

Vysoká škola ekonomická v Praze

Národohospodářská fakulta

Obor: Ekonomie



VYPLATÍ SE MÍT SOUROZENCE?
VLIV POČTU SOUROZENCŮ A JEJICH POŘADÍ
NA STUDIJNÍ VÝSLEDKY V ČR.

bakalářská práce

Autorka: Ilona Němečková

Vedoucí práce: Ing. Marek Hudík, Ph.D.

Rok: 2012

Prohlašuji na svou čest, že jsem diplomovou práci vypracovala samostatně a s použitím uvedené literatury.

Ilona Němečková

V Praze, dne 19. 8. 2012

Poděkování:

Ráda bych poděkovala vedoucímu této práce Ing. Markovi Hudíkovi, Ph.D. za cenné připomínky k této práci. Dále chci poděkovat panu Ladislavu Tomáškoví a paní RNDr. Janě Palečkové z České školní inspekce za ochotu a poskytnutí cenných informací ohledně datasetu a mým blízkým za trpělivost po dobu psaní této práce.

Abstrakt

Práce zkoumá vliv počtu sourozenců a jejich pořadí narození na studijní výsledky v České republice, vyjádřené výsledkem čtenářské gramotnosti a průměru známek. Za použití metody nejmenších vážených čtverců a datasetu z průzkumu PISA 2009, byl zjištěn negativní vliv rostoucího počtu sourozenců, který však hraje roli až v rodinách s čtyřmi a více dětmi. Dále se ukázalo, že je lepší být ze sourozenců relativně starší, avšak tento efekt se prokázal pouze u čtenářské gramotnosti. Významně se také prokázaly ostatní proměnné, týkající se socio-ekonomického statusu.

Abstract

This work analysis the influence of a number of siblings and their birth order on educational attainment in Czech Republic expressed by reading literacy and average grade. Using the weighted least squares method and data-set based on survey PISA 2009, was found a negative sibling size effect. Nevertheless this effect became significant in families with four or more children. Furthermore, it seems to be better a relatively older, but this effect was significant only in the case of reading literacy. Other variables concerning socio-economic status were found significant too.

Klíčová slova

počet sourozenců, pořadí narození, vzdělání, regresní analýza

Key words

number of siblings, birth order, education, regression analysis

JEL klasifikace/ JEL Classification

D13, I21, J13

Obsah

Úvod	1
1. Souhrn literatury	2
2. Data	5
3. Model	5
4. Proměnné	8
4.1. Závislé proměnné	8
4.1.1. Čtenářská gramotnost (CTENGRAM).....	8
4.1.2. Průměr (PRUM12)	9
4.2. Vysvětlující proměnné	10
4.2.1. Počet sourozenců.....	10
4.2.2. Pořadí sourozenců	11
4.2.3. Počet knih.....	11
4.2.4. Domácí vzdělávací zdroje (HEDRES).....	12
4.2.5. Kulturní majetek (CULTPOSS).....	12
4.2.6. Bohatství (WEALTH).....	12
4.2.7. Nejvyšší pracovní postavení (HISEI).....	12
4.2.8. Pohlaví	12
4.2.9. Neúplná rodina (NEUPLNAROD)	13
4.3. Nezahrnuté proměnné.....	13
4.4. Popisná statistika zahrnutých proměnných	14
5. Regresní analýza	17
5.1. Model počtu sourozenců	18
5.2. Model pořadí narození.....	22
5.3. Diskuze výsledků	25
Závěr	27
Použitá literatura	28

Úvod

Nejen ekonomové zkoumají jaké faktory ovlivňují jedincův úspěch, měřený například výdělkem či dosaženým vzděláním. Kromě vzdělání rodičů či například socioekonomického statusu, může mít na jedincův úspěch vliv i s jakými sourozenci vyrůstá.

Zatímco v zahraniční literatuře je možno nalézt mnoho empirických prací zabývajících se vlivem velikosti rodiny a složením sourozenců nejen na studijní úspěch či výdělek, v České republice jsou práce na toto téma výjimkou. Přitom mohou poskytovat nejen empirické informace pro ekonomii, ale zároveň mohou být zajímavou informací pro tvůrce školní a sociální politiky, především pak v oblasti různých sociálních programů či podpor.

Cílem této práce je pomocí regresní analýzy odhadnout, zda a jakým způsobem má počet sourozenců a jejich pořadí vliv na studijní výsledky v České republice. Mimo vliv sourozenců jsou zkoumány vlivy i dalších proměnných, týkající se například prostředí, v kterém žáci žijí. Bylo prokázáno, že existuje negativní vliv rostoucího počtu sourozenců a že pořadí narození „hraje roli“ alespoň ve výsledcích čtenářské gramotnosti.

Tato práce je členěna následovně: v první kapitole je představena odborná literatura, věnující se tématu; datový soubor, který je použit pro analýzu, je pak popsán v kapitole druhé. Třetí kapitola se zabývá konstrukcí modelu, čtvrtá pak zkoumá jednotlivé proměnné zahrnuté do modelu, důvody jejich zahrnutí a jejich možný vliv. Samotná regresní analýza, její výsledky a diskuze těchto výsledků, jsou zahrnuty v páté kapitole a poslední kapitola, závěr, pak shrnuje poznatky celé práce.

1. Souhrn literatury

Becker v roce 1960 (dále 1974) formuloval teorii „kvantity a kvality“, ve které popisuje, jak se rodiče rozhodují o tom kolik dětí a jak „kvalitních“ vyprodukují; kdy kvalita odráží investice (náklady) do každého dítěte. Z této teorie můžeme předpokládat, že existuje negativní vztah mezi počtem dětí a investicemi do nich vložených, například prostřednictvím vzdělání.

Becker a Tomes (1979,1986) také poukazují na to, že rodiče nemusí investovat do každého dítěte stejné množství zdrojů do lidského kapitálu; to souvisí s nadáním dětí, které může být různé u každého dítěte; rodiče mohou kompenzovat méně nadané dítě (tzv. „efekt bohatství“) nebo mohou podporovat ty děti, které jsou více nadané a zvýšit tak jejich potencionální výnosy („cenový efekt“, „maximalizační efekt“). Podpora více nadaných dětí (cenový efekt) pak, dle Beckera s Tomesem, převládá a zpravidla probíhá skrze zdroje do lidského kapitálu, zatímco méně nadané děti rodiče spíše kompenzují jinými dary (či časem a zdroji, které primárně nevedou ke zvýšení lidského kapitálu). Cenový efekt pak převládá i u chudších rodičů, kteří mohou předpokládat, že se více nadané děti budou starat o své sourozence.

Významný negativní efekt počtu sourozenců na počet let vzdělání našli ve Spojených státech Behrman a Taubman (1986). Nalezli také významný pozitivní efekt pro prvorozené, avšak pouze na počet let vzdělání, na výdělky při započtení vlivu rodiny nemělo pořadí statisticky významný vliv.

Behrman a Taubman také uvádí několik důvodů, jimiž mohou být způsobeny efekty pořadí. Později narození se zpravidla rodí starším rodičům a se zvyšujícím se věkem matky zpravidla narůstá šance různých defektů při narození. Dvojčata si nekonkurují jenom po porodu, ale už i v děloze, což zpravidla způsobuje nižší porodní váhu a opět zvyšuje pravděpodobnost různých vrozených onemocnění. Zároveň narození dvojčat může pro rodinu znamenat zvýšení její velikosti nad očekávanou a rodiče tak musí snížit své výdaje na každého potomka.

Dalším důvodem, proč očekávat dle Behrmana a Taubmana efekt pořadí, je, že mladší děti ve velkých rodinách mohou mít nevýhodu v tom, že jejich starší sourozenci vyčerpají rodinné zdroje a úspory v době, kdy se rozhodují mezi dalším studiem a prací a budou tak nuceni zvolit práci. Na druhou stranu mohou této volby dosáhnout v době, kdy jsou jejich rodiče blízko svého mzdového

profilu. To souvisí nejen s rozestupem mezi sourozenci, ale také s věkem rodičů. Mladší sourozenci také mohou těžit z toho, že rodiče jsou více vyspělí, naučili se „rodičovství“ na starších dětech a mohou být zkušenější nejen co se týče vzdělávání svých potomků.

