

Univerzita Karlova v Praze
Fakulta sociálních věd

Institut ekonomických studií

Bakalářská práce

2005

Adam Tomis

**Univerzita Karlova v Praze
Fakulta sociálních věd**

Institut ekonomických studií

BAKALÁŘSKÁ PRÁCE

**Robustní regresní metody a jejich využití pro modelování
tranzitivních ekonomik**

(případová studie na základě dat o českém exportu do EU v 90. letech)

Vypracoval: Adam Tomis

Vedoucí: Doc. RNDr. Jan Ámos Víšek, CSc.

Akademický rok: 2004/2005

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a použil pouze uvedené prameny a literaturu.

V Praze dne

Adam Tomis

Robustní regresní metody a jejich využití pro modelování tranzitivních ekonomik

(případová studie na základě dat o českém exportu do EU v 90. letech)

Abstrakt:

Cílem práce je seznámit čtenáře se základními poznatky a technikami robustních regresních metod. Jejich nespornou výhodou proti klasickým odhadům reprezentovaným především metodou obyčejných nejmenších čtverců je schopnost čelit kontaminaci dat. Právě kontaminace, resp. heterogenita dat je z podstaty celého procesu ekonomické transformace vlastní datům z tranzitivních ekonomik. Na regresním modelu exportu ČR do Evropské unie v letech 1993 – 1999 je demonstrováno, že odhady pořízené robustními procedurami, jmenovitě odhady *least trimmed squares* a *least median of squares*, mají lepší vlastnosti než odhady pořízené metodou obyčejných nejmenších čtverců. Navíc se ukazuje, že heterogenita dat, kterou robustní odhady odhalí, má svou z ekonomické teorie vycházející logiku.

Robust regression methods and their utilization for modeling economies in transition

(Case study based on the Czech export data into the EU in the nineties)

Abstract:

The goal of this thesis is to make a reader familiar with the essential findings and techniques of the robust regression methods. Their undoubted advantage comparing to classical estimates represented mainly by the method of ordinary least squares is their ability to cope with contamination of the data. Because of the process of the economic transformation, contamination or more likely data heterogeneity is inherent to the data emerging from transition economies. The regression model of the Czech export into the EU during 1993 – 1999 period is used to demonstrate that robust estimates namely the *least trimmed squares* and the *least median of squares* achieve better characteristics than estimates acquired by the method of ordinary least squares. Moreover, it turns out that data heterogeneity detected by robust techniques has its economic theory based logic.

Obsah

1. Úvod.....	2
2. Robustní regrese	4
2.1 Co je robustní regrese?	4
2.2 Parametrické modely a typy odchylek.....	5
2.3 Cíle robustní regrese.....	9
2.4 Úskalí spojená s použitím robustních metod.....	11
2.5 Základní teoretické koncepce	12
2.6 Historické příklady robustních odhadů.....	17
2.7 Odhady s vysokým bodem selhání	20
2.8 Rozmanitost odhadů	24
3. Empirická analýza	27
3.1 Specifikace modelu	27
3.2 Globální model	33
3.3 Modely pro jednotlivé roky – metoda LTS	36
3.4 Modely pro jednotlivé roky – metoda RLS	40
3.5 Teoretické zdůvodnění rozdělení dat.....	41
3.6 Globální model na základě očištěných dat	45
4. Závěr.....	47
5. Reference.....	48
6. Přílohy.....	50
Příloha I. Odhad globálního modelu	50
Příloha II. Výpočet modelů metodou LTS	52
Příloha III. Výpočet modelů metodou RLS	65
Příloha IV. Vyloučená pozorování – porovnání metod.....	78
Příloha V. Klasifikace SITC 02	79

1. Úvod

Politické a potažmo ekonomické reformy na přelomu osmdesátých a devadesátých let s sebou přinesly pro země bývalého východního bloku, Českou republiku nevyjímaje, změny nevídaného rozsahu. Bezprecedentní přechod od jedné z nejrigidnějších centrálně plánovaných ekonomik světa k ekonomice tržní s sebou nesl veliká úskalí. Unikátnost celého procesu transformace tudíž představovala a do jisté míry stále ještě představuje pro soudobou ekonomickou obec předmět širokého zájmu.

Z jedinečného charakteru celého procesu plyne nutnost užívání nestandardních a specifických metod nejenom v oblasti politické, národohospodářské či zákonodárné, ale svou povahou si transformace vynucuje použití specifických metod i v oblasti svého vlastního zkoumání. Právě teoretické obeznámení s několika takovýmito specifickými metodami, týkajícími se vybrané partie matematického modelování a ukázkou jejich praktické aplikace na reálných datech o českém exportu si klade za cíl tato práce.

Jmenovitě jde o regresní odhady *least trimmed squares* a *least median of squares*, které mohou představovat, především v případech, kdy je zkoumaný vzorek dat heterogenní povahy, tj. sdružující dvě nebo více autonomních populací, velmi efektivní alternativy k nejběžněji používaným odhadům reprezentovaným především metodou obyčejných nejmenších čtverců či metodou maximální věrohodnosti.

Proč právě v datech týkajících se exportu České republiky do EU v 90. letech lze očekávat heterogenitu? S procesem ekonomické transformace velice úzce souvisí i liberalizace zahraničního obchodu, která svým rozsahem představovala pro české podniky nevídanou a neopakovatelnou příležitost. Podmínkou úspěšného vstupu na doposud zapovězené západoevropské trhy však byla efektivní restrukturalizace a především schopnost rychle změnit administrativně určenou strukturu výroby na strukturu flexibilní, poptávkově orientovanou, kterou si vyžadují vyspělé západní trhy. Jelikož tento proces probíhal v rozdílných odvětvích s rozdílnou rychlostí a intenzitou, je pravděpodobné, že data týkající se exportu jsou vnitřně nesourodá.

Dynamika českého zahraničního obchodu je úctyhodná. Objem exportu do Evropské unie vzrostl z 8 mld. USD v roce 1993 na 18,4 mld. USD v roce 1999, což představuje průměrný roční růst 16,3 %, zatímco do ostatních světových zemí to byl růst jen zhruba 4% (Benáček, Víšek (2002b)). Vzhledem k tomu, že Česká republika je malou

otevřenou ekonomikou, má analýza exportu, jakožto významné složky podílející se na celkovém hrubém domácím produktu, zásadní význam.

Snahou regresní analýzy bude odhalit, na základě dat o exportu České republiky do Evropské unie v letech 1993 – 1999, hlavní determinanty, které ovlivňovaly jeho vývoj. Bohužel kvalita odhadů pořízená metodou obyčejných nejmenších čtverců není valná. Důvodem je právě heterogenita dat.

Ukážeme, že robustní metody jsou schopny se s touto situací vypořádat podstatně lépe. Na rozdíl od metody obyčejných nejmenších čtverců se nesnaží aplikovat model na všechna data. Jsou schopny identifikovat odchylující se pozorování, eliminovat jejich vliv a na takto očištěných datech pak získat podstatně kvalitnější odhady.

Navíc ukážeme, že heterogenita dat má svůj ekonomický smysl, který má kořeny v podstatě celého procesu ekonomické transformace. Robustní analýza nám tak umožní nahlédnout, ve kterých odvětvích restrukturalizace rychle a úspěšně proběhla a naopak, ve kterých odvětvích více méně přetrvaly původní, s tržním hospodářstvím se neslučující způsoby řízení a organizace výroby.

Vzhledem k tomu, že robustní metody nepatří mezi standardní náplň ekonometrických kurzů a neřadí se ani mezi běžně užívané nástroje ekonomického výzkumu, je práce rozdělena do dvou hlavních částí.

První část seznamuje čtenáře se základními pojmy a výsledky robustních metod zpracování dat. Je rozdělena do několika kapitol, které mapují nejpodstatnější témata teorie robustní regrese. Ve stručnosti nastiňuje cíle robustní regrese a prostředky, jakými se těchto cílů více, či méně úspěšně dosahovalo a dosahuje. Zároveň upozorňuje na některé problémy a úskalí, která mohou především při nedostatečně teoreticky podložené aplikaci nastat.

Druhá část se věnuje specifikaci modelu, praktické demonstraci použití robustních regresních technik na empirických datech a následně též porovnání jejich výsledků s výsledky získanými pomocí metody obyčejných nejmenších čtverců.

2. Robustní regrese

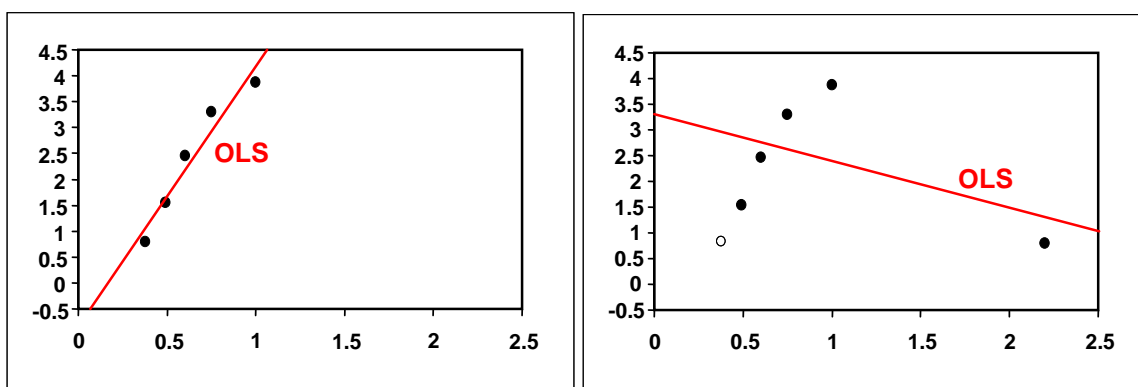
2.1 Co je robustní regrese?

Regresní analýza představuje velice efektivní a podotkněme, že v některých případech též velice pozoruhodný způsob zpracování dat. Možnosti její aplikace jsou velmi široké, své uplatnění tudíž nachází v mnoha vědních oborech. Základní procedury pro její výpočet nalezneme ve větší či menší míře v každém statistickém programu. Výrobci statistických programů uvádějí, že téměř 90 % úloh zpracovávaných statistickými softwary je regresního charakteru (Víšek (2003)). Vysoká dostupnost této bezpochyby úžasné techniky zpracování dat s sebou na jedné straně přináší možnost, aby si téměř každý počítačově gramotný jedinec provedl svou vlastní regresní analýzu na datech rozličného původu, avšak na straně druhé přináší riziko v podobě zpracovávání dat bez znalosti potřebné teorie, které může v extrémním případě vést až k naprosto bezcenným výsledkům.

Nejznámější klasické regresní odhady, reprezentované především metodou obyčejných nejmenších čtverců (OLS) a metodou maximální věrohodnosti (ML) jsou úzce spjaty s tradičními předpoklady jako linearita regrese, nezávislost pozorování, homoskedasticita a normalita disturbancí. Přestože pro většinu těchto předpokladů existuje celá řada aposteriorních diagnostik a testů, testování jejich splnění je z různých důvodů velice často opomíjeno s tím, že se předpokládá, že většina z nich je alespoň částečně splněna. V praxi se, bohužel, ukazuje, že tomu tak není a že i velmi nepatrné odchylky od základních předpokladů modelu mohou mít fatální dopad na takto získané výsledky. Dříve než se začneme podrobněji zabírat teorií robustní regrese, podívejme se na malý motivační příklad na obrázku 2.1.

Intuitivně asi cítíme, že model v levé části obrázku vystihuje data, resp. jejich převažující charakter, i v případě, že jejich geometrie vypadá jako na obrázku vpravo. Metoda nejmenších čtverců nám však regresní model na data umístí podstatně jinak a svým způsobem tak dává zavádějící výsledky.

2.1. Citlivost odhadu OLS na odlehlá pozorování



Obrázek 2.1. názorně ukazuje, co může nastat v případě odhadu pořizovaného metodou obyčejných nejmenších čtverců (OLS), obsahuje-li soubor pozorování data, která jsou z nějakého důvodu vychýlena proti své „skutečné“ pozici, tedy kontaminována. Odchylna může být způsobena mnoha vlivy: například chybou měření, či špatným přepisem dat do počítače nebo i tím, že soubor dat obsahuje podsoubory rozdílného charakteru. Jak je z obrázku patrné, pouhé jedno jediné pozorování může odhad naprosto znehodnotit.

V souvislosti s výše uvedeným zjištěním nás tedy přirozeně zajímají dvě otázky:

- (i) Do jaké míry jsou klasické statistické postupy použitelné a za jakých podmínek si ještě zachovávají svou optimalitu?
- (ii) Existují nějaké jiné postupy či metody, které nejsou tolik vázány na splnění určitých základních podmínek?

V širším neformálním smyslu lze tedy říci, že robustní statistika se potýká s faktem, že mnohé, často zásadní předpoklady klasických statistických modelů jsou porušovány nebo alespoň nejsou naplněny beze zbytku (Hampel & kol. (1986)).

2.2 Parametrické modely a typy odchylek

Klasické statistické postupy mají typicky parametrický charakter: model je plně určen až na hodnoty několika parametrů nabývajících reálných nebo vektorových hodnot, které odhadujeme a testujeme. Výhodou tohoto přístupu je schopnost modelu koncentrovat obrovské množství informací uložených ve stovkách a tisících individuálních dat v několika málo parametrech a uspořádat tedy celkovou informaci do jasné stochastické struktury, která nám umožňuje učinit o datech více méně jednoznačný závěr. Nevýhodou tohoto přístupu je fakt, že dobré vlastnosti a závěry z těchto modelů plynoucí jsou podmíněny splněním některých základních předpokladů. Zpravidla se jedná především o vazbu na určitou distribuci.

Většina nejběžněji používaných statistických postupů je odvozena za předpokladu, že pozorování byla generována na základě modelu řídicího se normálním rozdělením. Nepochybnou výhodou tohoto přístupu je jeho algebraická jednoduchost, a proto bývá používán všude tam, kde je nosičem hustoty celá reálná přímka, přičemž se na předpoklad normality jaksi „zapomíná“.¹

Fakt, že řada klasických postupů jako Snedecorův F-test, odhad metodou nejmenších čtverců, či v menší míře například Studentův t-test je značně citlivá na odchylky od normálního rozdělení vedl, především od čtyřicátých let 20. století, ke snaze zavést postupy, které nejsou tak citlivé na splnění alespoň některých požadavků kladených na klasické metody.

Často se tak můžeme setkat s *neparametrickými postupy*, které chápou celou hustotu, případně celou regresní funkci, jako neznámý parametr, a nejsou tudíž tolik závislé na tvaru základního pravděpodobnostního rozdělení, přičemž si zachovávají některé dobré vlastnosti pro co nejširší třídu distribučních funkcí. Dalšími postupy, se kterými se v této souvislosti můžeme setkat, jsou *semiparametrické postupy*, *adaptivní statistické postupy*, *částečně adaptivní postupy* či *robustní postupy*.

Vzhledem k tomu, že všechny tyto teorie vznikaly paralelně v průběhu relativně nedávné minulosti, není mezi nimi ostrá hranice a některé cíle a pojmy se vzájemně překrývají.

Robustní statistické postupy můžeme v tomto smyslu chápat jako určitá vylepšení klasických metod, která si zachovávají svou optimalitu v okolí nějakého rozdělení pravděpodobností, např. normálního. Jsou to modifikace klasických metod, které jsou optimální v okolí daného rozdělení vzhledem k určité vzdálenosti a vzhledem k určitému kritériu optimality. Jako takové jsou vydatnější než neparametrické postupy, které za svou funkčnost pro široký model platí určitou ztrátou vydatnosti (Jurečková (2001), str. 2).

