

Odvození parametrů rovnice mzdové dynamiky pro dynamický mikrosimulační model důchodového systému MPSV

ZÁVĚREČNÁ
ZPRÁVA



MINISTERSTVO PRÁCE
A SOCIÁLNÍCH VĚCÍ

WWW.TREXIMA.CZ

Obsah:

Úvod.....	3
1. Příprava historických dat na SQL serveru.....	3
1.1 Harmonizovaná data Informačního systému o průměrném výdělku.....	3
1.2 Specifické úpravy dat za roky 2010 - 2014	4
1.3 Kompletní tvorba harmonizovaných dat za roky 2008 - 2009.....	4
1.4 Kontrola historických dat	4
1.5 Odstranění vlivu legislativních změn z položek prémie a odměny.....	5
1.6 Modelování očištěných mezd o vliv daňových změn a úroveň minimální mzdy, očištění dat o nominální nárůsty	7
1.7 Odstranění vlivu změn úrovně minimální mzdy	7
1.8 Postup tvorby pseudo-panelu zaměstnanců za roky 2008 - 2014	8
2. Přístupy k modelování výše mzdy používané v jiných zemích	10
2.1 Švédský model (SESIM III)	10
2.2 Francouzský model (Destinie I).....	10
2.3 Belgický, německý a italský model (MIDAS).....	11
2.4 Britský model (PenSim2).....	11
2.5 Americký model (CBOLT)	11
2.6 Srovnání modelů a výběr modelu pro ČR	12
3. Tvorba mzdových rovnic a testování jejich stability	13
3.1 Analýza a kategorizace faktorů	13
3.2 Redukce množství faktorů a kategorií.....	13
3.3 Další parametry mikrosimulačního modelu	14
3.4 Popis vstupních dat a faktorů	14
3.4.1 Mzdové rozdělení a design effect.....	14
3.4.2 Vstupní faktory.....	15
3.5 Kategorizace proměnných a požadavek na stabilitu mzdových rovnic.....	23
3.6 Interakce faktorů.....	23
3.7 Tvorba mzdových rovnic	24
3.7.1 Základní model	24
3.7.2 Model 1	28
3.7.3 Model 2	30
3.8 Schopnost predikce modelů a rezidua	33
3.9 Výběr modelu	42
4. Výpočty parametrů variability mezd a jejich implementace do mikrosimulačního modelu ...	42
4.1 Permanentní a transitorní šoky (variabilita mzdového růstu).....	42
4.2 Permanentní mzdový rozdíl (individuální variabilita ve mzdách)	44
Závěr	48
Použitá literatura	49

Úvod

Závěrečná zpráva popisuje tvorbu mzdové rovnice dle Smlouvy o odvození parametrů rovnice mzdové dynamiky pro mikrosimulační model důchodového systému MPSV uzavřené dle § 1746 odst. 2 zákona č. 89/2012 Sb., občanský zákoník (dále jen „občanský zákoník“) a zákona č. 137/2006 Sb., o veřejných zakázkách, ve znění pozdějších předpisů mezi smluvními stranami Česká republika – Ministerstvo práce a sociálních věcí jako objednatel a TREXIMA, spol. s r.o. jako poskytovatelem. Zpráva popisuje jednotlivé části tvorby mzdové rovnice a je rozdělena do 4 kapitol.

Kapitola 1 popisuje postup přípravy dat a jejich specifické úpravy pro účely tvorby mzdové rovnice. Kapitola 2 se zabývá analýzou zahraničních modelů, které se zabývají touto problematikou a také volbou zahraničního modelu, kterým se model v ČR nejvíce inspiroje. Kapitola 3 popisuje vlastní proces tvorby a výsledky mzdové rovnice. Závěrečná kapitola popisuje postup a výsledky výpočtu parametrů variability mezd a jejich implementace do mikrosimulačního důchodového systému MPSV.

1. Příprava historických dat na SQL serveru

1.1 Harmonizovaná data Informačního systému o průměrném výděлку

Pro odvození parametrů rovnice mzdové dynamiky je potřeba vytvořit datovou základnu na základě co nejdelší časové řady dat šetření Informačního systému o průměrném výděлку (ISPV), která by ovšem zároveň reflektovala úroveň a strukturu mezd zaměstnanců ČR.

ISPV reflektuje celou zaměstnaneckou populaci od roku 2011 [18], kdy došlo i k harmonizaci metodiky ISPV s Českým statistickým úřadem (ČSÚ). K dosažení tohoto cíle byli nejdříve v letech 2008 a 2009 osloveni respondenti ze sektoru fyzických osob-podnikatelů, v roce 2010 respondenti ze sektoru neziskových institucí a z šetření malých podniků s méně než 10 zaměstnanci (šetření probíhá jedenkrát za 4 roky - v roce 2007 proběhlo šetření za referenční rok 2006, v roce 2011 za referenční rok 2010 a v roce 2015 proběhne šetření za referenční rok 2014).

Ke zpřesnění statistických odhadů je použita metoda odvětvové poststratifikace na údaje ze mzdového výkaznictví ČSÚ. Odbor metodiky realizace statistických zpracování ČSÚ předává na základě dohody mezi ČSÚ a MPSV zpracovateli ISPV údaje se mzdami a počty zaměstnanců zjištěnými v rámci podnikového výkaznictví ČSÚ. Toto šetření se vztahuje na řádově větší počet subjektů než ISPV a výsledky na agregátní úrovni mají mnohem menší variabilitu než je tomu v ISPV.

Pomocí vážení jsou dopočteny počty zaměstnanců v třídění podle institucionálních sektorů a sekcí odvětvové klasifikace CZ-NACE. Následně jsou dopočteny objemy mezd v jednotlivých sekcích odvětvové klasifikace CZ-NACE.

Pro potřeby Eurostatu a jeho šetření Structure of Earning Survey (SES) byla vytvořena **harmonizovaná zdrojová data i za rok 2010**. Bylo však potřeba prozkoumat, nedošlo-li v uvedeném období ke změně číselníků u faktorů, které se mohou využít při odvození parametrů rovnice mzdové dynamiky. Blíže popsáno v části 1.2.

Speciálně pro účely odvození parametrů rovnice mzdové dynamiky byla vytvořena **harmonizovaná datová základna za roky 2008 a 2009**. Zde se kromě případné změny číselníků řešila tvorba dat jako celku, zejména doplnění chybějících sektorů ekonomických subjektů. Tvorba datové základny za roky 2008 a 2009 je podrobněji popsána v části 1.3.

Část 1.4 popisuje kontrolu historických dat. Jedná se zejména o identifikaci extrémních výkyvů a její konfrontaci s ekonomickou a zejména legislativní situací příslušného období.

Část 1.5 popisuje nedostatky a úskalí, ke kterým se dospělo při řešení první etapy a doporučení pro další etapy zejména v souvislosti s tím, k čemu se bude potřeba v následujících etapách vrátit.

1.2 Specifické úpravy dat za roky 2010 - 2014

Pro odvození parametrů rovnice mzdové dynamiky se předpokládá využití následujících faktorů: pohlaví, věk, vzdělání, doba zaměstnání, místo výkonu práce, státní občanství, zaměstnání, odvětví, kolektivní smlouva, finanční kontrola.

Pro výše uvedené faktory a výše uvedené roky nedošlo k významným změnám v číselnících, které by měly za následek složitější úpravu dat.

1.3 Kompletní tvorba harmonizovaných dat za roky 2008 - 2009

Tvorba harmonizovaných dat za roky 2008 – 2009 probíhala následujícím způsobem. Vycházelo se z dat ISPV sesbíraných za referenční období příslušného roku. Na rozdíl od standardního zpracování však musela být zahrnuta (imputována) i data ze sektoru neziskových institucí, která byla sbírána až od roku 2010. Ta byla aktualizována obdobným způsobem jako data ekonomických subjektů s méně než 10 zaměstnanci, a to pomocí indexů z Výkaznictví ČSÚ. Dodavatel měl na základě dílčí dohody s ČSÚ k dispozici odpovídající data, na základě kterých byly příslušné indexy počítány.

Z hlediska tvorby datové základny bylo nutné vyřešit problémy týkající se změn klasifikace odvětví z OKEČ na CZ-NACE a změny klasifikace zaměstnání z KZAM-R na CZ-ISCO. Změnu klasifikace odvětví vyřešil ČSÚ ve svých databázích (registr ekonomických subjektů, Výkaznictví), změna klasifikace KZAM-R na CZ-ISCO byla řešena pomocí převodníku.

1.4 Kontrola historických dat

Kontrola historických dat pro účely odvození parametrů rovnice mzdové dynamiky spočívá zejména ve výpočtu agregovaných hodnot, jejích indexů, identifikaci neočekávaných hodnot indexů, zjištění příčin takto neočekávaných hodnot a případně návrhu na úpravu zdrojových dat takovým způsobem, aby nemohlo dojít ke zkreslení parametrů rovnice mzdové dynamiky.

Největší vliv pro parametry rovnice mzdové dynamiky mají **změny ve zdanění v letech 2008 - 2013**, jak ukazují následující tabulky (pro srovnatelnost je vypočtena skutečná daň) [35] :

Tab. 1 Zdanění pro rok 2008

Základ daně		Skutečná daň	Ze základu přesahujícího	Sociální a zdravotní pojištění	Daň ze superhrubé mzdy
od [Kč]	do [Kč]				
0	1 034 880	20,25 %		35%	15%
1 034 880	a více	209 563 Kč + 15 %	1 034 880 Kč		

Tab. 2 Zdanění pro rok 2009

Základ daně		Skutečná daň	Ze základu přesahujícího	Sociální a zdravotní pojištění	Daň ze superhrubé mzdy
od [Kč]	do [Kč]				
0	1 130 640	20,10 %		34%	15%
1 130 640	a více	227 259 Kč + 15 %	1 130 640 Kč		

Tab. 3 Zdanění pro rok 2010

Základ daně		Skutečná daň	Ze základu přesahujícího	Sociální a zdravotní pojištění	Daň ze superhrubé mzdy
od [Kč]	do [Kč]				
0	1 707 848	20,10 %		34%	15%
1 707 848	a více	343 277 Kč + 15 %	1 707 848 Kč		

Tab. 4 Zdanění pro rok 2011

Základ daně		Skutečná daň	Ze základu přesahujícího	Sociální a zdravotní pojištění	Daň ze superhrubé mzdy
od [Kč]	do [Kč]				
0	1 781 280	20,10%		34%	15%
1 781 280	a více	358 037 Kč + 15 %	1 781 280 Kč		

Tab. 5 Zdanění pro rok 2012

Základ daně		Skutečná daň	Ze základu přesahujícího	Sociální a zdravotní pojištění	Daň ze superhrubé mzdy
od [Kč]	do [Kč]				
0	1 206 576	20,10%		34%	15%
1 206 576	1 809 864	242 522 Kč + 16,35%	1 206 576 Kč	9%	15%
1 809 864	a více	341 159 Kč + 15 %	1 809 864 Kč		

Tab. 6 Zdanění pro rok 2013

Základ daně		Skutečná daň	Ze základu přesahujícího	Sociální a zdravotní pojištění	Daň ze superhrubé mzdy
od [Kč]	do [Kč]				
0	1 242 432	20,10%		34%	15%
1 242 432	a více	249 729 Kč + 23,35%	1 242 432 Kč	9%	15%
zrušen strop ZP		(16,35 % + 7% solidární daň)			

Největší vliv na mzdu a tím pádem i na rovnici mzdové dynamiky mají změny v daních z příjmů, které nastaly od 1.1. 2008 (zavedení superhrubé mzdy spolu se stropem na zdravotní a sociální pojištění) a od 1.1. 2013 (zrušení stropu na zdravotní pojištění + zavedení solidární daně). Menší vliv má změna, která nastala od 1.1. 2010 (zvedly se stropy na zdravotní a sociální pojištění ze 48 na 72 násobek průměrné měsíční mzdy).

1.5 Odstranění vlivu legislativních změn z položek prémie a odměny

Jak již bylo uvedeno v předchozí části, docházelo mezi roky 2008 – 2013 k poměrně významným změnám v oblasti daní. Změny se projeví zejména v oblasti odměn. Tohoto institutu využívaly firmy pro optimalizaci daní při legislativních změnách. Podle druhu změny byly odměny vyplaceny dříve, byla-li avizovaná legislativní změna daňově nevýhodná, nebo naopak později, byla-li avizovaná legislativní změna daňově výhodná.

Zkušenost ze zpracování Informačního systému o průměrném výdělku (ISPV) z posledních let ukazuje, že legislativní změny se nejvíce projeví odvětvově, a to ve mzdové sféře. Platová sféra odměňuje podle platových tabulek a proto je při vyrovnání odměn důležité vzít v úvahu také sféru.

Vyrovnání odměn proběhlo následujícím způsobem:

- 1) Data se rozdělila podle střední úrovně třídění NACE (dále podsekce) a sféry mzdové, nebo platové,
- 2) sloučily se podsekce s nedostatečným počtem záznamů, vytvořily se odvětvové skupiny,
- 3) za každou odvětvovou skupinu se vypočítaly průměrné hodnoty odměn za každý rok a spočítaly se průměrné růsty odměn mezi roky 2008 a 2014 pomocí geometrického průměru,

- 4) podílem teoretických a reálných hodnot v letech 2009-2013 byly vynásobeny všechny jednotlivé záznamy v odpovídajících kategoriích, čímž došlo k vyrovnání legislativních změn v odměnách.

Uvedený postup byl vyzkoušen jak na nepravidelných odměnách, tak i na celkových odměnách. Na základě analýzy výsledných koeficientů růstu mezd a odměn se ukázala jako vhodnější metoda vyrovnání celkových odměn.

Následující tabulka ukazuje koeficienty růstu celkových odměn za jednotlivé roky, zvolené odvětvové skupiny a sféry. Nejzajímavějším příkladem je odvětvová sekce K – peněžnictví, kde jsou vidět obrovské výkyvy mezi lety 2011 a 2014. Index mezi roky 2012 a 2011 je vysoký z důvodu předčasného vyplacení odměn, které by jinak byly vyplaceny až v roce 2013. Ze stejného důvodu došlo k poklesu v meziročním indexu mezi roky 2013 a 2012 a zároveň, díky nízké základně roku 2013 k nárůstu meziročního indexu mezi roky 2014 a 2013. I tento extrémní případ ilustruje vhodnost zvolené metody vyrovnání odměn.

Tab. 7 Koeficienty růstu odměn

Sféra	Sektor NACE	Průměrný koeficient růstu	Meziroční indexy neočištěných odměn					
			09/08	10/09	11/10	12/11	13/12	14/13
mzdová	A	1,044	0,921	1,082	1,047	1,076	1,035	1,118
mzdová	B	1,003	0,919	1,106	1,125	1,014	0,935	0,938
mzdová	CA	1,003	1,049	0,957	1,002	1,006	0,976	1,032
mzdová	CB	1,053	0,952	1,172	1,140	1,057	0,950	1,070
mzdová	CC	1,006	0,945	1,045	0,906	1,090	1,035	1,025
mzdová	CD	0,959	0,784	1,272	1,044	0,878	1,072	0,794
mzdová	CE	1,017	0,903	1,233	1,044	0,899	0,984	1,079
mzdová	CF	1,062	1,343	1,021	0,971	0,997	1,096	0,984
mzdová	CG	1,008	0,969	1,045	1,088	0,945	0,961	1,051
mzdová	CH	1,043	1,008	1,103	1,009	1,089	0,992	1,061
mzdová	CIJ	1,017	0,849	1,039	1,145	1,010	1,053	1,032
mzdová	CK	1,048	1,006	0,990	1,157	1,001	1,064	1,081
mzdová	CL	1,023	0,983	1,096	1,003	0,970	1,063	1,027
mzdová	CM	1,051	1,196	1,079	1,080	1,040	0,943	0,987
mzdová	D	1,021	0,938	1,087	1,033	1,021	1,028	1,026
mzdová	E	0,997	1,065	0,950	0,981	0,959	0,927	1,114
mzdová	F	1,002	0,891	1,033	1,054	0,995	0,976	1,075
mzdová	G	1,000	0,965	0,948	1,015	1,044	1,028	1,001
mzdová	H	1,015	0,845	1,109	1,068	0,962	1,021	1,113
mzdová	I	1,010	1,074	0,942	1,163	0,814	1,045	1,058
mzdová	J	0,990	0,893	1,080	1,066	1,056	0,810	1,068
mzdová	K	0,999	0,985	0,957	1,118	1,249	0,632	1,196
mzdová	L	1,022	1,169	0,766	1,104	1,026	1,127	0,995
mzdová	M	0,989	0,985	0,902	1,154	0,993	0,937	0,982
mzdová	N	0,996	1,026	0,943	1,103	0,993	1,052	0,875
mzdová	P	1,045	1,078	1,067	1,078	1,080	0,998	0,976
mzdová	QO	1,047	1,216	0,957	0,982	1,010	1,034	1,105
mzdová	R	1,020	1,117	1,025	1,037	1,033	0,811	1,132
mzdová	S	1,048	1,102	0,981	1,212	1,062	0,991	0,962
platová	O	0,982	0,930	0,781	1,041	1,067	1,076	1,033
platová	P	1,013	1,291	0,640	1,168	0,807	1,196	1,163
platová	QA	1,020	0,913	0,892	1,122	1,007	1,170	1,047
platová	QB	1,019	1,109	0,886	1,101	1,079	0,896	1,068
platová	R	1,010	0,997	0,874	1,043	0,964	1,098	1,105
platová	Ostatní	1,027	1,017	0,842	1,071	1,017	1,086	1,158

1.6 Modelování očištěných mezd o vliv daňových změn a úroveň minimální mzdy, očištění dat o nominální nárůsty

Očištěné mzdy byly získány následujícím způsobem:

- 1) Ke mzdám bez odměn se přičetly vyrovnané odměny, jejichž způsob výpočtu je uveden v části 1.5.
- 2) Následně byly nastaveny minimální hodnoty mzdy na úroveň minimální mzdy podle platné legislativy pro daný rok. Díky případnému snížení odměn může totiž dojít k tomu, že vyrovnaná mzda bude nižší než minimální mzda stanovená legislativou. V letech 2008 - 2013 byla minimální hodnota očištěné mzdy stanovena na 8000 Kč, v roce 2014 na 8500 Kč.
- 3) Na vypočtených vyrovnaných mzdách byly vypočteny koeficienty nominálních růstů/poklesů za celou ČR.

Všechny jednotlivé záznamy byly meziročními indexy převedeny na úroveň roku 2014. Následující tabulka ukazuje indexy, kterými se data v jednotlivých letech převáděla na úroveň roku 2014.

Tab. 8 Očištění hrubých měsíčních mezd

rok	Nominální úroveň hrubé měsíční mzdy		Hrubá měsíční mzda ošetřená o odměny		Hrubá měsíční mzda ošetřená o minimální mzdu		Hrubá měsíční mzda úroveň 2014	
	HMM	Meziroč. index HMM	HMM očištěná	Meziroč. index HMM očištěné	HMM min. mzda	Meziroč. index HMM min. mzda	HMM 2014	Meziroč. index HMM 2014
2008	23 876 Kč		23 876 Kč		23 910 Kč		26 911 Kč	
2009	24 446 Kč	1,024	24 558 Kč	1,029	24 588 Kč	1,028	26 911 Kč	1,000
2010	24 807 Kč	1,015	25 002 Kč	1,018	25 030 Kč	1,019	26 911 Kč	1,000
2011	25 558 Kč	1,030	25 574 Kč	1,023	25 599 Kč	1,023	26 911 Kč	1,000
2012	26 118 Kč	1,022	26 114 Kč	1,021	26 137 Kč	1,021	26 911 Kč	1,000
2013	26 245 Kč	1,005	26 401 Kč	1,011	26 423 Kč	1,011	26 911 Kč	1,000
2014	26 886 Kč	1,024	26 886 Kč	1,018	26 911 Kč	1,018	26 911 Kč	1,000

1.7 Odstranění vlivu změn úrovně minimální mzdy

Změna úrovně minimální mzdy má u zaměstnanců, kteří se pohybují v její blízkosti bezprostřední vliv na její nominální růst. Změna úrovně minimální mzdy z hodnoty 8000 Kč na 8500 Kč kupříkladu představuje nominální nárůst o 6,25 %. Pokud by se neodstranil vliv změny úrovně minimální mzdy, mohl by model přinášet chybné závěry v tom smyslu, že největší vliv na mzdový nárůst má zaměstnání, které má výdělek na úrovni minimální mzdy. Ve skutečnosti se však stále jedná o nízkovýdělková zaměstnání s výdělkem na úrovni minimální mzdy, která jsou pro růst mezd u těchto skupin zaměstnání určující. Typickým příkladem jsou zaměstnání z odvětví pohostinství. Aby nedocházelo k výše uvedenému zkreslení, byla stanovena minimální úroveň očištěné mzdy dle úrovně roku 2014 na 8500 Kč. Díky tomu došlo v letech 2011 - 2013 k mírně navýšené úrovni mezd, jak ukazuje následující tabulka.

Tab. 9 Odstranění vlivu minimálních mezd

rok	Hrubá měsíční mzda úroveň 2014		Minimální měsíční mzda	
	HMM 2014	index HMM 2014	nominální hodnota	úroveň 2014
2008	26 911 Kč		8 000 Kč	9 006 Kč
2009	26 911 Kč	1,000	8 000 Kč	8 760 Kč
2010	26 911 Kč	1,000	8 000 Kč	8 598 Kč
2011	26 912 Kč	1,000	8 000 Kč	8 415 Kč
2012	26 914 Kč	1,000	8 000 Kč	8 249 Kč
2013	26 915 Kč	1,000	8 000 Kč	8 162 Kč
2014	26 911 Kč	1,000	8 500 Kč	8 500 Kč

Jak ukazuje tabulka 9, do roku 2010 byla úroveň minimální mzdy nad 8500 Kč ve vyjádření úrovně mezd roku 2014, od roku 2011 byla úroveň minimální mzdy pod 8500 Kč.

1.8 Postup tvorby pseudo-panelu zaměstnanců za roky 2008 - 2014

Byla vytvořena zdrojová databáze, ve které jsou data z panelu zaměstnanců i data za zaměstnance, kteří v panelu nejsou.