Blakeová (1981) uvádí další důvody, proč mohou profitovat více některá pořadí narození. Z výše nejmenovaných je to například ten, že nejstarší sourozenec může být v některých případech nucen opustit školu dříve, aby pomohl mladším sourozencům. Starší sourozenci také mohou těžit z „učení“ svých mladších sourozenců a tím se vzdělávat a zdokonalovat. Nevýhodu mají zase naopak v tom, že neexistuje nikdo starší, kdo by je „učil“ a vzdělával. Blakeová však ve své práci zdůrazňuje, že žádný efekt pořadí sourozenců neexistuje; důkazy v empirické literatuře jsou prý jen důsledkem vychýleného vzorku pozorování.

Hanushek (1992) také zkoumal vliv pořadí narození ve Spojených státech, avšak v jeho práci se neprokázal být významným. Dle něj je výhoda být prvorozeným především v tom, že prvorození mají větší pravděpodobnost být v malé rodině, což znamená výhodu samo o sobě. Oproti Beckerovi s Tomasem (1979, 1986), kteří předpokládají, že se rodiče nejčastěji snaží o maximalizaci úspěchu potomků, však maximalizační efekt ve své práci nenašel, dle jeho výsledků rodiče spíše kompenzují méně nadané děti (tím že s nimi tráví více soukromého času, tedy toho, jež může být stráven pouze s jedním dítětem, než s více nadanými sourozenci) nebo vykazují tzv. nediskriminační alokaci času (soukromý čas rozdělují mezi děti rovnoměrně). Je však nutno poznamenat, že Hanushek zkoumal efekty na černých Američanech z lowy, kteří pocházeli především z nízkopříjmových skupin.

Kromě Hanusheka našli negativní efekt počtu sourozenců také Wolter a Vellacott (2002). Ti zkoumali vliv sourozenců ve Švýcarsku na čtenářské gramotnosti (výzkum PISA). Negativní efekt našli především u rodin s nižším socioekonomickým statusem a zahraničním původem. Naopak pro rodiny s „domácím“ původem a vyšším socioekonomickým statusem, které jsou co do velikosti malé (méně než 4 děti), našli pozitivní efekt. Nejlepších výsledků při porovnávání efektu pořadí narození dosáhli nejstarší sourozenci, a to i tehdy, když byl efekt zkoumán pro danou velikost rodiny. Wolter s Vellacotem také našli významný negativní efekt „jedináčka“. Jedináčci, dle jejich výsledků, prospívají hůře než děti v menších rodinách (méně než 4 děti). Poukazují také na to, že boj o rodinné zdroje může být nejsilnější u prostředních sourozenců (tento

efekt bývá často nazýván efektem „prostředního dítěte“); zatímco nejstarší děti mohou těžit z periody, kdy jsou jedináčci a nemladší z periody, kdy jsou starší sourozenci již dospělí, „prostřední“ děti žádnou takovou periodu nezažívají. Tento efekt však také závisí na rozestupu mezi sourozenci, o němž však autoři vzhledem k datům neměli informace.

K podobným výsledkům došel také Black et al. (2005) a Booth a Kee (2009), kteří také zkoumali vliv počtu sourozenců a pořadí narození. V obou pracích se prokázal negativní efekt počtu sourozenců, po zahrnutí počtu pořadí se však velikost rodiny neproказuje významně; oba také našli negativní efekt pro později narozené. Black et al. (2011) také zkoumali vliv pořadí narození na IQ. Jejich výsledky ukazují, že později narození dosahují nižšího průměrného IQ; rozdíl mezi prvorozeným a druhorozeným činí dokonce 3 body na stupnici IQ.

Efekt pořadí sourozenců byl zkoumán například i na vyživovacím statusu dětí (Horton, 1988). Ta ve své práci také uvádí několik důvodů, které mohou způsobovat rozdíly mezi různě starými sourozenci. Mohou to být kulturní důvody, díky kterým rodiče upřednostňují různá pořadí či mnoho socioekonomických důvodů (pokud se chtějí rodiče zajistit v pozdějším věku, mohou favorizovat ty děti, které se stanou dříve ekonomicky nezávislými).

Hyunjoon (2005) pak trochu z jiného pohledu poukazuje, že efekt počtu sourozenců souvisí s politikou v dané zemi. Ve své srovnávací studii našel méně negativní efekt pro země se silnou veřejnou politikou, definované jako sociálně demokratické režimy.

Pokud můžeme shrnout výsledky studií zabývajících se efekty pořadí sourozenců, tak i přesto, že nalezneme argumenty zvýhodňující jak nejstarší, tak nejmladší sourozence, většina empirických studií našla pozitivní efekt pro prvorozené; výjimku tvořil již zmíněný Hanushek (1992), který právě poukazuje na tyto výsledky s tím, že prvorození mají větší pravděpodobnost býti v menší rodině. Počet sourozenců se pak především v pozdější době a ve vyspělých zemích začíná stávat v některých studiích nevýznamným.

Přesto, že téma vlivu složení sourozenců na studijní úspěch je poměrně častým tématem zahraničních vědeckých prací, nepodařilo se mi najít žádnou empirickou práci, která by zkoumala vliv počtu sourozenců a jejich pořadí na studijní výsledky pro Českou republiku Podobně jako u

Volterra s Vellacottem (2002) bude tento vliv zkoumán na datech z průzkumu PISA, nicméně na nejnovějším dostupném datasetu z roku 2009.

2. Data

V této práci jsou použity data z průzkumu OECD, zvaného PISA 2009 (Programme for International Student Assessment). Tento projekt již od roku 2000 opakovaně srovnává úspěchy 15-ti letých žáků v čtenářské, matematické a přírodovědné gramotnosti napříč různými zeměmi. Kromě testování jednotlivých gramotností sleduje výzkum i charakteristiky týkající se škol, především však studentů, jejich rodinného zázemí či přístupu ke studiu.

V roce 2009 se průzkumu zúčastnilo 65 zemí; v České republice pak 290 škol a k testování bylo vybráno 7 500 žáků (Palečková a kol., 2010). Celkový dataset pak zahrnuje pouze 6064 pozorování a to především z důvodu toho, že se někteří žáci výzkumu nezúčastnili či jejich hodnoty byly jiným způsobem neplatné.

Čeští žáci dosáhli v roce 2009 průměrného skóre z čtenářské gramotnosti 478 bodů, čímž se nacházejí pod průměrem zemí OECD, který činil 493 bodů. (OECD, 2010)

Dataset pro Českou republiku je volně dostupný na stránkách bývalého Ústavu pro informace ve vzdělávání (PISA, 2009), společně s dotazníky, které žáci vyplňovaly. V této práci používáme žákovský dataset, který obsahuje informace o dosaženém skóre, o otázkách vyplněných v žákovském dotazníku a další proměnné, především indexy vypočítané dle metodiky OECD.

3. Model

V této práci budou odhadovány dva modely, které jsou částečně inspirovány modelem Wolterra a Vellacotta (2002). Narozdíl od jejich modelu, kde zahrnuli proměnné pro počet sourozenců a dummy proměnné vyjadřující zda je žák nejmladším či prostředním sourozencem, je v této práci odhad rozložen na modely dva.

První model bude testovat hypotézu o vlivu počtu sourozenců na studijní výsledky, respektive zda rostoucí počet sourozenců ovlivňuje negativně tyto výsledky. Jelikož ale může být tento vliv různý (jak ukazuje např. právě Wolter s Vellacottem, 2002), pro různý počet sourozenců, bude zařazena dummy proměnná pro jednotlivé počty sourozenců (více v kapitole 4.2).

Změna oproti modelu Woltera a Vellacotta je především v druhém modelu, který zahrnuje proměnné počtu mladších, starších a stejně starých sourozenců, který se bude snažit otestovat, zda má na studijní výsledky vliv i pořadí, v jakém se sourozenci narodili. Jak již bylo řečeno, Wolter a Vellacott (2002) zahrnují v jednom modelu kromě počtu sourozenců i dummy proměnné určující nejmladšího, prostředního a nejstaršího sourozence a jsou tak schopni přímo říci, zda je “lepší býti” nejmladším, prostředním či nejstarším sourozencem. Nastává však problém s tím, že pokud je jedinec prostředním, musí mít zákonitě alespoň dva sourozence. Pokud zahrneme do modelu počet mladších, starších a stejně starých sourozenců, budeme schopni dle koeficientů říci, jestli se žákovi zhorší¹ výsledek testu více pokud mu “přibude” mladší nebo starší sourozenec. Dalším rozdílem oproti Wolterovi s Vellacottem (2002) je nezahrnutí některých proměnných, konkrétně proměnných vyjadřujících čas strávený žákem s rodiči. Tyto data nemáme k dispozici z toho důvodu, že žakovské dotazníky obsahují národní část, jejíž obsah si volí každá země sama.