Vzhledem k tomu, že robustní metody „vylepšují“ klasické parametrické metody ve smyslu zvýšení odolnosti modelů proti odchylkám od jejich základních předpokladů, je užitečné rozebrat, k jakým odchylkám a z jakých důvodů obvykle dochází. Typy odchylek

¹ Například odhad OLS je velice úzce spjat s normálním rozdělením disturbancí. V případě, že jsou disturbance rozděleny podle normálního modelu (a za předpokladu, že jsou splněny ostatní předpoklady modelu) je OLS nejlepším nestranným odhadem mezi *všemi* nestrannými odhady, pokud však disturbance normálně rozděleny nejsou, je podle známé Gauss-Markovovy věty OLS odhad *pouze* nejlepším nestranným odhadem mezi *všemi lineárními* odhady, což je podstatně užší třída odhadů.

můžeme v návaznosti na jejich příčiny rozdělit do následujících tří skupin ((Hampel & kol. (1986)):

- (i) výskyt hrubých chyb,
- (ii) zaokrouhlování a agregace,
- (iii) model je aproximací skutečnosti.

Ad (i), důvod starý pravděpodobně jako statistika sama. Typickým příkladem jsou překlepy, špatné umístění desetinné čárky, chyba měřicího přístroje, chyba v přepisu dat nebo třeba zemětřesení, které zkreslí přístrojem naměřenou hodnotu.² Jedná se o případy, kdy v souboru pozorování vzniknou výrazně odchylené body (angl. outlier problem³).⁴

Zdálo by se, že takovéto body musí být při grafické analýze „podezřelé“ svojí velkou odchylkou od hlavního shluku dat, tak jako je tomu na obrázku 2.1. Pro malé soubory pozorování s omezenou dimenzionalitou lze skutečně tyto body pečlivou grafickou analýzou odhalit. Situace se ovšem dramaticky komplikuje s rostoucím počtem pozorování a především s rostoucí dimenzionalitou zkoumaného modelu. Vlivné body se v oblaku dat tzv. maskují a i s použitím velmi kvalitních grafických editorů umožňujících všemožné projekce a natáčení dat je v podstatě nemožné je odhalit.

Dodejme ještě, že komplexita celého problému je o to větší, uvědomíme-li si, že ne každý vlivný bod představuje špatně zaznamenané pozorování, které je třeba vymazat nebo v lepším případě opravit. Odlehlá pozorování mohou být často naopak tím nejcennějším, co v celém souboru pozorování máme, neboť právě tyto body jsou nositeli největší informace.⁵ Mohou totiž poukazovat na nějakou zajímavou strukturu či odlišnost, která v datech není na první pohled patrná. Například mohou napomoci odhalit, že se soubor dat skládá ze směsi dvou nebo více různých populací.⁶ Problém detekce vlivných bodů je tedy

² Ne nadarmo je statistika označována jako přesná věda o nepřesných číslech.

³ V anglicky psané literatuře se rozlišují dva typy odlehlých pozorování. Typem prvním jsou odlehlá pozorování v prostoru vysvětlované proměnné (resp. vysvětlovaných proměnných) tzv. *outliers*. Druhým typem jsou pozorování odlehlá v jedné nebo více souřadnicích faktorového prostoru vysvětlujících proměnných, tzv. *leverage points*. Uspokojivý překlad do češtiny především u druhého typu pravděpodobně neexistuje, bývají proto souhrnně označovány jako vlivné body (Víšek (1997)).

⁴ Podotkněme na tomto místě, že vzhledem k tomu, že téměř veškerá literatura týkající se teorie robustní statistiky je i v našich zeměpisných šířkách publikována anglicky, neustálilo se zatím české názvosloví. Použité termíny jsou tedy volným překladem anglických ekvivalentů.

⁵ Tato myšlenka se projevuje i u metody nejmenších čtverců. Rezidua umocněná na druhou přiřazují odlehlým pozorováním větší váhu. Z tohoto důvodu nelze k odhalení vlivných bodů použít analýzu reziduí OLS.

⁶ Zajímavým příkladem v tomto směru je například objevení argonu v důsledku rozdílných atomových hmotností „dusíku“ získaného ze vzduchu a jiných chemických sloučenin (Hampel & kol. (1986)).

jedním z největších úkolů teorie robustní statistiky. Závažnost problému potvrzují i údaje uvedené v následujícím odstavci:

„Pečlivé studie provedené v posledních letech poukazují na fakt, že v podstatě každý soubor dat v sobě nese jistou dávku kontaminace. V medicíně je to okolo 8-12 %, v zemědělství kolem 6-7 %, data ze seismografů vykazují 5-7% a technická data zhruba 10% obsah vlivných pozorování. Dokonce i úmrtnostní tabulky, považované za jeden z nespolehlivějších zdrojů informací, obsahují 7-10 % chyb. V současné době panuje shoda, že kontaminace pod 1 % je i v pečlivě připravených a kontrolovaných experimentech v podstatě nedosažitelná.“ Při měření veličin ve společenských vědách, ekonomii nevyjímaje, je kontaminace ze zřejmých důvodů obvykle ještě vyšší.

„V případě tranzitivních ekonomik jsou už z podstaty věci celého procesu data kontaminována a v některých případech složena alespoň ze dvou rozdílných populací, jak uvidíme později.“ (obě citace: Víšek (1999), str. 427, překlad AT)

Vzhledem k zaměření této práce je bod (i) asi nejzásadnější. Shrňme jej tedy tvrzeními: (1) postačuje jen jedno velmi odlehlé pozorování k tomu, aby značně poškodilo či znehodnotilo celou statistickou analýzu a (2) určitému procentu kontaminace se téměř nelze vyhnout.

Ad (ii), dalším důvodem proč předpoklady modelů obvykle nejsou splněny, je fakt, že jakékoli měření dat má svoji omezenou přesnost a tudíž data jsou tak ve své podstatě diskrétního charakteru. Navíc jsou nejružnějším způsobem zaokrouhlena a agregována. Jejich nahrazení spojitou distribucí tudíž není vždy ideální. Hrubé měření dat tak může vést k systematickým nepřesnostem, které mohou v určitých případech vést k podstatným poklesům eficeince (více viz: Hampel & kol. (1986)).

Ad (iii), i sebelepší model je pouhou aproximací reality. Předpoklady jako normalita, linearita, nezávislost, či limitní chování tak představují pouze zjednodušení a idealizaci skutečného světa. Empirická data se však podle těchto předpokladů zcela přesně nechovají. Nejobvyklejším příkladem tohoto typu jsou odchylky od normality. Centrální limitní věta sice zaručuje konvergenci k normálnímu rozdělení, nicméně neříká nic o tom, jak daleko se ve skutečnosti od této limity nacházíme. Ukazuje se, že v praxi normalita obvykle dosažena není:

„Například i velké a kvalitní soubory pozorování, v nichž je celá informace soustředěna v kompaktním oblaku dat, neobsahujícím vlivné body, tíhnou sice k malým, ale stále ještě podstatným odchylkám od normálního modelu. Proti klasické distribuci vykazují

hustota obvykle těžší chvosty.“ (Hampel & kol. (1986)), str. 22, překlad AT). To se samozřejmě opět projevuje především ve vydatnosti získaných odhadů.

Robustní statistika tedy zkoumá chování klasických parametrických modelů při odchylkách od jejich předpokladů a snaží se nalézt takové postupy a metody, které budou schopné dávat uspokojivé výsledky i v případech, kdy jsou tyto předpoklady porušeny, resp. v případech, kdy je soubor dat kontaminován.

Robustní regresní metody vysoce odolné ke kontaminaci (označované jako metody s vysokým bodem selhání⁷ - definice viz kapitola 2.5.) jsou v současné době schopny čelit až 50% kontaminaci dat. To je nejvyšší dosažitelná hranice, neboť v případě vyšší kontaminace vyvstává reálný problém, jak určit, která pozorování přísluší mezi „ta správná“ a která nikoli.

2.3 Cíle robustní regrese

Známe-li druhy a příčiny odchylek, kvůli kterým klasické regresní postupy selhávají, rozeberme stručně cíle, které si v souvislosti s nimi klade robustní statistika. Opět je možné je shrnout do několika bodů (Hampel & kol. (1986)):

- (i) Popsat strukturu nejlépe vystihující data.
- (ii) Identifikovat odchylující se pozorování („outliers“) nebo odchylující se substruktury pro následné zacházení s nimi.
- (iii) Identifikovat a varovat před vysoce vlivnými body („leverage points“).
- (iv) Pokud jsou navíc v modelu předpokládány nějaké korelační struktury, tak nalézt i případné odchylky od těchto struktur.

Ad (i), stejně jako v případě „běžné“ regrese předpokládáme nějaký parametrický model, jehož koeficienty se snažíme odhadnout a otestovat. Navíc též explicitně počítáme s tím, že model může být „deformovaný“ a část pozorování (menšina) mu nemusí vůbec odpovídat. Závěry jsou tedy podmíněné; snažíme se model ponechat jako aproximaci pro většinu dat s tím, že pravdivost závěrů je zaručena jenom pro situace, ve kterých je v datech nejvýše určité omezené procento kontaminace.

⁷ angl.: *high breakdown point methods*.

Ad (ii), rezidua z robustního odhadu automaticky odhalí odlehlá pozorování, což například rezidua OLS nedokáží. Samostatnou otázkou zůstává, jak s vlivnými body naložit.

Principiálně by měly tyto body být studovány separátně. Studium těchto pozorování může napomoci odhalit nějakou speciální strukturu nebo neočekávaný fenomén. Ovšem dokonce i v případech, kdy se jedná čistě o hrubé chyby, může analýza těchto bodů nejdnou napomoci odhalit důvod jejich vzniku a přispět tak k jejich opravě (například posunutá desetinná čárka či přehozené pořadí číslic). Jedině pokud se jedná o dobře etablovaný model a zajímá nás popis *pouze* většinové části dat, je oprávněné zamítnout tato data bez jakéhokoli dalšího šetření.

Ad (iii), zatímco pro vyvážené nebo skoro vyvážené soubory pozorování je možné najít eficientní robustní metody, tak v případech, kdy jeden nebo několik málo bodů vzniklých za vysoce informativních podmínek, má výrazný vliv na odhad, se situace podstatně komplikuje. Typickým příkladem je jednoduchá regrese na obrázku 2.1.

Existence takovýchto pozorování s sebou nese následující konflikt: pokud tato pozorování vymažeme, ačkoli jsou relevantní, pak náš odhad ztratí velké množství eficiency. Na druhou stranu pokud tato pozorování v souboru ponecháme, i když se jedná o hrubé chyby, naše výsledky tím můžeme podstatně poškodit nebo úplně znehodnotit. Jedinou rozumnou (i když jak uvidíme v kapitole 2.8 rozhodně nikoli samospasitelnou) možností se v tomto případě jeví aplikace dvou robustních odhadů. Jeden na soubor dat obsahující tyto body a druhý na soubor těchto bodů zbavený. Pokud se výsledky odhadů liší (ve statistickém smyslu slova), pak je alespoň nějaký z bodů buď hrubá chyba nebo je model špatný.

Ad (iv), jedná se o nejnovější oblast ze všech čtyř výše jmenovaných. Jestliže v minulosti byly hlavním cílem robustní statistiky především odchylky od určité distribuce (body (ii) a (iii)), tak v současnosti se ukazuje, že kromě vlivných pozorování se často mezi daty vyskytují i nejrůznější formy korelací. A to dokonce i v takových oborech, jako je astronomie, fyzika, či chemie, kde lze nezávislost pozorování oprávněně předpokládat. Odchylky od nezávislosti zužují konfidenční intervaly jednotlivých parametrů. Snahou robustních statistiků v této oblasti je odhad korelačního koeficientu a namodelování příslušného procesu, který data generoval tak, aby bylo s jejich pomocí možno opravit délku těchto intervalů.

2.4 Úskalí spojená s použitím robustních metod

Proč tedy, když robustní odhady nabízejí zjevné výhody, je stále nejběžněji používanou metodou odhad pomocí obyčejných nejmenších čtverců? Vždyť už sami její duchovní otcové C. F. Gauss a A. M. Legendre na počátku devatenáctého století, kdy byla metoda nejmenších čtverců navržena, tvrdili, že se jedná o nejprimitivnější a svým způsobem vágní způsob pořízení odhadů. Tehdy dostala tato metoda přednost především z výpočetních důvodů, jelikož jednoznačné řešení lze získat relativně snadno aplikováním jednoduché maticové algebry.

Co tedy způsobuje, že i v době počítačů se robustní odhady přes své nepochybné výhody jen obtížně prosazují proti tradičním odhadům reprezentovaným především metodami OLS a maximální věrohodnosti? Tak, jako každá mince má dvě strany, i robustní odhady mají „stinné“ stránky představující daň, kterou musíme platit za výhodu robustnosti. Naštěstí pro ně je tato daň v porovnání s výhodami, které nabízejí, relativně malá. Největší úskalí lze shrnout do následujících bodů:

- (i) nižší eficeience,
- (ii) výpočetní náročnost,
- (iii) komplikovanější interpretace výsledků.

Ad (i), často zmiňovaným nedostatkem odhadů s vysokým bodem selhání je jejich nízká, resp. proti klasickým metodám relativně nižší eficeience (či chcete-li vydatnost). Toto tvrzení ovšem přirozeně platí *pouze* pro nekontaminované soubory dat, kde je OLS skutečně nejlepším ze všech nestranných odhadů. Tento pokles eficeience je však zanedbatelný v porovnání se ztrátou výkonnosti, která nastane u klasických statistických metod aplikovaných na kontaminovaná data. Připomeňme na tomto místě motivační příklad uvedený v kapitole 2.1. Další šokující příklady neeficeience většiny běžně užívaných odhadů parametrů polohy a variance lze nalézt v Hampel & kol. (1986). Navíc citace uvedená v kapitole 2.2 jasně ukazuje, že nekontaminovaná data v praxi v podstatě neexistují a kontaminace na úrovni kolem 10 % je spíše pravidlem než výjimkou.

Ad (ii), mnohem větší problém než úroveň eficeience představuje u robustních odhadů jejich výpočet. Jak uvidíme později, stejně jako standardní odhady jsou i robustní odhady definovány jako extrémální úlohy. Vzhledem k povaze optimalizovaných funkcí však není obvykle možné vyjádřit explicitní formuli pro jejich výpočet. Řešení je tedy nutné hledat numericky.

Přestože v některých případech jsou známy algoritmy pro nalezení přesného řešení, nelze je většinou z důvodů abnormální časové náročnosti pro rozsáhlejší vzorky rozumně použít. Obvykle tedy hledáme řešení přibližné, které bývá založeno na nějakém matematickém triku, díky němuž je vyřešení úlohy z hlediska výpočetního času méně náročné. Nalezení takového triku, jeho implementace a otestování oprávněnosti jeho použití představuje velice odborný, náročný a zdlouhavý proces.

Lze tedy říci, že nalezení spolehlivého výpočetního algoritmu představuje největší problém praktické aplikace robustních metod. Ovšem rozhodně se nejedná o problém nepřekonatelný, neboť pokud je tento jakkoli komplikovaný algoritmus již předem zabudován do nějakého statistického programu, nemusí složitost jeho nalezení běžného uživatele příliš trápit (Víšek (1999)).

Ad (iii), posledním úskalím, které je spojeno s užitím robustních metod a které bychom rozhodně neměli opominout, je podstatně zdlouhavější analýza dat a s ní související složitější interpretace výsledků. Ta vyžaduje výrazně hlubší teoretickou znalost a v neposlední řadě také větší pečlivost a trpělivost při analyzování dat a utváření závěrů.

2.5 Základní teoretické koncepce

Přístupů k teorii robustní statistiky existuje v současné době celá řada. Nejstarší myšlenky sahají až do prehistorie statistiky.⁸ Prvními, kteří si výrazněji povšimli citlivosti průměru a rozptylu k odlehlým pozorováním, byli astronomové a fyzikové, kteří se při svých pozorováních pokoušeli stanovit hodnoty nejrůznějších fyzikálních, geofyzikálních, či astronomických konstant jakožto průměrů naměřených hodnot (Jurečková (2001)). Pověstí šlo však spíše o ad hoc řešení jednotlivých problémů než o systematictější práce.

Intenzivnější studie začaly vznikat až od 40. let dvacátého století. Do tohoto období spadá například princetonská skupina kolem J. W. Tukeyho, která se systematictěji zabývala studiem alternativ k metodě OLS.