Panel dat za uvedené roky se tvořil pomocí meziročních panelů, které se na závěr vzájemně spojily. Při tvorbě meziročního panelu se spojují zaměstnanci mezi jednotlivými lety na základě shodného identifikačního kódu, které používá zaměstnavatel pro konkrétního zaměstnance. Často však identifikační kód nestačí. Zaměstnavatelé občas nedodržují pravidla, kterými by se měli řídit při určování identifikačního čísla zaměstnance (identifikační číslo zaměstnance v ekonomickém subjektu by mělo být jedinečné a v čase stálé). Proto se přistoupilo ke spojování na základě roku narození, pohlaví, doby zaměstnání, klasifikace zaměstnání (KZAM-R nebo CZ-ISCO), kdy na základě jedinečnosti ve dvou po sobě jdoucích obdobích je tímto způsobem zaměstnanec identifikován. Algoritmicky lze postup tvorby meziročního panelu zjednodušeně popsat následujícím způsobem:

- 1) Označí se záznamy, u kterých je četnost kombinací položek IDZAM (identifikace zaměstnance), ROKNAR (rok narození zaměstnance) a POHLAVI (pohlaví zaměstnance) vyšší než 1, a to v obou sledovaných letech,
- 2) neoznačené záznamy, tj. záznamy s četností kombinací 1 v obou letech se podle položek IDZAM, ROKNAR a POHLAVI spojí,
- 3) speciálním kódem se označí záznamy, které jsou podle IDZAM, ROKNAR a POHLAVI nejednoznačné, ale po přidání položky CZICSE (kód postavení zaměstnání) už mají tyto kombinace četnost 1,
- 4) neoznačené záznamy + záznamy označené speciálním kódem, tj. záznamy s četností kombinací 1 v obou letech se podle položek IDZAM, ROKNAR, POHLAVI a CZICSE spojí,
- 5) analogicky k bodu 3 se postupně vytvoří speciální kódy a analogicky k bodu 4 se spojí záznamy s četností kombinací 1 po postupném přidání položek KZAM-R či CZ-ISCO (klasifikace zaměstnání), VZDELANI (vzdělání zaměstnance), DOBAZAM (doba zaměstnání zaměstnance). Při spojování podle doby zaměstnání se shodný zaměstnanec určí podle rozdílu doby zaměstnání 1 rok,
- 6) body 1 – 5 se zopakují bez využití položky IDZAM, protože leckdy není dodržena stálost tohoto kódu v čase, zejména při změně mzdového SW ve firmě.

Dále byla vytvořena zdrojová databáze. Zdrojovou databázi tvoří data jednotlivých let, která se mezi sebou rozliší pomocí nově zavedené položky ROK. Ve zdrojové databázi byli shodní zaměstnanci identifikováni podle nově vytvořené položky IDENTIFIKATOR. Položka identifikátoru se tvořila následujícím způsobem:

- 1) Za základ se vzala identifikace záznamu z roku 2008, kdy se identifikace upravila takovým způsobem, aby začínala čísly „2008“,
- 2) pokud byl zaměstnanec v meziročním panelu mezi roky 2008 a 2009, převzala položka IDENTIFIKATOR v roce 2009 hodnotu z roku 2008,
- 3) pokud zaměstnanec nebyl v meziročním panelu, položka IDENTIFIKATOR vznikla úpravou identifikace záznamu z roku 2009 takovým způsobem, aby začínala čísly „2009“,
- 4) analogicky se postupovalo i v dalších letech až do roku 2014.

Podle položky IDENTIFIKATOR v datech roku 2014 lze zjistit podle počátečních 4 čísel, kolik záznamů tvoří panel přes všechny roky, kolik jenom přes některé roky a kolik záznamů představuje pouze rok 2014. Z přibližně 2,5 milionu záznamů v roce 2014, tvoří cca. 800 tis. záznamů panel zaměstnanců za roky 2008 – 2014.

Z dat ISPV v zásadě není možné jednoznačně vytvořit meziroční panel u zaměstnanců, kteří nepracují u stejného ekonomického subjektu. Jednotlivé zahraniční modely, které využívají panelu dat, však panel neomezují pouze na zaměstnance, kteří jsou celou dobu zaměstnáni v jednom ekonomickém subjektu. Z tohoto důvodu jsme u nespojených záznamů přistoupili k vytvoření pseudo-panelu, který meziročně spojí podstatnou část záznamů vykazující maximální podobnost u těchto faktorů (seřazeno dle zvolené důležitosti pro pseudo-panel): pohlaví, rok narození, státní občanství, vzdělání, místo výkonu práce (úroveň okresu), zaměstnání (5 místné CZ-ISCO), sféra a odvětví (6 místný kód NACE).

Nově vytvořený pseudo-panel seskupil zaměstnance do nově zvolených kategorií, které obsahovaly maximálně 50 záznamů, skupiny obsahující více záznamů nebyly uvažovány. V dalším kroku byla tolerance pseudo-panelu snížena u nejméně důležitého faktoru (5 místný kód NACE) a znovu vytvořen další pseudo-panel s maximálně 50 záznamy. Takto tvorba pseudo-panelů pokračovala až k úrovni občanství, u které se v zásadě nepředpokládá změna u zaměstnanců. Téměř 90 % záznamů, které se spojily, byly ve skupinách do počtu 5 záznamů (2/3 záznamů je v pseudo-panelech obsahující maximálně 2 záznamy). Konečný krok bylo náhodné, avšak jednoznačné, meziroční spojení záznamů v rámci každého pseudo-panelu, čímž byl vytvořen jednoznačný, meziroční pseudo-identifikátor pro významnou část záznamů, jak ukazuje následující tabulka.

Tab. 10 Pseudo-panel zaměstnanců – procento spojených záznamů

	Pseudo-panel zaměstnanců 2008 – 2014					
	1letý	2letý	3letý	4letý	5letý	6letý
Spojeno v rámci firmy	75%	61%	51%	43%	37%	32%
Spojeno mezi firmami	13%	21%	26%	30%	33%	38%
Celkem spojeno	88%	81%	76%	73%	70%	69%

2. Přístupy k modelování výše mzdy používané v jiných zemích

2.1 Švédský model (SESIM III)

Data a datové zdroje

Hlavním datovým zdrojem pro švédský model SESIM [12] je LINDA. LINDA obsahuje data, která pokrývají 3,5 % švédské populace. V roce 1999 bylo náhodně vybráno 308 tis. jedinců, k nimž byli přidáni ostatní členové rodiny. Celkový vzorek v databázi byl o velikosti 786 tis. osob. LINDA je tvořena panelovými daty, protože jedinci ve vzorku jsou sledováni jak do budoucnosti, tak do minulosti. Databáze je kompletně založena na datech z administrativních registrů.

Databáze pro SESIM je tvořena z náhodného výběru 104 tis. jedinců z databáze LINDA a 8 tis. jedinců z registru národního sociálního pojištění. Dodatečný vzorek obsahuje informaci o osobách žijících mimo Švédsko s nárokem na starobní důchod.

Modelování mezd

Mzdová rovnice je založena na panelových datech za období 1992 až 1996. Vzorek obsahuje 49 040 jedinců ve věku 18 až 64 let. Pětiletý panel je tvořen z celkem 245 200 pozorování. K odhadu velikosti mzdy je využito modelu náhodných efektů. Vysvětlovanou proměnnou v modelu je logaritmus hodinové mzdy. Vysvětlujícími proměnnými v modelu je věk, kvadrát věku, vzdělání a pohlaví. Mzdová rovnice kromě vysvětlujících proměnných obsahuje i stochastickou proměnnou zachycující nepozorovaný individuální efekt (permanentní šok), který je v čase konstantní a dále náhodný člen (transitorní šok), který se pro každého jedince v čase mění. Pomocí modelu fixních efektů odhadují parametry pro vysvětlující proměnné, rozptyl permanentního šoku a rozptyl transitorního šoku. Na začátku projekce mezd se simuluje permanentní šok, který je po celou dobu projekce stejný. Transitorní šok se simuluje v každém kroku projekce mezd.

2.2 Francouzský model (Destinie I)

Data a datové zdroje

Model Destinie I [3] [4] je založen na individuálních datech získaných v roce 1991 ze šetření domácností, který provedl francouzský statistický úřad (INSEE). Data byla převážena na celou francouzskou populaci na základě údajů ze sčítání lidu z roku 1990. Databáze pro Destinie obsahovala 15 tis. domácností a 37 tis. jedinců.

Modelování mezd

Ve mzdové rovnici je vysvětlovanou proměnnou logaritmus roční superhrubé mzdy. Mzdové rovnice jsou tvořeny zvlášť pro muže a ženy, ale také pro každý sektor (veřejný, soukromý). Vysvětlujícími proměnnými v modelu je věk ukončení studia relativní (vypočtený jako rozdíl věku ukončení studia daného jedince a jeho generace), počet let odborné praxe, kvadrát počtu let odborné praxe, počet let odborné praxe vynásobený věkem ukončení studia relativní. Stochastickou proměnnou představující individuální efekt jednotlivce modelují zvlášť a využívají k tomu modelu fixních efektů. K určení rozptylu náhodného efektu využívají panelu dat z ročních hlášení sociálních údajů za období 1984 až 1989 (Déclarations Annuelles de Données Sociales). Více viz [8].

V současné době je model aktualizován daty z roku 2003 a je vyvinuta nová verze mikrosimulačního modelu Destinie II.

2.3 Belgický, německý a italský model (MIDAS)

Belgický model

Vysvětlovanou proměnnou v modelu je logaritmus hodinové mzdy, který je odhadován odděleně pro muže a ženy a také pro sektor (veřejný, soukromý). Model je založen na datech z roku 2002 (databáze PSBH). Logaritmus mzdy je odhadován na základě klasické lineární regrese. Vysvětlujícími proměnnými v modelu je věk, kvadrát věku, dummy proměnná pro střední vzdělání s maturitou (jako poslední dosažené), dummy proměnná pro vysokoškolské vzdělání a dummy proměnná pro částečný úvazek (pouze u žen).

Německý model

Pro modelování mezd byla použita databáze SOEP z roku 2002. Pro modelování mezd je využit model náhodných efektů. Vysvětlovanou proměnnou je logaritmus hodinové mzdy, který je modelován zvlášť pro muže a ženy. Vysvětlujícími proměnnými je dummy proměnná pro střední vzdělání s maturitou (jako poslední dosažené), dummy proměnná pro vysokoškolské vzdělání, praxe, kvadrát praxe, dummy proměnná pro ženatý/vdaná, velikost firmy, dummy proměnná pro počet dětí (ve věku 0-11), dummy proměnná pro chronicky nemocné (pouze u mužů), doba zaměstnání (pouze muži), dummy proměnná pro veřejný sektor (pouze ženy), dummy proměnná pro státního úředníka (pouze ženy), dummy proměnná pro počet dětí (ve věku 12-15, pouze ženy).

Italský model

Pro modelování mezd byla použita databáze ECHP z roku 2001. Pro modelování mezd je využit model náhodných efektů. Vysvětlovanou proměnnou je logaritmus hodinové mzdy, který je modelován zvlášť pro muže a ženy. Vysvětlujícími proměnnými je dummy proměnná pro střední vzdělání s maturitou (jako poslední dosažené), dummy proměnná pro vysokoškolské vzdělání, praxe, kvadrát praxe, dummy proměnná pro veřejný sektor, dummy proměnná pro smlouvu na dobu neurčitou, doba zaměstnání (pouze muži). Více viz [10]

2.4 Britský model (PenSim2)

Data pro modelování mezd jsou využita z panelového šetření domácností (British Household Panel Survey). Mzda je odhadována na základě modelu fixních efektů, který je založen na panelových datech. Více viz [11]

2.5 Americký model (CBOLT)

Data a datové zdroje

Pro predikci mzdy v modelu CBOLT [31] je využito dvou datových zdrojů. První datový zdroj má název March Current Population Survey (CPS). Sbírá údaje od 60 tis. domácností každý rok. Data se sbírají od roku 1976. Tento datový zdroj nemá charakter panelových dat. Druhý datový zdroj se nazývá Social Security Administration's Continuous Work History Sample (CWHHS) a představuje 1 procento z celkového počtu klientů sociálního zabezpečení. Do CBOLTu bylo imputováno přibližně 300 tis. pozorování pro každý rok od roku 1984 až po rok 2009. Tento datový zdroj má charakter panelových dat.

Modelování mezd

Mzdová rovnice je založena na datech CPS a to za období 1975 až 2009. Vysvětlovanou proměnnou v modelu je logaritmus roční mzdy a odhaduje se odděleně pro muže a ženy. Vysvětlujícími proměnnými jsou dummy proměnná pro věk, interakce dummy proměnné pro věk a vzdělání, kohorta podle narození (desetiletý interval), interakce kohorty a vzdělání, dummy proměnná pro status v sociální zabezpečení, interakce statusu v sociální zabezpečení a vzdělání.

Pro projekci mezd se v americkém modelu využívá kromě predikované mzdy na základě mzdové rovnice i náhodných složek označených jako permanentní a transitorní šoky. Konstrukce těchto složek vychází z teorie, kterou popsal Carroll [7] a nejsou součástí mzdové rovnice, ale odhadují se odděleně. Permanentní šok (trvalý) měří dlouhodobý výkyv ve mzdách mezi jednotlivcem a průměrnou mzdou za sledovanou skupinu (podle věku a pohlaví). Tyto výkyvy jsou způsobeny např. povýšením, dosažením vyššího stupně vzdělání, atd. Transitorní šok (přechodný) měří jakékoliv další, ale dočasné kolísání mzdy u jednotlivce. Přechodný ráz je způsoben např. získáním bonusu nebo chybějící práci v důsledku nemoci. Výpočet permanentního a transitorního šoku je založen na panelových datech CWSHS.

Pro projekci mezd je dále konstruován permanentní mzdový rozdíl (PED) jako míra individuální variability ve mzdách. PED představuje pětiletý průměr difference mezi mzdou jednotlivce za období 2005 až 2009 a predikované mzdy na základě stejných charakteristik plynoucích z vysvětlujících proměnných použitých ve mzdové rovnici. Pro konstrukci PED je využito panelových dat z CWSHS a mzdové rovnice, která je založena na datech z CPS. Vychází se z předpokladu, že rozdíl mezi mzdou a predikovanou mzdou je součtem permanentních a transitorních šoků pro daný rok, za podmínky nulové hodnoty transitorního šoku. Pětiletý průměr by měl dát lepší (stabilnější) mírou permanentního rozdílu než rozdíl založený pouze na jednom období. Více viz Congressional Budget Office (2013) [32].

2.6 Srovnání modelů a výběr modelu pro ČR

Následující tabulka ukazuje srovnání zahraničních modelů mezi sebou.

Tab. 11 Vzájemné srovnání zahraničních modelů

Model	Data	Mzdová rovnice	Náhodná složka	Poznámka k náhodné složce
SESIM III	panelová	model fixních efektů	permanentní a transitorní šok	součást modelu
Destinie	průřezová	klasický lineární regresní model	permanentní a transitorní šok	odděleně – panelová data, model fixních efektů
MIDAS Belgie	průřezová	klasický lineární regresní model		
MIDAS Německo	panelová	model fixních efektů	permanentní a transitorní šok	součást modelu
CBOLT	průřezová	klasický lineární regresní model	permanentní a transitorní šok	odděleně – panelová data, model rozptylu diferencí

Vzhledem k povaze dat, jež budou pro model ČR použity (pseudo-panel), nelze pro tvorbu mzdové rovnice využít zahraniční modely, které jsou založené na panelových datech. Naopak je možné využít modely, které využívají průřezová data za více období. Z tohoto pohledu se jako nejvhodnější jeví americký model CBOLT, který pro tvorbu mzdové rovnice nevyužívá panelová data. Pro výpočet permanentních a transitorních šoků ale panelová data využívá, a stejným způsobem je možno využít pseudo-panel z ISPV.

3. Tvorba mzdových rovnic a testování jejich stability

3.1 Analýza a kategorizace faktorů

K analýze mezd se dá ze šetření ISPV využít jak faktorů přináležících podnikům, tak zaměstnanců. Následující tabulka ukazuje výsledek analýzy rozptylu. Nejpodrobnějším tříděním daných faktorů bylo většinou vysvětleno nejvíce variability závislé proměnné (logaritmus hrubé měsíční mzdy).

Tab. 12 Analýza rozptylu u podnikových proměnných

Proměnné podniku	Max. vysvětlenost	Min. vysvětlenost
Odvětví NACE	28,7%	16,6%
Velikost ES	17,9%	11,0%
Kolektivním smlouva	4,4%	
Sféra	0,3%	
Region (sídlo)	4,0%	
Finanční kontrola	2,2%	

Tab. 13 Analýza rozptylu u osobnostních charakteristik

Proměnné zaměstnance	Max. vysvětlenost	Min. vysvětlenost
CZISCO	49,6%	32,5%
Vzdělání	26,1%	22,8%
Vedoucí	7,0%	
Doba zaměstnání	6,8%	
Region	6,2%	4,6%
CZ-ICSE	4,1%	
Pohlaví	3,5%	
Věk	1,9%	
Úvazek	1,7%	
Praxe	1,5%	
Státní občanství	1,4%	

Uvedené tabulky ukazují vztah faktorů k výdělkové úrovni za celou ČR, v případě podrobnějších třídění se význam některých faktorů zvýší.

3.2 Redukce množství faktorů a kategorií

Na základě diskuze se zadavatelem o využití faktorů pro mikrosimulační model došlo k následujícím úpravám:

- a) Charakteristiky podniku nebudou do modelu zahrnuty.
 - Poměrně snadno lze naplnit stávající stav, ale bylo by obtížné zvolit takové kategorie faktorů, které by byly v čase stabilní. Předpokládá se, že v budoucnu mohou mít vliv na mzdy úplně jiné podnikové charakteristiky (zejména odvětví) než v současnosti.
- b) Osobnostní charakteristiky využitelné v modelu
 - **CZ-ISCO** – v současném mikrosimulačním modelu není dostupné, ale má takový vliv, že zadavatel zváží implementaci tohoto faktoru. Co se týče počtu kategorií – maximálně hlavní třídy CZ-ISCO, případně ještě redukované.

- **Vzdělání** – postačuje na 4 základní kategorie, v simulovaných datech zadavatele se jedná o odvozenou veličinu – je znám věk při ukončení studia. V případě významného vlivu se do modelu zařadí i jiná než základní kategorie vzdělání.
- **Doba zaměstnání u posledního zaměstnavatele** – bude využita přímo, nebo kategorizovaně.
- **Pohlaví** – bude využito jako významný faktor, nebo budou vytvořeny mzdové rovnice zvlášť pro muže a ženy.
- **Věk** – bude zahrnut do modelu.
- **Úvazek** – bude zahrnut do modelu jako faktor pro plný a neplný (částečný) úvazek. Úvazky zadavatel do svého modelu dodá.
- **Státní občanství** – bude využito jako faktor (občanství české/cizí).

c) Osobnostní charakteristiky nevyužitelné v modelu

- **Vedoucí** – nevyužitelné v modelu zadavatele.
- **Region** – má-li vliv, lze do modelu zakomponovat, ale mzdovou rovnici je potřeba sestavit i bez tohoto faktoru. Integrace uvažována spíše do budoucích modelů.
- **Kód postavení v zaměstnání** – nevyužitelné v modelu zadavatele.
- **Praxe** – odvozená úzce korelovaná veličina se vzděláním a věkem. Bude do modelu integrována pouze v omezené míře ve formě kategorií, které s věkem nebudou korelovat.

3.3 Další parametry mikrosimulačního modelu

Na základě diskuze se zadavatelem byly přijaty tyto postupy:

- Pro modelování bude využita hrubá měsíční mzda ISPV, tj. mzdy přepočtené na plný úvazek, není potřeba normalizovat na 170 hodin.
- V modelu se bude předpokládat nehybnost minimální mzdy, tak aby nepůsobila jako faktor – dosavadní způsob očištění mezd to reflektuje.
- Očištění o nominální nárůsty bude probíhat formou jednoho indexu na jeden rok, nominální nárůst bude do simulací dodán zadavatelem ve formě tzv. scénáře.
- Parametr PED z modelu CBOLT bude vypočten pro každého zaměstnance zvlášť, do modelu se spočítají hodnoty za kategorizované faktory, které budou kompatibilní se mzdovou rovnicí.

3.4 Popis vstupních dat a faktorů

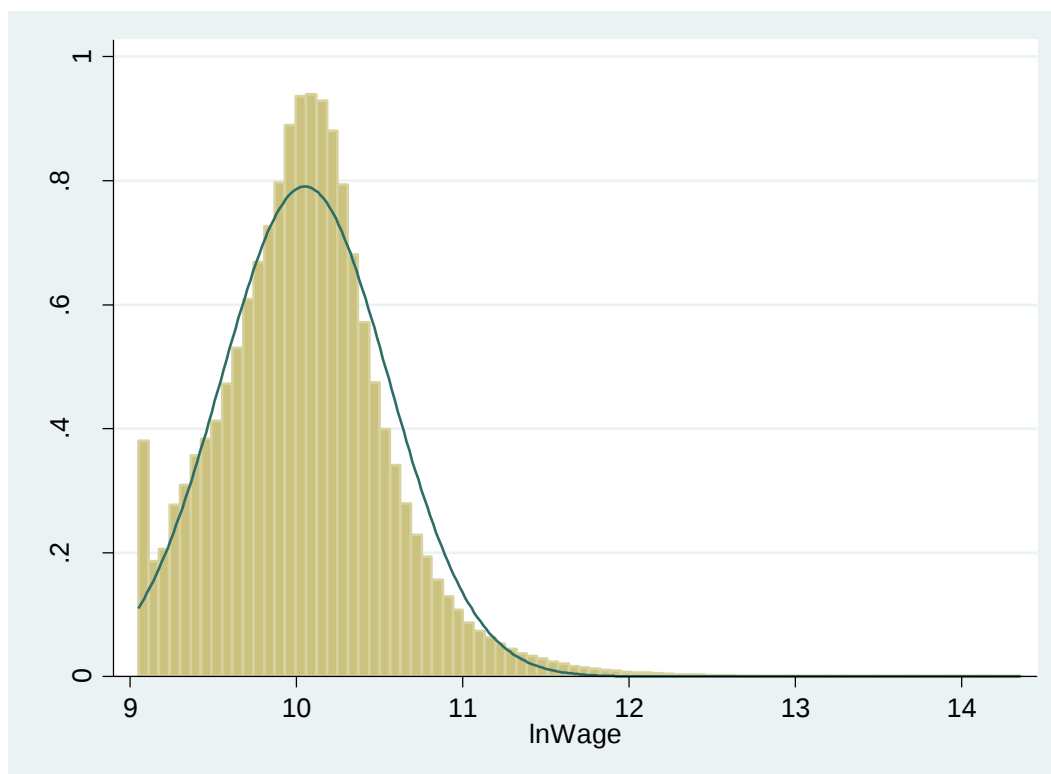
3.4.1 Mzdové rozdělení a design effect

Mzdové rozdělení

Celkový soubor dat je tvořen šesti soubory dat ISPV v letech 2008 – 2014, celkový počet záznamů přesahuje hodnotu 15,3 mil.

Při tvorbě mzdových rovnic je nutné vzít v úvahu předpoklad, že mzdy jsou log-normálně rozdělené. Následující graf ukazuje rozdělení mezd zaměstnanců České republiky po jejich očištění o vliv legislativních změn a o nominální růsty za sledované období (2008 - 2014). Z grafu je patrné, že pro tvorbu mzdové rovnice je log-normální rozdělení vhodné. Zvýšená hustota na levém okraji histogramu je způsobena existencí minimální mzdy.

Graf 1: Rozdělení pravděpodobnosti logaritmu hrubé měsíční mzdy, 2008-2014



Design effect

Výběr dat do šetření ISPV je definován jako jednostupňový, stratifikovaný, pravděpodobnostní, systematický výběr skupin s vrácením s nestejnými pravděpodobnostmi na víceletém rotačním panelu. Z povahy šetření ISPV je třeba při výpočtech variability odhadů uvážit především stratifikaci šetření a jejich výběrové podíly, není možné s daty počítat jako s prostým náhodným výběrem.