Je třeba poznamenat, že nejsme schopni v naší práci přesně říci proč a z jakého důvodu se efekt pořadí projeví (či neprojeví). Jak již naznačila první kapitola, existuje mnoho důvodů pro existenci tohoto efektu, ale proto, abychom stanovili důvod přesně, nám chybí určitě ještě mnoho informací (ať už se jedná o věk rodičů, rozestup mezi sourozenci a další). To abychom byli schopni říci, jakým způsobem rozdělují rodiče zdroje mezi potomky (zda kompenzují, podporují nadané, rozdělují zdroje rovnoměrně či preferují určité pořadí bez ohledu na vrozené schopnosti), záleží zajisté i na tom, zda určitá pořadí prosperují například z biologického hlediska; rozdíl mezi sourozenci pak může být způsoben i jen vlivem samotných sourozenců, kteří si navzájem znalosti předávají. (viz například Blakeová, 1981)

Jako měřítko studijních výsledků (úspěchu) bude pak do modelu zahrnuta, jako u mnoha prací, čtenářská gramotnost (např. u Woltera a Vellacotta, 2002). Přestože je tato schopnost do jisté míry

¹ Pokud se potvrdí vliv počtu sourozenců v prvním modelu, je dle teorie i předchozích studií pravděpodobné, že bude negativní, proto předpokládám, že i u těchto proměnných bude v tom případě mínusové znaménko.

vrozená, rodiče a čas s nimi strávený (hraje zcela jistě roli právě ve čtení, kdy mohou tuto schopnost ovlivnit v žákově dětství (například čtením spolu s dítětem v raném dětství apod.). Oba modely budu odhadovat i na průměru známek, aby se prokázalo, zda i na známky, další ukazatel studijních výsledků, působí efekt stejně. Obecně můžeme model, který vyjadřuje vztah závislosti mezi počtem sourozenců a studijními výsledky, vyjádřit jako:

$$\text{CTENGRAM}(\text{PRUM12}) = f(\text{POCETSOUROZENCU}, \text{DEVCE}, \text{HEDRES}, \text{CULTPOSS}, \text{WEALTH}, \text{HISEI}, \text{NEUPLNAROD}, \text{POCETKNIH})$$

Po zahrnutí vlivu pořadí sourozenců:

$$\text{CTENGRAM}(\text{PRUM12}) = f(\text{PORADISOUROZENCU}, \text{DEVCE}, \text{HEDRES}, \text{CULTPOSS}, \text{WEALTH}, \text{HISEI}, \text{NEUPLNAROD}, \text{POCET KNIH})$$

V následující tabulce je pak stručný přehled toho, co každá proměnná představuje. Podrobný popis konstrukce proměnných, důvody jejich zařazení do modelu a jejich možný vliv na vysvětlované proměnné je pak shrnut v kapitole následující, stejně jako důvody nezařazení ostatních proměnných.

Tabulka 1: Přehled proměnných

Proměnná	Vysvětlení
CTENGRAM	skóre čtenářské gramotnosti
PRUM12	průměr známek za poslední pololetí
POCETSOUROZENCŮ	proměnné vyjadřující počet sourozenců
PORADISOUROZENCU	proměnné vyjadřující pořadí sourozenců
DEVCE	pohlaví
HEDRES	index vlastnictví vzdělávacích zdrojů
CULTPOSS	index vlastnictví kulturních zdrojů
WEALTH	index bohatství
HISEI	nejvyšší pracovní postavení matky nebo otce
NEUPLNAROD	proměnná pro žáky, kteří nežijí s dvěma
POCETKNIH	přibližný počet knih v domácnosti

V této práci tedy bude odhadován model dvakrát, poprvé jako “model počtu sourozenců” a podruhé jako “model pořadí sourozenců”. Každý model bude odhadován jak pro čtenářskou gramotnost, tak pro průměr známek; celkem tedy budou provedeny čtyři odhady.

4. Proměnné

Tato kapitola podrobněji rozebírá proměnné zařazené do našeho modelu. Nejdříve popisuje závislé proměnné, poté vysvětlující proměnné a v další subkapitole se věnuje popisu vynechaných proměnných. V poslední subkapitole jsou pak uvedeny popisné statistiky zahrnutých proměnných.

4.1. Závislé proměnné

4.1.1. Čtenářská gramotnost (CTENGRAM)

Výzkum PISA 2009 definuje čtenářskou gramotnost jako „schopnost porozumět psanému textu, zabývat se jím, přemýšlet o něm a používat ho k dosažení vlastních cílů, k rozvoji vlastních vědomostí a potenciálu a k aktivní účasti ve společnosti”. (Palečková a kol., 2010) Čtenářská gramotnost tak nejen odráží znalosti žáků v této disciplíně, ale především reflektuje, nakolik jsou žáci schopni s textem pracovat, což bývá často obsahem různých testů studijních předpokladů. Čtenářská gramotnost tak tedy může znázorňovat i schopnost dalšího studia.

Skóre čtenářské gramotnosti je ve výzkumu vyjádřeno pomocí pěti „pravděpodobných hodnot“, tzv. „plausible values“. Student může být v den testu ovlivněn mnoha faktory, které mohou zapříčinit, že dosažené skóre z testu gramotnosti nebude plně odrážet jeho skutečné schopnosti či znalosti. „Plausible values“ jsou pak zkonstruovány tak, že se z matematicky vypočteného rozdělení okolo dosaženého skóre přidělí každému žákovi pět náhodných čísel. (více OECD, 2009)

OECD (2009) doporučuje provádět všechny výpočty (resp. regrese) pětkrát, pro každou „plausible value“ (dále jen PV) zvlášť, a následné výsledky zprůměrovat (koeficienty, směrodatné chyby apod.), přestože významný rozdíl oproti použití jedné (či zprůměrované) PV nastává především při malém vzorku; u vzorku s vysokým počtem pozorování je rozdíl zanedbatelný. Zanedbatelný

rozdíl koeficientů, směrodatných odchylek a ostatních parametrů regrese prokázaly v našem případě i zkušební odhady, kdy bylo použito jak průměru PV, tak pěti regresí pro každou PV zvlášť. Proto budeme v této práci nadále pracovat se zprůměrovanou hodnotou PV, jelikož to naše výsledky a závěry nijak zásadně neovlivní a velmi to zjednoduší celý proces odhadu a především testování modelu.

4.1.2. Průměr (PRUM12)

Jako další vysvětlovaná proměnná je použit průměr ze známek žáka v prvním pololetí (resp. v posledním pololetí před testováním). Čtenářská gramotnost může odrážet jistou schopnost studovat, nicméně žák je ve škole hodnocen dle svých znalostí, které mohou být od gramotnosti rozdílné.

V databázi je sledováno hodnocení ze sedmi předmětů a to z českého jazyka, matematiky, cizího jazyka, fyziky, chemie, přírodopisu (biologie) a zeměpisu. Žáci vyplňují známku pouze z předmětu, který byl v daném pololetí na jejich škole vyučován (v ostatních případech vyplňují nulu).

Jelikož se jedná o známky za jedno pololetí a známka nemusí plně odrážet žákovy znalosti, je jako vysvětlovaná proměnná použit právě průměr z těchto známek. Vzhledem k chybějícím hodnotám je průměr spočten pouze ze dvou předmětů, a to z českého jazyka a matematiky.

U této proměnné však očekávám, že bude mít model menší vypovídací schopnost (bude vysvětlena menší část rozptylu této závislé proměnné), než v případě čtenářské gramotnosti. Čtenářská gramotnost je sice také testována v jednom okamžiku, ale domnívám se, že známku může ovlivňovat mnohem více proměnných, které nemáme k dispozici, a zároveň známka nemusí být objektivní; každý učitel zpravidla posuzuje každého žáka individuálně.