⁸ Zajímavým historickým příkladem je bezesporu způsob jímž starověcí Řekové určovali délku obléhacích žebříků při dobývání měst; sčítali vrstvy cihel v hradebních zdech a ke stanovení průměrné výšky opevnění z těchto jednotlivých součtů nepočítali aritmetický průměr, nýbrž *medián*, který je sám o sobě velmi robustní metodou (Hampel & kol. (1986)).

Prvním, kdo v této souvislosti použil označení *robustní*, byl G. E. P. Box (Box (1953)), který jím označoval *odhad málo citlivý ke změnám parametrů rušivých, ale vydatný, tj. citlivý k parametrům, které jsou středem jeho zájmu* (Jurečková (2001), str. 31).

Za první, opravdu ucelenou formalizovanou práci shrnující teorii robustních odhadů lze považovat Huberovy články „Robust estimation of a location parameter“ z roku 1964 a „The behaviour of maximum likelihood estimates under nonstandard conditions“ z roku 1967 (Huber (1964) a (1967)). Základní myšlenkou byl přístup pomocí minimaxu. Huber totiž arbitrárně nepředpokládal jediné rozdělení dat v modelu, ale zaobíral se vždy celou třídou příbuzných rozdělení. Jedno rozdělení (obvykle normální) bylo považováno za tzv. centrální a zmíněná třída rozdělení pak tvořila okolí (nikoliv nutně v topologickém smyslu)⁹ tohoto centrálního rozdělení. Při předpokládaném podílu kontaminace dat pak Huber hledal odhad minimalizující maximální možný rozptyl ve zmíněném okolí centrální distribuce. Jinými slovy se snažil minimalizovat nejhorší možný stav, měřeno asymptotickým rozptylem odhadu, který by při dané kontaminaci mohl teoreticky nastat. To dalo vzniknout takzvaným M-odhadům,¹⁰ které se záhy staly poměrně populárním nástrojem.

Druhou hlavní koncepcí je Hampelův infinitezimální přístup založený na tzv. influenčních funkcích (Hampel & kol. (1986)). „*Jeho hlavní výhodou je možnost generalizace na jakýkoli parametrický model, což Huberova strategie minimaxu neumožňuje*“ (Čížek (2001), str. 9, překlad AT). Koncepce se opírá o tři klíčové pojmy: kvalitativní robustnost, influenční funkce a bod selhání¹¹.

K ilustraci jejich významu lze nalézt hezkou analogii v matematické analýze. Představíme-li si robustní odhad (odhad je zde myšlen jakožto náhodná veličina, tj. statistika, která je definována na prostoru všech distribučních funkcí)¹² jako reálnou funkci, pak kvalitativní robustnost je analogická její spojitosti. Influenční funkci lze přirovnat

⁹ Označíme-li centrální rozdělení F a alternativní rozdělení G . Okolím rozdělení F (v „netopologickém“ smyslu) pak můžeme rozumět například množinu O , která je lineární kombinací (tzv. Huberův model kontaminace) distribučních funkcí F a G , tedy $O = \lambda F + (1 - \lambda)G$, kde $\lambda \in (0;1)$.

¹⁰ Název M-estimator (M-odhad) je odvozen z anglického „generalized maximum likelihood“ (Hampel & kol. (1986)).

¹¹ Anlg. *breakdown point*. (Do češtiny se tento termín někdy též překládá jako *bod zvratu*, vzhledem k možným asociacím se však přidržme překladu *bod selhání*.)

¹² Anglicky psaná literatura rozlišuje v tomto případě dva termíny. *Estimator*, který znamená právě onu zmíněnou náhodnou veličinu a *estimate*, který je konkrétním odhadem regresních koeficientů pro konkrétní data. Čeština si, žel bohu musí vystačit pouze se slovem odhad, pečlivý čtenář však obvykle rozdíl vyrozumí z kontextu.

k první derivaci v určitém bodě a bod selhání jako vzdálenost z tohoto bodu k nejbližší singularitě. Zkusme na celý problém nahlédnout poněkud formálněji.

Uvažujme statistiku T počítanou na vzorku \mathbf{X} ($\mathbf{X} = (x_1, \dots, x_n)$), kde $x_i \in \mathfrak{R}^p$ (\mathfrak{R}^p naznačuje, že se jedná o p -rozměrný euklidovský prostor) pro všechna $i=1, \dots, n$). Kvalitativní robustnost je definována jako spojitost statistiky T v distribuci s tím, jak se mění distribuční funkce, která generovala naše pozorování. Takto definovaná robustnost se nazývá kvalitativní, neboť pouze říká, zda je, či není daný statistický funkcionál robustní.

Protože však chceme srovnávat jednotlivé statistické funkcionály mezi sebou, potřebujeme robustnost nějakým způsobem kvantifikovat, vyjádřit ji nějak číselně. Takových způsobů je celá řada, avšak nahrazení tak komplexního pojmu jako robustnost jedním číslem je vždy svým způsobem zjednodušující a jednostranné. Mezi nejznámější charakteristiky robustnosti patří *influenční funkce a bod selhání*.

Influenční funkce popisuje, přibližně a standardizovaně, chování statistiky T pro velmi malé změny souboru pozorování (přidání jednoho pozorování $x \in \mathfrak{R}^p$). Obvykle je definována jako první derivace statistiky T na nekonečně rozměrném prostoru distribučních funkcí ve směru distribuční funkce, která dává bodu x jednotkovou pravděpodobnost, tj. tato distribuční funkce je degenerovaná v bodě x .

Má-li T být robustní, měla by mít ohraničenou influenční funkci, což je vlastnost, která přímo neplyne z její spojitosti, tj. kvalitativní robustnosti (Jurečková (2001)). Formálně lze influenční funkci (IF) definovat například následovně:

$$IF(\mathbf{X}; T, F) = \lim_{t \rightarrow 0} \frac{T((1-t)F + tG_{\mathbf{X}}) - T(F)}{t}, \quad (2.1)$$

kde T je zkoumaná statistika, F je předpokládané rozdělení dat, \mathbf{X} je soubor pozorování $\mathbf{X} = (x_1, \dots, x_n)$ a $G_{\mathbf{X}}$ je rozdělení vzorku \mathbf{X} vypočtené na základě empirické distribuční funkce. Jak je z definice patrné, nejedná se o „směrovou“ derivaci, tak jak je definována například běžná parciální derivace, nýbrž jde o derivaci, která popisuje chování statistiky T v souvislosti s tím, jak se blížíme od jedné distribuce k druhé (v tomto případě tedy od G k F). Směr derivace nahrazuje zkoumaný vzorek dat \mathbf{X} , který představuje souřadnice v nekonečně rozměrném prostoru všech možných distribucí, „čímž udává, odkud a kam se bude derivovat“.

Lokálně tedy můžeme statistiku T pomocí influenční funkce linearizovat, což nám dává jednoduchý a účinný nástroj pro studování jejích vlastností. Pokud nás tedy například zajímá dopad dodatečného pozorování na statistiku T , můžeme IF použít následovně.

Označme $F_{\mathbf{X}_n}$ empirickou distribuční funkci vypočtenou na základě vzorku \mathbf{X}_n (dolní index označuje, že se jedná o vzorek čítající n pozorování tedy $\mathbf{X}_n = (x_1, \dots, x_n)$) a $F_{\mathbf{X}_{n-1}}$ jako empirickou distribuční funkci vypočtenou na základě vzorku $\mathbf{X}_{n-1} = (x_1, \dots, x_{n-1})$. Po dosazení do (2.1) dostaneme

$$IF(\mathbf{X}_{n-1}; T, F) = \lim_{t \rightarrow 0} \frac{T((1-t)F_{\mathbf{X}_{n-1}} + tF_{\mathbf{X}_n}) - T(F_{\mathbf{X}_n})}{t}. \quad (2.2)$$

Pokud nyní dosadíme za t člen $1/n$ dostaneme

$$IF(\mathbf{X}_{n-1}; T, F) = \lim_{\frac{1}{n} \rightarrow 0} \frac{T((1 - \frac{1}{n})F_{\mathbf{X}_{n-1}} + \frac{1}{n}F_{\mathbf{X}_n}) - T(F_{\mathbf{X}_n})}{\frac{1}{n}}, \quad (2.3)$$

což je hodnota IF pro soubor x_1, \dots, x_n s množstvím kontaminace velikosti $t = 1/n$.

Jako aproximaci asymptoticky definované influenční funkce lze využít tzv. křivku sensitivity (SC) (Tukey (1970-71)), která pro konečně velký soubor pozorování obvykle k influenční funkci konverguje. Pro odhad T je definována následovně:

$$SC(\mathbf{X}_{n-1}) = n \cdot (T_n(x_1, \dots, x_{n-1}, x_n) - T_{n-1}(x_1, \dots, x_{n-1})) \quad (2.4)$$

odhadu T_n počítaného na vzorku x_1, \dots, x_{n-1} a stejně jako (2.3) měří efekt dodatečného pozorování. Na konceptu influenční funkce je založena celá řada dalších charakteristik. Za všechny jmenujme alespoň *globální a lokální citlivost* (více viz. Jurečková (2001)).

Jak naznačuje obrázek 2.1. v první kapitole, i jedno jediné odlehle pozorování může v případě OLS celý odhad výrazně vychýlit a v krajním případě ho úplně „zničit“. Na druhou stranu, jak uvidíme záhy, existují metody, které jsou schopné spolehlivě odolat určitému množství kontaminace dat. K měření této vlastnosti jednotlivých bodových odhadů slouží globální měřítko robustnosti – koncept bodu selhání.

Zde uvedená definice je převzata z knihy P. J. Rousseeuwa a A. M. Leroye (Rousseeuw, Leroy (1987), kap. 1.2.). Platí sice jen pro konečně velký vzorek, ale na druhou stranu na rozdíl od Hampelovy definice (Hampel & kol. (1986), kap. 2.2.), která je sice obecnější, není tato asymptotická, a je tudíž poněkud průhlednější.

Vezměme libovolný soubor čítající n pozorování

$$\mathbf{Z} = \{(x_{11}, \dots, x_{1p}, y_1), \dots, (x_{n1}, \dots, x_{np}, y_n)\}. \quad (2.5)$$

Nechť T je odhad, který aplikováním na vzorek \mathbf{Z} dá vektor regresních koeficientů $\hat{\boldsymbol{\theta}}$, tedy

$$T(\mathbf{Z}) = \hat{\boldsymbol{\theta}}. \quad (2.6)$$

Uvažujme nyní všechny možné poškozené vzorky \mathbf{Z}' , které získáme záměnou libovolných m pozorování za arbitrární hodnoty. To nám umožňuje tvorbu libovolně vzdálených pozorování. Označme veličinou *vychýlení*($m; T, \mathbf{Z}$) maximální možnou výchylku, kterou je možné takto získat

$$\text{vychýlení}(m; T, \mathbf{Z}) = \sup_{\mathbf{Z}'} \|T(\mathbf{Z}') - T(\mathbf{Z})\|. \quad (2.7)$$

Pokud se *vychýlení*($m; T, \mathbf{Z}$) blíží k nekonečnu, je zřejmé, že m vlivných bodů má výrazný vliv na odhad T . Lze říci, že při takovéto kontaminaci odhad selhal. Bod selhání je tedy definován následovně

$$\varepsilon_n^*(T, \mathbf{Z}) = \min \left\{ \frac{m}{n}; \text{vychýlení}(m; T, \mathbf{Z}) = \infty \right\}. \quad (2.8)$$

Jedná se tedy o takový nejmenší podíl kontaminace, který způsobí, že odhad T „selže“. Poněkud přesněji lze říci, že bod selhání představuje takovou vzdálenost od předpokládaného rozdělení modelu, za níž se už statistika T stává nepoužitelnou a dává zavádějící výsledky. Například pro odhad metodou OLS to nastane už pro jediné pozorování (viz. obr. 2.1.), tedy

$$\varepsilon_n^*(T, \mathbf{Z}) = \frac{1}{n}, \quad (2.9)$$

což je pro $n \rightarrow \infty$ rovno 0. Říkáme tedy, že OLS má 0% (asymptotický) bod selhání. Čímž jsme jenom formálně potvrdili již několikrát zmiňovanou vysokou citlivost této metody na vlivné body.

2.6 Historické příklady robustních odhadů

S tím, jak se historicky vyvíjely přístupy k teorii robustní regrese, vyvíjely se i konkrétní odhady. Podívejme se tedy alespoň na nejznámější z nich.

Uvažujme lineární regresní model

$$Y_i = \sum_{j=1}^p x_{ij} \beta_j^0 + \varepsilon_i; \quad i = 1, \dots, n, \quad (2.10)$$

kde Y_i je vysvětlovaná veličina (pro i -té pozorování), x_{ij} je hodnota j -té vysvětlující proměnné (opět pro i -té pozorování), β_j^0 je j -tá souřadnice (skutečného) vektoru regresních koeficientů a konečně ε_i jsou disturbance – tedy stejně rozdělené nezávislé náhodné veličiny. Pro libovolné β ($\beta \in \mathfrak{R}^p$) označme i -té reziduum $r_i(\beta)$

$$r_i(\beta) = Y_i - \sum_{j=1}^p x_{ij} \beta_j \quad i = 1, \dots, n. \quad (2.11)$$

Metodu OLS můžeme tedy v tomto formalizmu vyjádřit například následovně

$$\hat{\beta}^{OLS} = \arg \min_{\beta \in \mathfrak{R}^p} \sum_{i=1}^n r_i^2(\beta) \quad (2.12)$$

Jak jsme si výše ukázali, odhady pořízené metodou OLS jsou velice citlivé na odlehlá pozorování. V době, kdy vznikala metoda OLS, byla již téměř 160 let (Galileo Galilei (1632), Roger Joseph Boscovich (1757), Pierre Simon Laplace (1793)) známa metoda jiná, robustnější, avšak početně výrazně náročnější. Totiž metoda minimalizující součet absolutních odchylek.

$$\hat{\beta}^{L_1} = \arg \min_{\beta \in \mathfrak{R}^p} \sum_{i=1}^n |r_i(\beta)| \quad (2.13)$$

Tato metoda bývá často označována L_1 , zatímco metodu OLS označujeme někdy též jako L_2 . Heuristika, která vedla ke konstrukci tohoto odhadu, je založena na myšlence zmenšení váhy odlehlých pozorování tím, že jejich rezidua vstupují do celkového součtu pouze v první mocnině. Tento druh odhadu je robustní k odlehlým pozorováním ve vysvětlované proměnné (outliers), bohužel nikoli však k pozorováním odlehlým v některé

¹³ Zápísem *argmin* je naznačeno, že β je argument minimalizující hodnotu výrazu stojící za tímto zápísem.

z vysvětlujících proměnných (leverage points). Jak je z definice patrné, ve smyslu bodu selhání nedosahuje tato metoda lepších vlastností než OLS. Má tedy opět pouze 0% bod selhání. „Navzdory předsudkům o její nízké eficienci lze dokázat, že pokud data generovaná přesně normálním modelem obsahují na každých 1000 pozorování dvě poškozená (kontaminovaná) pozorování, je tato metoda výkonnější než metoda OLS“ (Víšek (1997), str. 9).

Dalším vylepšením metody OLS v tomto směru byly tzv. M-odhady. Ty jsou založeny na myšlence nahradit rezidua umocněná na druhou ($r_i^2(\beta)$) nějakou jinou vhodnou funkcí reziduí

$$\hat{\beta}^M = \arg \min_{\beta \in \mathcal{R}^p} \sum_{i=1}^n \rho(r_i(\beta)), \quad (2.14)$$

kde $\rho(r_i(\beta))$ je nějaká symetrická (sudá) funkce s jediným minimem v bodě 0. Derivováním výrazu (2.14) vzhledem k regresním koeficientům $\hat{\beta}_i^M$ dostaneme

$$\sum_{i=1}^n \psi(r_i(\beta)) \mathbf{x}_i = \mathbf{0}^{14}, \quad (2.15)$$

kde ψ je derivace funkce ρ , \mathbf{x}_i je řádkový vektor vysvětlujících proměnných pro i -té pozorování a $\mathbf{0}$ je nulový vektor. Jedná se o soustavu p rovnic, jejíž řešení není vždy triviální nelézt a obvykle vyžaduje užití iteračních metod.