3.4.2 Vstupní faktory

Vybrané faktory můžeme rozdělit na faktory numerické a na faktory kvalitativní. U numerických faktorů je potřeba rozhodnout, budou-li se pro zařazení do mzdové rovnice kategorizovat. U kategoriálních proměnných je zase důležité, na jaké kategorie se pro mzdovou rovnici rozdělí. Následující tabulka ukazuje rozdělení vybraných faktorů podle typu proměnných.

Tabulka 14: Vstupní faktory do modelu

Faktor	Vstupní typ	Poznámka
CZ-ISCO	kategoriální	redukované hlavní třídy
Vzdělání	kategoriální	základní, střední bez maturity, s maturitou, VŠ
Doba zaměstnání u posledního zaměstnavatele	spojitá	předpoklad kategorizace
Pohlaví	kategoriální	
Věk	kategoriální	předpoklad další kategorizace
Úvazek	kategoriální	plný, částečný
Státní občanství	kategoriální	české, cizí

Při tvorbě mzdové rovnice je nutno zanedbat podnikové proměnné, které vysvětlují přibližně stejnou část variability jako proměnné uvedené výše. Důvodem nezařazení je aktuální složitost implementace do mikrosimulačního modelu důchodového systému.

Následující grafy ukazují tvar funkce průměru logaritmu hrubé měsíční mzdy podle věku a dalších faktorů, které budou využity ve mzdové rovnici.

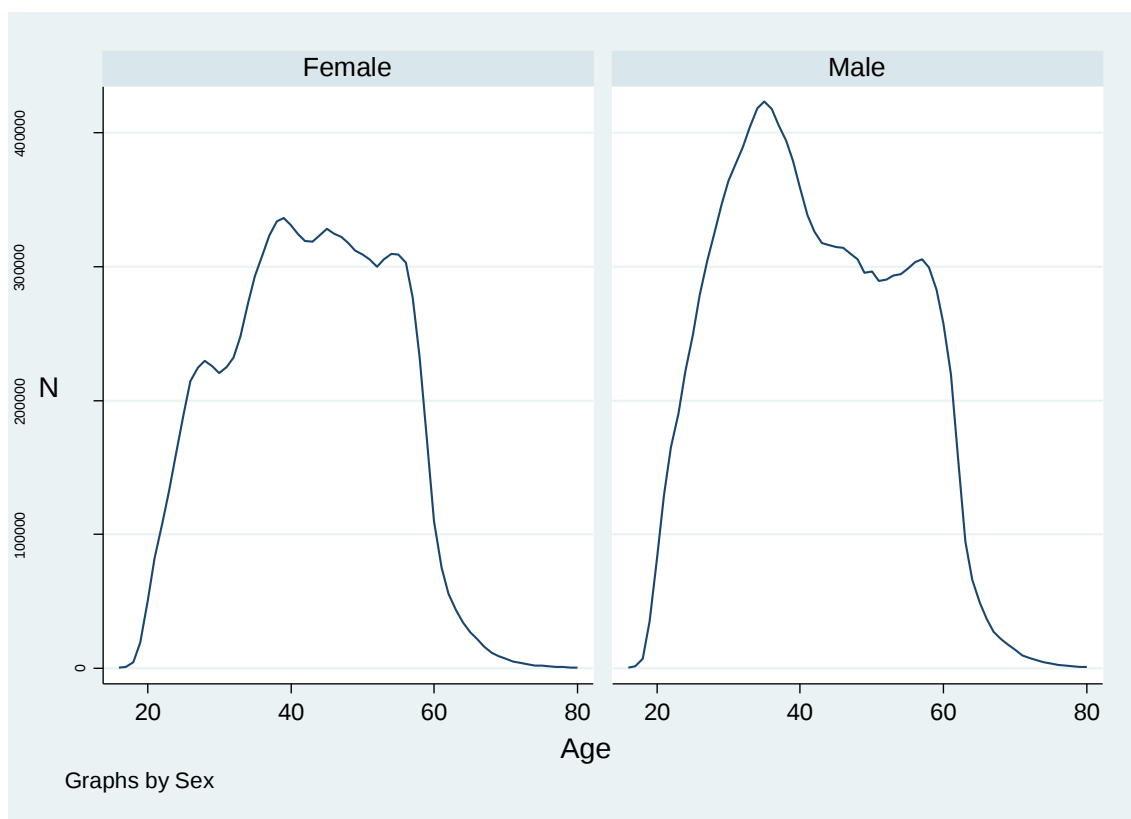
Věk

Graf rozdělení logaritmu mzdy podle věku a pohlaví nabízí využití kvadrátu věku, nebo věk kategorizovat. V zahraničních modelech se s faktorem věk zachází různě. Např. americký CBOLT věk kategorizuje na jednotlivé roky a využívá věkových kohort. Švédský model zase pracuje s věkem jako spojitou veličinou a pro model využívá i kvadrát věku. Věkové kohorty jsou vzhledem k délce časových řad nepoužitelné. Tvary křivek mezd i počtů pro muže a ženy se významně liší, proto bude zřejmě vhodnější přistoupit ke tvorbě dvou mzdových rovnic, zvlášť pro muže a ženy. Zařazení kombinace faktoru věku s dalšími faktory a kategoriemi je popsána v části 3.6.

Graf 2: Průměrný logaritmus mzdy podle věku a pohlaví, 2008-2014

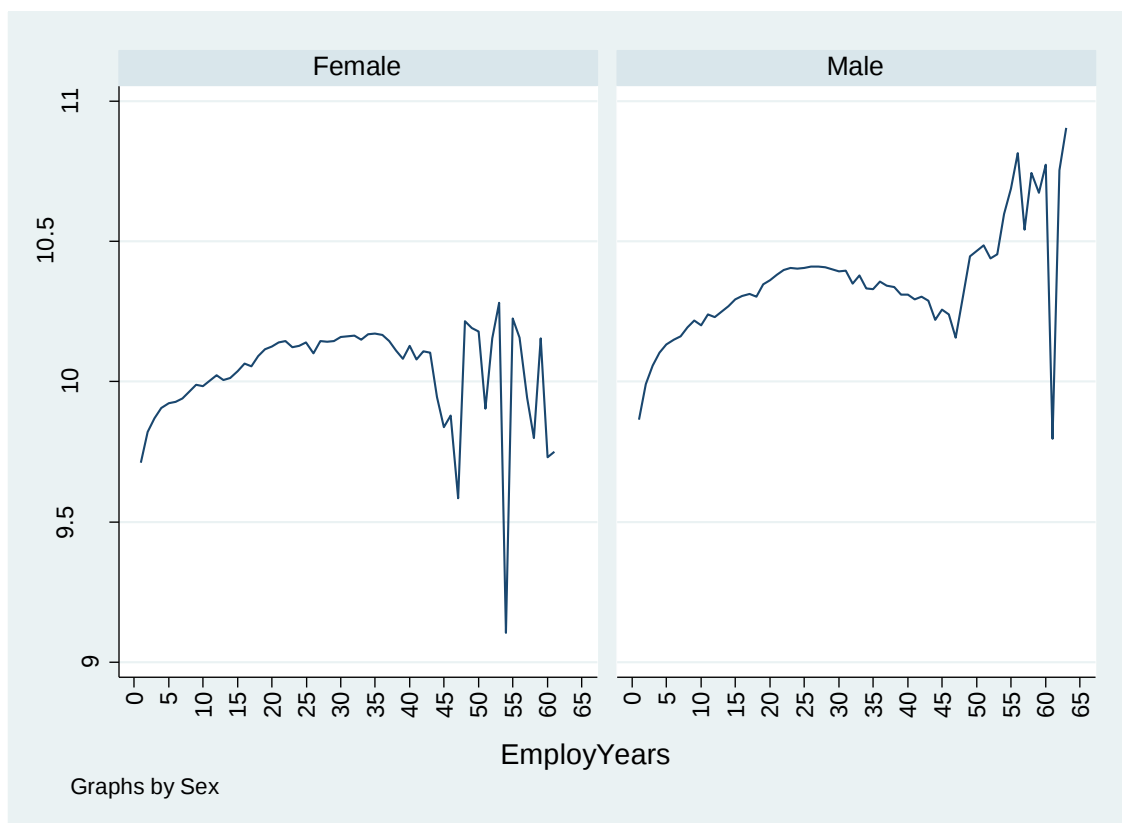


Graf 3: Vážený počet zaměstnanců podle věku a pohlaví, 2008-2014



Doba zaměstnání

Graf 4: Průměrný logaritmus mzdy podle doby zaměstnání a pohlaví, 2008-2014



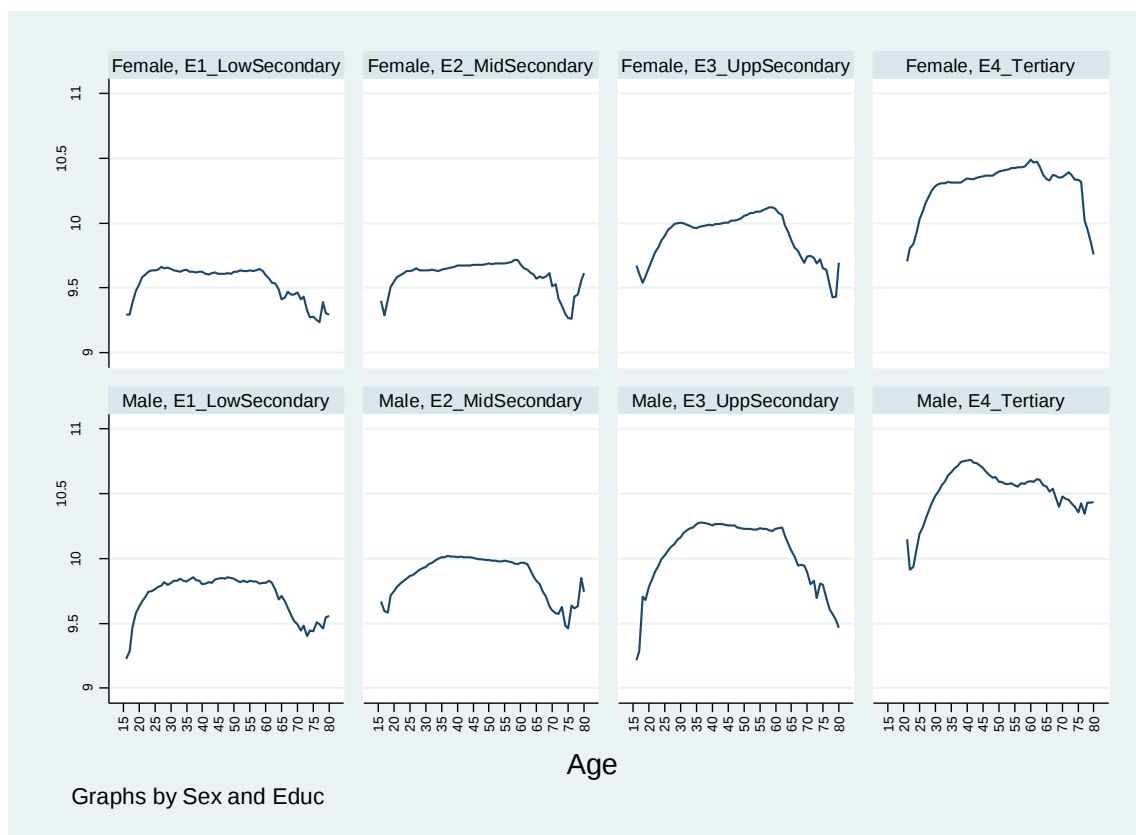
Podobně jako faktor věk je numerickým faktorem i doba zaměstnání u posledního zaměstnavatele. V grafu je vidět nelineárnost průběhu funkce podle daného faktoru. S daným faktorem nelze tudíž pracovat jako se spojitou veličinou a je nutné přistoupit ke kategorizaci dané proměnné. Pro nejvhodnější kategorizaci se použil postup, který je detailněji popsán v kapitole 3.5.

Vzdělání

Pro faktor vzdělání byla použita následující kategorizace: základní vzdělání (LowSecondary), střední vzdělání bez maturity (MidSecondary), střední vzdělání s maturitou (UppSecondary), vysokoškolské vzdělání (Tertiary).

Grafy ukazují značně rozdílný průběh funkce podle vzdělání. Z tohoto důvodu bylo nutné přistoupit k zařazení kombinace faktoru vzdělání s jinými faktory do modelu, více viz kapitola 3.6.

Graf 5: Průměrný logaritmus mzdy podle věku, vzdělání a pohlaví, 2008-2014



Klasifikace zaměstnání CZ-ISCO

U faktoru zaměstnání podle klasifikace CZ-ISCO se vycházelo z hlavních tříd CZ-ISCO, jejichž volba a struktura vychází z diskuse nad proměnnými popsané v části 3.2. Z hlediska variability je tato proměnná nejvýznamnější. Hlavní třída 6 není dostatečně zastoupená vzhledem ke své formální významnosti, proto byla sloučena s hlavní třídou 8., která má nejpodobnější mzdové a vzdělanostní charakteristiky. Následující grafy ukazují průběh funkce podle klasifikace CZ-ISCO a věku u žen a mužů.

Složení hlavních tříd CZ-ISCO:

- ISCO_0 – Zaměstnanci v ozbrojených silách,
- ISCO_1 – Zákonodárci a řídící pracovníci,

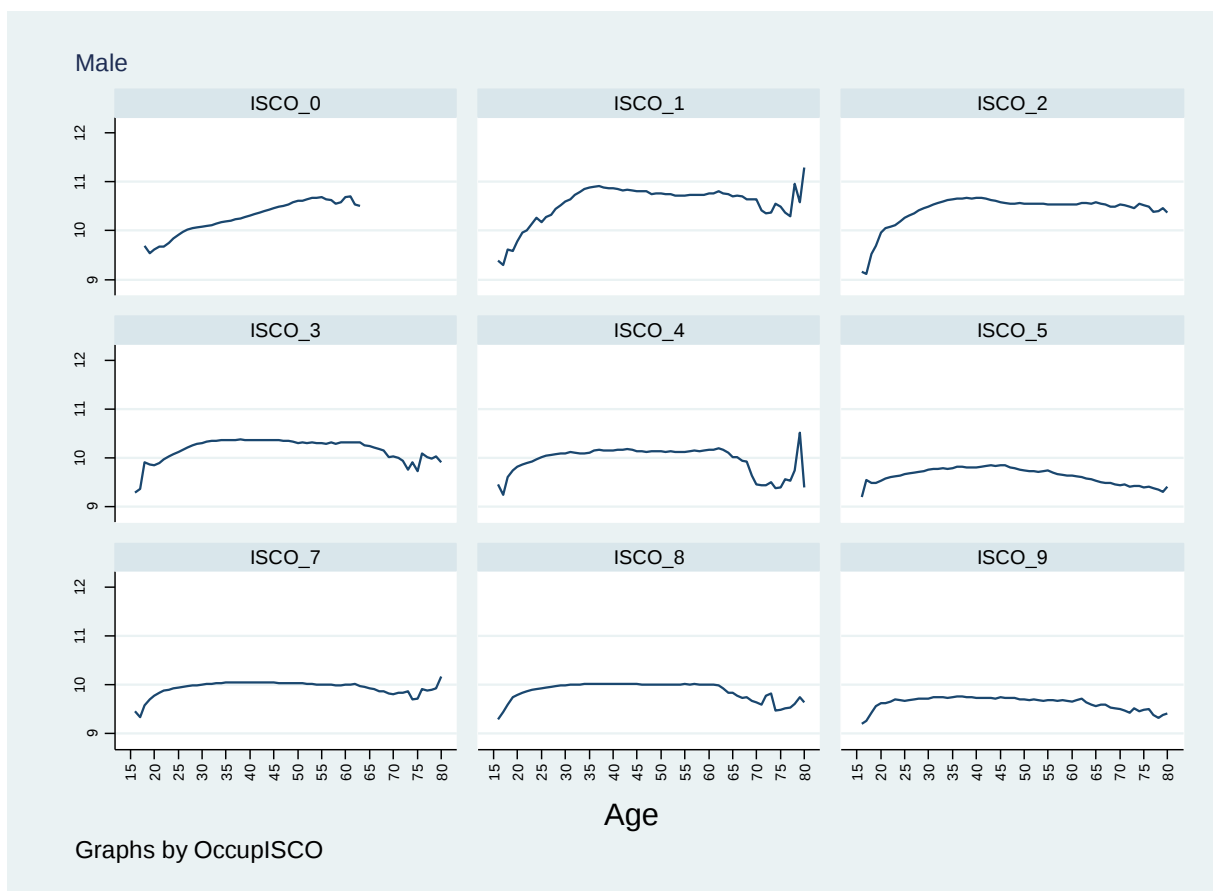
- ISCO_2 – Specialisté,
- ISCO_3 – Techničtí a odborní pracovníci,
- ISCO_4 – Úředníci,
- ISCO_5 – Pracovníci ve službách a prodeji,
- ISCO_7 – Řemeslníci a opraváři,
- ISCO_8 – Obsluha strojů a zařízení, montéři a kvalifikovaní pracovníci v zemědělství, lesnictví a rybářství,
- ISCO_9 – Pomocní a nekvalifikovaní pracovníci.

Graf 6: Průměrný logaritmus mzdy podle klasifikace CZ-ISCO, věku a pohlaví – ženy, 2008-2014



Na grafech 6 a 7 lze vidět, že tvary křivek jak pro ženy, tak pro muže jsou rozdílné také pro hlavní třídy zaměstnání klasifikace CZ-ISCO. Mužům zaměstnancům v ozbrojených silách (ISCO_0) roste mzda od 19 do 63 let téměř lineárně, kdežto u ostatních zaměstnání je zaznamenána spíše kvadratická funkce logaritmu mzdy s menšími vzájemnými rozdíly.

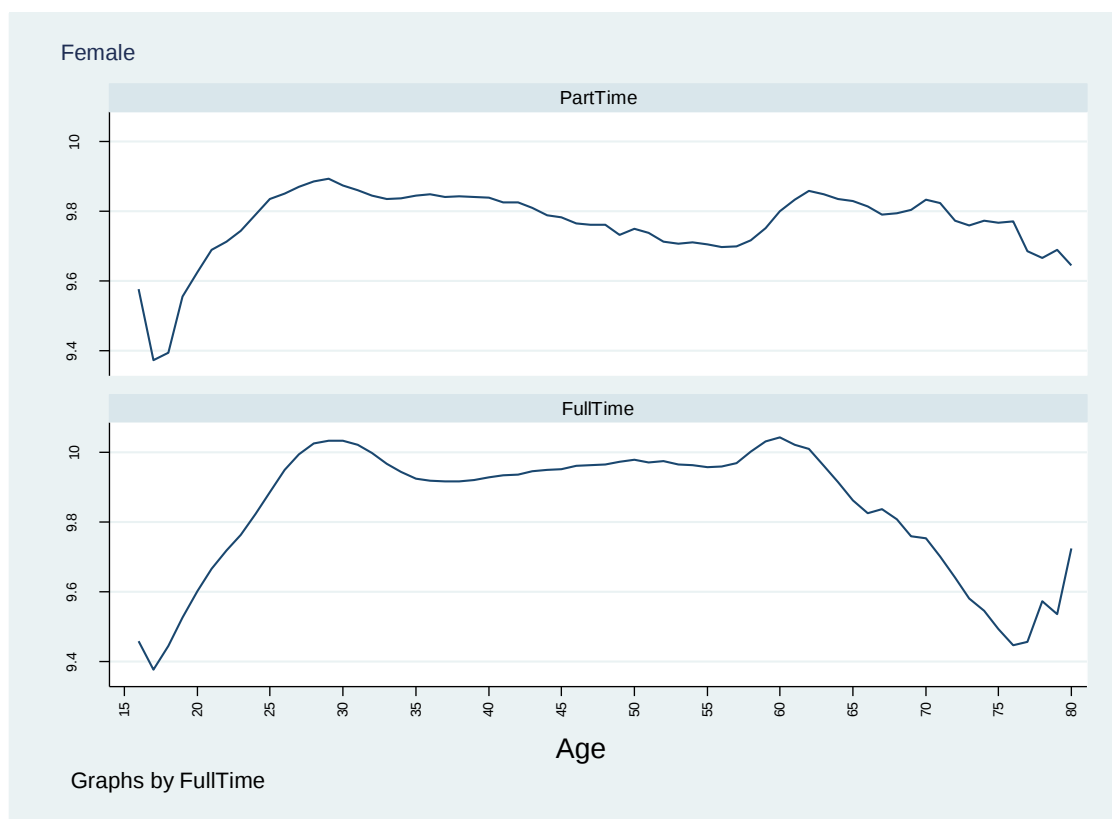
Graf 7: Průměrný logaritmus mzdy podle klasifikace CZ-ISCO, věku a pohlaví – muži, 2008-2014



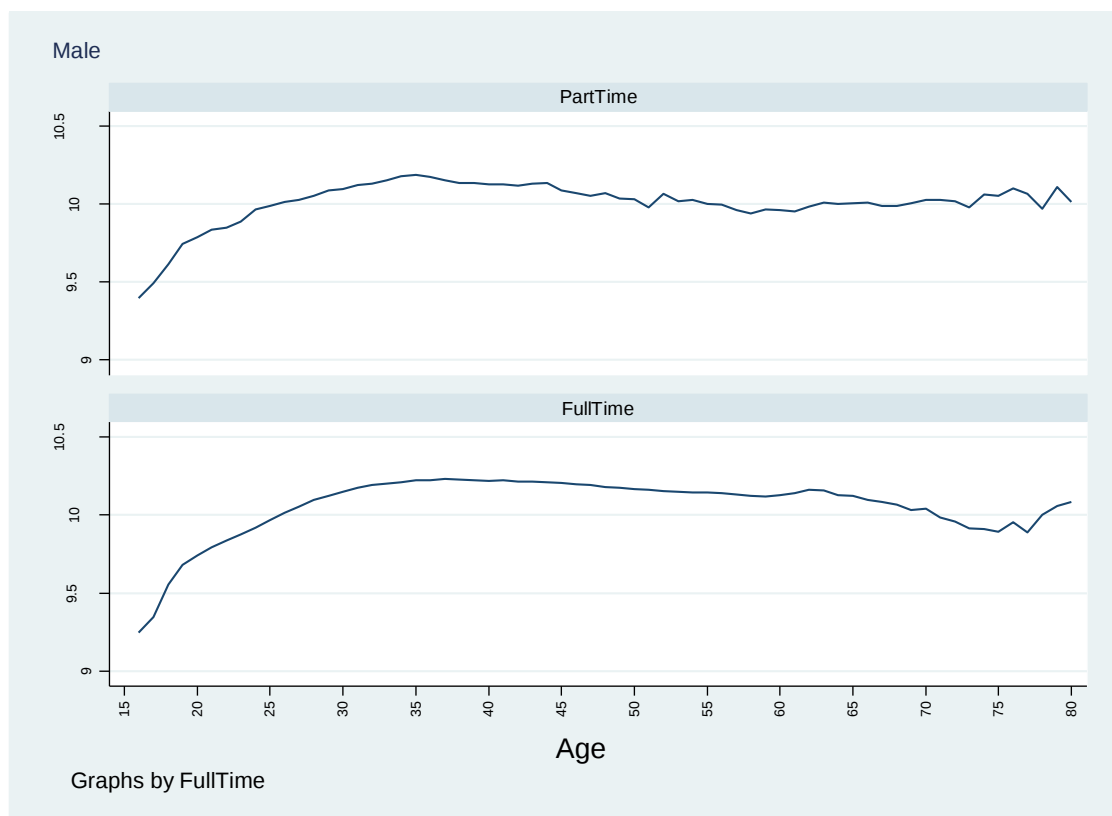
Úvazek

Úvazek byl zvolen jako binární proměnná – plný a částečný úvazek. V datech ISPV je zřejmý zlom v rozdělení dat na úrovni 95% úvazku, proto hodnoty pod 95% úvazkem jsou identifikovány jako úvazky částečné. Následující grafy ukazují rozdílný průběh funkce pro plný a pro zkrácený úvazek, stejně jako rozdílný průběh pro muže a ženy. U mužů se tvar funkce podle typu úvazku příliš neliší. U žen je rozdíl ve tvaru funkce logaritmu mzdy zřetelnější.