4.2. Vysvětlující proměnné

4.2.1. Počet sourozenců

V žákovském dotazníku žáci odpovídají na tři otázky týkající se počtu sourozenců a to kolik mají mladších, starších a stejně starých sourozenců, nezáleží však na tom, zda sourozenci žijí společně.

V modelu jsou pak zahrnuty dummy proměnné pro jednotlivé počty sourozenců a to:

SOUR1 – pro žáky s jedním sourozencem,

SOUR2 – pro žáky s dvěma sourozenci,

SOUR3 – pro žáky s třemi sourozenci a

SOUR4 – pro žáky s čtyřmi a více sourozenci.

Dummy proměnné používám v tomto případě nejen kvůli srozumitelnější interpretaci, ale také, aby byl ošetřen případný nelineární vztah mezi počtem sourozenců a čtenářskou gramotností (průměrem).

V tabulce 2 pak vidíme, kolik získali průměrně bodů z testu čtenářské gramotnosti žáci s jednotlivými počty sourozenců. Jak je vidět, nejlépe dopadli žáci s jedním sourozencem, jež mají v průměru 519 bodů, to je o více než čtrnáct bodů více než u jedináčků, což může naznačovat, že by se v regresi mohl objevit negativní efekt jedináčka, který našel Wolter a Vellacott (2002). Dále se projevuje již klesající skóre.

Tabulka 2: Průměrné skóre čtenářské gramotnosti dle počtu sourozenců, chybějící pozorování vynechána, průměr nevážených dat

Počet sourozenců	Průměrné skóre z čtenářské gramotnosti
0	504,8507
1	519,1013
2	504,5547
3	490,9858
4 a více	465,1518

4.2.2. Pořadí sourozenců

Jak již bylo řečeno v třetí kapitole, některé studie (např. Wolter a Vellacott, 2002) zahrnují do modelu pořadí sourozenců jednoduše jako dummy proměnnou pro to, zda je žák nejmladším či prostředním dítětem. S tím však nastává problém, jelikož prostřední dítě musí mít již minimálně dva sourozence.

V této práci jsou pro určení vlivu pořadí sourozenců zahrnuty tři proměnné a to:

počet mladších sourozenců **(POCETMLASICH),**
počet starších sourozenců **(POCETSTARSICH)** a
počet stejně starých sourozenců **(POCETSTEJNESTARYCH).**

Na základě koeficientu pak budeme schopni říci, zda je horší “přírůstek” mladších nebo starších sourozenců, jinými slovy zda je lepší být relativně starším nežli relativně mladším sourozencem. Proměnná nám tedy sice neříká nic o krajních sourozencích (nejmladší či nejstarší), nicméně i tak poskytuje poměrně dobrý ukazatel relativního pořadí sourozenců.

4.2.3. Počet knih

Na tuto otázku žáci odpovídali zaškrtnutím přibližného počtu knih v domácnosti. Do modelu jsou zahrnuty dummy proměnné vyjadřující daný rozsah knih v rodině žáka:

DPOCETKNIH_2 pro žáky, v jejichž domácnosti se nachází 11-25 knih,
DPOCETKNIH_3 pro žáky, v jejichž domácnosti se nachází 26-100 knih,
DPOCETKNIH_4 pro žáky, v jejichž domácnosti se nachází 101-200 knih,
DPOCETKNIH_5 pro žáky, v jejichž domácnosti se nachází 201-500 knih a konečně
DPOCETKNIH_6 pro žáky, v jejich domácnosti je 501 knih a více.

Tuto proměnnou zahrnují do svých modelů mnohé studie zabývající se vysvětlením determinant ovlivňujících vzdělání dětí, proto i zde očekáváme pozitivní vliv (čím více knih, tím lepší výsledky). Proměnná do určité míry vyjadřuje úroveň vzdělání rodičů, ale také dostupnost vzdělávacích zdrojů (což především pro čtenářskou gramotnost knihy jsou) v domácnosti žáka.

4.2.4. Domácí vzdělávací zdroje (HEDRES)

Index HEDRES (Home Educational Resources) je složen ze sedmi proměnných, které vyjadřují, zda má žák doma k dispozici věci potřebné ke studiu. Jedná se o psací stůl, tiché místo pro studium, počítač, který může využít ke studiu, výukové programy, knihy potřebné ke studiu a manuály či návody. (str. 288, OECD, 2012). Slouží tak jako jeden z dalších indikátorů bohatství rodiny a zároveň může vyjadřovat i přístup rodičů ke studiu svého dítěte. Dá se tedy předpokládat, že čím vyšší bude index, tím vyšších výsledků bude žák dosahovat.

4.2.5. Kulturní majetek (CULTPOSS)

Podobně jako proměnná HEDRES vyjadřuje index CULTPOSS kulturní vlastnictví rodiny, které je složeno z vlastnictví klasické literatury, básnických sbírek či uměleckých děl. (str. 288, OECD, 2012)

4.2.6. Bohatství (WEALTH)

Tato proměnná vyjadřuje vlastnictví statků, jež mají vyjadřovat status bohatství. Jedná se o vlastní pokoj, připojení k internetu, myčku nádobí, DVD přehrávač, vlastní laptop, videokameru a domácí kino. Dále množství mobilních telefonů, televizí, počítačů, aut a koupelen. (str. 288, OECD, 2012). Je tak další proměnnou vyjadřující socio-ekonomický status žáka.

4.2.7. Nejvyšší pracovní postavení (HISEI)

Zde je zařazen mezinárodní socioekonomický index ISEI (International Socio-Economic Index), který odráží vztah mezi vzděláním, pracovním postavením a příjmem, kdy čím vyšší má jedinec pracovní postavení, tím výše se nachází v žebříčku socioekonomického indexu. (více Ganzeboom et. al., 1992) Proměnná HISEI (Highest International Socio-Economic Index) pak udává nejvyšší hodnotu ISEI u matky či otce (dle toho, který z nich dosahuje vyšší hodnoty).

4.2.8. Pohlaví

Jak vyplývá z výsledků průzkumu, ve všech zúčastněných zemích si dívky v testech čtenářské gramotnosti vedou lépe. V zemích OECD dívky dosahují o 39 bodů lepšího skóre, v České

republice pak o 48 bodů více než chlapci. (Palečková a kol., 2010) Nezahrnutí této proměnné by znamenalo nezahrnutí pravděpodobně velmi významné proměnné, proto bude v modelu použita dummy proměnná pro dívku (**DEVCE**) a je očekáván významný, kladný koeficient.

4.2.9. Neúplná rodina (NEUPLNAROD)

Tuto proměnou zahrnuji především ze dvou důvodů: Za prvé, pokud dítě vyrůstá s jedním či žádným rodičem, je zde větší pravděpodobnost, že jeho socioekonomický status bude nižší. A za druhé můžeme předpokládat, že dítě vyrůstající pouze s jedním rodičem je ochuzeno o čas, který by mu mohl věnovat rodič druhý; nemluvě o situaci, kdy potomek vyrůstá zcela bez rodičů.

Lze totiž předpokládat, že existuje pozitivní vztah mezi časem stráveným s rodiči a studijními výsledky. Čím více času tráví rodič s dětmi, tím větší lze očekávat pravděpodobnost, že tráví více času i nad učením s potomkem a dalšími činnostmi ovlivňující jeho prospěch či gramotnost.

Proměnná je do modelu zařazena jako nula-jednotková proměnná, kdy hodnota proměnné jedna značí, že žák nežije s dvěma rodiči. V průzkumu se pak nerozlišuje mezi vlastními či nevlastními rodiči, důležité je, zda rodiče žijí s potomkem, což se zdá i pro tuto práci významnějším.

4.3. Nezahrnuté proměnné

Tato subkapitola se zabývá proměnnými, jež nejsou v modelu z datasetu zařazeny či je nemáme vůbec k dispozici.

Dataset obsahuje poměrně velké množství proměnných zabývajících se vztahem žáka ke škole či k učiteli, jeho přístupem ke studiu či tím, jak naopak vnímá přístup učitele k žákům. Tyto proměnné nezařazuji především z toho důvodu, že subjektivní hodnocení žáka může být značně zkresleno právě díky jeho schopnostem. To také může platit i o samotném učiteli, který může k žákům přistupovat právě dle jeho výkonů.