Největší nevýhodou M-odhadů je fakt, že řešení (2.14) není ekvivariantní vzhledem k měřítku¹⁵ (tj. vzhledem ke změnám osy y). Rezidua proto musí být standardizována (studentizována) nějakým k regresi invariantním a k měřítku ekvivariantním odhadem rozptylu disturbancí $\hat{\sigma}$, který ovšem musí být pořízen simultánně s odhadem regresních koeficientů, což celou situaci ještě více komplikuje.

¹⁴ Poznamenejme na tomto místě, že rovnice (2.14) a (2.15) nemusí být vždy ekvivalentní, nicméně použití tvaru (2.15) je obvykle vhodnější pro nalezení řešení problému (2.14) (Hampel & kol. (1986)).

¹⁵ Slovem *ekvivariantní* označujeme statistiku, která se při transformaci dat mění proporcionálně s touto transformací. Podle typu transformace dat rozdělujeme ekvivariance na několik typů.

Řekneme, že odhad regresních koeficientů $\hat{\beta}^{(n)}(X, Y)$ je *ekvivariantní vzhledem k měřítku*, jestliže $\forall c \in \mathcal{R}^+$, $\mathbf{Y} \in \mathcal{R}^n$ a designovou matici \mathbf{X} (typu $(n \times p)$) platí: $\hat{\beta}^{(n)}(c\mathbf{Y}, \mathbf{X}) = c \cdot \hat{\beta}^{(n)}(\mathbf{Y}, \mathbf{X})$. Řekneme, že odhad regresních koeficientů $\hat{\beta}^{(n)}(X, Y)$ je *ekvivariantní vzhledem k regresi*, jestliže $\forall \mathbf{b} \in \mathcal{R}^+$, $\mathbf{Y} \in \mathcal{R}^n$ a \mathbf{X} (typu $(n \times p)$) platí: $\hat{\beta}^{(n)}(\mathbf{Y} + \mathbf{X}\mathbf{b}, \mathbf{X}) = \hat{\beta}^{(n)}(\mathbf{Y}, \mathbf{X}) + \mathbf{b}$. Naproti tomu *invariantní* statistiky se při transformacích dat nemění.

Ekvivariance odhadů vzhledem k měřítku a vzhledem k regresi jsou vlastnosti, které pramení z přirozeného požadavku, že odhady by neměly záviset na jednotkách a poloze os souřadnicového systému, ve kterém měříme vysvětlující a vysvětlované veličiny, a do jisté míry by neměly záviset ani na tom, zda a případně jak moc (v rámci nějakých rozumných hranic) jsou jednotlivé vysvětlující veličiny mezi sebou korelovány (Víšek (2003)).

I když M-odhady dosahují, při normálním rozdělení reziduí, větší eficientnosti než odhad L_1 (Rousseeuw, Leroy (1987)), jejich bod selhání není vyšší než 0 %, opět především kvůli vlivu bodů odlehlých ve směru některé z vysvětlujících proměnných.

Ve snaze omezit vliv bodů odlehlých ve faktorovém prostoru vysvětlujících veličin a získat tak odhady s vyšším než 0% bodem selhání, byly M-odhady obohaceny o váhovou funkci $w(\mathbf{x}_i)$. To dalo vzniknout zobecněným (generalizovaným) M-odhadům, tedy GM-odhadům¹⁶

$$\sum_{i=1}^n w(\mathbf{x}_i) \psi(r_i(\beta)/\hat{\sigma}) \mathbf{x}_i = \mathbf{0}. \quad (2.16)$$

I při optimálním nastavení funkcí w a ψ maximalizujícím velikost bodu selhání se však ukázalo, že ten nemůže překročit určitou hodnotu, která je, bohužel, klesající funkcí p . Dimenzionalita modelu tedy stále představovala značný problém, především vezmeme-li v potaz, že právě vícerozměrná data jsou náchylnější k výskytu nebezpečných vlivných bodů.

Neúspěchy při hledání odhadů, které by byly schopny odolávat vysoké kontaminaci dat, vyvolaly debatu o tom, zda vůbec je robustní regrese možná. Kladnou odpověď přinesl Siegel navržením odhadu pomocí opakovaného mediánu (Siegel (1982)). Ten skutečně dosahoval asymptoticky 50% bodu selhání. Jeho nevýhodou však byla velká časová náročnost. V praxi se tedy příliš neprosadil, nicméně svoji úlohu sehrál především v tom, že ukázal, že robustní regrese je možná (Rousseeuw, Leroy (1987)). Nechme tedy opakovaný medián jeho historickému osudu a podívejme se podrobněji na odhady s vysokým bodem selhání, které jsou v praxi mnohem používanější.

¹⁶ Vzhledem ke snaze omezit vliv bodů odlehlých ve faktorovém prostoru vysvětlujících veličin jsou tyto odhady též souhrnně označovány jako *bounded-influence estimators*.

2.7 Odhady s vysokým bodem selhání

Všechny výše uvedené metody se snažily získat robustní odhad vylepšením metody OLS ve smyslu změny tvaru funkce, se kterou vstupují rezidua do odhadu. Nikdo se však nezabýval myšlenkou, že s rezidui by se dalo provést i něco jiného než je sčítat. Teprve v roce 1984 přišel P. Rousseeuw s odhadem *least median of squares* (LMS), který stojí na myšlence nahradit sumační znaménko mezi umocněnými rezidui mediánem, který je sám velmi robustní metodou. Odhad je definován následujícím způsobem

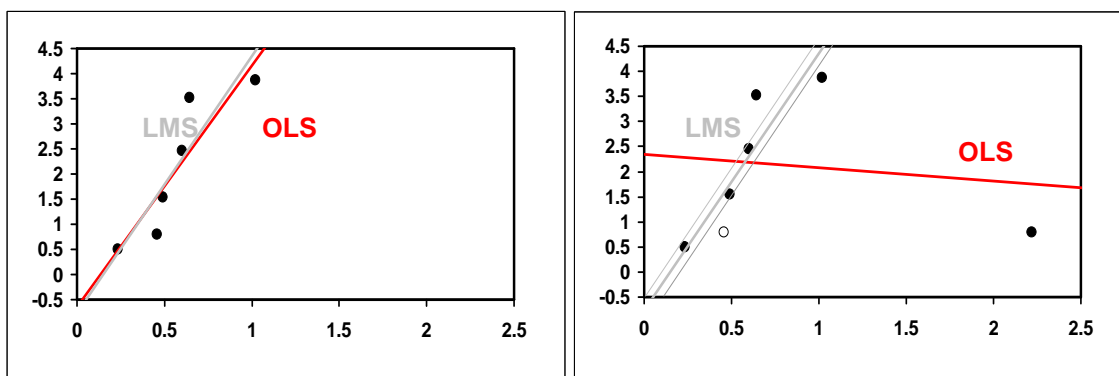
$$\hat{\beta}^{LMS} = \arg \min_{\beta \in \mathcal{R}^p} \text{med}_i \left(r_i^2(\beta) \right), \quad (2.17)$$

alternativně lze odhad LMS vyjádřit též

$$\hat{\beta}^{LMS,h} = \arg \min_{\beta \in \mathcal{R}^p} r_{(h:n)}^2(\beta)^{17}, \quad (2.18)$$

kde $r_{(h:n)}^2(\beta)$ je h -tá pořádková statistika mezi $r_1^2(\beta), r_2^2(\beta), \dots, r_n^2(\beta)$. Slovní spojení pořádková statistika naznačuje, že pro rezidua platí $r_{(1:n)}^2(\beta) \leq r_{(2:n)}^2(\beta) \leq \dots \leq r_{(n:n)}^2(\beta)$. Odhad tedy minimalizuje v pořadí h -tou nejmenší hodnotu mezi $r_1^2(\beta), r_2^2(\beta), \dots, r_n^2(\beta)$.

2.2. Robustnost (resp. citlivost) odhadů LMS a OLS v přítomnosti „leverage points“



Obrázek 2.2. porovnává robustnost (resp. citlivost) odhadů pořízených metodou LMS (resp. OLS) vzhledem k bodům odlehlým ve směru vysvětlovací veličiny („leverage points“). Pravý obrázek zároveň naznačuje geometrickou reprezentaci odhadu LMS, která spočívá v hledání nejužšího pásu který obsahuje polovinu (resp. h) pozorování.

¹⁷ Vzorec (2.17) je speciálním případem obecnějšího zápisu (2.18), který bývá někdy též nazýván *least-quantile squared* (LQS) odhadem (Erickson, Har-Peled, Mount (2004)). Je patrné, že oba zápisy jsou ekvivalentní, pokud $h = [n/2]+1$ pro n liché nebo $h = (n+1)/2$ pro n sudé.

Takto vzniklý odhad je jasně ekvivariantní vzhledem k lineárním transformacím dat, neboť do odhadu vstupují pouze rezidua. Odolává jak odlehlým pozorováním ve vysvětlované proměnné, tak i pozorováním odlehlým ve faktorovém prostoru vysvětlujících proměnných (viz. obr. 2.2.). Bod selhání asymptoticky dosahuje 50 %, tedy nejvyšší možné hodnoty. Skutečná hodnota bodu selhání je maximální pro $h = [n/2] + [(p+1)/2]$ ¹⁸ a je rovna $n^{-1}([(n-p)/2] + 1)$ (Rousseeuw, Leroy (1987)).¹⁹

I přes prokazatelnou existenci a jednoznačnost řešení problému (2.18) (Rousseeuw, Leroy (1987)) není hledání konkrétního řešení vůbec jednoduchou záležitostí. Existuje hned několik algoritmů pro hledání přibližného řešení. Jako nejefektivnější metoda se v současné době jeví algoritmus navržený P. Bočkem a P. Lachoutem (Boček, Lachout (1995)), který se opírá o duální podobu simplexové metody lineárního programování. Podobný algoritmus založený také na metodě lineárního programování je popsán i v Erickson, Har-Peleld, Mount (2004). Studie provedené na příkladech skutečných dat (Víšek (1996) a (2000a)) ukazují, že alespoň v některých případech dosahuje lepších výsledků než původní algoritmus navržený autory P. J. Rousseeuwem a A. M. Leroyem (Rousseeuw, Leroy (1987)).

I když je výpočetní stránka LMS problematická, jeho největší nevýhodu představují asymptotické vlastnosti. Vzhledem k abnormálně pomalé míře konvergence dosahuje LMS velmi slabých výsledků ve smyslu asymptotické efience. LMS konverguje stejně jako $n^{-1/3}$, naproti tomu OLS i LTS (bude definován níže) konvergují jako $n^{-1/2}$. Nicméně pokud chceme použít LMS pouze jako diagnostickou metodu, tj. k odhalení vlivných pozorování, nemusí nás tento fakt příliš znepokojovat. Na této myšlence je například založen odhad *reweighted least squares* (RLS) (Rousseeuw, Leroy (1987)), který na základě reziduální analýzy odhadu LMS určí a odstraní vlivná pozorování a na takto „očištěná“ data pak uplatní odhad metodou obyčejných nejmenších čtverců.

Snaha vylepšit neuspokojivé asymptotické vlastnosti odhadu LMS inspirovala P. Rousseeuwa k odhadu *least trimmed squares* (LTS), který je definován následovně:

$$\hat{\beta}^{LTS,h} = \arg \min_{\beta \in \mathfrak{R}^p} \sum_{i=1}^h r_{(i:n)}^2(\beta), \quad (2.19)$$

¹⁸ Hranaté závorky označují celou část.

¹⁹ Hodnota bodu selhání LMS definovaného vzorcem (2.17) je rovna $([n/2] - p + 2)/2$ (Rousseeuw, Leroy (1987)).

kde $r_{(in)}^2(\beta)$ je opět i -tá pořádková statistika mezi $r_1^2(\beta), r_2^2(\beta), \dots, r_n^2(\beta)$ přičemž pro parametr h platí $n/2 < h \leq n$. Odhad je tedy založen na myšlence minimalizace součtu h nejmenších reziduí. Zbývajících $n-h$ bodů je „odříznuto“. Jinými slovy odhad hledá takovou podmnožinu (resp. vzorek dat) velikosti h mezi všemi n rezidui umocněnými na druhou, na niž aplikace obyčejných nejmenších čtverců (OLS) dá pro konkrétní β nejmenší součet čtverců reziduí. Parametr h je arbitrárně zvolen tak, aby odpovídal předpokládané kontaminaci datového souboru. Pokud je $h = n$, pak $LTS = OLS$ a bod selhání je tedy 0%. Naopak, pokud $h = n/2$, odhad LTS asymptoticky dosahuje 50% bodu selhání. Velikostí parametru h lze tedy efektivně řídit velikost bodu selhání. Praktické určení parametru h bude demonstrováno ve výpočetní části.

Jedná se tedy opět o regresně ekvivariantní odhad s bodem selhání asymptoticky dosahujícím 50 % a s mírou konvergence rovnou $n^{-1/2}$. Výpočetní algoritmus a praktické příklady aplikací na reálných souborech dat lze opět nalézt např. v člancích Víšek (1996) a (2000a).

Připomeňme na tomto místě, že odhad LTS je často nesprávně zaměňován s podobně se jmenujícím, nicméně co do logiky konstrukce podstatně odlišným, odhadem *trimmed least squares* (TLS). Ten je sice také založen na minimalizaci součtu čtverců h reziduí (a v tomto smyslu je tedy taktéž robustní procedurou), nicméně těchto h pozorování je vybráno na základě nějakého pravidla (obvykle diagnostického, založeného např. na geometrii dat) ještě před vlastní aplikací metody odhadu metodou nejmenších čtverců na vybraných h bodů. Naproti tomu metoda LTS si oněch h „nekontaminovaných“ bodů vybere implicitně sama, což je způsob sofistikovanější, neboť nedochází ke zbytečné ztrátě informace, především u souborů obsahujících vlivné body odlehle ve směru některé vysvětlující proměnné (leverage points). Tato dobrá vlastnost je ovšem „vykoupena“ podstatně vyšší výpočetní náročností.

Jestliže u LMS je slabou stránkou asymptotická konvergence, tak u odhadu LTS lze spatřovat slabinu v omezení se na někdy až příliš limitovanou část vzorku s tím, že vliv zbývajících dat je zcela potlačen. LTS vezme v potaz pouze h bodů a zbývajících $n - h$ bodů zcela zamítne, čímž především v případě, že některý ze zamítnutých bodů nepředstavuje kontaminaci ve smyslu hrubých chyb, může dojít k nežádoucí ztrátě informace. Typicky tato situace nastává v případech, kdy je předpokládaná úroveň kontaminace vyšší než její skutečná hodnota.