Graf 8: Průměrný logaritmus mzdy podle věku, úvazku a pohlaví – ženy, 2008-2014



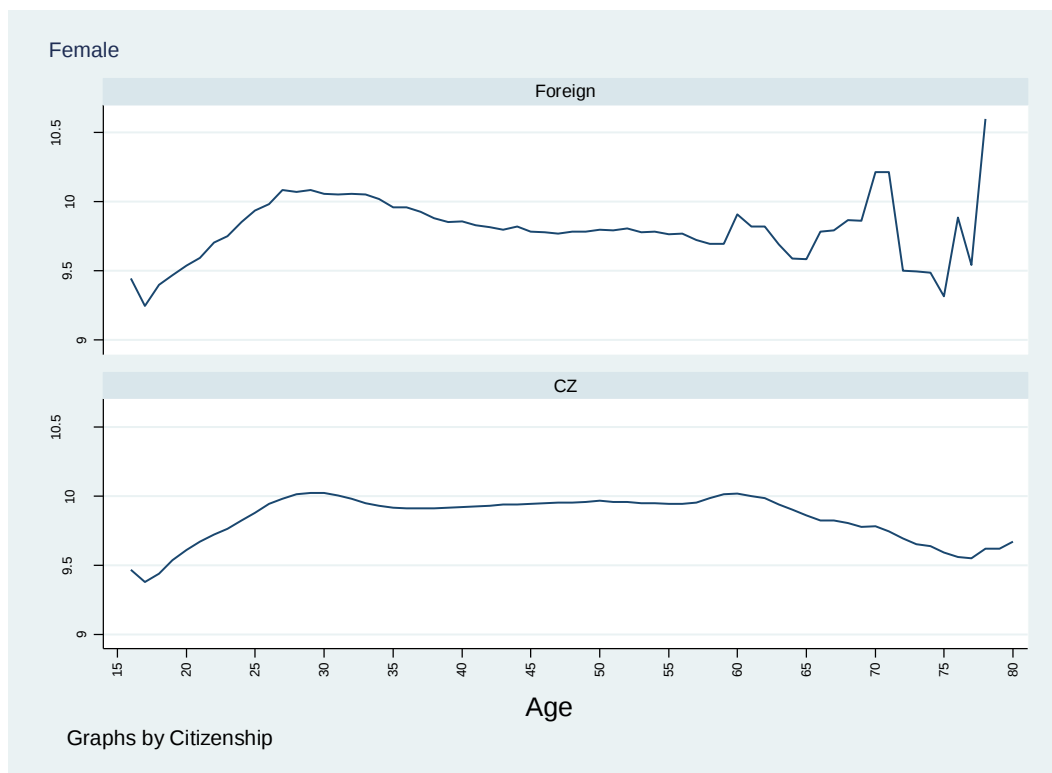
Graf 9: Průměrný logaritmus mzdy podle věku, úvazku a pohlaví – muži, 2008-2014



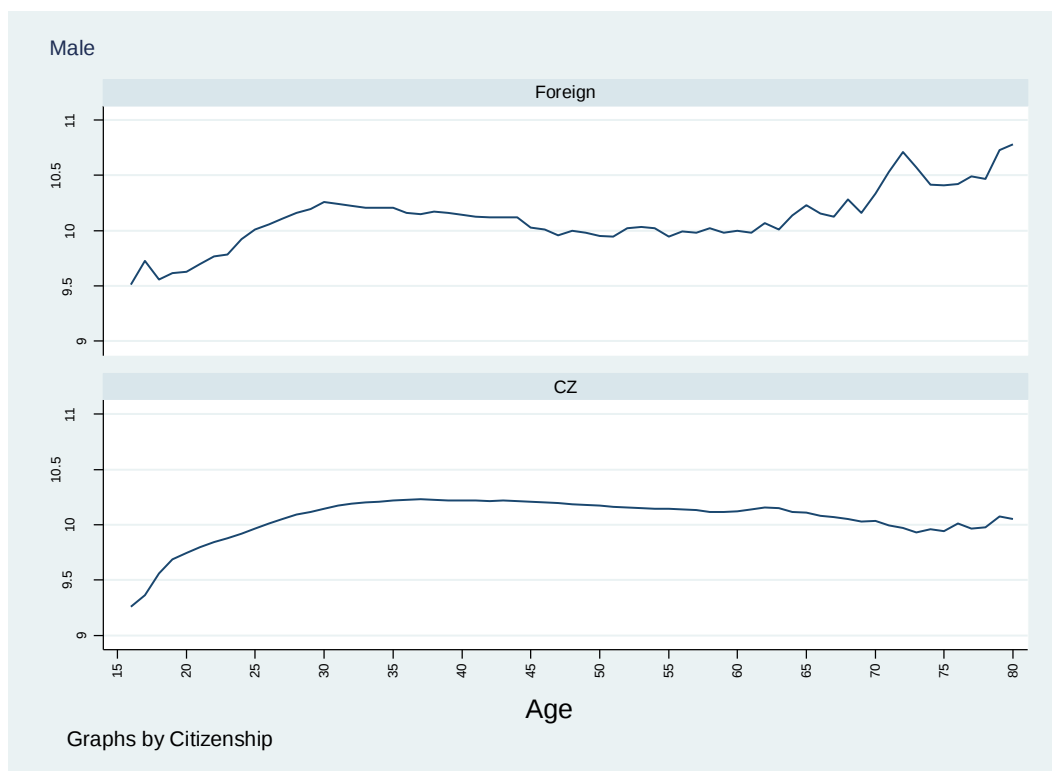
Státní příslušnost

Značně rozdílný průběh funkce pro české a cizí státní příslušníky implikuje také zařazení tohoto faktoru do modelu.

Graf 10 – Průměrný logaritmus mzdy podle věku, státního občanství a pohlaví – ženy, 2008-2014



Graf 11: Průměrný logaritmus mzdy podle věku, státního občanství a pohlaví – muži, 2008-2014



3.5 Kategorizace proměnných a požadavek na stabilitu mzdových rovnic

V případě faktorů věku a doby zaměstnání jsme přistoupili ke kroku kategorizace do několika skupin zejména z důvodů nelineární či nekadratické povahy proměnných. Pro maximální schopnost kategorizace byl použit jeden z nástrojů evolučních algoritmů, tj. universální metoda s označením diferenciální evoluce.[29]

Metoda genetického programování s označením diferenciální evoluce je založena na náhodné tvorbě rodičů (jako dostatečný počet je pro naše potřeby 10 rodičů), tvorbě dalších jedinců (potomků) vždy kombinací a mutací náhodně vybraných tří rodičů a řešeného rodiče a následné vyhodnocení přínosu potomka dle vložené účelové funkce, v našem případě přínosu nového složení spojitých křivek ve schopnosti popsat reálné hodnoty (regrese). Při výpočtech docházelo vždy ke konvergenci ke stejnému řešení během maximálně 100 generací.

Bez kategorizace proměnných není možné zajistit stabilitu modelu v každém roce, model by obsahoval velké množství koeficientů, které se značně liší např. mezi rokem 2008 a rokem 2014, což jsou počáteční a konečný rok datové základny. Po zvolené kategorizaci jsou modely dostatečně stabilní, více viz Průběžné modely kapitola 3.7.1.

Doba zaměstnání

Maximální schopnost vysvětlení variability závislé proměnné (lnWage) modelem s pouze jednou vysvětlující proměnnou doba zaměstnání byla 7 %, po nalezení optimálního počtu 5 kategorií (0-1, 1-3, 3-7, 7-15, 15 a více let zaměstnání u stejného zaměstnavatele) evoluční metodou se schopnost vysvětlení variability touto proměnnou téměř nezměnila (opět téměř 7 %), ale výrazně se, díky snížení počtu kategorií, zvýšila stabilita modelu.

Věk

Přístup k proměnné Věk byl kombinací švédského modelu (spojitá proměnná, kvadrát) a modelu CBOLT, který přistupoval k modelu pro každý rok samostatným koeficientem. V našem případě byla provedena prvotní analýza rozdělení proměnné věk do 6 kategorií a interakcí v regresi spojena opět s proměnnou věk, ale tentokrát však jako spojitou veličinou. Tato regrese byla určena jako účelová funkce do nástroje evolučního algoritmu, který našel kvalitnější složení křivek věku a rozdělil věk do 6 věkových intervalů zvlášť pro muže a ženy. Konečné věkové intervaly pro ženy jsou 15-28, 29-35, 36-52, 53-58, 58 a více let a muži 15-28, 29-34, 35-45, 46-60, 60 a více let.

Přístup „spojitých kategorií“ věku zajistil, že predikce průběhu věku aproximují skutečné hodnoty (viz Kapitola 3.8) a současně množství koeficientů není vysoké, což významně zvýšilo stabilitu modelů v čase.

3.6 Interakce faktorů

Z hlediska analýzy faktorů, schopnosti predikce a stavu reziduí testovaných modelů se jako významný požadavek ukázaly interakce faktorů vzdělání a zaměstnání (CZ-ISCO) s věkem. Konečná volba padla na dvě interakce tzv. spojitých kategorií.

Do první interakce vstupuje věk jako spojitá veličina, kategorie věku a současně kategorie vzdělání. Současně bylo také nutné do modelu zahrnout zvlášť faktor vzdělání, výsledky interakce viz kapitoly 3.7.2 a 3.7.3 - Model 1 a Model 2.

Do druhé uvažované interakce faktorů vstupuje opět věk jako spojitá veličina, kategorie věku a kategorie zaměstnání (CZ-ISCO). Současně však kategorie zaměstnání jako samostatný faktor do modelu nevstupuje z důvodu statisticky nevýznamných koeficientů, výsledky interakce viz kapitola 3.7.3 Model 2.

3.7 Tvorba mzdových rovnic

Pro modelování mezd byl použit klasický lineární model, který předpokládá normální rozdělení závislé proměnné přirozeného logaritmu hrubé měsíční mzdy ($\ln Wage$). Vysvětlujícími (nezávislými) proměnnými v modelu je typ úvazku ($FullTime$, FT) rozdělený na částečný ($PartTime$) a plný ($FullTime$), státní občanství ($Citizenship$, Cs) rozdělené na české (CZ) a cizí ($Foreign$), doba zaměstnání u posledního zaměstnavatele ($EmployYears$, EY) rozdělené do pěti kategorií, hlavní třídy zaměstnání CZ-ISCO ($OccupISCO$, OI) rozdělené do osmi kategorií, vzdělání ($Educ$, Ed) rozdělené do čtyř kategorií a věk (Age). Mzdy budou modelovány zvlášť pro muže a ženy.

3.7.1 Základní model

V průběhu tvorby modelů jsme testovali velké množství faktorů a jejich kombinací, které vedly k následujícímu základnímu modelu.

Věk jsme zpočátku do modelu zařadili jako kategoriální proměnnou pro jednotlivé roky. Základní model je pro obě pohlaví stejný a má tvar:

$$(1) \quad \ln Wage_i = \alpha_0 + \alpha_1 FT_{FullTime} + \alpha_2 Cs_{CZ} + \sum_{j=2}^5 \beta_j EY_j + \sum_{j=2}^9 \gamma_j OI_j + \sum_{j=2}^4 \delta_j Ed_j + \sum_{j=17}^{80} \varepsilon_j Age_j,$$

kde $i = 1, 2, \dots, n$. U kategoriálních faktorů jsou první kategorie vždy považovány za redundantní proměnné.

Obrázek 1: Výsledky základního modelu - muži

Survey: Linear regression

Number of strata = 139
Number of PSUs = 26,402

Number of obs = 15,379,721
Population size = 292,521,094
Subpop. no. obs = 7,615,707
Subpop. size = 161,583,441
Design df = 26,263
F(81, 26183) = 124.43
Prob > F = 0.0000
R-squared = 0.3788

$\ln Wage$	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
FullTime	.0499822	.0144256	3.46	0.001	.0217073	.0782571
Citizenship CZ	-.0059295	.0238703	-0.25	0.804	-.0527165	.0408575
EmployYears						
Em2_01-03	.1023604	.0098434	10.40	0.000	.0830668	.1216541
Em3_03-07	.1865236	.0105376	17.70	0.000	.1658693	.2071779
Em4_07-15	.2406521	.0122125	19.71	0.000	.2167148	.2645893
Em5_15-99	.3526192	.0156583	22.52	0.000	.321928	.3833104
OccupISCO						
ISCO_1	.5741938	.0167576	34.26	0.000	.5413479	.6070397
ISCO_2	.3392412	.0127145	26.68	0.000	.3143201	.3641623
ISCO_3	.1993455	.0104278	19.12	0.000	.1789065	.2197846
ISCO_4	.0605537	.0164108	3.69	0.000	.0283877	.0927197
ISCO_5	-.2331857	.0247855	-9.41	0.000	-.2817667	-.1846048
ISCO_7	.0408512	.0124863	3.27	0.001	.0163774	.065325
ISCO_8	.0364295	.0137776	2.64	0.008	.0094246	.0634344
ISCO_9	-.1776724	.0147121	-12.08	0.000	-.206509	-.1488358
Educ						
E2_MidSecondary	.0782591	.0111394	7.03	0.000	.0564254	.1000929
E3_UppSecondary	.183102	.0132285	13.84	0.000	.1571734	.2090306
E4_Tertiary	.387329	.0171754	22.55	0.000	.3536642	.4209938

Age						
17	.0470044	.0597098	0.79	0.431	-.0700301	.1640389
18	.1349432	.0371795	3.63	0.000	.0620694	.207817
19	.2076105	.0343751	6.04	0.000	.1402335	.2749875
20	.2192688	.0349895	6.27	0.000	.1506874	.2878501
21	.2325173	.035205	6.60	0.000	.1635135	.301521
22	.2457638	.0354447	6.93	0.000	.1762903	.3152373
23	.2529489	.0354873	7.13	0.000	.1833919	.3225059
24	.2512329	.0355004	7.08	0.000	.1816503	.3208156
25	.2586746	.035564	7.27	0.000	.1889673	.3283819
26	.262518	.0355635	7.38	0.000	.1928115	.3322245
27	.273278	.0356473	7.67	0.000	.2034073	.3431487
28	.2894546	.0357807	8.09	0.000	.2193225	.3595867
29	.3029731	.0358668	8.45	0.000	.2326722	.3732739
30	.3155274	.0359744	8.77	0.000	.2450156	.3860391
31	.3344018	.035945	9.30	0.000	.2639477	.404856
32	.3454215	.0359366	9.61	0.000	.2749837	.4158593
33	.353679	.0359195	9.85	0.000	.2832749	.4240831
34	.3620806	.0358486	10.10	0.000	.2918155	.4323457
35	.3677622	.0358005	10.27	0.000	.2975912	.4379331
36	.3682681	.0355943	10.35	0.000	.2985013	.4380349
37	.3692971	.0355553	10.39	0.000	.2996067	.4389875
38	.3628245	.0356305	10.18	0.000	.2929868	.4326622
39	.3573339	.0355638	10.05	0.000	.2876269	.427041
40	.3495534	.0356848	9.80	0.000	.2796093	.4194975
41	.3511965	.0356443	9.85	0.000	.2813318	.4210612
42	.3428584	.0356016	9.63	0.000	.2730773	.4126396
43	.3405298	.0359015	9.49	0.000	.270161	.4108986
44	.3361864	.0358823	9.37	0.000	.2658552	.4065176
45	.3266247	.0358696	9.11	0.000	.2563183	.3969311
46	.3184217	.0358124	8.89	0.000	.2482275	.3886159
47	.3075507	.0357817	8.60	0.000	.2374168	.3776847
48	.3002394	.035774	8.39	0.000	.2301205	.3703584
49	.2964047	.0357994	8.28	0.000	.2262359	.3665735
50	.2894539	.0358922	8.06	0.000	.2191032	.3598045
51	.2833293	.0359245	7.89	0.000	.2129154	.3537432
52	.2801316	.036014	7.78	0.000	.2095422	.350721
53	.2759916	.0359538	7.68	0.000	.2055203	.346463
54	.2742377	.0359668	7.62	0.000	.2037408	.3447346
55	.2742988	.0357676	7.67	0.000	.2041923	.3444054
56	.2665941	.0357711	7.45	0.000	.1964807	.3367074
57	.2662728	.035695	7.46	0.000	.1963086	.3362369
58	.257365	.0357347	7.20	0.000	.187323	.327407
59	.2570372	.0356976	7.20	0.000	.1870679	.3270064
60	.2596386	.0357493	7.26	0.000	.189568	.3297091
61	.2553375	.0358748	7.12	0.000	.1850209	.3256541
62	.2647887	.0361072	7.33	0.000	.1940167	.3355607
63	.2532315	.0365696	6.92	0.000	.181553	.3249099
64	.2250925	.0373387	6.03	0.000	.1519065	.2982784
65	.2126105	.0377453	5.63	0.000	.1386278	.2865933
66	.1933975	.0385795	5.01	0.000	.1177795	.2690155
67	.178483	.0405035	4.41	0.000	.099094	.257872
68	.1545158	.042505	3.64	0.000	.0712037	.237828
69	.0993999	.0448128	2.22	0.027	.0115643	.1872354
70	.1042282	.0457387	2.28	0.023	.014578	.1938785
71	.0522673	.0498551	1.05	0.294	-.0454513	.1499859
72	.064646	.0546487	1.18	0.237	-.0424683	.1717604
73	.0518357	.0628612	0.82	0.410	-.0713756	.175047
74	.0410535	.058163	0.71	0.480	-.0729492	.1550561
75	.0115181	.0633121	0.18	0.856	-.112577	.1356132
76	.0789109	.0524614	1.50	0.133	-.0239163	.1817382
77	.0246845	.0540532	0.46	0.648	-.0812627	.1306317
78	.0639649	.0568396	1.13	0.260	-.0474438	.1753736
79	.0889053	.0689214	1.29	0.197	-.0461843	.223995
80	.0723727	.0847438	0.85	0.393	-.0937298	.2384752
_cons	9.328577	.0438908	212.54	0.000	9.242549	9.414606

Pro muže vychází téměř všechny faktory statisticky významné (na 5 % hladině významnosti). Statisticky nevýznamné je státní občanství a věkové kategorie pro věk 17 let a 71 a starší. Variability závislé proměnné (lnWage) bylo vysvětleno z 37,88 %.

Ženy mají všechny proměnné kromě věku statisticky významné. Ovšem kategorie věku jsou téměř všechny statisticky nevýznamné. Variabilita logaritmu mzdy je vysvětlena ze 43,89 %. Kategoriaální faktor věku se v této podobě zdá být pro základní model problematický. Proto pro další modely využijeme kombinace spojitého věku a předem zvolených kategorií věku.

Obrázek 2: Výsledky základního modelu - ženy

Survey: Linear regression

Number of strata = 139
Number of PSUs = 26,402

Number of obs = 15,379,721
Population size = 292,521,094
Subpop. no. obs = 7,764,014
Subpop. size = 130,937,654
Design df = 26,263
F(81, 26183) = 162.77
Prob > F = 0.0000
R-squared = 0.4389

lnWage	Linearized		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
FullTime	.042163	.0082893	5.09	0.000	.0259156	.0584104
Citizenship						
CZ	-.0522906	.0166386	-3.14	0.002	-.0849031	-.019678
EmployYears						
Em2_01-03	.0894934	.0080499	11.12	0.000	.0737151	.1052718
Em3_03-07	.1539026	.0081794	18.82	0.000	.1378706	.1699347
Em4_07-15	.1933286	.0091205	21.20	0.000	.1754519	.2112053
Em5_15-99	.2426262	.0114477	21.19	0.000	.2201881	.2650642
OccupISCO						
ISCO_1	.3335832	.0196112	17.01	0.000	.2951441	.3720223
ISCO_2	.083505	.0098371	8.49	0.000	.0642238	.1027863
ISCO_3	-.0036578	.0089427	-0.41	0.683	-.021186	.0138705
ISCO_4	-.1423228	.0111535	-12.76	0.000	-.1641843	-.1204614
ISCO_5	-.3510345	.013459	-26.08	0.000	-.3774149	-.3246541
ISCO_7	-.2683035	.015436	-17.38	0.000	-.2985589	-.2380481
ISCO_8	-.1817511	.0127857	-14.22	0.000	-.2068117	-.1566905
ISCO_9	-.4528769	.0122178	-37.07	0.000	-.4768244	-.4289293
Educ						
E2_MidSecondary	.0185714	.0056579	3.28	0.001	.0074816	.0296612
E3_UppSecondary	.1438789	.0080085	17.97	0.000	.1281817	.1595761
E4_Tertiary	.3240662	.0119391	27.14	0.000	.3006648	.3474676