Z jiného důvodu pak v modelu není zahrnuta proměnná označující imigranty či proměnná označující žáky, jež doma mluví jiným, než českým jazykem. V celém datasetu je pouze 96 žáků, jež pocházejí z druhé generace imigrantů a 64 z první generace. Žáků, jež doma hovoří jiným, než českým jazykem je pouze 100, z tohoto důvodu do modelu nezahrnu ani jednu z těchto dvou

proměnných. Je však zřejmé, že tato proměnná může významně ovlivňovat nejen dosažené skóre z čtenářské gramotnosti, ale také samotný vliv počtu sourozenců (či jejich pořadí), z důvodu různých kulturních a rodinných zvyklostí.

Jednou z dalších proměnných, nezahrnutých do modelu, je proměnná týkající se typu školy. Přestože žáci z různých škol dosahují různých hodnot jak ve výsledcích čtenářské gramotnosti v průměrné známce, není možno určit, zda má žák dobré výsledky, jelikož je na gymnáziu, či zda je na gymnáziu, jelikož má dobré výsledky (respektive je schopen dobrých výsledků dosáhnout).

Pro testování toho, zda čím více dětí rodiče mají, tím méně času mohou věnovat každému z nich a zda svůj čas mezi potomky rozdělují rovnoměrně by bylo zajisté vhodné zahrnutí proměnných, jež jakýmsi způsobem vyjadřují zájem rodičů o vzdělání dětí a především proměnných, jež vyjadřují čas strávený rodiči s daným dítětem a tuto hypotézu otestovat. Tyto data však bohužel nemáme k dispozici.

Úspěch žáka ve škole (a tím částečně i v čtenářských testech) může být ovlivněn i dalšími faktory, jako je úspěch ostatních spolužáků, v rámci efektu sourozenců i úspěch sourozenců. Je možno předpokládat, že sourozenec (především mladší) bude za jinak nezměněných okolností mít lepší prospěch v předmětech, ve kterých prospívá jeho sourozenec (především starší). Avšak ani tyto proměnné nejsou v datasetu k dispozici.

Je nutno poznamenat že významnou roli mohou hrát i další faktory, které nejsou v datasetu či dokonce nejsou měřitelné a pozorovatelné. Ať už se jedná o různé vrozené vady týkající se poruchy čtení (jako je například dysgrafie), ale také o samotné “nadání” žáka, vrozenou inteligenci a ostatní nepozorovatelné faktory, které mohou ovlivňovat to, že konkrétní dítě dosahuje lepších výsledků v dané oblasti než jiné, *ceteris paribus*. Z těchto a výše uvedených důvodů lze očekávat, že náš model vysvětlí pouze menší část z vysvětlované proměnné.

4.4. Popisná statistika zahrnutých proměnných

V tabulce 2 jsou uvedeny popisné statistiky použitých proměnných a to počet platných pozorování (tedy počet pozorování snížený o pozorování chybějící), průměr, směrodatná odchylka, minimum, medián a maximum.

Nejvíce chybějících pozorování je u dummy proměnných vyjadřujících počet sourozenců (SOUR1-4), bez chybějících hodnot je pak proměnná DEVCE a CTENGRAM.

Průměrný počet bodů v testu čtenářské gramotnosti činí 496.11 bodů, což je rozdíl -18.11 bodu oproti oficiálně prezentovaným výsledkům. Tento rozdíl však může být způsoben především tím, že popisná statistika obsahuje nevážená data (více o váhách datasetu v kapitole 5) a zároveň odlišným způsobem výpočtu, jak již bylo vysvětleno v subkapitole 4.1. Nejmenší dosažená hodnota z testu je pak 172.61 bodu a největší 768.64.

Průměrná známka z matematiky a z českého jazyka je 2,6212, nejčastější hodnota (medián), je pak přesně 2,5.

Překvapivě působí maximální hodnoty u počtu mladších, starších a stejně starých sourozenců; především pak hodnota 10 u proměnné stejně starých sourozenců působí velmi neobvykle, vzhledem k tomu, že by žák měl pocházet z jedenácti stejně starých sourozenců. Celkem se hodnota deset u této proměnné objevuje dvakrát a dvakrát také hodnota 9. Je pravděpodobné, že se jedná o chybu v datech. Přestože by vzhledem k počtu pozorování neměla výrazně ovlivnit výsledky našeho zkoumání, je možné že může tato „chyba“ zkreslit koeficienty u této proměnné, proto v modelu pořadí sourozenců budou tyto čtyři hodnoty vynechány.

Tabulka 3: Souhrné statistiky proměnných modelu (nevážená data)

Proměnná	Platných pozorování	Průměr	Sm. odch.	Min	Medián	Max
CTENGRAM	6064	496.11	94.541	172.61	501.49	768.64
PRUM12	5503	2.6212	0.89849	1	2.5	5
CULTPOSS	5999	-0.3208	0.97203	-1.813	-0.0649	0.888
DEVCE	6064	0.4863	0.49985	0	0	1
DPOCKETNIH_1	6015	0.0786	0.26919	0	0	1
DPOCKETNIH_2	6015	0.1244	0.33001	0	0	1
DPOCKETNIH_3	6015	0.3262	0.46885	0	0	1
DPOCKETNIH_4	6015	0.2101	0.40744	0	0	1
DPOCKETNIH_5	6015	0.1706	0.37617	0	0	1
DPOCKETNIH_6	6015	0.0901	0.28636	0	0	1
HEDRES	6037	0.1067	0.81905	-4.053	0.7785	0.819
HISEI	5880	50.052	14.293	16	51	90
NEUPLNAROD	5777	0.1819	0.38582	0	0	1
POCETMLADSICH	5543	0.7765	0.98802	0	1	14
POCETSTARSICH	5532	0.7912	0.97999	0	1	14
POCETSTEJNESTAR	5112	0.0362	0.33346	0	0	10
SOUR1	4963	0.5547	0.49705	0	1	1
SOUR2	4963	0.2071	0.40529	0	0	1
SOUR3	4963	0.0737	0.26138	0	0	1
SOUR4	4963	0.0498	0.21749	0	0	1
WEALTH	6053	-0.1953	0.78115	-5.661	-0.2968	2.952

5. Regresní analýza

Tato část obsahuje vícenásobné regrese jednotlivých variant modelu. Všechny jsou odhadnuty za pomoci metody vážených nejmenších čtverců, kdy je jako váha použita proměnná W_FSTUWT neboli finální váha studenta, která vyjadřuje rozdílnou pravděpodobnost zařazení do průzkumu (více OECD, 2012).

Nejdříve je odhadován model, který obsahuje dummy proměnné pro jednotlivé počty sourozenců, v další části pak model obsahující počty jednotlivých sourozenců. Oba modely jsou odhadovány jak pro závislou proměnnou čtenářská gramotnost, tak pro průměr známek.

Všechny modely byly testovány na heteroskedasticitu pomocí Whiteova testu. U všech modelů jsme zamítli nulovou hypotézu homoscedasticity ve prospěch významné heteroskedasticity ($\chi^2_{0,05}(q) < nR^2_e$). Při existenci heteroskedasticity jsou odhady stále nestranné a konzistentní, je však porušen jeden z Gauss-Markovových předpokladů konstantního a konečného rozptylu, odhady standardních chyb jsou tudíž vychýlené a může tak být zkreslena významnost parametrů. Zde již proto uvádím regresi s robustním odhadem směrodatných chyb, které jsou již konzistentní.

Pro testování multikolinearity bylo využito statistik VIF (Variance Inflation Factors), kdy „hodnoty VIF větší než 10 signalizují přítomnost silné multikolinearity vysvětlujících proměnných zařazených v modelu” (Kába, 1999). Hodnoty byly u všech proměnných v každém modelu velmi nízké, nejvyšší hodnoty nebyly větší než 5, tudíž předpokládáme, že se v našich modelech multikolinearita nevyskytuje.