Vylepšení v tomto směru představují odhady *least weighted squares* (LWS) (Víšek (2000b), (2003)) či *smoothed least trimmed squares* SLTS (Čížek (2001)), které představují vylepšení metody LTS v tom, že ještě navíc obsahují váhovou funkci, která řídí míru vlivu odlehlých pozorování. Jedná se tedy o jakýsi kompromis mezi metodou LTS a M-odhady. Odhad SLTS je definován následovně:

$$\hat{\beta}^{SLTS, \mathbf{w}} = \arg \min_{\beta \in \mathfrak{R}^p} \sum_{i=1}^n w_i r_{(i:n)}^2(\beta), \quad (2.20)$$

kde $r_{(i:n)}^2(\beta)$ představuje opět i -tou pořádkovou statistiku mezi $r_1^2(\beta), r_2^2(\beta), \dots, r_n^2(\beta)$ a \mathbf{w} označuje vektor vah, pro který platí: $w_1 \geq w_2 \geq \dots \geq w_n \geq 0$. Odhad je podobný odhadu WLS – *weighted least squares*, který minimalizuje $\sum_{i=1}^n w_i r_i^2(\beta)$ ovšem s tím podstatným rozdílem, že v případě SLTS jsou váhy přiřazeny pořádkovým statistikám umocněných reziduí, nikoli reziduím přímo. Jedná se o analogickou situaci jako v případě odhadů LTS a TLS. Je zřejmé, že na vlastnosti odhadu SLTS má rozhodující vliv charakter vektoru \mathbf{w} . Podívejme se tedy, jakým způsobem vektor \mathbf{w} ovlivňuje odhad:

- (i) $w_1 = w_2 = \dots = w_n = 1$: SLTS je ekvivalentní OLS.
- (ii) $w_1 = w_2 = \dots = w_h = n/h$ pro $\frac{n}{2} < h \leq n$ a $w_{h+1} = \dots = w_n = 0$: SLTS je ekvivalentní LTS_h .
- (iii) $w_i = f(\frac{i}{n}; \omega_1, \dots, \omega_m)$, $\forall i = 1, \dots, n$ kde $f(x; \omega_1, \dots, \omega_m)$ je reálná funkce definovaná na $\langle 0, 1 \rangle$ parametrizovaná vektorem $\omega_1, \dots, \omega_m \in \mathfrak{R}^m$. Příkladem takové funkce je třeba $f(x; \omega) = \frac{1}{1 + e^{\omega(x-1/2)}}$, $\forall x \in \langle 0, 1 \rangle$ a jeden parametr $\omega \in \langle 0, \infty \rangle$.

Výhodou tohoto odhadu je schopnost zachovat si vysoký bod selhání (při nastavení vah jako ve druhém případě či pro $\omega \gg 1$ v případě třetím). Pokud váhy zvolíme tak, že všechna pozorování budou zahrnuta do regrese tedy: $w_1 \geq w_2 \geq \dots \geq w_n > 0$, nebude docházet ke ztrátě informace (ačkoli efekt vzdálených pozorování budeme moci stále dostatečně omezit) a navíc můžeme odhad dobře použít i pro modely s dummy či kategorickými proměnnými. Jak je patrné, správná volba vah je rozhodujícím faktorem pro úspěšné pořízení odhadu. Její určení však sahá mimo rámec této práce (více viz. Čížek (2001)).

2.8 Rozmanitost odhadů

Intenzivní výzkumy v teorii odhadů za posledních 30 – 40 let přivedly na svět celou řadu metod, a to nejen robustních, z nichž každá má své silné i slabé stránky. Pro praktické použití to znamená dvě věci. Na straně jedné to přináší možnost vybrat si metodu, která je přímo „kovaná“ na konkrétní data, ale na straně druhé tím vyvstává reálný problém, jak vybrat v tak široké nabídce metodu optimální.

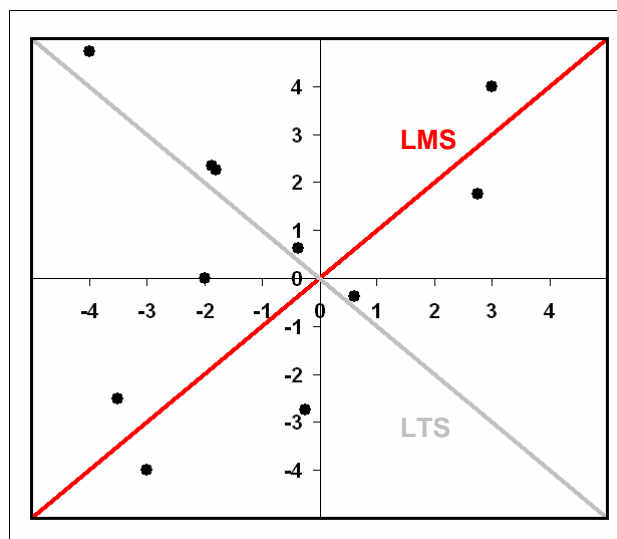
Jako rozumné se jeví aplikovat na data nejdříve nějakou robustní metodu, abychom dostali alespoň základní ideu o „skutečném“ modelu, který za data stojí, a na jejím základě následně zvolit konkrétní metodu.

V praxi však může a to nikoli výjimečně nastat situace, v níž dva odhady (oba dokonce s vysokým bodem selhání!) budou pro totožná data dávat diametrálně rozdílné výsledky. Bývá to zpravidla způsobeno specifickou strukturou zkoumaného vzorku.

Typicky tato situace nastává v případě souborů pozorování složených ze dvou nebo více různých populací, obsahujících oblast, kde se tyto populace mísí (tedy jakýsi pomyslný průsečík). Právě v této oblasti vyvstává problém jak rozpoznat, které body přísluší ke které populaci. Tato situace je typická u dat z tranzitivních ekonomik – tedy i v našem případě. Z pochopitelných důvodů je patrné, že celá situace se samozřejmě výrazně komplikuje s rostoucí velikostí a částečně také s rostoucí dimenzionalitou zkoumaného vzorku.

Obecně mohou být data skutečně takové povahy, že je lze uspokojivě vysvětlit dvěma, i více modely, avšak díky nejednoznačnosti v oblasti průsečíku jednotlivých populací nastane situace, kdy každá procedura vzhledem k tomu, jak je konstruována, považuje různé části vzorku za kontaminaci. Překvapivé je, že ani zvýšení počtu pozorování nemusí, žel bohu, nutně vést ke zlepšení této situace (matematická formalizace – viz. Víšek (2000a)).

2.3. Rozmanitost odhadů



Obrázek 2.3. je názorným příkladem struktury dat, na které i vysoce robustní metody jako jsou odhady LMS a LTS mohou dávat diametrálně rozdílné výsledky. Byť se jedná o akademický příklad, který je konstruován tak, aby odhady vyšly navzájem kolmé, je zřejmé, že tento fenomén představuje závažný problém. Výše uvedená struktura dat navíc dává velmi rozdílné hodnoty odhadů i pro jiné metody. Viz Tabulka 2.1

Zdroj: Víšek (2000a)

Tabulka 2.1 Hodnoty odhadů pro různé metody

Koeficienty	Metoda				
	LS	LTS	LMS	L_1	TLS
β_0	0,829	-0,013	0,001	0,736	0,593
β_1	0,293	-1,025	1,000	0,368	0,420

Zdroj: Víšek (2000a)

Fakt, že nastane takováto situace, neznamená, že by odhady byly špatně sestrojeny, problém je jinde. Jeho jádro tkví v tom, na jaké heuristice je celý koncept bodu selhání postaven. Odhady vždy správně rozdělí data na „dobrá“ pozorování a kontaminaci. Mýlka je ovšem v implicitní avšak naprosto mylné domněnce, že správná data a kontaminace jsou vždy a jednoznačně určeny. To však obecně v praxi nemusí být pravda. Tím lze zároveň i vyvrátit široce rozšířený názor, že robustní metody sice dosahují nižší eficiency, nicméně vždy správně poukazují na „skutečný“ model, který generoval data. Stoprocentní spolehnutí se na robustní metody může být v tomto směru zničující.

Pro praktické použití to znamená, že pokud analyzujeme data, je vhodné použít maximální počet metod, které jsou dostupné, neboť rozmanitost odhadů není pouze

akademickým případem a v praxi může velice často nastat a to pro jakkoli velký soubor dat. Pokud obdržíme shodné výsledky vše je v pořádku (nasvědčuje to faktu, že v souboru dat není kontaminace přítomna a nebo má alespoň, co se jejího jednoznačného určení týče, „rozumnou“ strukturu tak jak bylo popsáno výše) a pravděpodobně se v tomto případě rozhodneme pro metodu, která je považována za nejefektivnější.

Obdržíme-li výsledků více, tj. jsou-li výsledky rozdílné ve statistickém smyslu slova, je nutné jeden z nich vybrat. Tím implicitně říkáme, který model a potažmo i která data považujeme za „správná“. Problém v tu chvíli ovšem neřešíme čistě matematickými nástroji, nýbrž i na základě nějaké další, obvykle z praxe daného oboru vycházející heuristiky.

Tento závěr ani v nejmenším neznamena, že je s robustní statistikou něco v nepořádku, pouze varuje před jejím neuvážlivým použitím a nabádá nás k velice pečlivému analyzování získaných výsledků.

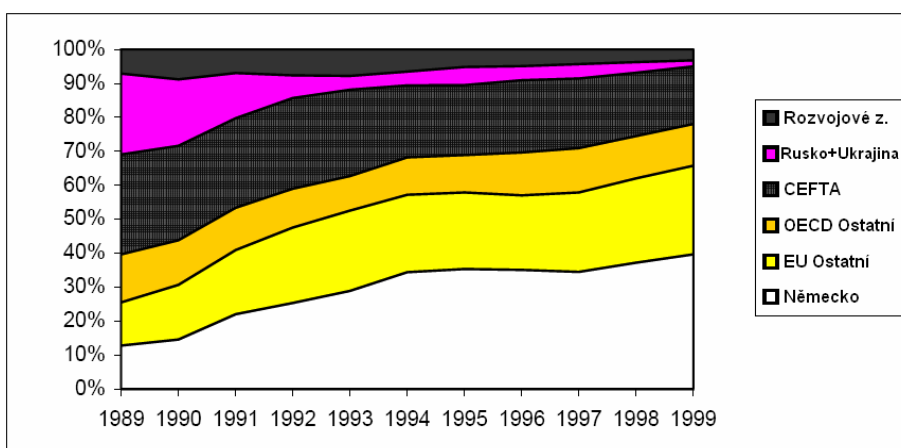
3. Empirická analýza

3.1 Specifikace modelu

Jak již bylo řečeno v úvodu, rozvoj exportu během transformace měl zásadní vliv na růst české ekonomiky a zkoumání faktorů, které jej ovlivňují, je zajímavé nejen z pohledu ekonomického výzkumu, ale především může představovat vodítko pro sestavování odpovídající hospodářské politiky.

Rozvoj českého zahraničního obchodu byl po celá devadesátá léta velice silně poznamenán rozpadem RVHP, v jejímž institucionálním rámci do té doby relativně intenzivně probíhal. Po zhroucení RVHP a s postupující liberalizací zahraničního obchodu se pro české podniky otevřelo okno do té doby nedostupných příležitostí v podobě vyspělých především západoevropských trhů.

Graf 3.1. Vývoj podílu exportu do jednotlivých zemí



Graf 3.1. názorně ukazuje změnu podílů exportů do jednotlivých zemí vzhledem k celkovému exportu z ČR během 90. let. Jak je patrné nejvíce ztrácelo Rusko a Ukrajina a naopak nejvíce se rozvíjel export do Německa a ostatních států EU.

Zdroj: Benáček, Víšek (2002)

S masivním přeorientováním vývozu směrem k vyspělým zemím OECD a především EU docházelo současně k výrazné změně ve struktuře obchodovaného zboží. Zatímco v centrálně plánované ekonomice byla rozhodnutí týkající se zahraničního obchodu prováděna administrativně především na makro úrovni, a lze tedy oprávněně pochybovat, že jeho struktura odpovídala ekonomickým kritériím efektivnosti, je na vyspělých, vysoce konkurenčních trzích struktura obchodu určována tržně, především na bázi interakcí mezi jednotlivými ekonomickými agenty.

Z důvodů rozsáhlých strukturálních změn (nejen) v rámci zahraničního obchodu, které mají své silné mikroekonomické kořeny, není tedy principiálně možné vysvětlovat zahraniční obchod *pouze* na bázi makroekonomických agregátů. Ty totiž nutně nemusí vypovídat o situaci v ekonomice správně. Na makro úrovni se mohou jednotlivé strukturální změny stírat v tom smyslu, že úpadek jednoho odvětví je kompenzován rozkvetem jiného.²⁰

Abychom byli schopni zachytit maximum těchto překotných změn, je třeba v modelu zohlednit i mikroekonomické a institucionální faktory, mající vliv na zahraniční obchod. Níže prezentovaná konstrukce je drobnou modifikací modelů z článků autorů V. Benáčka, J. Á. Víška a L. Prokopa (Benáček, Prokop, Víšek, (2002), Benáček, Víšek (2002) a Víšek (2002)). Základní podoba modelu je následující:

$$X_{it}^{EU} = f(GDP_t^{EU}, REER_t^{EU}, PX_{it}^{EU} / PM_{it}^{EU}, K_{it} / L_{it}, FDI_{it}, Tar_{it}, De_{it} / VA_{it}, VA_{it}, \varepsilon_{it}), \quad (3.1)$$

kde:

$i =$	Index označující komoditní skupiny (01,...,89) tříděné podle dvouciferné klasifikace SITC – celkem 61 odvětví ²¹
$t =$	Index označující jednotlivé roky (1993 – 1999)
$X_{it}^{EU} =$	Export (i -tého odvětví v roce t) do EU (v nominálních USD)
$GDP_t^{EU} =$	Hrubý domácí produkt EU15 v běžných cenách vyjádřený paritou kupní síly – PPP (v mld. USD)
$REER_t^{EU} =$	Efektivní reálný směnný kurz se zeměmi EU15 deflovaný CPI
$PX_{it}^{EU} / PM_{it}^{EU} =$	Podíl ceny exportovaného a importovaného zboží (v nominálních USD)
$K_{it} / L_{it} =$	Podíl kapitálu (K) vyjádřeného v cenách roku 1994 a práce (L)

²⁰ Viz. např. Egert (2002), který ve své studii dokázal vysvětlit pouze 5 - 20 % reálné apreciacie v zemích jako je Polsko, Maďarsko, Česko nebo Slovensko. V některých případech jeho odhady dokonce nebyly signifikantní vůbec.

²¹ Pozorného čtenáře pravděpodobně zarazí, že ačkoli $i = 01, \dots, 89$, je skupin pouze 61. Tento fakt je dán způsobem kódování klasifikace SITC 02.

$FDI_{it} =$	Přímé zahraniční investice (v CZK)
$Tar_{it} =$	Průměrná cla uvalovaná zeměmi EU15 na český export (v %)
$De_{it}/VA_{it} =$	Implicitní dotace vyjádřené jako podíl dluhu daného odvětví a přidané hodnoty (v %)
$VA_{it} =$	Výstup odvětví měřený přidanou hodnotou upravenou o růst HDP a růst cen výrobců – PPI (v ‰ HDP)
$\varepsilon_{it} =$	Náhodné fluktuace

Rozeberme nyní stručně, co popisují jednotlivé proměnné a z jakých důvodů byly do modelů zahrnuty. Lze je rozdělit do tří hlavních skupin:

- (i) Proměnné popisující zahraniční obchod na úrovni makroekonomických agregátů.
- (ii) Veličiny popisující zahraniční obchod na základě mikroekonomických faktorů.
- (iii) Proměnné odrážející institucionální rámec zahraničního obchodu.

Ad (i), tato skupina zahrnuje standardní makroekonomické veličiny, s jejichž pomocí popisují ekonomické teorie zahraniční obchod. Keynesiánci vysvětlují úroveň exportu především na základě agregátní poptávky. V případě exportu se jedná o agregátní poptávku v cílových zemích, jejíž úroveň v EU15 odrážejí proměnné GDP_t^{EU} a $REER_t^{EU}$. V souladu s teorií lze očekávat pozitivní vliv proměnné GDP_t^{EU} na úroveň exportu (s jejím růstem roste absorpční schopnost cílové ekonomiky) a na druhou stranu negativní vliv $REER_t^{EU}$ (s apreciací koruny se české exporty stávají relativně dražšími).

Na rozdíl od keynesiánství považuje neoklasická teorie zahraničního obchodu úroveň exportu za funkci relativních jednotkových cen, jejichž index $PX_{it}^{EU} / PM_{it}^{EU}$ dává do přímého vztahu ceny domácích a zahraničních. Pro praxi tím však vyvstávají dva problémy (Benáček, Víšek, (2002)).

Prvním z nich je otázka, jaké ceny použít do čitatele a jmenovatele těchto zlomků? Ideálem je postavit proti sobě vnitřní ceny zboží doma s vnitřními cenami téhož zboží v

zahraničí, leč takové statistiky, bohužel, nejsou k dispozici. Jako vhodné se tedy jeví použít domácí ceny exportovaného zboží a domácí ceny zboží importovaného. Tím je ovšem problém vyřešen pouze zčásti. Indexy relativních cen proti sobě totiž staví ceny jednotkové. Ty, vzhledem k tomu, že zahraniční obchod probíhá mezi velkým množstvím účastníků, taktéž nejsou souhrnně k dispozici. Druhou nejlepší možností se tedy jeví použití jako proxy veličiny ceny za 1 tunu vyváženého, resp. dováženého zboží.

