Detection*. New Jersey : Wiley, 2003. ISBN 0-471-48855-0.
- ROUSSEEUW, P. J.; VAN ZOMEREN, B. C.. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*. 1990, vol. 85, no. 411, s. 633–639.
- RUPPERT, D.; CARROLL, R. J. Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*. 1980, vol. 75, no. 372, s. 828–838.
- S-PLUS. S-PLUS 6 Robust Library User's Guide.
- SAS. SAS 9.1. 3 Help and documentation.
- SAMARDŽIJA, V.; BUTKOVIĆ, H. (eds.). *From the Lisbon Strategy to Europe 2020*. Institute for International Relations, 2010.
- SORIANO, F. H.; MULATERO, F. Knowledge policy in the EU: From the Lisbon strategy to Europe 2020. *Journal of the Knowledge Economy*. 2010, vol. 1, no. 4, s. 289–302.
- STAGNARO, C. Europe 2020: an Alternative Proposal. Update 2009.
- STUBBS, P.; ZRINŠČAK, S. Social Protection and Social Inclusion from Lisbon to Europe 2020. *From the Lisbon Strategy to Europe 2020*.
- SWALLOW, W. H.; KIANIFARD, F. Using robust scale estimates in detecting multiple outliers in linear regression. *Biometrics*. 1996, vol. 52, s. 545–556.
- TIŠMA, S.; ČERMAK, H. Environmental Dimension of the Lisbon Strategy and Europe 2020. *From the Lisbon Strategy to Europe 2020*, 7.
- YOHAI, V. J. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*. 1987, vol. 15, no. 20, s. 642–656.

## ANALYSIS OF EUROPE 2020 INDICATORS USING REGRESSION ANALYSIS

**Abstract:** Europe 2020 is a set of eight indicators used by the European Commission for monitoring headline targets of the Strategy for Smart, Sustainable and Inclusive Growth, which is considered to be the successor of the Lisbon Strategy. Values of these indicators vary among the European countries. Because some countries can be identified as outliers, robust regression as an acceptable analytic tool was applied. The aim of the paper is to construct relevant regression models for each Europe 2020 indicator as a dependent variable. The targets of the Europe 2020 indicators can be achieved by targeting some specific economic, social and environmental indicators.

**Keywords:** Europe 2020 indicators, regression, LS regression, robust regression, outliers

**JEL Classification:** O57, O10, C490