5.1. Model počtu sourozenců

Zde odhaduji model, kde je jednou z vysvětlujících proměnných dummy proměnná vyjadřující počet sourozenců; v prvním modelu je závislou proměnnou čtenářská gramotnost, v druhém pak průměr z českého jazyka a matematiky. Model můžeme zapsat takto:

$$CTENGRAM/PRUM12_i = \beta_0 + \beta_1 CULTPOSS_i + \beta_2 DEVCE_i + \beta_{3-6} DPOCKETKNIH(2-6)_i + \beta_7 HEDRES_i + \beta_8 NEUPLNAROD_i + \beta_{9-12} SOUR(1-4)_i + \beta_{13} WEALTH_i + u_i$$

Tabulka 4: Model 1, WLS, za použití pozorování 1-6064 (n = 4841)

Vynechané chybějící pozorování: 1223

Závislá proměnná: CTENGRAM

Robustní směrodatné chyby, varianta HC1

Proměnná použita jako váha: W_FSTUWT

	koeficient	st. chyba	t-poměr	p-hodnota	
const	350.017	7.39276	47.3459	<0.00001	***
CULTPOSS	4.76092	1.39284	3.4181	0.00064	***
DEVCE	35.2243	2.27397	15.4903	<0.00001	***
DPOCKETKNIH_2	16.0855	5.37141	2.9946	0.00276	***
DPOCKETKNIH_3	34.5561	4.95999	6.9670	<0.00001	***
DPOCKETKNIH_4	60.507	5.42676	11.1498	<0.00001	***
DPOCKETKNIH_5	70.4428	5.62962	12.5129	<0.00001	***
DPOCKETKNIH_6	82.3901	6.5009	12.6736	<0.00001	***
HEDRES	9.08277	1.62257	5.5978	<0.00001	***
HISEI	1.70197	0.0921334	18.4729	<0.00001	***
NEUPLNAROD	-11.5998	3.08701	-3.7576	0.00017	***
SOUR1	4.83183	3.91583	1.2339	0.21729	
SOUR2	-2.48502	4.38631	-0.5665	0.57105	
SOUR3	-10.1771	5.36624	-1.8965	0.05795	*
SOUR4	-25.9808	6.30498	-4.1207	0.00004	***
WEALTH	-10.3508	1.62376	-6.3746	<0.00001	***
Koeficient determinace				0.316306	
Korigovaný koeficient determinace				0.314180	
F(15, 4825)				148.816	
P-value(F)				0.000000	

Model vysvětluje okolo 31% rozptylu závislé proměnné, což není mnoho, ale v porovnání s ostatními studii na toto téma, je koeficient determinace srovnatelný. Dle p-hodnoty celkového F testu zamítáme hypotézu o nulové hodnotě regresních koeficientů, vysvětlovaná proměnná tedy závisí na lineární kombinaci vysvětlujících proměnných.

Všechny proměnné jsou významné na 1% hladině významnosti, výjimku tvoří proměnná SOUR1 a SOUR2, proměnná SOUR3 je pak významná na 10% hladině významnosti.

Z tabulky vidíme, že vyšší počet sourozenců, za jinak nezměněných okolností, negativně ovlivňuje dosažený výsledek v čtenářské gramotnosti, což je v souladu s naší hypotézou. Nejhuře dopadají žáci, kteří mají čtyři a více sourozenců; mají oproti jedináčkům o téměř 26 bodů méně. Podobně žáci s třemi sourozenci dosahují o přibližně 10 bodů méně, i když je tato proměnná méně statisticky významná. Z koeficientu u proměnné SOUR1 by vyplývalo, že žáci s jedním sourozencem dosahují o téměř pět bodů lepších výsledků, avšak tato proměnná je statisticky nevýznamná, podobně jako proměnná SOUR2, představující žáky s dvěma sourozenci. Nemůžeme tedy říci, že se v tomto modelu vyskytuje negativní efekt jedináčka jako u Woltera a Vellacota (2002).

Býti děvčetem znamená ceteris paribus dosáhnout o 35 bodů více, být z neúplné rodiny pak snižuje skóre žáka o přibližně 11,6 bodu.

S pozitivním znaménkem, dle očekávání, vyšly i proměnné CULTPOSS, HEDRES, HISEI a POCETKNIH. Nejvyšší koeficienty jsou u proměnných vyjadřujících počet knih, což i vzhledem k tomu, že je závislou proměnnou čtenářská gramotnost, není žádným překvapením. Z tabulky můžeme vidět, že se koeficient u dummy proměnných týkajících se počtu knih postupně zvyšuje; žáci s více než 500-ti knihami dosahují (ceteris paribus) dokonce o 82 bodů více, než jejich vrstevníci s méně než deseti knihami.

Pokud se zvýší index kulturního vlastnictví (CULTPOSS) o jednotku, zvýší se skóre o přibližně 8 bodů, u indexu domácích vzdělávacích zdrojů (HEDRES) pak o 9 bodů. Postup jednoho z rodičů o jednotku výš v rámci indexu ISEI pak znamená o 1,7 bodu více.

Velmi zajímavým je pak záporné znaménko proměnné WEALTH. Pokud se zvýší index bohatství o jednotku (opět ceteris paribus), sníží se celkové skóre gramotnosti o více než 10 bodů. Může to

být způsobeno tím, že žáci, jež mají k dispozici více elektroniky (jak již bylo řečeno, proměnná sleduje například vlastnictví televizí, DVD přehrávače, počítačů aj.) věnují více času jiným aktivitám (například počítačovým hrám) než kupříkladu čtení či samostudiu.

Tabulka 5: Model 2, WLS, za použití pozorování 1-6064 (n = 4693)

Vynechaná chybějící pozorování: 1371

Závisle proměnná: PRUM12

Robustní směrodatné chyby, varianta HC1

Proměnná použitá jako váha: W_FSTUWT

	koeficient	st. Chyba	t-poměr	p-hodnota	
const	3.61005	0.0875048	41.2555	<0.00001	***
DEVCE	-0.273593	0.0277849	-9.8468	<0.00001	***
DPOCKETNIH_2	-0.160915	0.0652066	-2.4678	0.01363	**
DPOCKETNIH_3	-0.308679	0.0587993	-5.2497	<0.00001	***
DPOCKETNIH_4	-0.440757	0.062526	-7.0492	<0.00001	***
DPOCKETNIH_5	-0.498939	0.065356	-7.6342	<0.00001	***
DPOCKETNIH_6	-0.585093	0.0730754	-8.0067	<0.00001	***
HEDRES	-0.0739463	0.0194358	-3.8046	0.00014	***
HISEI	-0.0108658	0.00110551	-9.8288	<0.00001	***
NEUPLNAROD	0.221536	0.0382686	5.7890	<0.00001	***
SOUR1	0.00230785	0.0485446	0.0475	0.96208	
SOUR2	0.0603919	0.0534008	1.1309	0.25815	
SOUR3	0.152024	0.0665925	2.2829	0.02248	**
SOUR4	0.290049	0.0707381	4.1003	0.00004	***
WEALTH	0.0445333	0.0200297	2.2234	0.02624	**
Koeficient determinace				0.128242	
Korigovaný koeficient determinace				0.125633	
F(14, 4678)				49.15479	
P-hodnota (F)				4.0e-128	

Koeficient determinace u druhého modelu je podstatně nižší než u předchozího modelu (v souladu s očekáváním) model vysvětluje pouze necelých 13% rozptylu závislé proměnné. Znaménka koeficientů vycházejí stejně jako u modelu se závislou proměnnou, resp. opačně, jelikož závislou proměnnou je nyní průměr známek a tedy snížení závislé proměnné znamená zlepšení žákovy situace.

Podobně jako v předchozím modelu vyšly i statistické významnosti jednotlivých proměnných, z hlediska statistické nevýznamnosti byla vyloučena pouze proměnná CULTPOSS. Hladina významnosti klesla z 1% na 5% u proměnné DPOCETKNIH_2 a WEALTH, naopak se zvýšila u proměnné SOUR3.

Můžeme tedy vidět, že v případě, pokud má jedinec 4 a více sourozenců, zvýší se jeho průměr ze známek z matematiky a českého jazyka o necelých 0,3 bodu. I v tomto modelu se tedy potvrdila hypotéza, že zvyšující se počet sourozenců působí na dosažené výsledky negativně. Pokud se podíváme i na ostatní koeficienty, lze vidět, že mají vliv na průměr známek velmi podobný vliv jako na čtenářskou gramotnost v předchozím modelu. Koeficient u proměnné WEALTH zůstává stále záporný a významný na pětiprocentní hladině.