Age						
17	-.0774595	.0194766	-3.98	0.000	-.1156347	-.0392843
18	-.0432199	.0320744	-1.35	0.178	-.1060875	.0196476
19	-.0074559	.0338115	-0.22	0.825	-.0737282	.0588164
20	.0027449	.0341357	0.08	0.936	-.0641628	.0696527
21	.0126274	.034531	0.37	0.715	-.0550552	.08031
22	.0196068	.0346777	0.57	0.572	-.0483634	.087577
23	.0107165	.0349242	0.31	0.759	-.0577368	.0791698
24	.0136955	.0348762	0.39	0.695	-.0546639	.0820548
25	.0213046	.0349824	0.61	0.543	-.0472629	.0898721
26	.0388229	.0349519	1.11	0.267	-.0296847	.1073305
27	.0556895	.0349778	1.59	0.111	-.0128688	.1242479
28	.0744767	.0350671	2.12	0.034	.0057434	.1432101
29	.0858688	.0350143	2.45	0.014	.0172389	.1544987
30	.0950819	.0349953	2.72	0.007	.0264892	.1636746
31	.0929963	.0349559	2.66	0.008	.0244809	.1615117
32	.0866286	.0349869	2.48	0.013	.0180525	.1552048
33	.0751353	.0349873	2.15	0.032	.0065582	.1437124
34	.0651119	.0348993	1.87	0.062	-.0032927	.1335165
35	.056996	.0348459	1.64	0.102	-.0113037	.1252958
36	.0543041	.0348338	1.56	0.119	-.0139721	.1225802
37	.0519766	.0347783	1.49	0.135	-.0161907	.1201439
38	.0544398	.0347628	1.57	0.117	-.0136972	.1225768
39	.0575015	.0347579	1.65	0.098	-.0106259	.1256289
40	.0556152	.0347063	1.60	0.109	-.0124111	.1236415
41	.0566967	.0347099	1.63	0.102	-.0113365	.1247299
42	.054081	.03468	1.56	0.119	-.0138937	.1220558
43	.0554493	.0346944	1.60	0.110	-.0125535	.1234522
44	.0538247	.0347109	1.55	0.121	-.0142106	.1218601
45	.0513487	.0347213	1.48	0.139	-.016707	.1194044
46	.0536847	.0347562	1.54	0.122	-.0144393	.1218087
47	.0531018	.0347836	1.53	0.127	-.0150759	.1212796
48	.0541429	.0347766	1.56	0.120	-.0140211	.1223069
49	.0591939	.0347889	1.70	0.089	-.0089942	.127382
50	.0685099	.0347619	1.97	0.049	.0003747	.136645
51	.0648174	.0347485	1.87	0.062	-.0032916	.1329264
52	.070367	.0347458	2.03	0.043	.0022633	.1384707
53	.0691187	.0347712	1.99	0.047	.0009652	.1372721
54	.0711629	.0347833	2.05	0.041	.0029857	.1393401
55	.071714	.0348063	2.06	0.039	.0034917	.1399362
56	.0726654	.0348399	2.09	0.037	.0043773	.1409535
57	.0748724	.0348455	2.15	0.032	.0065734	.1431714
58	.0863718	.0349359	2.47	0.013	.0178955	.154848
59	.0954552	.0351167	2.72	0.007	.0266245	.1642859
60	.0990904	.0353796	2.80	0.005	.0297446	.1684363
61	.0861541	.0358413	2.40	0.016	.0159032	.156405
62	.0808571	.0363983	2.22	0.026	.0095145	.1521997
63	.0469945	.0377249	1.25	0.213	-.0269483	.1209374
64	.0142199	.0392566	0.36	0.717	-.0627251	.0911649
65	-.0238283	.0408124	-0.58	0.559	-.1038228	.0561662
66	-.0479495	.0411291	-1.17	0.244	-.1285648	.0326658
67	-.0328288	.04136	-0.79	0.427	-.1138966	.0482389
68	-.0423377	.0462394	-0.92	0.360	-.1329695	.0482941
69	-.0492187	.0494087	-1.00	0.319	-.1460623	.047625
70	-.0361134	.0502849	-0.72	0.473	-.1346746	.0624477
71	-.0399028	.0518795	-0.77	0.442	-.1415895	.0617839
72	-.0528107	.0565543	-0.93	0.350	-.1636602	.0580388
73	-.1018193	.0578085	-1.76	0.078	-.215127	.0114885
74	-.1245696	.0592698	-2.10	0.036	-.2407415	-.0083976
75	-.1877511	.0647353	-2.90	0.004	-.3146358	-.0608664
76	-.1956767	.0698556	-2.80	0.005	-.3325976	-.0587559
77	-.2394534	.0907948	-2.64	0.008	-.4174161	-.0614907
78	-.2567441	.1057474	-2.43	0.015	-.4640148	-.0494734
79	-.2478724	.1110142	-2.23	0.026	-.4654663	-.0302784
80	-.3168599	.1594979	-1.99	0.047	-.6294845	-.0042353
_cons	9.733319	.0404902	240.39	0.000	9.653956	9.812682

3.7.2 Model 1

Model 1 pro ženy a muže obsahuje kromě faktorů použitých v prvním modelu i interakci předem zvolených kategorií věku pro ženy (Age_F) nebo muže (Age_M), vzdělání (Ed) a spojitého věku (Age). Kombinaci věku a vzdělání jsme použili na základě grafického znázornění průměrného logaritmu mzdy podle věku a vzdělání (graf 5), kde byly vidět rozdílné tvary křivek pro jednotlivá vzdělání. Využitím této interakce dojdeme k rozdílným křivkám ve vzdělání, které právě indikuje uvedený graf.

Muži

$$(2) \quad \ln \widehat{Wage}_i^{Male} = \alpha_0 + \alpha_1 FT_{FullTime} + \alpha_2 CS_{CZ} + \sum_{j=2}^5 \beta_j EY_j + \sum_{j=2}^9 \gamma_j OI_j + \sum_{j=2}^5 \varepsilon_j Age_M_j + \varphi_i Age_i + \sum_{j=2}^{20} \epsilon_j Age_M_j Ed_j Age_i,$$

kde $i = 1, 2, \dots, n$.

Obrázek 3: Výsledky Modelu 1 – muži

Survey: Linear regression

Number of strata = 139
Number of PSUs = 26,402

Number of obs = 15,379,721
Population size = 292,521,094
Subpop. no. obs = 7,615,707
Subpop. size = 161,583,441
Design df = 26,263
F(38, 26226) = 254.73
Prob > F = 0.0000
R-squared = 0.3813

lnWage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
FullTime	.049319	.0144128	3.42	0.001	.0210693	.0775688
Citizenship CZ	-.007463	.024018	-0.31	0.756	-.0545395	.0396135
EmployYears						
Em2_01-03	.1026414	.0098263	10.45	0.000	.0833813	.1219014
Em3_03-07	.1824379	.0105492	17.29	0.000	.1617609	.2031149
Em4_07-15	.2355706	.0122208	19.28	0.000	.2116171	.2595241
Em5_15-99	.3492667	.0157119	22.23	0.000	.3184704	.3800629
OccupISCO						
ISCO_1	.5723464	.0167955	34.08	0.000	.5394263	.6052664
ISCO_2	.3411875	.012825	26.60	0.000	.3160497	.3663253
ISCO_3	.2023065	.0104234	19.41	0.000	.1818761	.2227369
ISCO_4	.0635183	.0165662	3.83	0.000	.0310477	.0959889
ISCO_5	-.2328937	.0248019	-9.39	0.000	-.2815068	-.1842807
ISCO_7	.0421127	.0124306	3.39	0.001	.017748	.0664774
ISCO_8	.0379632	.0137295	2.77	0.006	.0110527	.0648736
ISCO_9	-.1790327	.0146606	-12.21	0.000	-.2077683	-.150297
Age_Male						
29-34	.081957	.0594299	1.38	0.168	-.0345289	.1984429
35-45	.6266933	.0491273	12.76	0.000	.5304011	.7229854
46-60	.6083595	.0501523	12.13	0.000	.5100584	.7066607
60+	1.427922	.1490934	9.58	0.000	1.135691	1.720153
Age	.0106304	.0014329	7.42	0.000	.0078219	.0134389
Age_Male#Educ#c.Age						
15-28#E2_MidSecondary	.0019436	.0007206	2.70	0.007	.0005312	.0033561
15-28#E3_UppSecondary	.0048634	.0007833	6.21	0.000	.0033281	.0063987
15-28#E4_Tertiary	.0082017	.0008984	9.13	0.000	.0064409	.0099626
29-34#E1_LowSecondary	-.0045386	.0021498	-2.11	0.035	-.0087524	-.0003248
29-34#E2_MidSecondary	-.0019666	.0021724	-0.91	0.365	-.0062247	.0022914
29-34#E3_UppSecondary	.0018699	.0021591	0.87	0.386	-.002362	.0061017
29-34#E4_Tertiary	.0078616	.0022115	3.55	0.000	.003527	.0121962
35-45#E1_LowSecondary	-.0199852	.0016888	-11.83	0.000	-.0232953	-.0166751
35-45#E2_MidSecondary	-.0173944	.001673	-10.40	0.000	-.0206735	-.0141153
35-45#E3_UppSecondary	-.0143629	.0016782	-8.56	0.000	-.0176522	-.0110736
35-45#E4_Tertiary	-.0074085	.0016932	-4.38	0.000	-.0107273	-.0040897
46-60#E1_LowSecondary	-.0176795	.00161	-10.98	0.000	-.0208353	-.0145238
46-60#E2_MidSecondary	-.0161784	.0016095	-10.05	0.000	-.0193331	-.0130237
46-60#E3_UppSecondary	-.0142856	.0016206	-8.82	0.000	-.017462	-.0111092
46-60#E4_Tertiary	-.0110535	.0016072	-6.88	0.000	-.0142037	-.0079033
60+ #E1_LowSecondary	-.0302395	.0027043	-11.18	0.000	-.03554	-.0249389
60+ #E2_MidSecondary	-.0290044	.0027371	-10.60	0.000	-.0343692	-.0236395
60+ #E3_UppSecondary	-.0279314	.0027394	-10.20	0.000	-.0333009	-.022562
60+ #E4_Tertiary	-.0240832	.0027311	-8.82	0.000	-.0294363	-.01873
_cons	9.385815	.0439405	213.60	0.000	9.29969	9.471941

Ženy

$$(3) \quad \ln \widehat{Wage}_i^{Female} = \alpha_0 + \alpha_1 FT_{FullTime} + \alpha_2 CS_{CZ} + \sum_{j=2}^5 \beta_j EY_j + \sum_{j=2}^9 \gamma_j Ol_j + \sum_{j=2}^5 \varepsilon_j Age_F_j + \varphi_i Age_i + \sum_{j=2}^{20} \epsilon_j Age_F_j Ed_j Age_i,$$

kde $i=1,2,...,n$.

Obrázek 4: Výsledky Modelu 1 - ženy

Number of strata	=	139	Number of obs	=	15,379,721
Number of PSUs	=	26,402	Population size	=	292,521,094
			Subpop. no. obs	=	7,764,014
			Subpop. size	=	130,937,654
			Design df	=	26,263
			F(38, 26226)	=	363.29
			Prob > F	=	0.0000
			R-squared	=	0.4413

lnWage	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
FullTime	.0415138	.0082235	5.05	0.000	.0253953	.0576323
Citizenship CZ	-.0508556	.0166396	-3.06	0.002	-.0834701	-.0182411
EmployYears						
Em2_01-03	.0890841	.0079872	11.15	0.000	.0734287	.1047394
Em3_03-07	.1491446	.0081311	18.34	0.000	.1332072	.165082
Em4_07-15	.1888458	.009083	20.79	0.000	.1710426	.206649
Em5_15-99	.2366918	.0115649	20.47	0.000	.2140239	.2593598
OccupISCO						
ISCO_1	.3171902	.0199088	15.93	0.000	.2781679	.3562125
ISCO_2	.0717537	.0102577	7.00	0.000	.051648	.0918593
ISCO_3	-.0093626	.009012	-1.04	0.299	-.0270267	.0083014
ISCO_4	-.1499319	.0111736	-13.42	0.000	-.1718327	-.1280312
ISCO_5	-.3615284	.0134246	-26.93	0.000	-.3878414	-.3352154
ISCO_7	-.2779395	.0154364	-18.01	0.000	-.3081956	-.2476834
ISCO_8	-.1926444	.0127595	-15.10	0.000	-.2176537	-.1676352
ISCO_9	-.4603918	.0121297	-37.96	0.000	-.4841667	-.4366169
Age_Female						
29-35	.6610669	.0562418	11.75	0.000	.55083	.7713038
36-52	.403154	.0376428	10.71	0.000	.3293721	.4769359
53-58	.2947512	.060059	4.91	0.000	.1770322	.4124701
58+	1.49951	.1395377	10.75	0.000	1.226008	1.773011
Age	.0145642	.0013103	11.11	0.000	.0119959	.0171325
Age_Female#Educ#c.Age						
15-28#E2_MidSecondary	-.0005809	.0005221	-1.11	0.266	-.0016043	.0004426
15-28#E3_UppSecondary	.0036188	.0005549	6.52	0.000	.0025312	.0047063
15-28#E4_Tertiary	.006228	.0006457	9.64	0.000	.0049624	.0074937
29-35#E1_LowSecondary	-.0240781	.0020042	-12.01	0.000	-.0280065	-.0201497
29-35#E2_MidSecondary	-.0248605	.0019965	-12.45	0.000	-.0287737	-.0209473
29-35#E3_UppSecondary	-.0202202	.0019835	-10.19	0.000	-.024108	-.0163324
29-35#E4_Tertiary	-.0143587	.0019954	-7.20	0.000	-.0182699	-.0104476
36-52#E1_LowSecondary	-.0165054	.0013945	-11.84	0.000	-.0192387	-.0137721
36-52#E2_MidSecondary	-.0161158	.0014026	-11.49	0.000	-.018865	-.0133665
36-52#E3_UppSecondary	-.013607	.0013896	-9.79	0.000	-.0163308	-.0108833
36-52#E4_Tertiary	-.0089942	.001399	-6.43	0.000	-.0117364	-.0062521
53-58#E1_LowSecondary	-.0143634	.0015974	-8.99	0.000	-.0174943	-.0112325
53-58#E2_MidSecondary	-.0139276	.0016003	-8.70	0.000	-.0170643	-.0107909
53-58#E3_UppSecondary	-.0111089	.0015931	-6.97	0.000	-.0142313	-.0079864
53-58#E4_Tertiary	-.0076707	.0016117	-4.76	0.000	-.0108296	-.0045117
58+ #E1_LowSecondary	-.0340642	.0025772	-13.22	0.000	-.0391156	-.0290128
58+ #E2_MidSecondary	-.0335227	.0025783	-13.00	0.000	-.0385764	-.0284691
58+ #E3_UppSecondary	-.0317483	.0026354	-12.05	0.000	-.0369138	-.0265829
58+ #E4_Tertiary	-.0273488	.002586	-10.58	0.000	-.0324175	-.0222801
_cons	9.491293	.0379521	250.09	0.000	9.416905	9.565681

Model 1 pro ženy až na kategorii charakterizující technické a odborné pracovníky (ISCO_3) a interakci popisující vyučené ženy ve věku od 15 do 28 let má všechny použité proměnné statisticky významné. Variabilita závislé proměnné (lnWage) je vysvětlena ze 44,13 %. Model 1 pro ženy je tedy vhodnější než základní model, viz 3.7.1.

3.7.3 Model 2

V Modelu 2 byly zohledněny rozdílné tvary křivek pro kategorie indikující hlavní třídu zaměstnání CZ-ISCO.

Muži

$$(4) \ln \widehat{Wage}_i^{Male} = \alpha_0 + \alpha_1 FT_{FullTime} + \alpha_2 CS_{CZ} + \sum_{j=2}^5 \beta_j EY_j + \sum_{j=2}^5 \epsilon_j Age_M_j + \varphi_i Age_i + \sum_{j=2}^{20} \epsilon_j Age_M_j Ed_j Age_i + \sum_{j=2}^{45} \omega_j Age_M_j Ol_j Age_i,$$

kde $i=1,2,...,n$.

Obrázek 5: Výsledky Modelu 2 - muži

Survey: Linear regression

Number of strata	=	139	Number of obs	=	15,379,721
Number of PSUs	=	26,402	Population size	=	292,521,094
			Subpop. no. obs	=	7,615,707
			Subpop. size	=	161,583,441
			Design df	=	26,263
			F(70, 26194)	=	354548.77
			Prob > F	=	0.0000
			R-squared	=	0.3816

lnWage	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
FullTime	.0476166	.0143387	3.32	0.001	.019512	.0757212
Citizenship CZ	-.0056091	.0241468	-0.23	0.816	-.052938	.0417199
EmployYears						
Em2_01-03	.103509	.009796	10.57	0.000	.0843083	.1227097
Em3_03-07	.1835079	.0105002	17.48	0.000	.1629268	.2040889
Em4_07-15	.2356771	.0121678	19.37	0.000	.2118275	.2595267
Em5_15-99	.3473465	.0157596	22.04	0.000	.3164567	.3782362
Age_Male						
29-34	.2029028	.0585736	3.46	0.001	.0880954	.3177102
35-45	.7565511	.0489968	15.44	0.000	.6605148	.8525875
46-60	.7224279	.0489961	14.74	0.000	.6263928	.818463
60+	1.535709	.1516758	10.12	0.000	1.238416	1.833002
Age	.0069229	.001521	4.55	0.000	.0039417	.0099041
Age_Male#Educ#c.Age						
15-28#E2_MidSecondary	.0023257	.0007089	3.28	0.001	.0009363	.0037152
15-28#E3_UppSecondary	.0062031	.0007824	7.93	0.000	.0046697	.0077366
15-28#E4_Tertiary	.010384	.0008871	11.71	0.000	.0086452	.0121228
29-34#E1_LowSecondary	-.010047	.0021947	-4.58	0.000	-.0143487	-.0057452
29-34#E2_MidSecondary	-.0074417	.0022234	-3.35	0.001	-.0117998	-.0030836
29-34#E3_UppSecondary	-.0037136	.0022637	-1.64	0.101	-.0081506	.0007235
29-34#E4_Tertiary	.0020656	.0023736	0.87	0.384	-.0025868	.0067179
35-45#E1_LowSecondary	-.0237038	.0017631	-13.44	0.000	-.0271596	-.020248
35-45#E2_MidSecondary	-.0211403	.0017543	-12.05	0.000	-.0245789	-.0177017
35-45#E3_UppSecondary	-.0183028	.001777	-10.30	0.000	-.0217859	-.0148198
35-45#E4_Tertiary	-.0115443	.0018118	-6.37	0.000	-.0150956	-.0079931
46-60#E1_LowSecondary	-.0195433	.0016531	-11.82	0.000	-.0227834	-.0163032
46-60#E2_MidSecondary	-.0181243	.001658	-10.93	0.000	-.0213742	-.0148745
46-60#E3_UppSecondary	-.0162348	.0016826	-9.65	0.000	-.0195328	-.0129367
46-60#E4_Tertiary	-.0128743	.0017048	-7.55	0.000	-.0162157	-.0095328
60+ #E1_LowSecondary	-.0308633	.0027789	-11.11	0.000	-.0363102	-.0254165
60+ #E2_MidSecondary	-.0298467	.0028164	-10.60	0.000	-.0353669	-.0243264
60+ #E3_UppSecondary	-.0292616	.0028121	-10.41	0.000	-.0347735	-.0237496
60+ #E4_Tertiary	-.0261076	.0028351	-9.21	0.000	-.0316645	-.0205507

Age_Male#OccupISCO#c.Age						
15-28#ISCO_1	.0146865	.0017839	8.23	0.000	.01119	.018183
15-28#ISCO_2	.0146597	.0006613	22.17	0.000	.0133635	.0159558
15-28#ISCO_3	.0098532	.0004274	23.05	0.000	.0090155	.0106909
15-28#ISCO_4	.0062179	.0006827	9.11	0.000	.0048798	.007556
15-28#ISCO_5	-.0057385	.0010732	-5.35	0.000	-.0078422	-.0036349
15-28#ISCO_7	.0051834	.0005654	9.17	0.000	.0040751	.0062916
15-28#ISCO_8	.005524	.00056	9.86	0.000	.0044264	.0066217
15-28#ISCO_9	-.0024673	.0008158	-3.02	0.002	-.0040663	-.0008682
29-34#ISCO_0	.0048002	.0006475	7.41	0.000	.0035311	.0060693
29-34#ISCO_1	.0231108	.001042	22.18	0.000	.0210684	.0251532
29-34#ISCO_2	.0169804	.0009143	18.57	0.000	.0151883	.0187724
29-34#ISCO_3	.0124349	.0007584	16.40	0.000	.0109485	.0139213
29-34#ISCO_4	.0067555	.0011175	6.05	0.000	.0045652	.0089458
29-34#ISCO_5	-.0017298	.0010683	-1.62	0.105	-.0038238	.0003641
29-34#ISCO_7	.0067998	.0006843	9.94	0.000	.0054585	.008141
29-34#ISCO_8	.0070337	.0006742	10.43	0.000	.0057123	.0083551
29-34#ISCO_9	0	(omitted)				
35-45#ISCO_0	.0048657	.0005054	9.63	0.000	.0038751	.0058564
35-45#ISCO_1	.0197646	.0008512	23.22	0.000	.0180961	.021433
35-45#ISCO_2	.0132525	.0006685	19.83	0.000	.0119423	.0145627
35-45#ISCO_3	.0095784	.0005633	17.00	0.000	.0084742	.0106826
35-45#ISCO_4	.0059828	.0006749	8.87	0.000	.00466	.0073056
35-45#ISCO_5	-.0008499	.0008024	-1.06	0.290	-.0024227	.000723
35-45#ISCO_7	.0054789	.0004676	11.72	0.000	.0045625	.0063954
35-45#ISCO_8	.0051623	.0005029	10.26	0.000	.0041766	.0061481
35-45#ISCO_9	0	(omitted)				
46-60#ISCO_0	.0094429	.0004766	19.81	0.000	.0085087	.010377
46-60#ISCO_1	.0142026	.0004846	29.31	0.000	.0132527	.0151525
46-60#ISCO_2	.0098154	.0004395	22.33	0.000	.008954	.0106769
46-60#ISCO_3	.0075685	.0003515	21.53	0.000	.0068796	.0082574
46-60#ISCO_4	.005159	.0004754	10.85	0.000	.0042272	.0060908
46-60#ISCO_5	-.0008454	.000435	-1.94	0.052	-.001698	7.25e-06
46-60#ISCO_7	.0044939	.0002528	17.78	0.000	.0039984	.0049894
46-60#ISCO_8	.0044518	.0002741	16.24	0.000	.0039146	.0049891
46-60#ISCO_9	0	(omitted)				
60+ #ISCO_0	.0109466	.0017988	6.09	0.000	.0074207	.0144724
60+ #ISCO_1	.0131738	.0005901	22.32	0.000	.0120171	.0143305
60+ #ISCO_2	.009792	.0006381	15.34	0.000	.0085412	.0110428
60+ #ISCO_3	.0069161	.0005217	13.26	0.000	.0058935	.0079386
60+ #ISCO_4	.0042508	.0008501	5.00	0.000	.0025846	.0059171
60+ #ISCO_5	-.0014096	.0004371	-3.22	0.001	-.0022664	-.0005527
60+ #ISCO_7	.0040053	.0004076	9.83	0.000	.0032064	.0048041
60+ #ISCO_8	.003365	.0004696	7.17	0.000	.0024446	.0042854
60+ #ISCO_9	0	(omitted)				
_cons	9.376933	.0431205	217.46	0.000	9.292415	9.461452

Ženy

$$(5) \ln \widehat{Wage}_i^{Female} = \alpha_0 + \alpha_1 FT_{FullTime} + \alpha_2 CS_{CZ} + \sum_{j=2}^5 \beta_j EY_j + \sum_{j=2}^5 \epsilon_j Age_F_j + \varphi_i Age_i + \sum_{j=2}^{20} \epsilon_j Age_F_j Ed_j Age_i + \sum_{j=2}^{45} \omega_j Age_F_j Ol_j Age_i,$$

kde $i = 1, 2, \dots, n$.