Problémem druhým, který je spjatý s přístupem pomocí indexů relativních jednotkových cen, je jejich potenciální dvojsmyslná interpretace. Neoklasická teorie totiž uvažuje tyto indexy pro dokonale homogenní produkty, kde cena odráží čistě *cenovou* konkurenceschopnost, potažmo tedy efektivnost využití primárních zdrojů. Vzhledem k rozmanitosti obchodovaného zboží není tento předpoklad v praxi naplněn. Cenové indexy v sobě tedy nesou informace nejen o cenové konkurenceschopnosti, ale také o rozdílech v kvalitě a diferencovanosti jednotlivých výrobků. Lze tedy říci, že proměnná $PX_{it}^{EU} / PM_{it}^{EU}$ se pohybuje od prostoru indexů mezi dokonalou cenovou konkurenceschopností dvou identických výrobků, až po prostor indexů klasických směnných relací odvozených na základě barteru mezi dvěma naprosto rozdílnými, vzájemně nesoupeřícími výrobky. Interpretace odhadnutých koeficientů u této veličiny by tedy měla probíhat s maximální opatrností.

Ad (ii), již několikrát jsme zmínili, že ekonomická transformace je svou povahou mikroekonomický proces a pokud nás zajímá zahraniční obchod, který v této době procházel překotnými strukturálními změnami, není k jeho popisu vhodné a ani principiálně možné použít *pouze* veličiny měřené na makroúrovni. Druhou skupinou vysvětlujících veličin našeho modelu jsou tedy proměnné odrážející mikroekonomické faktory charakterizující nabídkovou stranu účastníků zahraničního obchodu.

První z nich je podíl kapitálu a práce (K_{it}/L_{it}), sloužící jako měřítko komparativních výhod na základě relativního faktorového vybavení jednotlivých odvětví. *Tradičně se předpokládá, že Česká republika je relativně lépe vybavena prací a tudíž i exportní produkce charakterizující vývozy do EU by měla být vychýlena směrem k produkci náročnější na vybavenost prací* (Benáček, Prokop, Víšek (2002), str. 14, překlad AT).

Druhou veličinou, výrazně ovlivňující úroveň zahraničního obchodu a nabídkovou stranu obecně, jsou přímé zahraniční investice (FDI_{it}). Jejich přítomnost v podniku (resp. v odvětví) lze do značné míry interpretovat jako vybavenost lidským kapitálem. Kromě samotného finančního kapitálu totiž do podniku zpravidla přichází (v rámci tzv. spillover

effects) i velice cenné know-how a zkušenosti zahraničního managementu, pokročilé informační technologie a metody řízení či schopnost proniknout na zahraniční trhy. Navíc se lze oprávněně domnívat, že zahraniční kapitál je alokován efektivně, tudíž že odvětví samo o sobě disponuje dostatečným potenciálem v podobě komparativních výhod, které zahraniční kapitál přilákaly.

Dalším pozitivním vlivem přímých zahraničních investic je sílící konkurenční tlak na rychlou a efektivní restrukturalizaci, a to jak přímý v rámci konkrétního podniku, tak i nepřímý, působící skrze konkurenční tlaky v rámci celého odvětví. S přílivem přímých zahraničních investic se zpravidla vylepší i institucionální systém daného odvětví, který v dlouhém období zlepšuje institucionální systém celé ekonomiky (Benáček (2000b)).

Ad (iii), poslední skupinou jsou proměnné popisující vliv institucionálních faktorů s přímým či nepřímým vlivem na zahraniční obchod. Nejstarším a nejběžněji používaným nástrojem tohoto druhu jsou cla. Protekcionismus v podobě uvalování cel na dovážené zboží uměle navyšuje ceny importu, čímž implicitně zvyšuje relativní konkurenceschopnost domácí produkce vzhledem k dováženému zboží, jehož úroveň v tomto důsledku klesá. Pro český export byla tedy v tomto smyslu rozhodující úroveň cel uvalovaných na české zboží dovážené do EU. Její procentuální výši odráží proměnná Tar_{it} , představující průměrnou výši cel uvalovanou na české exporty zeměmi Patnáctky.

Kromě cel, která podporují domácí výrobce přímo, lze k ovlivňování zahraničního obchodu využít i dalších, nepřímých nástrojů. Mezi ně patří především nejrůznější formy netarifních bariér, dotací a podpor. S výjimkou zemědělství byly v České republice přímé dotace z veřejných rozpočtů zpravidla velice nízké.

Naproti tomu implicitní quasi-dotace poskytované polostátním bankovním systémem byly obrovské. Skrývaly se v měkkých úvěrových linkách, u kterých se už *a priori* předpokládalo, že v době své splatnosti nebudou plně uhrazeny. V květnu 2000 bylo 34 % bankovních vkladů (odhadem 10 mld. USD, tedy asi 15 % HDP) klasifikováno (Benáček, Prokop, Víšek (2002)). Nesolventnost bankovního systému hrozila destabilizovat celou ekonomiku a tudíž byly špatné úvěry buď sekuritizovány, koupeny Konsolidační bankou nebo prominuty státními institucemi. Ministerstvo financí, Česká národní banka, či Fond národního majetku tak poskytl přímou podporu komerčním bankám a velkým institucím a nepřímo tedy i jejich dlužníkům.

Dalším zdrojem implicitních dotací byla zadluženost v rámci dodavatelsko-odběratelských vztahů. Situace v tomto případě vypadala zpravidla tak, že úspěšný podnik

skončil jako věřitel, zatímco neúspěšné podniky byly v pozici čistých dlužníků. Tato zadluženost představovala téměř 8 mld. USD (Benáček, Prokop, Víšek (2002)).

Posledním zdrojem implicitních dotací bylo neplacení daní, sociálního a zdravotního pojištění a privatizačních splátek. Za deset let transformace představují dohromady tyto tři zdroje quasi-podpor minimálně 30 mld. USD, tedy zhruba polovinu HDP (Benáček, Prokop, Víšek (2002)). „Příjemci“ těchto podpor se ze zřejmých důvodů rekrutovali převážně z řad podniků, které se buď z důvodů špatného managementu nedokázaly, nebo se v důsledku nepříznivých komparativních výhod zkrátka nemohly patřičně restrukturalizovat a úspěšně vstoupit na vyspělé trhy.

Aproximujme tedy tyto implicitní dotace proměnnou De_{it}/VA_{it} , představující poměr akumulovaného dluhu jednotlivých odvětví k přidané hodnotě v těchto odvětvích vytvořené, čímž je zohledněna různá velikost jednotlivých odvětví

Ekonomický smysl, proč byly zahrnuty jednotlivé proměnné do modelu, jsme tedy vysvětlili. Nyní ještě krátce rozeberme podobu funkce f . Model budeme uvažovat v mocninném tvaru s odhadovanými koeficienty jakožto mocniteli

$$X_{it}^{EU} = B_0 \cdot (GDP_t^{EU})^{\beta_1} \cdot (REER_t^{EU})^{\beta_2} \cdot (PX_{it}^{EU} / PM_{it}^{EU})^{\beta_3} \cdot (K_{it} / L_{it})^{\beta_4} \cdot (FDI_{it})^{\beta_5} \cdot (Tar_{it})^{\beta_6} \cdot (De_{it} / VA_{it})^{\beta_7} \cdot \varepsilon_{it} \quad (3.2)$$

Takovýto model lze logaritmickou transformací převést na lineární tvar

$$\ln(X_{it}^{EU}) = \beta_0 + \beta_1 \ln(GDP_t^{EU}) + \beta_2 \ln(REER_t^{EU}) + \beta_3 \ln(PX_{it}^{EU} / PM_{it}^{EU}) + \beta_4 \ln(K_{it} / L_{it}) + \beta_5 \ln(FDI_{it}) + \beta_6 \ln(Tar_{it}) + \beta_7 \ln(De_{it} / VA_{it}) + v_{it}, \quad (3.3)$$

který můžeme jednoduše odhadovat pomocí běžných regresních technik. Logaritmická funkcionální forma má navíc tu výhodu, že odhadnuté koeficienty je možné (při nezávislosti vysvětlovaných veličin) interpretovat přímo jako příslušné elasticity, tj. 1% nárůst proměnné X způsobí β_x % nárůst vysvětlované veličiny.

Jak se navíc ukazuje, je tato forma vhodná i z hlediska variability dat a kvality pořízených odhadů (Benáček, Prokop, Víšek (2002), Benáček, Víšek (2002) a Víšek (2002)).²²

²² Pochopitelně bychom mohli uvažovat i nejrůznější Box-Cox transformace, pravděpodobně s ještě lepšími výsledky, nicméně pro aplikaci robustních metod je tato podoba plně dostačující.

Jelikož kromě globálního modelu přes celé období let 1993 – 1999 budeme též odhadovat modely pro individuální roky, musíme zápis (3.3) lehce modifikovat. Vzhledem k tomu, že veličiny GDP_t^{EU} a $REER_t^{EU}$ mají roční periodicitu, resp. v případě $REER_t^{EU}$ jsou ročním průměrem, nemá jejich zahrnutí do modelů pro jednotlivé roky ani ekonomický smysl, ani vzhledem ke korelaci s interceptem není technicky možné. Proměnné GDP_t^{EU} a $REER_t^{EU}$ je tedy nutné z modelu pro individuální roky vypustit a jeho podoba je následující:

$$\ln(X_{it}^{EU}) = \beta_0 + \beta_1 \ln(PX_{it}^{EU} / PM_{it}^{EU}) + \beta_2 \ln(K_{it} / L_{it}) + \beta_3 \ln(FDI_{it}) + \beta_4 \ln(Tar_{it}) + \beta_5 \ln(De_{it} / VA_{it}) + \eta_{it}. \quad (3.4)$$

Ještě dříve než začneme s vlastní výpočetní částí, učiňme dvě malé, nicméně velice podstatné poznámky.

První se týká použitých dat. Ta byla nashromážděna Vladimírem Benáčkem a analyzována Janem Ámosem Víškem (Benáček, Víšek (2002) a Víšek (2002)). Jejich zaznamenání představovalo velice zdlouhavou a náročnou práci, bez které by následující výpočty nebyly možné. Autor této práce velmi děkuje za jejich ochotné poskytnutí a veškerou pomoc v souvislosti s nimi.

Druhá poznámka se týká použitého softwaru. Robustní metody nejsou standardní součástí většiny běžných statistických knihoven a jejich výpočty je nutné provádět ve speciálních, zpravidla vlastními silami vytvořených programech. Tyto byly zapůjčeny docentem Janem Ámosem Víškem. Za jejich zapůjčení proto autor práce také velice děkuje.

3.2 Globální model

Začneme odhadovat model zahrnující *všechny* dostupné vysvětlující veličiny, tedy:

$$\ln(X_{it}^{EU}) = \beta_0 + \beta_1 \ln(GDP_t^{EU}) + \beta_2 \ln(REER_t^{EU}) + \beta_3 \ln(PX_{it}^{EU} / PM_{it}^{EU}) + \beta_4 \ln(K_{it} / L_{it}) + \beta_5 \ln(FDI_{it}) + \beta_6 \ln(Tar_{it}) + \beta_7 \ln(De_{it} / VA_{it}) + v_{it}. \quad (3.3)$$

Postupným vylučováním proměnných s nejvyšší p-value (viz. Příloha I) dospějeme k modelu:

Tabulka 3.1.a.

Proměnná	Odhad koeficientu	Směrodatná odchylka	t-statistika	P-value
<i>Intercept</i>	29.4824	5.3706	5.490	[.000]
<i>log(VA)</i>	0.6846	0.0502	13.645	[.000]
<i>log(KL)</i>	-0.7935	0.0820	-9.677	[.000]
<i>log(De/VA)</i>	0.2775	0.0647	4.286	[.000]
<i>log(FDI)</i>	0.1093	0.0410	2.667	[.008]
<i>log(Tar)</i>	-0.6257	0.0662	-9.452	[.000]
<i>log(REER)</i>	-7.5083	2.6112	-2.875	[.004]

Tabulka 3.1.b.

Další charakteristiky modelu			
Mean of dep. var.	11.235	LM het. test	5.421 [.020]
Std. dev. of dep. var.	1.648	Durbin-Watson	0.178 [.000,.000]
SSR	512.817	White het. test	99.649 [.000]
Variance of residuals	1.221	Jarque-Bera test	252.860 [.000]
SE of regression	1.105	Shapiro-Wilk test	0.923 [.000]
R-squared	0.557	Ramsey's RESET2	4.087 [.044]
Adjusted R-squared	0.550	F (zero slopes)	87.861 [.000]
Log likelihood	-644.986	Schwarz B.I.C.	666.185

Proveďme základní diagnostiku (většinu výsledků lze nalézt v tabulce 3.1.b) a zkoumejme následující vlastnosti modelu:

- (i) autokorelaci,
- (ii) nezávislost vysvětlujících veličin,
- (iii) homoskedasticitu,
- (iv) normalitu reziduí.

Ad (i), přestože odhad byl proveden na panelových datech, a standardní Durbin-Watsonova statistika tudíž nedává zcela nevychýlené výsledky,²³ její hodnota bez jakýchkoli pochyb poukazuje na silnou autoregresní strukturu mezi rezidui.

Ačkoli jsou rezidua silně autokorelována, naivní odhady (tj. odhady pořízené metodou OLS nezohledňující autoregresní charakter reziduí) jsou stále ještě konzistentní a nestranné. Bohužel však nejsou eficientní. K opravení autokorelační struktury a získání vydatných odhadů je nutné použít nějakou transformaci dat. Jako vhodná se v tomto případě jeví např. transformace navržená J. Durbinem (Durbin (1960), též ve Víšek (2002)). Dva důvody však hovoří proti jejímu použití.

Za prvé je to fakt, že při odhadu modelu budeme používat kromě standardních metod i robustní techniky. Ve chvíli, kdy je na data aplikujeme, totiž zcela jistě nějaká pozorování ze souboru dat vypadnou, jelikož je robustní metody označí za kontaminaci.

²³ Jelikož data byla uspořádána do panelů tím způsobem, že jeden panel tvořily vždy údaje za jedno odvětví v letech 1993 – 1999, porovnává D-W test mezi každým 7. a 8. řádkem rezidua, která spolu věcně ani časově nesouvisí (rok 1999 odvětví X s rokem 1993 odvětví Y).

Durbin-Watsonova statistika bude tudíž dávat ještě více zavádějící výsledky, než tomu je v tabulce 3.1.b a těžko budeme moci vzájemné výsledky nějak porovnat. Avšak i kdybychom nějakým způsobem věděli, že i po aplikaci robustních metod jsou rezidua autokorelována, což pravděpodobně stále ještě budou, bude v podstatě nemožné najít nějakou vhodnou transformaci, která by autokorelační strukturu reziduí „opravila“. Všechny transformace totiž předpokládají alespoň nějakou formu symetrie dat, tj. že jsou například panely stejně velké, že v datech nejsou „díry“, apod. Nic z toho ovšem po aplikaci robustních odhadů nemusí nutně platit a zpravidla také neplatí.

Druhým důvodem hovořícím proti použití transformace dat je situace popsaná ve Víšek (2002) a Víšek (2003). V případě, že je odhadovaný model podurčený nebo jsou místo některých vysvětlujících veličin použity proxy proměnné, může totiž při odhadu *transformovaných* dat v důsledku silné autoregrese právě mezi proměnnými, které v modelu chybí, nastat situace, že nejlepším (nejvydatnějším) modelem bude regrese vysvětlované veličiny na zpožděnou hodnotu sebe samé

$$Y_{it} = \Theta \cdot Y_{i,t-1} + \vartheta_{it} \quad (3.5)$$

Tento model bude sice nejlepším prediktorem, tj. bude schopen nejspolehlivěji předpovídat budoucí hodnotu vysvětlované veličiny, nicméně nám příliš nepopíše dílčí vliv jednotlivých proměnných. Z těchto důvodů zůstaneme u modelu vypočteného v tabulce 3.1.a, který sice vzhledem k autokorelaci reziduí není dost dobře možné použít pro predikci, ale alespoň nám dává představu o vlivu jednotlivých proměnných na úroveň exportu.