5.2. Model pořadí narození

V této části odhaduji model, kde jsou mezi vysvětlující proměnné zařazeny proměnné určující počet mladších, starších a stejně starých sourozenců. Model 3 je odhadován pro závislou proměnnou CTENGRAM, model 4 pak opět pro závislou proměnnou PRUM12. Rovnici odhadu můžeme zapsat jako:

$$CTENGRAM/PRUM12_i = \beta_0 + \beta_1 CULTPOSS_i + \beta_2 DEVCE_i + \beta_{3-6} DPOCKETKNIH(2-6)_i + \beta_7 HEDRES_i + \beta_8 NEUPLNAROD_i + \beta_9 POCETMLADSICH_i + \beta_{10} POCETSTARSICH_i + \beta_{11} POCETSTEJNESTAR_i + \beta_{13} WEALTH_i + u_i$$

Tabulka 6: Model 3, WLS, za použití pozorování 1-6064 (n = 4837)

Chybějící vynechaná pozorování: 1227

Závislá proměnná: CTENGRAM

Robustní standardní chyby, variant HC1

Proměnná použitá jako váha: W_FSTUWT

	koeficient	st. chyba	t-poměr	p-hodnota	
const	359.543	6.83101	52.63	0.0000	***
CULTPOSS	5.22761	1.39809	3.739	0.0002	***
DEVCE	35.5397	2.27260	15.64	8.21e-054	***
DPOCKETKNIH_2	15.9877	5.39758	2.962	0.0031	***
DPOCKETKNIH_3	34.5344	4.97028	6.9482	<0.00001	***
DPOCKETKNIH_4	61.1774	5.43434	11.26	4.88e-029	***
DPOCKETKNIH_5	70.5376	5.63983	12.5180	2.40e-035	***
DPOCKETKNIH_6	82.3076	6.50709	12.65	4.22e-036	***
HEDRES	8.97958	1.62141	5.538	3.22e-08	***
HISEI	1.67957	0.0916476	18.33	1.41e-072	***
NEUPLNAROD	- 12.1177	3.05338	-3.969	7.33e-05	***
POCETMLADSICH	-3.05108	1.20072	-2.541	0.0111	**
POCETSTARSICH	- 8.33791	1.29110	6.458	1.17e-010	***
POCETSTEJNESTAR	- 9.54881	7.21163	-1.324	0.1855	
WEALTH	-10.2567	1.61608	-6.347	2.40e-010	***
Koeficient determinace				0.317630	
Korigovaný koeficient determinace				0.315649	
F(14, 4826)				160.3246	
P-value(F)				0.000000	

Koeficient determinace vyšel podobně jako u modelu 1, model tedy opět vysvětluje necelých 31% rozptylu vysvětlované proměnné. Vzhledem k tomu, že se pouze změnila interpretace počtu sourozenců, není to žádným překvapením.

Všechny proměnné vyšly statisticky významné na 1% hladině významnosti, výjimku tvoří proměnná POCETSTEJNESTAR, vyjadřující počet stejně starých sourozenců. To může být způsobeno především tím, že v celém datasetu je pouze 129 žáků, jež vykazují hodnotu větší než nula u této proměnné, přesto, především pro názornost, proměnnou v modelu ponechávám. Další výjimkou je proměnná POCETMLADSICH, kde p-hodnota mírně překročila 1% hranici, stále ale můžeme říci, že je proměnná významná pod 5% hladinou významnosti.

Koeficient u proměnných POCETMLADSICH a POCETSTARSICH vychází se záporným znaménkem, tudíž každý další sourozenec, ceteris paribus, snižuje počet dosažených bodů z testu čtenářské gramotnosti. Konkrétně starší sourozenec o přibližně 8,3 bodu a mladší sourozenec o 3 bodu, je tedy horší, pokud žákovi “přibude” starší sourozenec. Rozdíl tedy činí téměř 5,3 bodů. Jinými slovy můžeme říci, že je lepší nacházet se relativně níže v pořadí, respektive býti relativně starším sourozencem, nežli mladším.

Jelikož p-hodnota u proměnných POCETMLADSICH a POCETSTARSICH se vztahuje k těmto proměnným, které vyjadřují počet, je třeba ještě otestovat, zda rozdíl mezi těmito koeficienty je statisticky významný, tedy zda relativní pořadí hraje doopravdy roli. K tomu použijeme testu lineární restrikce koeficientů, kdy budeme testovat, zda jsou si koeficienty těchto dvou proměnných rovny, čili:

$$b \text{ [POCETMLADSICH]} - b \text{ [POCETSTARSICH]} = 0$$

p-hodnota testu vyšla přibližně 0.000379, tudíž pod 1% hladinou významnosti můžeme říci, že se koeficienty nerovnají. Považujeme tedy tento rozdíl za významný.

Koeficienty u ostatních proměnných vycházejí opět velmi podobně jako v prvním modelu, žádný se významně neodlišuje, i u tohoto modelu můžeme pozorovat zápornou hodnotu proměnné WEALTH.

Tabulka 7: Model 4: WLS, za použití pozorování 1-6064 (n = 4689)

Chybějící vynechaná pozorování: 1375

Závisle proměnná: PRUM12

Robustní standardní chyby, variant HC1

Proměnná použitá jako váha: W_FSTUWT

	koeficient	st. chyba	t-poměr	p-hodnota	
const	3.56167	0.0786943	45.26	0.0000	***
DEVCE	-0.275657	0.0278282	-9.906	6.60e-023	***
DPOCKETNIH_2	-0.160112	0.0652196	-2.455	0.0141	**
DPOCKETNIH_3	-0.309808	0.0588693	-5.263	1.48e-07	***
DPOCKETNIH_4	-0.445870	0.0625160	-7.132	1.14e-012	***
DPOCKETNIH_5	-0.501477	0.0654892	-7.657	2.29e-014	***
DPOCKETNIH_6	-0.586060	0.0730976	-8.017	1.35e-015	***
HEDRES	-0.0744384	0.0194446	-3.828	0.0001	***
HISEI	-0.0107004	0.00110796	-9.658	7.29e-022	***
NEUPLNAROD	0.22525	0.0376618	5.9809	<0.00001	***
POCETMLADSICH	0.0411889	0.0153299	2.687	0.0072	***
POCETSTARSICH	0.0716762	0.0148603	4.823	1.46e-06	***
POCETSTEJNESTAR	0.0656422	0.0800260	0.8203	0.4121	
WEALTH	0.0431313	0.0200072	2.156	0.0312	**
Koeficient determinace				0.127609	
Korigovaný koeficient determinace				0.125184	
F(13, 4679)				52.60298	
P-value(F)				3.7e-128	

Model je opět téměř totožný s modelem č. 2, opět se jedná jen o nepatrnou transformaci modelu. Můžeme vidět, že pokud se ceteris paribus “zvýší” počet mladších sourozenců o jednotku, zvýší se průměrná známka o 0,04 a u starších sourozenců pak o 0,07.

Pokud však provedeme lineární restrikcí koeficientů, vychází nám p-hodnota přibližně 0,11 a na 10% hladině pravděpodobnosti nemůžeme zamítnout hypotézu o rovnosti parametrů a předpokládáme, že rozdíl není významný.

5.3. Diskuze výsledků

Všechny modely byly odhadovány pomocí metody vážených nejmenších čtverců s robustním odhadem směrodatných chyb. Model 1 a 2 zkoumal vliv počtu sourozenců na čtenářskou gramotnost a studijní průměr, v modelu 3 a 4 byly zahrnuty proměnné vyjadřující počet mladších, starších a stejně starých sourozenců a byl tak zkoumán vliv relativního pořadí. V modelu 1 a 2 se potvrdila hypotéza, že počet sourozenců ovlivňuje negativně studijní výsledky, avšak tento vliv je významný až se vzrůstajícím počtem sourozenců, konkrétně od tří sourozenců dále. I když kladný koeficient pro jednoho sourozence naznačoval negativní efekt jedináčka, jako našli Wolter s Vellacottem (2002), byl statisticky nevýznamný.

V modelu 3 a 4 pak byl zkoumán vliv relativního pořadí sourozenců. I v těchto dvou modelech se prokázal negativní vliv rostoucího počtu sourozenců. V modelu 3 se potvrdil rozdílný vliv relativního pořadí sourozenců, kdy „přírůstek“ staršího sourozence odpovídá o 5,3 bodu horšímu výsledku z testu čtenářské gramotnosti, oproti „přírůstku“ mladšího sourozence. Je tedy lepší být relativně starším sourozencem, což odpovídá výsledkům většiny empirických prací, v našem případě dosáhli podobného závěru Behrman a Taubman (1986), Wolter a Vellacott (2002), Booth a Kee (2009) a Black et al. (2005, 2011).