Obrázek 6: Výsledky Modelu 2 - ženy

Survey: Linear regression

Number of strata = 139
Number of PSUs = 26,402

Number of obs = 15,379,721
Population size = 292,521,094
Subpop. no. obs = 7,764,014
Subpop. size = 130,937,654
Design df = 26,263
 $F(69, 26195)$ = .
Prob > F = .
R-squared = 0.4414

lnWage	Linearized					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
FullTime	.0407464	.0082111	4.96	0.000	.0246521	.0568406
Citizenship CZ	-.0498498	.0165455	-3.01	0.003	-.0822798	-.0174198
EmployYears						
Em2_01-03	.0902949	.0079454	11.36	0.000	.0747215	.1058683
Em3_03-07	.1508438	.0080754	18.68	0.000	.1350156	.166672
Em4_07-15	.1902931	.0090177	21.10	0.000	.172618	.2079683
Em5_15-99	.2348618	.0116405	20.18	0.000	.2120458	.2576778
Age_Female						
29-35	.769338	.056162	13.70	0.000	.6592573	.8794187
36-52	.4874359	.038137	12.78	0.000	.4126852	.5621865
53-58	.38372	.0603115	6.36	0.000	.2655062	.5019338
58+	1.568671	.1397041	11.23	0.000	1.294843	1.842498
Age	.0176683	.0013655	12.94	0.000	.0149918	.0203449
Age_Female#Educ#c.Age						
15-28#E2_MidSecondary	-.0003671	.0005146	-0.71	0.476	-.0013757	.0006415
15-28#E3_UppSecondary	.0045878	.0005821	7.88	0.000	.0034468	.0057287
15-28#E4_Tertiary	.0079159	.000735	10.77	0.000	.0064753	.0093566
29-35#E1_LowSecondary	-.0371431	.002025	-18.34	0.000	-.0411122	-.0331739
29-35#E2_MidSecondary	-.0377872	.0020274	-18.64	0.000	-.041761	-.0338135
29-35#E3_UppSecondary	-.0336112	.0020604	-16.31	0.000	-.0376497	-.0295726
29-35#E4_Tertiary	-.0278863	.0021085	-13.23	0.000	-.032019	-.0237536
36-52#E1_LowSecondary	-.0267349	.0014376	-18.60	0.000	-.0295527	-.023917
36-52#E2_MidSecondary	-.0264463	.0014516	-18.22	0.000	-.0292915	-.023601
36-52#E3_UppSecondary	-.0238271	.0014582	-16.34	0.000	-.0266854	-.0209689
36-52#E4_Tertiary	-.0191825	.0014986	-12.80	0.000	-.0221197	-.0162452
53-58#E1_LowSecondary	-.0234068	.0016355	-14.31	0.000	-.0266125	-.020201
53-58#E2_MidSecondary	-.0230816	.0016392	-14.08	0.000	-.0262946	-.0198685
53-58#E3_UppSecondary	-.0205967	.0016478	-12.50	0.000	-.0238265	-.017367
53-58#E4_Tertiary	-.0170839	.0016933	-10.09	0.000	-.0204029	-.0137649
58+ #E1_LowSecondary	-.0420166	.0025696	-16.35	0.000	-.0470531	-.03698
58+ #E2_MidSecondary	-.0416403	.0025509	-16.32	0.000	-.0466402	-.0366404
58+ #E3_UppSecondary	-.0402706	.0025553	-15.76	0.000	-.0452792	-.0352621
58+ #E4_Tertiary	-.0369739	.0025593	-14.45	0.000	-.0419903	-.0319576

Age_Female#OccupISCO#c.Age						
15-28#ISCO_1	.0070268	.0024853	2.83	0.005	.0021554	.0118982
15-28#ISCO_2	.0058964	.0005892	10.01	0.000	.0047416	.0070512
15-28#ISCO_3	.0038358	.0003872	9.91	0.000	.0030769	.0045947
15-28#ISCO_4	-.0002611	.0005601	-0.47	0.641	-.001359	.0008367
15-28#ISCO_5	-.0089557	.0007302	-12.26	0.000	-.010387	-.0075245
15-28#ISCO_7	-.0051333	.0009154	-5.61	0.000	-.0069276	-.003339
15-28#ISCO_8	-.0021633	.0005989	-3.61	0.000	-.0033372	-.0009895
15-28#ISCO_9	-.0113262	.0008344	-13.57	0.000	-.0129617	-.0096908
29-35#ISCO_0	.0132149	.0006098	21.67	0.000	.0120196	.0144101
29-35#ISCO_1	.0238643	.0010776	22.15	0.000	.0217521	.0259766
29-35#ISCO_2	.0165381	.0006575	25.15	0.000	.0152494	.0178268
29-35#ISCO_3	.0134371	.0005761	23.32	0.000	.0123079	.0145663
29-35#ISCO_4	.0097695	.000599	16.31	0.000	.0085954	.0109436
29-35#ISCO_5	.0020496	.000585	3.50	0.000	.000903	.0031961
29-35#ISCO_7	.004613	.0006373	7.24	0.000	.003364	.0058621
29-35#ISCO_8	.0073555	.0004784	15.37	0.000	.0064178	.0082933
29-35#ISCO_9	0	(omitted)				
36-52#ISCO_0	.0127436	.0003119	40.85	0.000	.0121322	.013355
36-52#ISCO_1	.0176119	.0005643	31.21	0.000	.0165058	.0187179
36-52#ISCO_2	.0117711	.0003682	31.97	0.000	.0110495	.0124928
36-52#ISCO_3	.0101542	.0003064	33.14	0.000	.0095536	.0107549
36-52#ISCO_4	.0065917	.000356	18.52	0.000	.0058939	.0072895
36-52#ISCO_5	.0022397	.0002941	7.62	0.000	.0016633	.0028161
36-52#ISCO_7	.0040853	.0003735	10.94	0.000	.0033532	.0048174
36-52#ISCO_8	.0062467	.0002671	23.39	0.000	.0057232	.0067702
36-52#ISCO_9	0	(omitted)				
53-58#ISCO_0	.012207	.0002487	49.09	0.000	.0117196	.0126944
53-58#ISCO_1	.0146355	.0005291	27.66	0.000	.0135986	.0156725
53-58#ISCO_2	.0099799	.0003409	29.28	0.000	.0093118	.0106481
53-58#ISCO_3	.0089299	.0002704	33.02	0.000	.0083999	.0094599
53-58#ISCO_4	.006067	.0002874	21.11	0.000	.0055037	.0066303
53-58#ISCO_5	.0020139	.000223	9.03	0.000	.0015768	.002451
53-58#ISCO_7	.0035423	.0002811	12.60	0.000	.0029914	.0040932
53-58#ISCO_8	.0047707	.0002027	23.53	0.000	.0043733	.005168
53-58#ISCO_9	0	(omitted)				
58+ #ISCO_0	.0105969	.0003292	32.19	0.000	.0099516	.0112423
58+ #ISCO_1	.0141145	.0008785	16.07	0.000	.0123927	.0158363
58+ #ISCO_2	.0106651	.0004582	23.28	0.000	.0097671	.0115632
58+ #ISCO_3	.0073951	.0004308	17.16	0.000	.0065506	.0082395
58+ #ISCO_4	.0053347	.0004476	11.92	0.000	.0044573	.0062121
58+ #ISCO_5	.0018021	.0002973	6.06	0.000	.0012194	.0023848
58+ #ISCO_7	.0034701	.0003301	10.51	0.000	.0028231	.0041172
58+ #ISCO_8	.0032997	.0002525	13.07	0.000	.0028047	.0037947
58+ #ISCO_9	0	(omitted)				
_cons	9.265113	.0363981	254.55	0.000	9.19377	9.336455

V Modelu 2 pro obě pohlaví je interakce věku jako spojité proměnné, kategorie věku a kategorie zaměstnání (CZ-ISCO) pro většinu kombinací statisticky významná.

3.8 Schopnost predikce modelů a rezidua

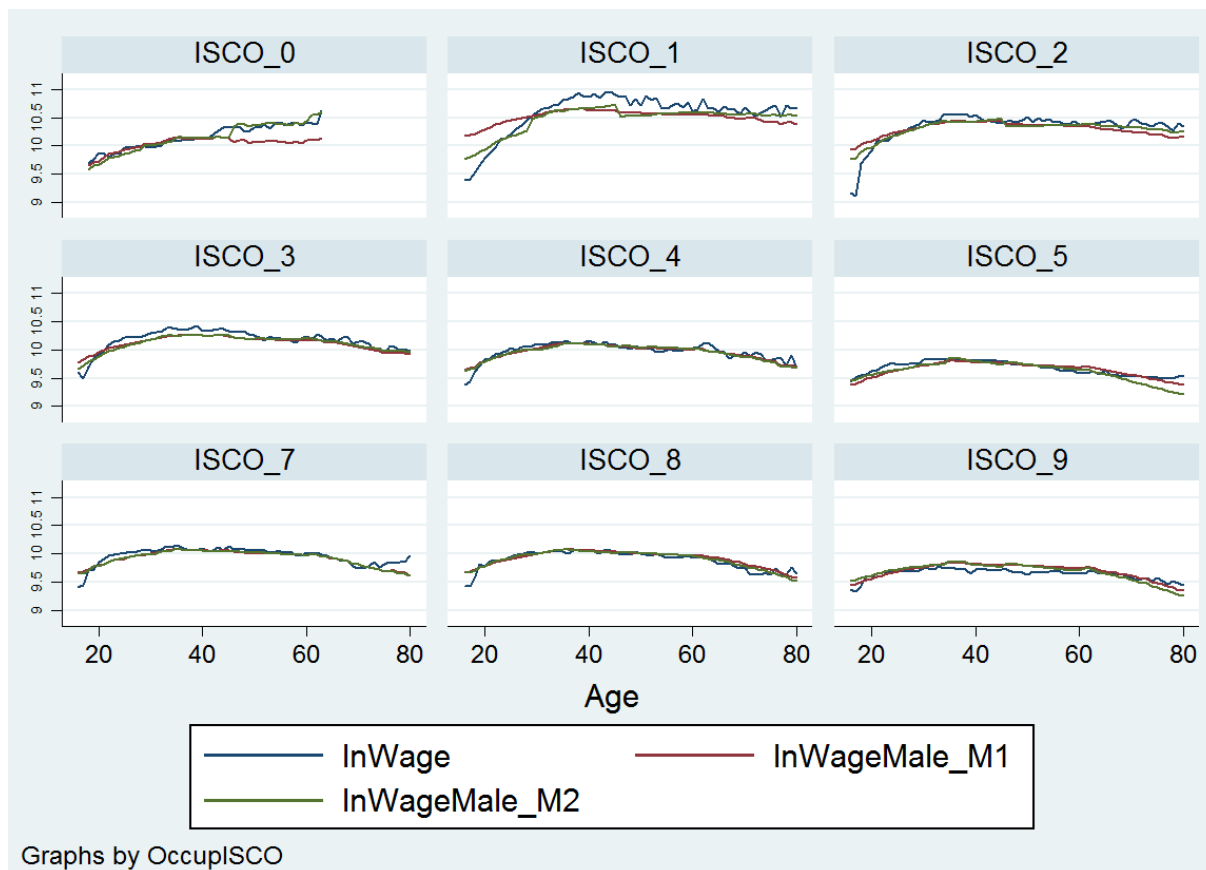
Náhled predikce Modelu 1 a Modelu 2 dle zaměstnání (CZ-ISCO)

Klasifikace zaměstnání CZ-ISCO je dle předchozích analýz nejvýznamnějším faktorem z hlediska variability, proto byla posouzena schopnost predikce vybraných modelů především u tohoto faktoru.

Schopnost predikce jednodušších modelů se ukazuje jako problematická zejména u zaměstnání hlavní třídy 0, třída 1 (příslušníci ozbrojených sil, zákonodárci a řídící pracovníci). Model 2, který uvažuje interakci věku a zaměstnání, naopak trend popisuje kvalitněji, viz grafy 12 a 13, kde pouze reálné hodnoty (lnWage) jsou z důvodu znázornění 15 mil. záznamů zprůměrovány.

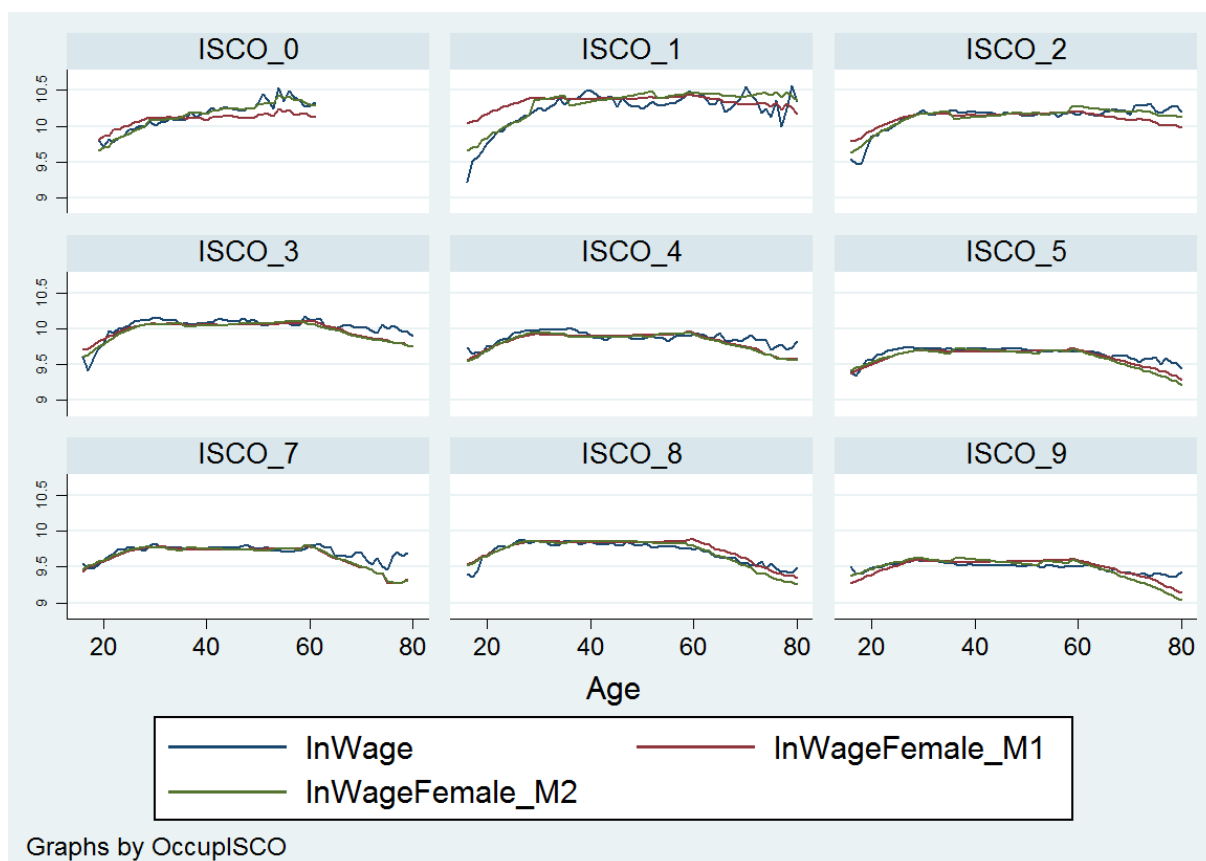
Predikce pro méně kvalifikačně náročná zaměstnání dodržují reálné trendy zaměstnanců dle věku ve vyšší míře než predikce u kvalifikačně náročnějších zaměstnání (ISCO_1, 2, 3). Model 2 se ukazuje jako vhodnější při pohledu na nejlépe hodnocenou skupinu zaměstnanců ISCO_1, kde je variabilita mezd nejvyšší. Model 2 se také ukazuje jako vhodnější i pro zaměstnance v ozbrojených silách (ISCO_0), kde lépe dodržuje nezvykle lineární trend mezi zaměstnanci v ČR v závislosti na věku.

Graf 12: Predikce Modelu 1 a Modelu 2 pro věk a klasifikaci zaměstnání (CZ-ISCO) - muži



Podobné výsledky ukazují i predikce pro ženy.

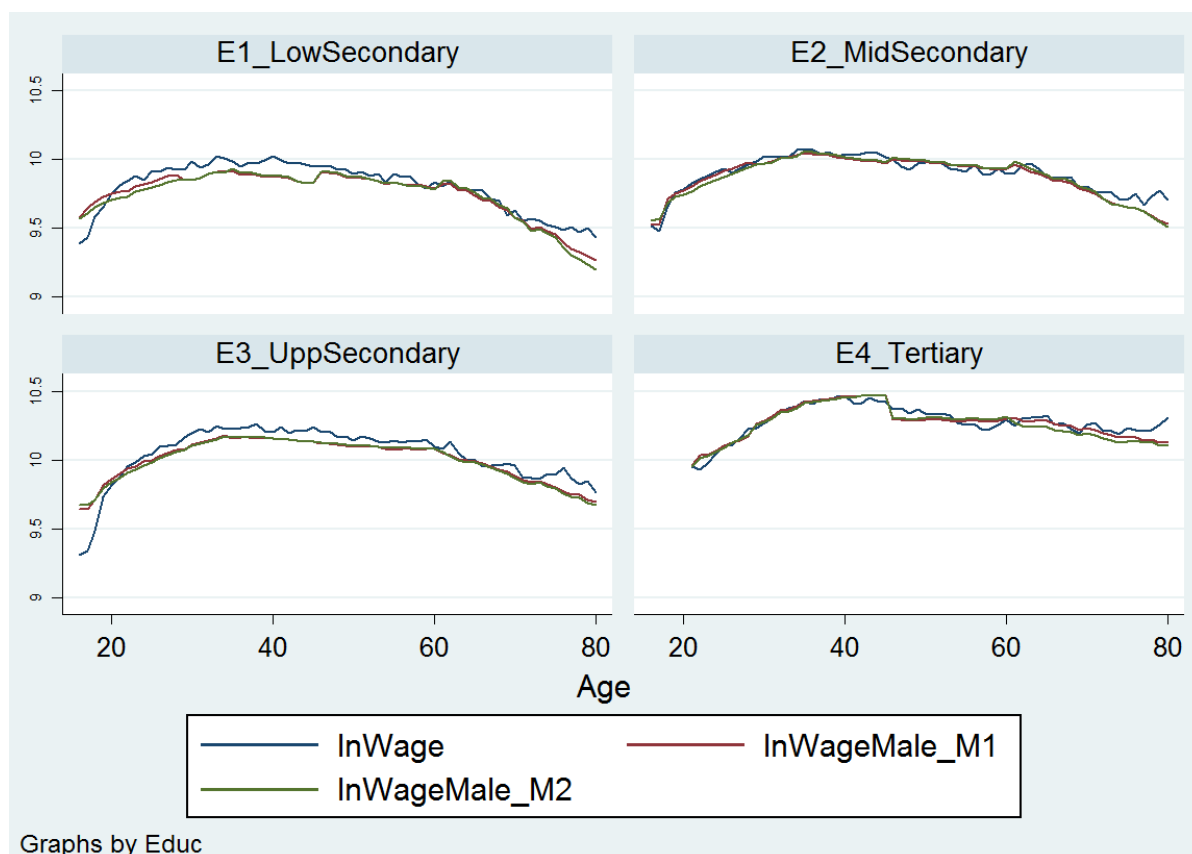
Graf 13: Predikce Modelu 1 a Modelu 2 pro věk a klasifikaci zaměstnání (CZ-ISCO) – ženy



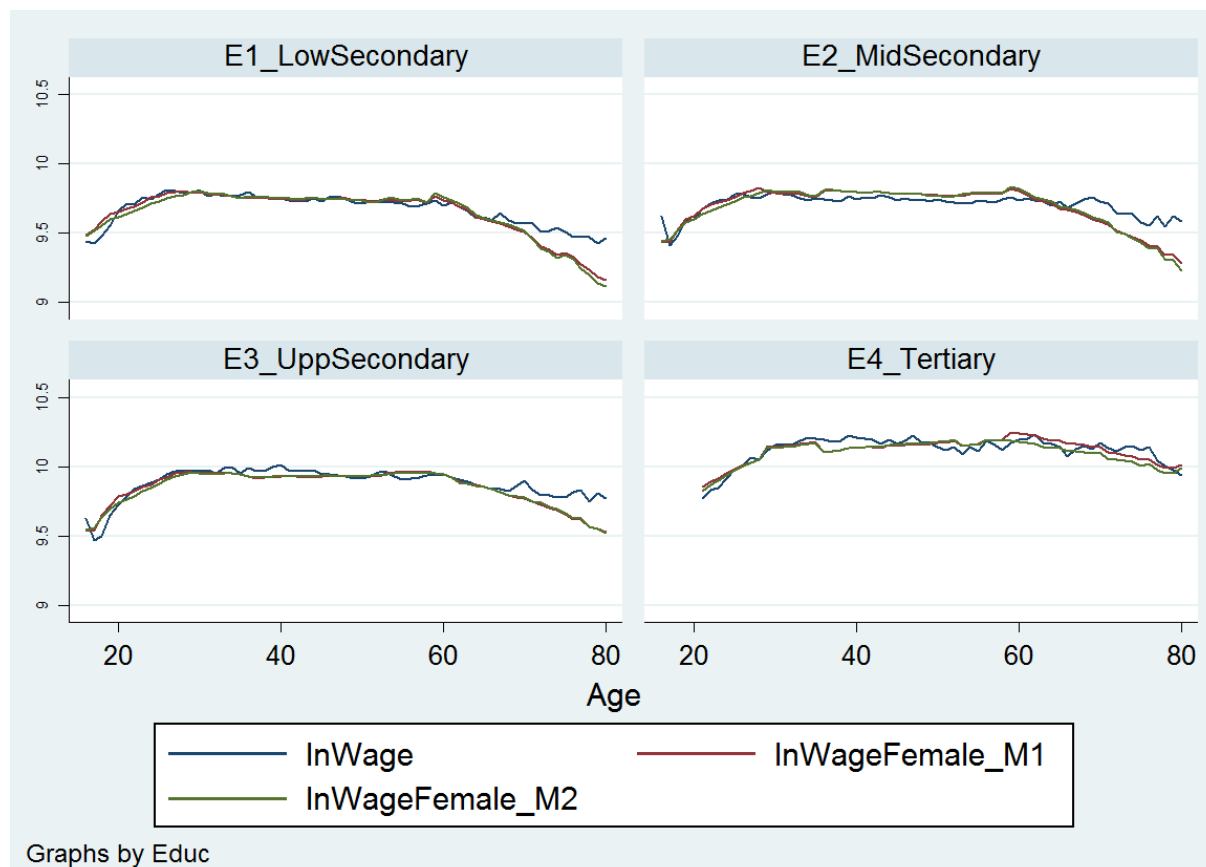
Náhled predikce Modelu 1 a Modelu 2 dle vzdělání

Predikce pro kombinace věku a vzdělání jsou dostačující, schopnosti Modelu 1 a Modelu 2 by neměly být odlišné v predikcích při pohledu na faktory vzdělání a věk.

Graf 14: Predikce Modelu 1 a Modelu 2 pro věk a vzdělání – muži



Graf 15: Predikce Modelu 1 a Modelu 2 pro věk a vzdělání – ženy



Náhled predikce Modelu 1 a Modelu 2 dle věku a pohlaví

Predikce dle věku naznačují dostačující složení kategorií věku do 5 skupin, viz **Graf 2** promítnutý s predikcemi do **Graf 16**. Způsob kategorizace proměnných byl diskutován v kapitole 3.5.