Ad (ii), abychom mohli interpretovat odhady parametrů přímo jako příslušné elasticity, je nutné, aby jednotlivé vysvětlující veličiny byly vzájemně nezávislé. Představu o závislostech mezi jednotlivými veličinami můžeme zjistit z korelační matice.

Vzhledem k tomu, že korelační koeficient zachycuje *pouze lineární* závislost, může se stát, že ačkoli bude závislost mezi proměnnými silná, avšak jiného charakteru než lineárního, bude korelační koeficient nízký. Pro analyzování závislostí vysvětlujících veličin je z tohoto důvodu vhodné zkoumat korelaci nejen mezi samotnými veličinami, ale též alespoň mezi několika jejich mocninami (Víšek (2002)). Z analýzy, jejíž výsledky shrnuje tabulka 3.2 je patrné, že jednotlivé vysvětlující veličiny lze považovat za vzájemně nezávislé.

Tabulka 3.2 Největší korelační koeficient (v absolutní hodnotě) mezi každými dvěma páry vysvětlujících veličin a jejich mocnin do 5. stupně

	$\log^2(FDI)$	$\log^2(De/VA)$	$\log^5(K/L)$	$\log^2(VA)$	$\log^5(REER)$	$\log^2(Tar)$
$\log(VA)$	0.600					
$\log(K/L)$		-0.374				
$\log(De/VA)$			-0.422			
$\log(FDI)$				0.476		
$\log(Tar)$					0.568	
$\log(REER)$						-0.365

Ad (iii), jak pomocí testu Lagrangeových multiplikátorů, tak pomocí Whiteova testu se zdá, že rezidua jsou silně heteroskedastická. Vzhledem k silné autoregresi reziduí jsou však výsledky obou testových statistiky velmi slabé, pokud ne zcela nevyhovující (Víšek (1997)) a nelze se na ně tudíž stoprocentně spoléhat.

Nicméně i přes tento fakt si můžeme dovolit tvrdit, že heteroskedasticita v reziduích pravděpodobně skutečně nastala. Jednotlivá odvětví se totiž výrazně liší svou velikostí. Je přirozené, že velká odvětví budou mít velký rozptyl na rozdíl od odvětví menších, kde bude ze zřejmých důvodů rozptyl menší. Podobné vlastnosti vykazují například národní data za jednotlivé regiony. Větší území tíhnou v důsledku řádově větších hodnot taktéž v absolutní hodnotě k větším odchylkám.

Ad (iv), hodnoty p-values Jarque-Berova i Shapiro-Wilkova testu jasně ukazují, že předpoklad normality reziduí není splněn. Důvodem je s největší pravděpodobností kontaminace dat. Především v této oblasti by aplikace robustních metod měla přinést výrazné vylepšení.

3.3 Modely pro jednotlivé roky – metoda LTS

Z předchozí kapitoly je patrné, že kvalita odhadů pořízených metodou obyčejných nejmenších čtverců není, především z důvodů kontaminace dat, nikterak vysoká. To spolu s faktem, že se jedná o data z tranzitivní ekonomiky, nasvědčuje několikrát zmiňované hypotéze, že soubor dat obsahuje pozorování příslušející do dvou nebo více rozdílných populací. S použitím robustních regresních metod bychom měli být schopni tuto hypotézu potvrdit.

Jelikož soubory obsahující více populací mohou být náchylné k tomu, že i odhady s vysokým bodem selhání mohou dávat vzájemně rozdílné výsledky, použijme v našich odhadech dvě rozdílné metody. Jestliže se jejich výsledky budou shodovat, je kontaminace

definována jednoznačně. V případě, že by se rozcházely, budou data vyžadovat další, podrobnější zkoumání.

Jak již bylo předesláno výše, použitými metodami budou odhady *least trimmed squares* (LTS) a odhad *reweighted least squares* (RLS), založený na diagnostice dat pomocí metody *least median of squares* (LMS), resp. *least-quantile squared* (LQS).

Provedení výpočtů a získání výsledků bylo časově velice náročné. Podrobné výpočty jsou přehledně uspořádány v přílohách II a III, přičemž do následujícího textu jsou zařazeny pouze souhrnné výsledky. Z nich je bezpochyby nejzajímavější údaj o velikosti „hlavní“ populace v jednotlivých letech. Za povšimnutí rovněž stojí relativní stabilita modelů v čase.

Rozhodujícím úkolem robustní analýzy bylo stanovení hodnoty parametru h , který by měl odpovídat počtu nekontaminovaných pozorování. Pro jeho určení neexistuje nějaký univerzálně platný postup. Pro soubor dat naší velikosti se jako rozumná jeví následující metoda (Víšek (2002)):

Stanovme $h_0 = \lceil n/2 \rceil + \lceil (p+1)/2 \rceil$, tedy na hodnotu maximalizující velikost bodu selhání. Velikost „hlavní“ (tj. nekontaminované) populace nyní hledáme tak, že postupně odhadujeme regresní model a v každém kroku zvětšíme h o jedničku (tj. přidáme vždy jedno pozorování), tedy $h = h_0, h_0 + 1, h_0 + 2, \dots$. S rostoucím h jsme se zpravidla setkávali s následujícím chováním modelů:

- (i) Pro několik prvních souborů, řekněme pro $h = h_0, h_0 + 1, h_0 + 2, \dots, h_1 - 1$, odhady modelů fluktovaly (v některých případech dokonce velmi výrazně). Skupiny vyloučených pozorování z jednotlivých kroků nevytvářely navzájem vnořené množiny.
- (ii) Následně se objevil interval, řekněme pro $h = h_1, h_1 + 1, h_1 + 2, \dots, h_2 - 1$, ve kterém se odhady jednotlivých koeficientů stabilizovaly a zároveň vyloučená pozorování začala vytvářet vzájemně vnořené množiny. Tj. každý následující podsoubor vyloučených pozorování byl celý nebo téměř celý obsažen v podsouboru předchozím.
- (iii) Pro velikost populace h_2 došlo ke skokové změně v odhadech parametrů modelu a rovněž k významné změně ve vyloučených pozorováních. Pro $h > h_2$ zpravidla začaly odhady znovu fluktuovat a zpravidla došlo také k

výrazné změně v ostatních testových statistikách – Shapiro – Wilkův a Jarque – Berův test či směrodatná chyba regrese.

Odhadujeme tedy metodou LTS model pro jednotlivé roky, který má následující tvar:

$$\ln(X_{it}^{EU}) = \beta_0 + \beta_1 \ln(PX_{it}^{EU} / PM_{it}^{EU}) + \beta_2 \ln(K_{it} / L_{it}) + \beta_3 \ln(FDI_{it}) + \beta_4 \ln(Tar_{it}) + \beta_5 \ln(De_{it} / VA_{it}) + \eta_{it} \quad (3.4)$$

Přístup přes modely individuálních let má dvě výhody. Jednak přináší možnost sledovat změny ve struktuře a velikosti kontaminace v čase. A za druhé, opakované vyhledávání optimální velikosti parametru h nám pomůže lépe demonstrovat chování modelů popsané v bodech (i) – (iii).

Tabulka 3.3.a Odhady modelů

Rok	1993	1994	1995	1996	1997	1998	1999
Velikost populace (h)	37	54	54	49	54	48	54
Intercept	14.232	17.195	13.403	16.718	11.842	12.771	12.268
log(PM)	-0.481	-0.659	-0.703	-0.546	-0.516	-0.588	-0.518
log(PX)	0.277	0.396	0.663	0.385	0.551	0.707	0.588
log(VA)	0.392	0.607	0.709	0.585	0.861	0.818	0.857
log(K/L)	-0.602	-0.906	-0.552	-0.868	-0.498	-0.663	-0.682
log(De/VA)	0.488	0.546	0.394	0.152	0.350	0.101	0.252
log(FDI)	0.473	0.403	0.229	0.290	0.164	0.275	0.210
log(Tar)	-0.477	-1.143	-0.650	-0.558	-0.643	-0.840	-0.741

Tabulka 3.3.b Další charakteristiky modelů

Rok	1993	1994	1995	1996	1997	1998	1999
Velikost populace (h)	37	54	54	49	54	48	54
R-squared	0.959	0.809	0.798	0.867	0.856	0.929	0.886
Adjusted R-squared	0.949	0.780	0.767	0.845	0.834	0.917	0.868
SSR	2.084	20.609	18.583	9.367	15.958	6.652	13.909
$\sum r_{(h,n)}^2(\beta)$	0.265	1.991	2.386	0.984	1.676	0.795	1.291
Variance of residuals	0.072	0.448	0.404	0.228	0.347	0.166	0.302
SE of regression	0.268	0.669	0.636	0.478	0.589	0.408	0.550
F (zero slopes)	97.018	27.912	25.989	38.289	39.108	74.966	50.874
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.009	2.107	1.550	2.287	1.976	1.858	1.700
P-value	[<.883]	[<.918]	[<.256]	[<.980]	[<.819]	[<.711]	[<.458]
Shapiro-Wilk test	1.230	0.970	0.982	0.973	0.987	0.994	0.962
P-value	[.541]	[.183]	[.572]	[.315]	[.810]	[.997]	[.084]
Jarque-Bera test	0.966	1.695	0.885	0.752	0.704	0.137	1.322
P-value	[.301]	[.428]	[.643]	[.686]	[.703]	[.934]	[.516]
LM het. Test	3.248	0.051	6.148	7.803	9.139	0.003	2.168
P-value	[.072]	[.821]	[.013]	[.005]	[.003]	[.987]	[.141]
White het. Test	37.000	36.640	40.321	40.972	39.125	21.883	45.465
P-value	[.377]	[.393]	[.247]	[.225]	[.290]	[.959]	[.111]

Tabulka 3.3.c Vyloučená pozorování

Rok	1993	1994	1995	1996	1997	1998	1999
Velikost populace (<i>h</i>)	37	54	54	49	54	48	54
1			1	1	1	1	1
4				4	4		
6				6			
7	7	7	7	7	7	7	7
9				9		9	9
10							
11			11				
12	12	12	12	12			
					13	13	13
16						16	
		19					
21							
23	23	23	23	23	23	23	23
24		24	24			24	
25				25	25	25	25
26				26			
27							
30						30	
						32	
33							
34	34			34		34	
36	36					36	36
38							
42							
44							
49	49	49	49	49	49		
						53	

3.4 Modely pro jednotlivé roky – metoda RLS

Postup byl analogický aplikaci metody LTS. Hlavním úkolem bylo opět nalezení maximálního nekontaminovaného podsouboru dat, tedy nalezení odpovídající hodnoty parametru h pro každý jeden rok.

Způsob hledání a jeho průběh byl v podstatě stejný jako u metody LTS. Jediným rozdílem byla technika odhadu, ten byl v tomto případě dvoustupňový. Nejdříve byla na data aplikována procedura LMS (resp. LQS), na jejímž základě byly rozlišeny kontaminace a „hlavní“ populace pro různé hodnoty parametru h . Kontaminovaná pozorování byla ze souboru dat vymazána a na takto „očistěná“ data byla následně aplikována metoda OLS. Souhrnné výsledky jsou obsaženy v následující tabulce.

Tabulka 3.4.a Odhady modelů

Rok	1993	1994	1995	1996	1997	1998	1999
Velikost populace (h)	37	54	52	48	53	49	54
<i>Intercept</i>	14.232	17.314	13.636	16.546	12.189	11.471	12.268
<i>log(PM)</i>	-0.481	-0.139	-0.648	-0.632	-0.475	-0.282	-0.518
<i>log(PX)</i>	0.277	-0.063	0.611	0.466	0.498	0.501	0.588
<i>log(VA)</i>	0.392	0.631	0.664	0.569	0.839	0.852	0.857
<i>log(K/L)</i>	-0.602	-0.734	-0.548	-0.834	-0.536	-0.576	-0.682
<i>log(De/VA)</i>	0.488	0.337	0.290	0.185	0.333	0.146	0.252
<i>log(FDI)</i>	0.473	0.264	0.234	0.319	0.177	0.192	0.210
<i>log(Tar)</i>	-0.477	-0.756	-0.553	-0.637	-0.603	-0.777	-0.741

Tabulka 3.4.b Další charakteristiky modelů

Rok	1993	1994	1995	1996	1997	1998	1999
Velikost populace (h)	37	54	52	48	53	49	54
R-squared	0.959	0.793	0.818	0.884	0.868	0.921	0.886
Adjusted R-squared	0.949	0.762	0.789	0.863	0.847	0.907	0.868
SSR	2.084	20.973	13.566	8.128	14.115	7.616	13.909
$\sum r_{(h_{it})}^2(\beta)$	0.265	1.757	2.209	0.889	1.493	0.700	1.291
Variance of residuals	0.072	0.456	0.308	0.203	0.314	0.186	0.302
SE of regression	0.268	0.675	0.555	0.451	0.560	0.431	0.550
F (zero slopes)	97.018	25.212	28.267	43.449	42.157	67.863	50.874
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.009	1.858	1.779	2.405	2.068	2.056	1.700
P-value	[<.883]	[<.683]	[<.586]	[<.993]	[<.895]	[<.892]	[<.458]
Shapiro-Wilk test	1.230	0.982	0.985	0.971	0.983	0.982	0.962
P-value	[.541]	[.581]	[.774]	[.274]	[.631]	[.640]	[.084]
Jarque-Bera test	0.966	1.617	0.557	1.833	0.921	1.079	1.322
P-value	[.301]	[.445]	[.757]	[.400]	[.631]	[.583]	[.516]
LM het. Test	3.248	2.930	3.691	9.502	12.167	0.042	2.168
P-value	[.072]	[.087]	[.055]	[.002]	[.000]	[.838]	[.141]
White het. Test	37.000	38.807	38.681	36.654	44.094	39.970	45.465
P-value	[.377]	[.302]	[.307]	[.392]	[.139]	[.259]	[.111]

Tabulka 3.4.c Vyloučená pozorování

Rok	1993	1994	1995	1996	1997	1998	1999
Velikost populace (h)	37	54	52	48	53	49	54
1			1	1	1	1	1
4	4			4	4		
6				6			
7	7	7	7	7	7	7	7
9	9	9	9	9	9	9	9
10							
11			11				
12			12	12			
					13	13	13
16				16			
21							
		22					
23	23	23	23	23	23	23	23
24			24	24		24	
25			25	25	25	25	25
26				26			
27							
30						30	
						32	
33							
34	34			34		34	
36						36	36
38							
42							
44							
49	49	49	49	49	49		
58							

3.5 Teoretické zdůvodnění rozdělení dat

Výsledky robustních regresí prezentované v kapitolách 3.3 a 3.4 a především v přílohách II a III dokazují, že užití robustních metod má na datech z tranzitivních ekonomik svůj smysl nejen proto, že jejich aplikací se podstatně vylepší vlastnosti modelů (zvýšení determinace modelu a výrazné vylepšení normality reziduí), ale především proto, že robustní analýza poměrně jasně potvrzuje několikrát zmiňovanou hypotézu, že data z tranzitivních ekonomik jsou, vzhledem k podstatě celého procesu ekonomické transformace, složena z alespoň ze dvou různých populací.