V modelu 4 napovídal rozdíl koeficientů podobnému výsledku, avšak nezamítli jsme hypotézu o rovnosti parametrů a předpokládáme tak, že rozdíl není významný. Čtenářská gramotnost může být dána více nejen vrozenými vlohami, ale i vztahem k četbě apod., tudíž vliv pořadí, způsobený ať již genetickými vlohami, či různou interakcí mezi sourozenci nebo odlišnými investicemi rodičů, u ní nelze tak snadno eliminovat jako u známek, kde je přeci jen větší možnost se danou látku naučit „ze dne na den“. Také to může odrážet fakt, že známky nemusí vždy odrážet plně schopnosti žáka, především v našem případě, kdy máme k dispozici pouze známky za jedno pololetí. Je však nutné poznamenat, že vypovídací hodnota tohoto modelu byla velmi nízká, koeficient determinace činil necelých 13 procent.

Významně ve všech modelech vyšla také proměnná vyjadřující pohlaví, dle očekávání dosahují dívky výrazně lepších výsledků a to nejen v čtenářské gramotnosti, ale také mají nižší průměrnou známku z českého jazyka a matematiky. Podobně prospívají výrazně hůře žáci z neúplných rodin, což naznačuje, že rodiče hrají ve studijním úspěchu velkou roli.

Ve všech modelech vyšly významně také proměnné určující socio-ekonomický status, s výjimkou proměnné CULTPOSS vyjadřující kulturní vlastnictví, která vyšla pro vysvětlovanou proměnnou PRUM12. Je zde vidět, že zájem o kulturu, klasickou literaturu či umění ovlivňuje čtení, ale na známky ve škole vliv zřejmě nemá.

Vzhledem k dostupným proměnným vyšel také poměrně nízký koeficient determinace, u modelů se závislou proměnnou čtenářské gramotnosti, bylo vysvětleno okolo 31%, což je v souladu například s Woltrem a Vellacottem (2002), ale u modelů, kde byl vysvětlován průměr známek, bylo pokaždé vysvětleno necelých 13 % rozptylu závislé proměnné, což se dá považovat za opravdu málo.

Závěr

Tato práce se pokusila odhadnout vliv počtu sourozenců a jejich pořadí na studijní výsledky v České republice pomocí regresní analýzy na datech z průzkumu PISA 2009.

V modelu počtu sourozenců se ověřila hlavní hypotéza a to, že počet sourozenců má negativní vliv na studijní výsledky. Tento vliv je však statisticky významný až pro rodiny se čtyřmi a více dětmi (pro žáky s třemi a více sourozenci), nebyl tedy nalezen ani žádný významný efekt jedináčka (i když by tomu koeficienty naznačovaly).

Model pořadí narození pak ukazuje, že existují rozdíly jak pro dosažené skóre čtenářské gramotnosti, tak pro studijní průměr co se týče pořadí narození. Je lepší být relativně starším sourozencem, což odpovídá výsledkům většiny zde uvedených studií. Po zahrnutí relativního vlivu pořadí narození stále přetrvává negativní efekt počtu sourozenců; námětem pro další zpracování by tak mohl být odhad modelu pro jednotlivé velikosti rodiny. V případě dostupnosti dalších dat, by bylo zajisté také zajímavé zkoumat vliv pohlavního složení sourozenců či rozestupu mezi nimi.

Všechny proměnné vyjadřující socioekonomický status studenta, jako index HISEI, množství vzdělávacích zdrojů HEDRES či počet knih vyšly významně (výjimku tvoří CULTPOSS u regresí se závislou proměnnou PRUM12). Je tedy možné udělat závěr, že socioekonomický status žáka významně ovlivňuje jeho studijní výsledky.

Dalším zajímavým výsledkem je negativní vliv proměnné WEALTH. Nemůžeme ale říci, čím je tento vliv způsoben. Zda se jedná o negativní vliv elektroniky, či zdase jedná o rodiče kompenzující nedostatek času dražšími dárky apod.. Zajisté by však stála podrobnější analýza toho vlivu za zvážení.

Přestože je průzkum PISA průzkumem velmi komplexním a obsahuje velké množství dat, podrobnější zkoumání vlivu počtu a pořadí sourozenců by si zasloužilo určitě více informací ohledně času stráveného s rodiči, jejich věku a dalších, pro analýzu těchto vlivů důležitých proměnných. Rozhodně zajímavá by byla existence panelových dat či časových řad podobného charakteru pro Českou republiku či data pro dospělé, pro zkoumání vlivu sourozenců například na výdělky.

Použitá literatura

BECKER, Gary S. An Economic Analysis of Fertility. *National Bureau of Economic Research*. Columbia University Press, 1960, 225 - 256.

BECKER, Gary S. a H. Gregg LEWIS. On the Interaction between the Quantity and Quality of Children. *The Journal of Political Economy*. 1973, Volume 81, Issue 2, Part 2: New Economic Approaches to Fertility, s. 279-288.

BECKER, Gary S. a Nigel TOMES. Child Endowments, and the Quantity and Quality of Children. *Journal of Political Economy*. University of Chicago Press, 1976, vol. 84, s. 143-62.

BECKER, Gary S. a Nigel TOMES. Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*. 1986, Vol. 4, no. 3, s. 1-39.

BEHRMAN, Jere R. a Paul TAUBMAN. Birth Order, Schooling, and Earnings. *Journal of Labor Economics: Part 2: The Family and the Distribution of Economic Rewards*. The University of Chicago Press, 1986, vol. 4, no. 3, s. 121-145.

BLACK, Sandra E., Paul J. DEVEREUX a Kjell G. SALVANES. Older and Wiser? Birth Order and IQ of Young Men. *Cesifo Economic Studies*. 2011, v. 57, iss. 1, s. 103-120. ISSN 1610241X.

BLACK, S. E.; DEVEREUX, P. J.; SALVANES, K. G. The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *Quarterly Journal of Economics*. 2005, 120 (2), s. 669-700.

BLAKE, Judith. Family Size and Achievement. Berkeley: University of California Press, 1989. 415 s.

BOOTH, A. L.; KEE, H. Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment. *Journal of Population Economics*. 2009, 22(2), s. 367-397.

GANZEBOOM, Harry B. G. et al A Standard International Socio-Economic Index of Occupational Status. *Social Science Research*. 1992, vol. 21, no. 1, s. 1-56. DOI: 0.1016/0049-089X(92)90017-B.

HORTON, Susan. Birth Order and Child Nutritional Status: Evidence from the Philippines. *Economic Development and Cultural Change*. 1988, 36(2), s. 341-354.

HYUNJOON, Park. Public Policy and the Effect of Sibship Size on Educational Achievement: A Comparative Study of 20 Countries. *Social Science Research* 37.3, 2002, s. 874-887

KÁBA, Bohumil. Soudobé techniky zjišťování multikolinearity v ekonometrických modelech. *Agrární WWW portál* [online]. 24.09.1999 [cit. 2012-07-14]. Dostupné z: <http://www.agris.cz/clanek/104154>

OECD (2009), *PISA Data Analysis Manual SAS® SECOND EDITION*, OECD Publishing. ISBN 978-92-64-05624-4.

OECD (2010), *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*, OECD Publishing. ISBN 978-92-64-09144-3.

OECD (2012), *PISA 2009 Technical Report*, PISA, OECD Publishing. ISBN 978-92-64-04018-2.

PALEČKOVÁ, Jana et al. *Hlavní zjištění výzkumu PISA: Umíme ještě číst?*. Praha: Ústav pro informace ve vzdělávání, 2010. ISBN 978-80-211-0608--6.

PISA 2009 – OECD PROGRAM PRO MEZINÁRODNÍ HODNOCENÍ ŽÁKŮ (PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT). *Ústav pro informace ve vzdělávání* [online]. 2011 [cit. 2012-07-14]. Dostupné z: <http://www.uiv.cz/clanek/607/1871>

WOLTER, Stefan C. a M. C. VELLACOTT. Sibling Rivalry: A Look at Switzerland with PISA Data. *IZA Discussion Paper*. 2002, no. 594.