Graf 16: Predikce Modelu 1 a Modelu 2 dle věku a pohlaví

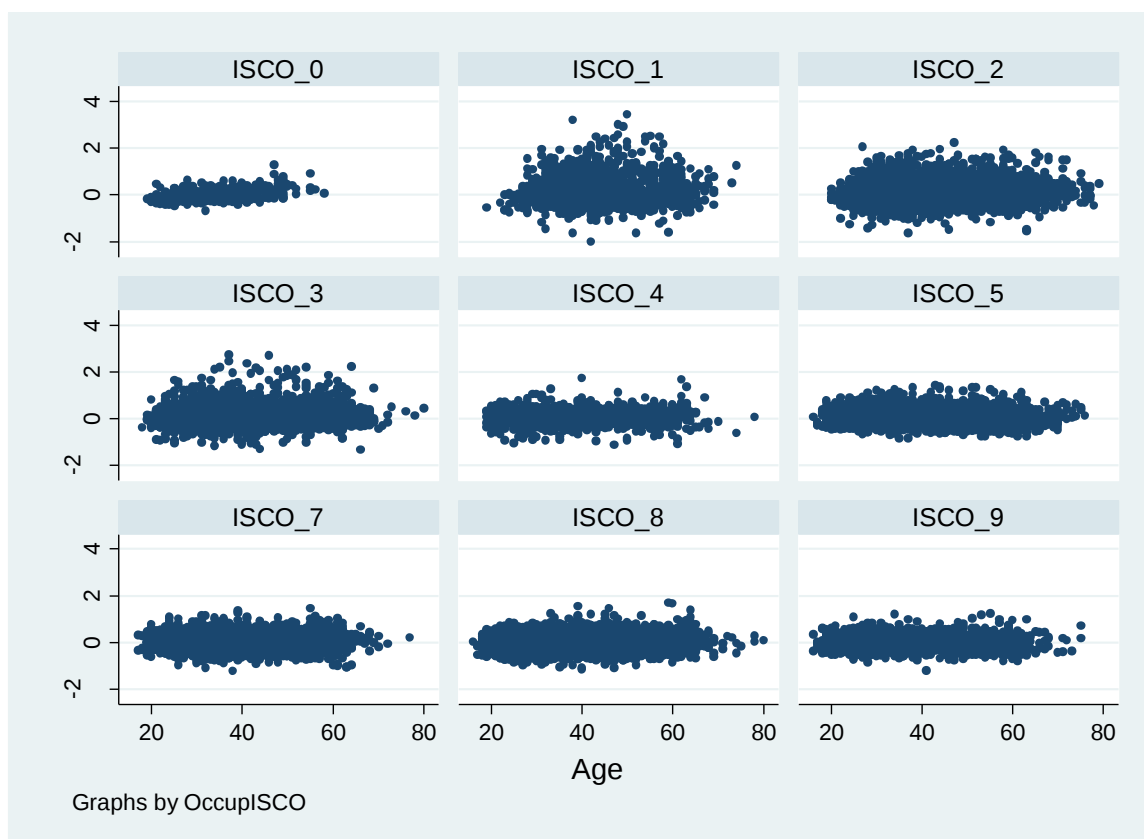


Náhled na rezidua Modelu 1 a Model 2 dle zaměstnání (CZ-ISCO)

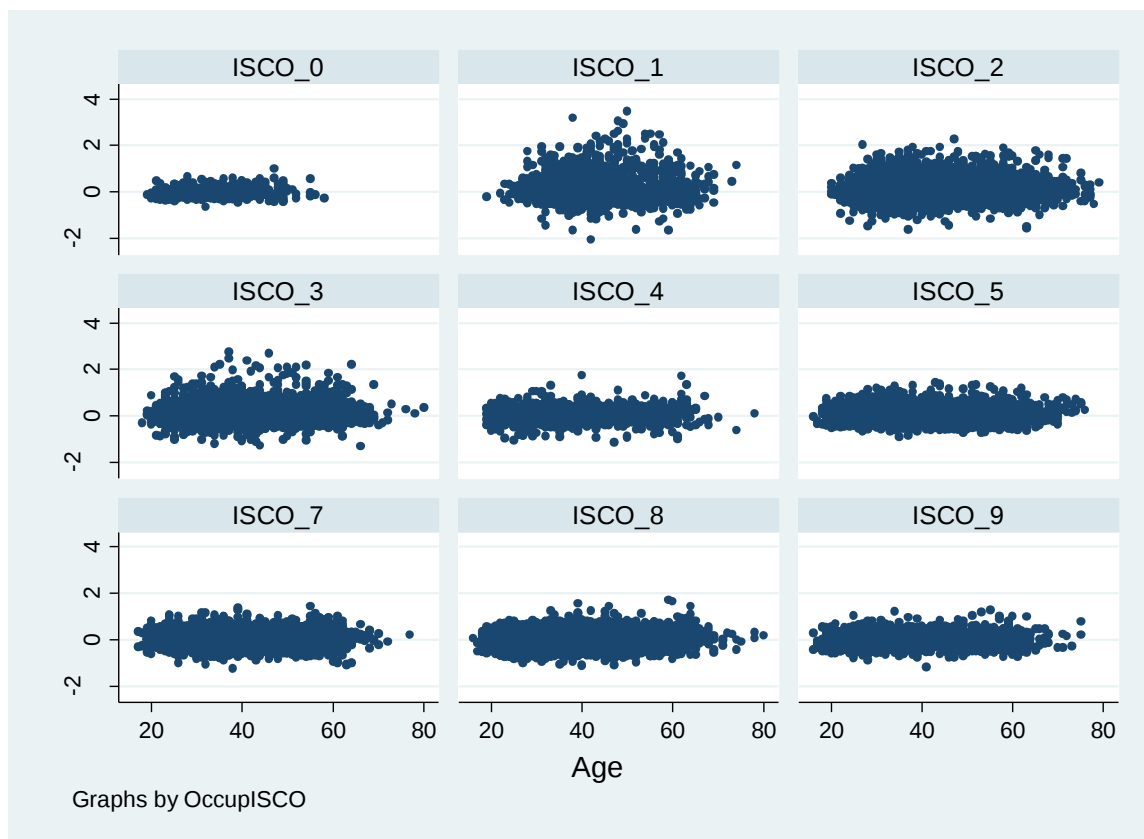
Z důvodu velkého množství dat a nemožnosti zahrnutí firemních charakteristik do modelů jsou veškeré testy (normální rozdělení, heteroskedasticita apod.) statisticky významné. Přesto pohled na náhodně vybraná rezidua značí náhodné a rovnoměrné rozdělení. Model 2 je z hlediska reziduí mírně vhodnější zejména ve vyšším věku u CZ-ISCO 0 a u nižšího věku CZ-ISCO 1, což potvrzuje předchozí pohled na predikce.

U klasifikace zaměstnání je při pohledu na rezidua patrné, že obecně predikce mezd u kvalifikačně náročnějších zaměstnání (ISCO_1, 2, 3) nejsou na takové úrovni jako predikce u méně odborných pozic zaměstnání. Následující grafy zobrazují velikosti reziduí podle věku a zaměstnání CZ-ISCO pro ženy a muže, lze si všimnout mírných trendů u CZ-ISCO 0 a CZ-ISCO 1 v Modelu 1. Rezidua u žen se chovají obdobným způsobem jako u mužů.

Graf 17: Rezidua Modelu 1 pro věk a klasifikaci zaměstnání (CZ-ISCO) – muži



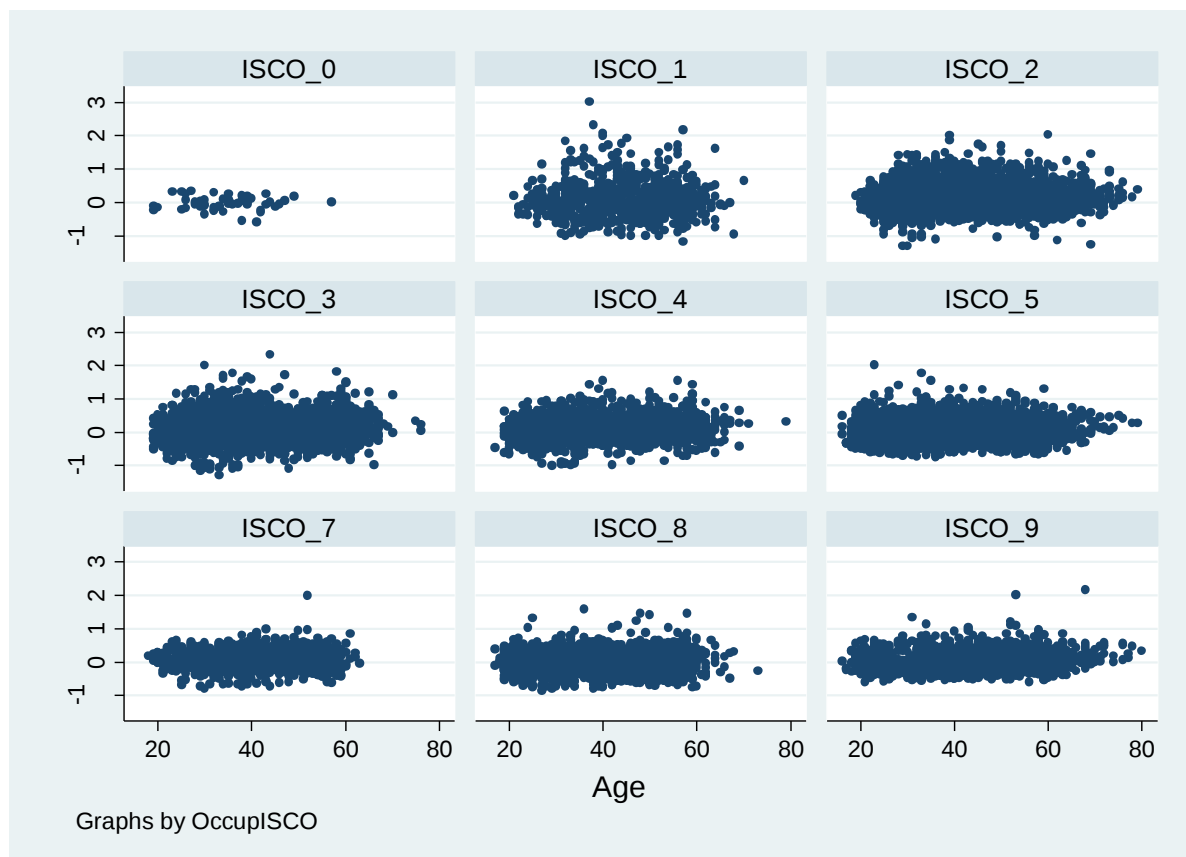
Graf 18: Rezidua Modelu 2 pro věk a klasifikaci zaměstnání (CZ-ISCO) – muži



Graf 19: Rezidua Modelu 1 pro věk a klasifikaci zaměstnání (CZ-ISCO) – ženy



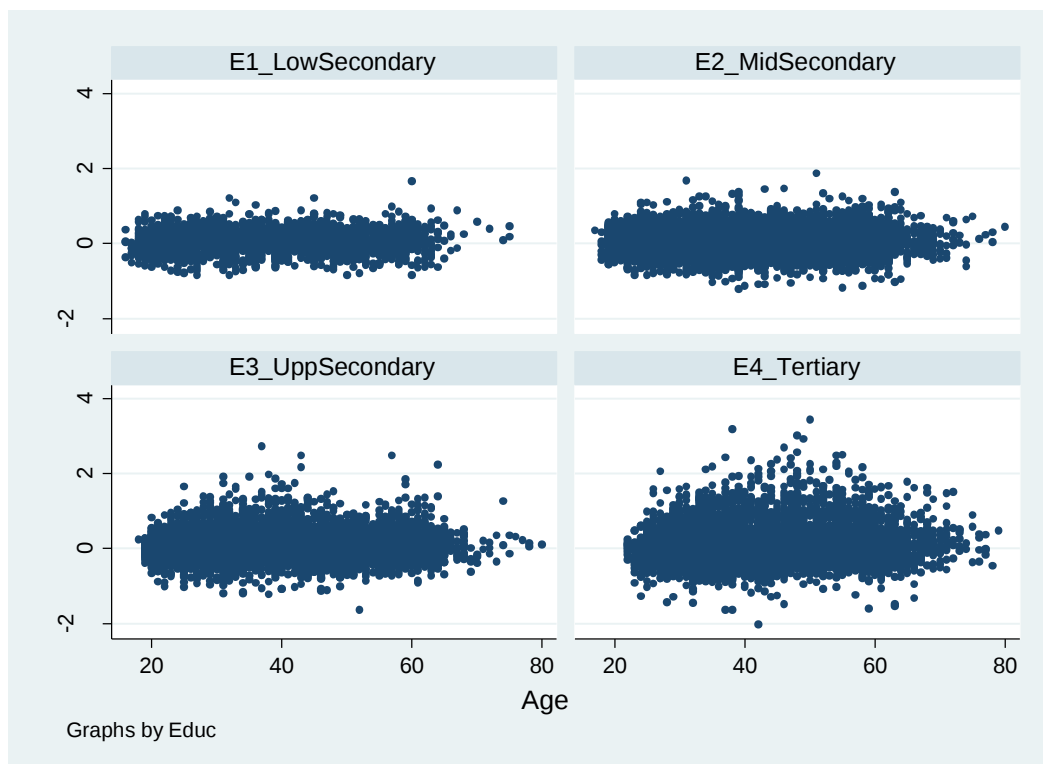
Graf 20: Rezidua Modelu 2 pro věk a klasifikaci zaměstnání (CZ-ISCO) – ženy



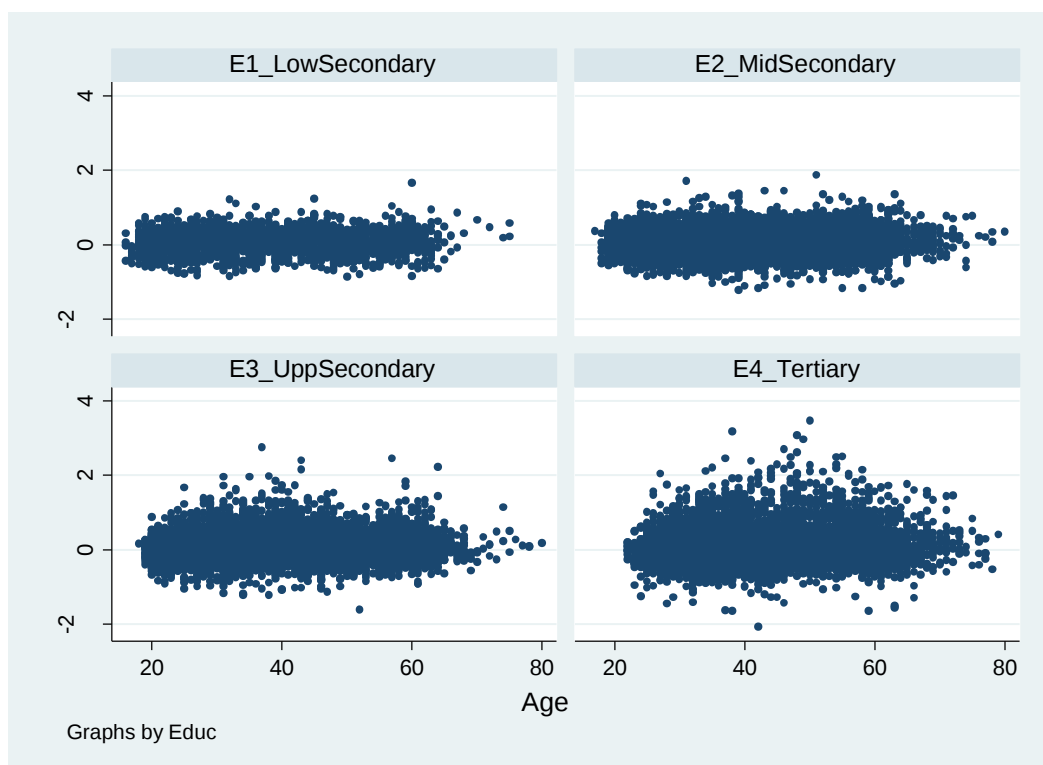
Náhled na rezidua Modelu 1 a Model 2 dle zaměstnání (CZ-ISCO)

Pohled na rezidua na kombinaci dle věku a vzdělání rovněž naznačuje náhodný a rovnoměrný výskyt reziduí jak u mužů, tak i u žen. Opět je však z reziduí patrná rozdílná úroveň predikce u vzdělanější části zaměstnanců, zaměstnanců s terciárním vzděláním.

Graf 21: Rezidua Modelu 1 pro věk a vzdělání – muži



Graf 22: Rezidua Modelu 2 pro věk a vzdělání – muži



Graf 23: Rezidua Modelu 1 pro věk a vzdělání – ženy



Graf 24: Rezidua Modelu 2 pro věk a vzdělání – ženy



3.9 Výběr modelu

Z hlediska schopnosti predikce se jako vhodnější ukazuje Model 2. Pokud by však komplikovanější struktura Modelu 2 byla problematická z hlediska implementace do mikrosimulačního důchodového systému, lze využít i Model 1.

Pro vysvětlenost variability mzdového modelu je nejvýznamnější proměnná klasifikace zaměstnání CZ-ISCO, proto doporučujeme Model 2 uvažující interakci CZ-ISCO se spojitými kategoriemi věku, viz kapitola 3.6.

4. Výpočty parametrů variability mezd a jejich implementace do mikrosimulačního modelu

4.1 Permanentní a transitorní šoky (variabilita mzdového růstu)

Při projekci mezd se kromě základní mzdové rovnice využívá i odhadů náhodné složky (Congressional Budget Office, 2013 [32]). Náhodnou složkou v modelu se odhaduje meziroční variabilita v permanentních a transitorních mzdách. Permanentní šoky představují individuální schopnosti jedince, které se v průběhu jeho života mění. Transitorní šoky představují nevysvětlenou část v modelu (dočasný výkyv mzdy). Modelem je potřeba odhadnout hodnoty permanentní a transitorní variability mezd pouze jednou (odhadují se odděleně – viz např. Carroll 1992 The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence [7]). Složka představující permanentní mzdu je v závěrečné rovnici závislá na hodnotách permanentní složky z minulých let. Další zahraniční modely předpokládají konstantní variabilitu mezd v čase a zabývají se odhadem těchto složek už v samotné základní mzdové rovnici. Náhodná složka se odhadne na počátku projekce a pak se nemění.

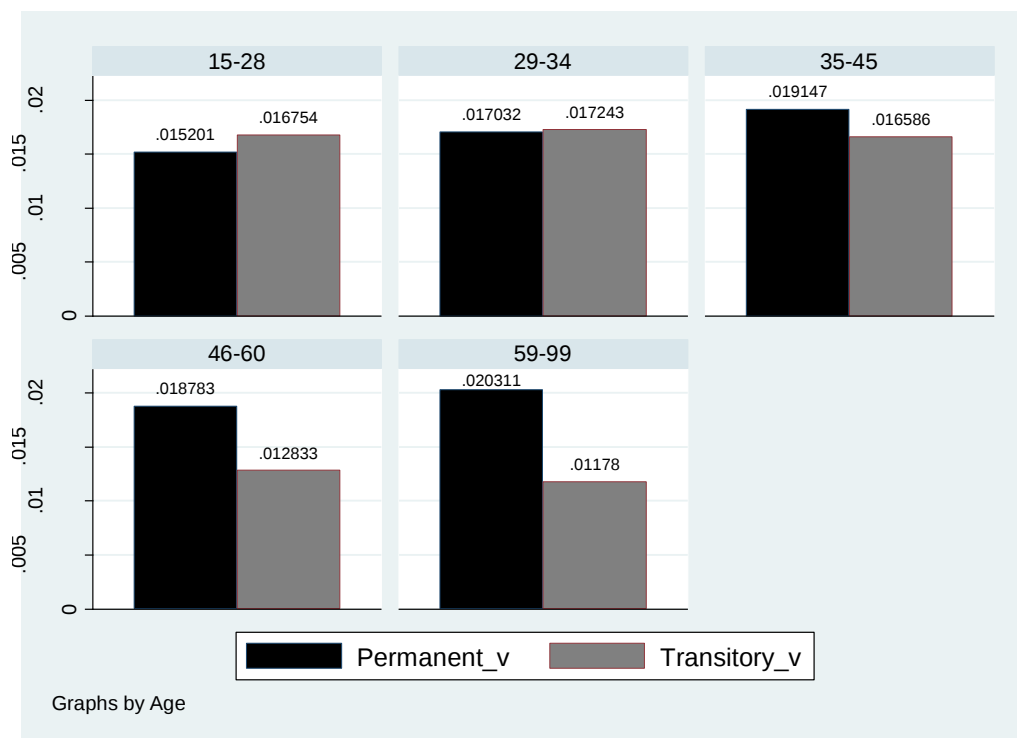
Model permanentních a transitorních šoků je založený na pseudo-panelu dat z ISPV za rok 2008 až 2014, viz 1.8. Pro každého jedince byly spočítány první, druhé až 6. difference z logaritmu mzdy. Dále byl vypočten rozptyl pro 1. diferenci, 2. diferenci až 6. diferenci, a to podle pohlaví a věkových kategorií pro muže a ženy stanovené stejně jako ve mzdové rovnici. K odhadnutí permanentního a transitorního šoku byl využit model, kde vysvětlovanou proměnnou je rozptyl diferencí a vysvětlující proměnnou je pořadí difference. Odhadnutá konstanta α_0 modelu představuje dvojnásobek transitorního rozptylu. Odhadnutý koeficient (sklon přímky) k pořadí difference α_1 představuje permanentní rozptyl.

Model rozptylu diferencí (variability růstu mezd) má následující podobu:

$$(6) \quad \hat{\sigma}_{r_m}^2 = \text{var}(\ln Wage_{t+m} - \ln Wage_t) = m\hat{\sigma}_{lnN}^2 + 2\hat{\sigma}_{lnV}^2,$$
$$\hat{\sigma}_{r_m}^2 = m\alpha_1 + \alpha_0,$$

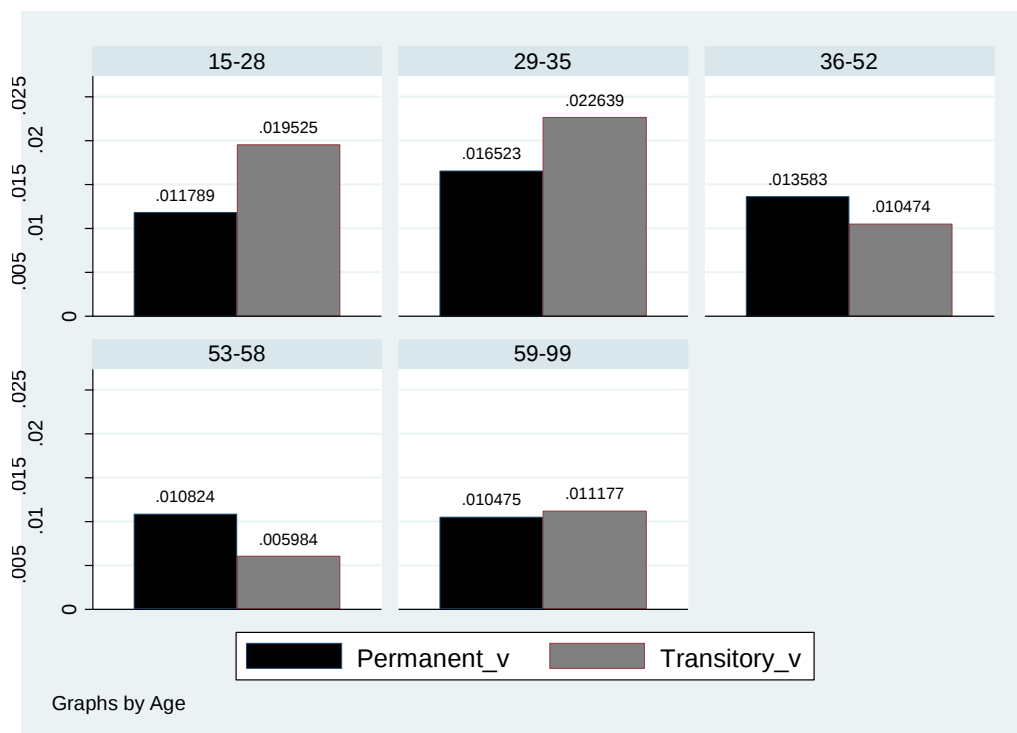
kde m je pořadí difference, $m = 1, 2, \dots, 6$, $\hat{\sigma}_{lnN}^2$ je odhad rozptylu permanentního šoku a $\hat{\sigma}_{lnV}^2$ je odhad rozptylu transitorního šoku, r_m je m -tá difference. Výsledky modelu rozptylu diferencí vidíme v následujících grafech.

Graf 25: Permanentní a transitorní rozptyl - muži



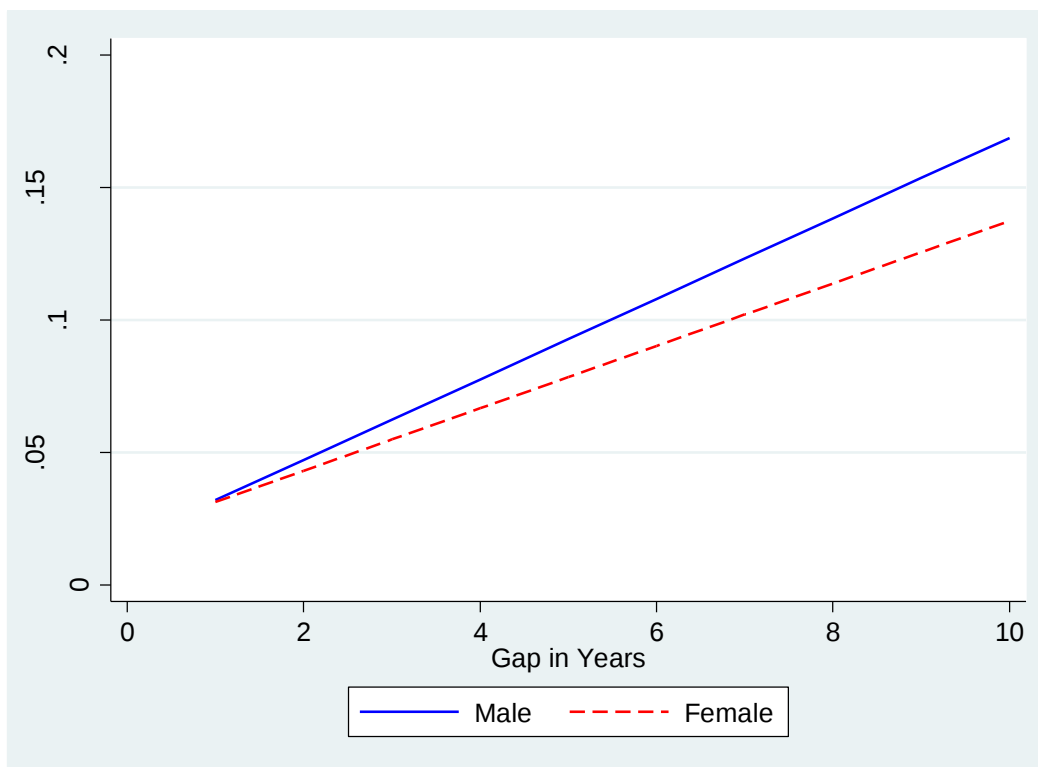
Permanentní rozptyl u mužů ve věku 15 až 28 let je zhruba 0,015 a transitorní rozptyl je zhruba 0,017. Tato věková kategorie u mužů dosáhla nejnižšího permanentního rozptylu, permanentní rozptyl dále roste až do věku 45 let. U transitorního rozptylu není vidět žádný trend, transitorní rozptyl pro mladší věkové kategorie od 15 do 45 let je vyšší než pro starší věkové kategorie od 46 let. Permanentní rozptyl pro muže ve věku 15 až 28 let je 1,1 krát větší, kdežto permanentní rozptyl pro muže starších 60 let je 1,7 krát menší.

Graf 26: Permanentní a transitorní rozptyl - ženy



Ženy ve věku 15 až 28 let mají permanentní rozptyl o velikosti zhruba 0,012 a transitorní rozptyl o velikosti zhruba 0,020. Transitorní rozptyl je u žen v této věkové kategorii 1,7 krát větší než permanentní rozptyl. Ovšem ženy ve věku 36 až 52 mají transitorní rozptyl 1,4 krát menší než permanentní rozptyl. Z výsledku modelu nevyplyvá jasný trend v růstu či poklesu permanentního nebo transitorního rozptylu podle věkové kategorie. Největší permanentní i transitorní rozptyl vykazují ženy ve věku 29 až 35 let.

Graf 27: Model variability růstu mezd pro zaměstnance ve věku 15 až 28 let



Předchozí graf ukazuje výsledek modelu rozptylu diferencí pro muže a ženy ve věku 15 až 28 let. Pouze tyto kategorie jsou srovnatelné vzhledem ke stanoveným věkovým kategoriím, které byly určeny pomocí evolučního algoritmu, viz 3.5. V grafu jsou zobrazeny vyrovnané hodnoty modelu rozptylu diferencí, sklon přímky pro muže je roven 0,015 a pro ženy 0,012. Odhadnutá konstanta z modelu je pro muže 0,034 a pro ženy 0,040, použije se polovina hodnoty konstanty (pro muže 0,017, pro ženy 0,020), která je považována jako transitorní rozptyl. Na grafu lze vidět, že variabilita mezd v čase roste rychleji mužům než ženám.

Odhadnuté permanentní rozptyly podle věkových kategorií a pohlaví se využijí při simulaci permanentních šoků v projekcích mezd. Předpokladem simulace permanentního šoku je, že pochází z normálního rozdělení s nulovou střední hodnotou a odhadnutým permanentním rozptylem. Obdobně se bude postupovat při simulacích transitorních šoků, tzn. předpoklad normálního rozdělení s nulovou střední hodnotou a odhadnutým transitorním rozptylem.

4.2 Permanentní mzdový rozdíl (individuální variabilita ve mzdách)

Permanentní mzdový rozdíl označovaný jako PED_i lze chápat jako průměrný rozdíl hodnot logaritmu mezd a vyrovnaných hodnot logaritmu mezd za několik posledních sledovaných období. Jedná se vlastně o průměrné reziduum. Permanentní mzdový rozdíl využívá americký model CBOLT [32]. U každého jedince se na začátku projekce mezd počítají PED_i (pětileté průměrné reziduum). PED_i je chápána jako míra variability ve mzdách, která je pro každého jedince jiná. Je zde předpoklad,

že rozptyl vysvětlované proměnné ($\ln Wage_i$) není konstantní, ale rozdílný. To znamená, že i náhodná chyba modelu mezd nemá stejné rozptyly pro všechna pozorování.

Vycházíme ze stejného předpokladu, tedy že rozptyl vysvětlované proměnné není pro všechny jedince stejný, a proto jsme přistoupili k modelování rozptylu náhodné chyby (odhadem je průměrné reziduum). Vyrovnané hodnoty $\ln \widehat{Wage}_{it}$ vychází z Modelu 2. Permanentní mzdový rozdíl PED_i získáme tak, že vypočteme průměrný rozdíl hodnot logaritmu mezd a vyrovnaných hodnot logaritmu mezd za posledních 5 let (2010 – 2014) na panelu a dodatečného pseudo-panelu dat zaměstnanců.

$$(7) \quad PED_i = \frac{\sum_{t=1}^5 (\ln Wage_{it} - \ln \widehat{Wage}_{it})}{5} = \bar{e}_i, i = 1, 2, \dots, n.$$

Předpokládáme multiplikativní efekt vysvětlujících proměnných na vysvětlovanou proměnnou (rozptyl náhodné chyby):

$$(8) \quad \hat{\sigma}_{i,PED}^2 = e^{z_i^T \alpha},$$

kde σ_i^2 je rozptyl náhodné chyby, z_i^T je transponovaný i-tý vektor proměnných způsobující rozdílný rozptyl pro i-tého jedince, α je vektor neznámých parametrů. Po úpravách dostaneme obecný model:

$$(9) \quad \ln \hat{\sigma}_{i,PED}^2 = \ln \bar{e}_i^2 = \alpha_0 + \sum_{j=1}^L \alpha_j z_{ij}, i = 1, 2, \dots, n,$$

kde $\bar{e}_i^2$ je kvadrát průměrných reziduí (PED_i).

Pro modelování rozptylu byla použita jako vysvětlující proměnná pouze kategorie hlavní třídy zaměstnání CZ-ISCO (OI). Ostatní proměnné použité ve mzdové rovnici vychází statisticky nevýznamné, proto nebyly do modelu jako vysvětlující proměnné zahrnuty.

Model individuálního rozptylu mezd - muži

$$(10) \quad \ln \hat{\sigma}_{i,PED}^{2,Male} = \ln \bar{e}_i^{2,Male} = \alpha_0 + \sum_{j=2}^9 \gamma_j OI_j, i = 1, 2, \dots, n.$$

Obrázek 7: Výsledky modelu individuálního rozptylu - muži

Survey: Linear regression

Number of strata	=	138	Number of obs	=	589,921
Number of PSUs	=	11,917	Population size	=	13,042,582
			Subpop. no. obs	=	589,921
			Subpop. size	=	13,042,582
			Design df	=	11,779
			F(8, 11772)	=	485.47
			Prob > F	=	0.0000
			R-squared	=	0.0381

lne2_M	Linearized		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
OccupISCO						
ISCO_1	2.434393	.0678803	35.86	0.000	2.301336	2.567449
ISCO_2	2.0216	.0547977	36.89	0.000	1.914188	2.129013
ISCO_3	1.533364	.0615898	24.90	0.000	1.412638	1.65409
ISCO_4	1.327331	.1265615	10.49	0.000	1.07925	1.575413
ISCO_5	1.940697	.0838558	23.14	0.000	1.776325	2.105068
ISCO_7	1.170193	.064764	18.07	0.000	1.043244	1.297141
ISCO_8	.9872008	.0664676	14.85	0.000	.8569134	1.117488
ISCO_9	1.213106	.0711129	17.06	0.000	1.073713	1.352499
_cons	-4.780114	.0089229	-535.71	0.000	-4.797604	-4.762623

Model individuálního rozptylu mezd - ženy:

$$(11) \quad \ln \hat{\sigma}_{i,PED}^{2,Female} = \ln \bar{e}_i^{2,Female} = \alpha_0 + \sum_{j=2}^9 \gamma_j OI_j, i = 1, 2, \dots, n.$$

Obrázek 8: Výsledky modelu individuálního rozptylu - ženy

Survey: Linear regression

Number of strata = 138
Number of PSUs = 16,430

Number of obs = 555,647
Population size = 9,998,766
Subpop. no. obs = 555,647
Subpop. size = 9,998,766
Design df = 16,292
F(8, 16285) = 256.09
Prob > F = 0.0000
R-squared = 0.0110

lne2_F	Linearized		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
OccupISCO						
ISCO_1	2.271986	.0838971	27.08	0.000	2.107538	2.436433
ISCO_2	1.719734	.0499858	34.40	0.000	1.621756	1.817712
ISCO_3	1.521808	.0629385	24.18	0.000	1.398442	1.645175
ISCO_4	1.618204	.1616521	10.01	0.000	1.301348	1.93506
ISCO_5	1.430457	.0673783	21.23	0.000	1.298389	1.562526
ISCO_7	1.570202	.1190885	13.19	0.000	1.336776	1.803629
ISCO_8	1.184162	.0768809	15.40	0.000	1.033467	1.334857
ISCO_9	1.072361	.0645341	16.62	0.000	.9458667	1.198855
_cons	-5.200101	.0257056	-202.29	0.000	-5.250487	-5.149715

Rozptyl získáme

$$(12) \quad \hat{\sigma}_{i,PED}^2 = \exp \left(\alpha_0 + \sum_{j=2}^9 \gamma_j OI_j \right), i = 1, 2, \dots, n.$$

Tabulka 15: Odhad individuálního rozptylu pro PED_i

OccupISCO	Muži	Ženy
	$\hat{\sigma}_{i,PED}^{2, Male}$	$\hat{\sigma}_{i,PED}^{2, Female}$
ISCO_0	0,0084	0,0055
ISCO_1	0,0958	0,0535
ISCO_2	0,0634	0,0308
ISCO_3	0,0389	0,0253
ISCO_4	0,0317	0,0278
ISCO_5	0,0585	0,0231
ISCO_7	0,0271	0,0265
ISCO_8	0,0225	0,0180
ISCO_9	0,0282	0,0161

Odhadnutý rozptyl při dané kombinaci vysvětlujících proměnných se pak použije při simulaci PED_i pro každého jedince na začátku projekce mezd. Předpokladem simulace průměrného rezidia je, že PED_i pochází z normálního rozdělení s nulovou střední hodnotou a odhadnutým rozptylem z modelu.

Projekce mezd

K projekci mezd pro každého jednotlivce se využije predikované mzdy ze mzdové rovnice, permanentního difereciálního rozdílu (PED_i) a permanentního a transitorního šoku. Mzda se jednotlivci stanoví podle:

$$(13) \quad \ln Wage_{it} = \ln \widehat{Wage}_{it} + PED_i + \sum_{s=1}^t \alpha_{is} + \beta_{it},$$

kde $PED_i \sim N(0, \hat{\sigma}_{i,PED}^2)$, $\alpha_{is} \sim N(0, \hat{\sigma}_{iNN}^2)$, $\beta_{it} \sim N(0, \hat{\sigma}_{iNV}^2)$, $i = 1, 2, \dots, n$, s je jednoletý krok projekce a t je horizont projekce. Každému jednotlivci se PED_i simuluje na začátku projekce a pro další kroky projekce zůstává stejný. Hodnoty permanentních a transitorních šoků se budou každému jedinci pro další kroky projekce měnit. Hodnoty permanentních šoků se po krocích projekce (po letech) sčítají.

Závěr

V jednotlivých kapitolách závěrečné zprávy je popsáno plnění Smlouvy o odvození parametrů rovnice mzdové dynamiky pro mikrosimulační model důchodového systému MPSV. Dle bodu 2.2 Smlouvy byla, na základě harmonizovaných dat z ISPV za roky 2008-2014, vytvořena datová základna pro účely tvorby mzdové rovnice a vytvořen panel a pseudo-panel zaměstnanců (kapitola 1). Po srovnání a vyhodnocení zahraničních modelů (kapitola 2) byl jako vhodný model pro ČR vybrán americký CBOLT, který na průřezových datech používá klasický lineární regresní model, přičemž permanentní a transitorní šoky počítá odděleně na panelových datech. Dále byla ověřena významnost faktorů, proběhla jejich kategorizace a na základě konzultace se zadavatelem i jejich redukce pro potřeby mikrosimulačního modelu. Jako vysvětlující faktory byly vybrány osobnostní charakteristiky zaměstnanců. Vlastní odvození koeficientů mzdové rovnice včetně zajištění jejich stability popisuje kapitola 3. Na základě podrobné analýzy, vícerozměrných statistických metod a evolučního programování byly vytvořeny dva modely, které nejlépe popisují mzdu pomocí mzdové rovnice s přijatelným množstvím parametrů. Téměř všechny faktory mzdové rovnice (včetně interakcí faktorů) jsou statisticky významné. Z důvodu odlišného vývoje růstu mezd se mzdy modelovaly zvlášť pro muže a ženy. Model 2 oproti Modelu 1 obsahuje navíc interakci CZ-ISCO se spojitými kategoriemi věku. Tato interakce vylepšuje predikční schopnosti mzdové rovnice zejména u hlavních tříd CZ-ISCO (rezidua nevykazují trend). Pro modelování mezd byl na základě zmíněných výhod doporučen Model 2. K modelování náhodných složek (permanentní mzdový rozdíl, permanentní šok a transitorní šok), které jsou součástí rovnice pro projekci mezd, se přistupovalo podle metodiky popsané v Carroll (1992). Byly odhadovány rozptyly náhodných šoků, na jejichž základě budou v jednotlivých krocích projekce mezd každému jedinci simulovány náhodné šoky ve mzdách. Predikovaná mzda z Modelu 2 a nasimulované šoky ve mzdách pro každého jedince budou následně implementovány do mikrosimulačního modelu důchodového systému.

K implementaci mzdové rovnice do mikrosimulačního modelu důchodového systému je potřeba zadavateli kromě parametrů mzdové rovnice dodat i rozptyly náhodných složek. Všechny parametry mzdové rovnice a rozptyly náhodných složek jsou součástí této závěrečné zprávy (kapitola 4). Přílohou závěrečné zprávy jsou také parametry Modelu 2 včetně rozptylů náhodných složek, a to ve formátu MS Excel a Stata verze 12.

Do budoucna by se dal model vylepšit zejména prodloužením časové řady šetření ISPV. Dalším výrazným posunem by bylo alespoň částečné zahrnutí podnikových proměnných. Podnikové proměnné nejméně v základním regionálním třídění Praha / ostatní, případně v třídění dle základních odvětvových skupin, které je obsaženo v některých zahraničních modelech. Významný přínos modelu z hlediska odvětvových skupin nabízí například dělení na finanční a nefinanční podniky. Model by se dal také vylepšit např. zahrnutím faktoru pravděpodobného věku odchodu do důchodu a předčasného důchodu. Některé zahraniční modely používají jako vysvětlující proměnnou i počet dětí, která však není součástí šetření ISPV.

I když je ISPV šetřením výběrovým, tak vzhledem k rozsahu (co do počtu šetřených zaměstnanců, sledovaných položek i období) a kvalitě je ISPV šetření, které je pro tvorbu rovnice mzdové dynamiky mikrosimulačního modelu důchodového systému velmi dobře využitelné.

Použitá literatura

- [1] Baltagi, B. H.: *Econometric Analysis of Panel Data*, 1995.
- [2] Bardaji J., Sédillot B., Walraet E.: *Un outil de prospective des retraites: le modèle de microsimulation Destinie*, 2003.
- [3] Blanchet D., Crenner E. a Le Minez S.: *The New Version of The Destinie Microsimulation Model*.
- [4] Blanchet D. a Le Minez S.: *Projecting Pensions and Retirement Age in France: some Lessons from the Destinie I Model*.
- [5] Blanchet D., Le Minez S.: *Joint macro/micro evaluations of accrued-to-date pension liabilities: an application to French reforms*, 2012.
- [6] Blundell R., Dias M. C., Meghir C. a Shaw J.: *The long-term effects of in-work benefits in a lifecycle model for policy evaluation*, 2012.
- [7] Carroll CH. D.: *The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence*, 1992.
- [8] Colin Ch.: *Modélisation des carrières salariales dans Destinie*, 1999.
- [9] Dahl M., DeLeire T., a Schwabish J.: *Trends in Earnings Variability Over the Past 20 Years*, 2007.
- [10] Dekkers G., Buslei H., Desmet R. Geyer, J., Hofmann D., Raitano M., Steiner T. Viktor, Tedeschi S. a Verschueren F.: *What are the consequences of the AWG-projections for the adequacy of social security pensions?; An application of the dynamic micro simulation model MIDAS for Belgium, Italy and Germany, Report of the Work Package 4 of The AIM project*, 2008.
- [11] Emmerson C., Reed H. a Shephard A.: *An Assessment of PENSIM2*.
- [12] Flood L., Jansson F., Pettersson T., Sundberg O. a Westerberg A.: *SESIM III - a Swedish dynamic micro simulation model*, 2005.
- [13] Haan P. a Prowse V.: *The Design of Unemployment Transfers: Evidence from a Dynamic Structural Life-cycle Model*.
- [14] Harris A. R., Sabelhaus J.: *Projecting longitudinal earnings patterns for long-run policy analysis*, 2003.
- [15] Harris A. R., Sabelhaus J. a Schwabish J. A.: *Projecting Labor Force Participation and Earnings in CBO's Long-Term Microsimulation Model*, 2006.
- [16] Hebák P., Hustopecký J, Malá I.: *Vícerozměrné statistické metody*, 2005.
- [17] Holmer M., Janney A. a Cohen B.: *PENSIM Overview*, 2014.
- [18] *Informační systém o průměrném výdělku, rok 2013, publikace výsledků šetření*.
- [19] Knoef M., Alessie R. a Kalwij A.: *Changes in the income distribution of the Dutch elderly between 1989-2020: a dynamic microsimulation*, 2010.
- [20] Keegan M. a Thurecht L.: *APPSIM – Modelling of Earnings*, 2008.
- [21] Klevmarken N. A.: *Behavioral modeling in micro simulation models. A survey.*, 1997.
- [22] Lawson T.: *Adding behaviour to microsimulation modelling by using agent-based modelling*, 2010. (European Workshop on Dynamic microsimulation modelling, Brusel).
- [23] Li J. a O'Donoghue C.: *Retirement Choice Simulation in Household Settings*, 2011.

- [24] Li J., a O'Donoghue C.: Retirement Choice Modelling in Dynamic Microsimulation, 2012.
- [25] Li J., O'Donoghue C.: A survey of dynamic microsimulation models: uses, model structure and methodology, 2013.
- [26] Manoli D. a Weber A.: The Effects of Increasing the Early Retirement Age on Social Security Claims and Job Exits, 2012.
- [27] Mazzaferro C., Morciano M. a Tedeschi S.: The Introduction of a Private Wealth Module in CAPP_DYN, 2010.
- [28] Novák, P., 2006: Analýza panelových dat. Diplomová práce, 2006.
- [29] Oplatková Z., Ošmera P., Šeda M., Včelař F., Zelinka I.: Evoluční výpočetní techniky – principy a aplikace, 2008.
- [30] Pylkkänen E.: Modeling Wages and Hours of Work, 2000.
- [31] Schwabish J. A.: Assigning Education Status in CBO's Long-Term Microsimulation Model, 2008.
- [32] Schwabish J. A. a Topoleski J. H.: Modeling Individual Earnings in CBO's Long-Term Microsimulation Model, 2013.
- [33] Sonsbeek van Jan-Maarten: Micro simulations on the effects of ageing-related policy measures: The Social Affairs Department of the Netherlands Ageing and Pensions Model, 2010.
- [34] Williamson P.: Optimal methods for mate-matching in a closed, discrete-time dynamic microsimulation model, 2010. (European Workshop on Dynamic microsimulation modelling, Brusel).
- [35] Zákony o dani z příjmu, vlastní výpočty.