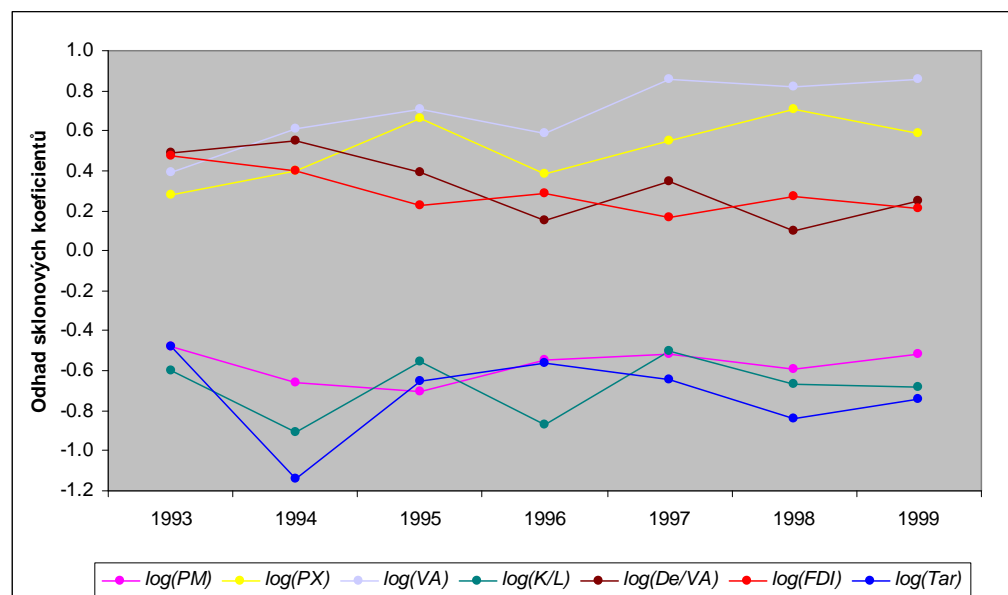
Pro fakt, že rozdělení dat na „hlavní“ a „komplementární“ populaci není jen čistě technického charakteru, ale že se skutečně jedná o dvě navzájem rozdílné populace, svědčí následující závěry:

- (i) Vysoká míra shody (téměř 85 %) mezi oběma metodami v tom, která pozorování přísluší k „hlavní“ a která ke „komplementární“ populaci,

svědčí o relativně vysoké homogenitě dat uvnitř obou populací. 72 z celkového počtu 85 pozorování bylo označeno oběma metodami zároveň jako kontaminace (viz Příloha IV).

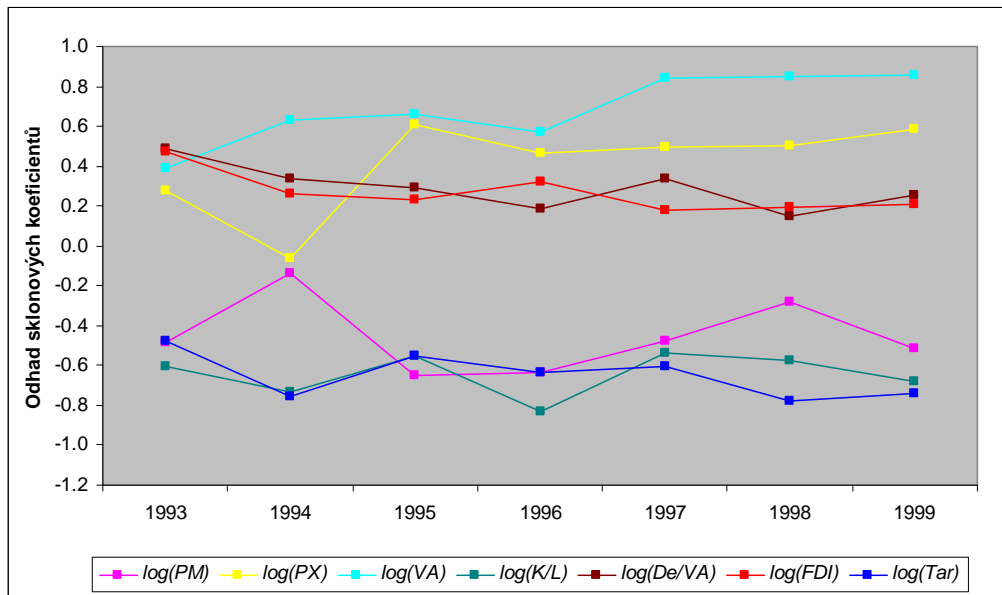
- (ii) Existence intervalu, v němž se s rostoucí velikostí „hlavní“ populace (tedy s rostoucím h) stabilizovaly odhady regresních koeficientů. Tento jev byl pozorovatelný u obou metod ve větší či menší míře v každém roce. (viz grafy II.93.a – III.99.a).
- (iii) Stabilita regresních koeficientů v čase – viz graf 3.2 a 3.2.a. Většina výraznějších výkyvů je způsobena především zahrnutím proměnných PX , PM a De/VA do všech modelů, ačkoli v některých případech nebyly signifikantní.

Graf 3.2. Stabilita koeficientů – metoda LTS



- (iv) Vyloučená pozorování v rámci jednotlivých let tvoří na intervalech, v nichž jsou stabilní odhady regresních koeficientů, vzájemně vnořené podsoubory. To znamená, že kontaminace je na tomto intervalu jasně patrná, dobře definovaná a stabilní (viz. tabulky II.93d – III.99.d).

Graf 3.2.a Stabilita koeficientů – metoda LMS



- (v) Rozdělení na „hlavní“ a „komplementární“ populaci je relativně stabilní v čase. Prozkoumáme-li vyloučená pozorování, která vznikla na základě „správných“ modelů z jednotlivých let (v tabulkách II.93.a–III.99.a označených tučně), vidíme sice, že vymazané podsoubory nejsou navzájem zcela úplně vnořené, nicméně některé „podezřelé“ body jsou i tak více než patrné (např. pozorování číslo 1, 7, 23, 49,...) – viz Příloha IV.

Ačkoli výše uvedená tvrzení prokazatelně potvrzují hypotézu o tom, že data jsou tvořena dvěma rozdílnými populacemi, podpořme je ještě jedním argumentem, jehož logika se spíše než o statistické techniky opírá o ekonomickou teorii.

Uvažujme standardní Cobb-Douglasovu produkční funkci

$$Q_i = \mu L_i^\lambda K_i^{1-\lambda}, \quad (3.6)$$

kde Q je množství výstupu, L je množství práce, K je množství kapitálu a μ a λ jsou parametry. Tento zápis můžeme upravit (Kmenta (1986)) na lineární model tvaru

$$\log \frac{Q_i}{L_i} = \log \mu + (1-\lambda) \log \frac{K_i}{L_i} + \varepsilon_i. \quad (3.7)$$

Dosadíme-li za L_i , K_i a $Q_i = X_i \cdot PX_i$, mění se úloha na odhadování modelu typu

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad (3.8)$$

který lze snadno odhadnout použitím běžné metody nejmenších čtverců.

Takto definovaný model vypočteme zvlášť pro obě skupiny dat. První z nich obsahuje data, která *obě* robustní procedury označila za „hlavní“ populaci, a druhá naproti tomu obsahuje data, která *alespoň* jedna z robustních procedur označila za kontaminaci.

Z výsledků jsou patrně nejzajímavější údaje o velikostech koeficientů determinace jednotlivých modelů a taktéž parametry produkční funkce přepočítané na základě koeficientů odhadnutých v regresním modelu (3.8).

Tabulka 3.5.a Koeficienty determinace

Populace	"hlavní"	"komplementární"
R-squared	0.478	0.173
Adjusted R-squared	0.477	0.163

Tabulka 3.5.b Parametry produkční funkce

Populace	"hlavní"	"komplementární"
μ	2675.793	6.200
λ	0.238	0.615

Ačkoli vlastnosti modelu jako normalita či homoskedasticita nejsou nejlepší, což není překvapující vzhledem k tomu, o jak jednoduchý model se jedná a vzhledem k tomu, že data nebyla primárně určena pro odhadování produkční funkce, některé základní tendence lze i na jeho základě dobře vypořádat.

Funkce Cobb-Douglasova tvaru jsou obecně považovány za produkční funkce, podle kterých se řídí výroba firem v dokonale konkurenčním prostředí. Vzhledem k tomu, že model Cobb-Douglasovy produkční funkce spočítaný pro data „hlavní“ populace je determinován podstatně lépe než model pro populaci „komplementární“, lze oprávněně tvrdit nejen, že rozdělení dat není náhodné a že se skutečně jedná o dvě navzájem rozdílné populace, ale navíc můžeme říci, co je jejich příčinou.

Data příslušející do „hlavní“ populace, která se „chová“ v souladu s předpoklady Cobb-Douglasovy produkční funkce, můžeme považovat za transformovaná odvětví. Podnikům v těchto odvětvích se, alespoň na úrovni klasifikace SITC 02, již podařilo přizpůsobit svou výrobu, způsoby řízení a využívání primárních zdrojů konkurenčním požadavkům, které jsou na ně kladeny na trzích EU.

Naproti tomu odvětví příslušející do „komplementární“ populace lze v tomto kontextu považovat za netransformovanou, či z důvodů probíhající transformace za natolik turbulentní, že na ně nelze vhodným způsobem Cobb-Douglasovu produkční funkci nasadit. Příčinou heterogenity dat je tedy rozdílná rychlost a úroveň ekonomické transformace jednotlivých odvětví.

Toto tvrzení podporují i hodnoty parametrů produkčních funkcí. Výrazně vyšší hodnota parametru μ svědčí o efektivnějším využívání primárních zdrojů v odvětvích „hlavní“ populace proti odvětvím populace „komplementární“. Hodnoty parametru λ svědčí o rozdílných proporcích využívání práce a kapitálu. Závěr hovořící pro to, že data komplementární populace představují netransformovaná odvětví, podporuje i fakt, že velikost „komplementární“ populace v čase klesá, a tudíž roste počet transformovaných odvětví.

3.6 Globální model na základě očištěných dat

Zkusme tedy na základě dat představujících transformovaná odvětví ještě jednou odhadnout globální model metodou obyčejných nejmenších čtverců, zkoumejme změny proti modelu odhadnutému v kapitole 3.2. Začneme opět se všemi vysvětlujícími proměnnými. Jejich postupným vylučováním na základě nejvyšších p-values dospějeme k následujícímu modelu:

Tabulka 3.6.a

Proměnná	Odhad koeficientu	Směrodatná odchylka	t-statistika	P-value
<i>Intercept</i>	34.7549	3.5881	9.686	[.000]
<i>log(VA)</i>	0.6585	0.0329	20.007	[.000]
<i>log(K/L)</i>	-0.5760	0.0523	-11.006	[.000]
<i>log(De/VA)</i>	0.2698	0.0455	5.933	[.000]
<i>log(FDI)</i>	0.2348	0.0280	8.374	[.000]
<i>log(Tar)</i>	-0.6002	0.0456	-13.174	[.000]
<i>log(REER)</i>	-10.8345	1.7562	-6.169	[.000]

Tabulka 3.6.b

Další charakteristiky			
Mean of dep. var.	11.622	LM het. test	17.231 [.000]
Std. dev. of dep. var.	1.369	Durbin-Watson	0.315 [.000,.000]
SSR	140.477	White het. test	69.487 [.000]
Variance of residuals	0.419	Jarque-Bera test	2.184 [.336]
SE of regression	0.648	Shapiro-Wilk test	0.995 [.291]
R-squared	0.780	Ramsey's RESET2	7.178 [.008]
Adjusted R-squared	0.776	F (zero slopes)	198.158 [.000]
Log likelihood	-333.127	Schwarz B.I.C.	353.549

Z výsledků je zřejmé, že model spočítaný na očištěných datech má lepší vlastnosti, než měl původní globální model v kapitole 3.2 počítaný na kompletních datech.

Došlo k výraznému zlepšení determinace modelu (z 0.550 na 0.776), patrné je vylepšení normality reziduí měřené jak Jarque-Berovým, tak Shapiro-Wilkovým testem, které nasvědčuje tomu, že data jsou podstatně homogennější povahy. Vylepšení můžeme rovněž pozorovat u směrodatné chyby regrese (z 1.105 pokles na 0.648).

Změny však můžeme pozorovat i u odhadů vlastních koeficientů modelu. Souhrnně o nich vypovídá následující tabulka:

Tabulka 3.7 Změny koeficientů globálního modelu

Proměnná	Odhady koeficientů		Změna (v %)
	Kompletní data	Očištěná data	
<i>Intercept</i>	29.4824	34.7549	17.88
<i>log(VA)</i>	0.6846	0.6585	-3.81
<i>log(KL)</i>	-0.7935	-0.5760	-27.41
<i>log(De/VA)</i>	0.2775	0.2698	-2.75
<i>log(FDI)</i>	0.1093	0.2348	114.85
<i>log(Tar)</i>	-0.6257	-0.6002	-4.09
<i>log(REER)</i>	-7.5083	-10.8345	44.30

Nejmarkantnější je především nárůst koeficientu u proměnné $\log(FDI)$, který opět potvrzuje závěry kapitoly 3.5, jelikož je zřejmé, že transformovaná odvětví přilákávají větší objemy zahraničního kapitálu a ten má tudíž i výraznější vliv na velikost exportu.

4. Závěr

Provedené analýzy jasně ukazují, že použití robustních regresních technik na datech z tranzitivních ekonomik má význam a smysl.

Z čistě technického, či chcete-li ekonometrického, úhlu pohledu přináší využití robustních metod v porovnání s odhady pořízenými metodou obyčejných nejmenších čtverců, která není schopna se uspokojivě vypořádat s heterogenním charakterem dat, podstatné vylepšení vlastností modelů. To je pochopitelně spjaté i se změnou v odhadech parametrů exportní funkce (v našem případě především v nárůstu vlivu proměnné $\log(FDI)$).

Ukazuje se však, že schopnost robustních technik rozdělit data na „hlavní“ populaci a kontaminaci má i svou z ekonomické teorie vycházející logiku. Jednoduchá analýza provedená na základě Cobb-Douglasovy produkční funkce potvrzuje hypotézu, že data jsou tvořena dvěma více méně autonomními populacemi. Jejich existence je spjata s procesy, kterými podniky v tranzitivní ekonomice procházejí a odráží tak rozdílnou úroveň transformace v jednotlivých odvětvích.

Pochopitelně není pravda, že v případě, že nějaké odvětví spadá do skupiny „nettransformovaných“ všechny podniky, které se v něm nacházejí, jsou nettransformované. Provedené analýzy vystihují pouze převažující charakter odvětví na úrovni klasifikace SITC 02. Bezpochyby zajímavý by byl výzkum těchto vlastností na úrovni jednotlivých podniků.

Ačkoli prezentované výsledky jsou přesvědčivé, je jejich platnost omezena pouze na Českou republiku. Zajímavým tématem pro další výzkum je i provedení stejné, nebo lehce modifikované analýzy i pro ostatní tranzitivní ekonomiky bývalého východního bloku, především však pro země „podobné“ České republice, jako jsou například Slovensko, Maďarsko či Slovinsko. Srovnání výsledků a jejich zasazení do hospodářsko-politického kontextu by mohlo jistě přinést nové pohledy na průběh transformací v těchto zemích.

Uzavřeme tedy tuto práci konstatováním, že robustní regresní techniky představují velmi efektivní nástroj zpracování dat, jehož aplikace může přinést mnohá, často netušená zjištění. Především v takových případech, kde se domníváme, že z nějakého důvodu mohou být data heterogenní povahy, bychom se vždy měli snažit provést alespoň elementární robustní analýzu.

5. Reference

- BENÁČEK, V. J. (2000a): The Usage of Factors in a Dual Economy: Specialization in the Czech Foreigner Trade. *Proceedings of the Conference „Macromodels‘99“ organized by Wroclaw University, 4 – 6 December, 1999, in Rydzyna, Poland, ISBN 83-86840-95-1, 201 – 209.*
- BENÁČEK, V. J. (2000b): Přímé zahraniční investice v české ekonomice. *Politická ekonomie č. 1, 2000.*
- BENÁČEK, V. J., PROKOP, L., VÍŠEK, J. Á (2002): Determining Factors of the Czech Foreign Trade Balance: Structural Issues in trade Creation. *Paper for the presentation and evaluation at the Czech National Bank, on 18 November, 2002.*
- BENÁČEK, V. J., VÍŠEK, J. Á (2002): Determining Factors of Competitiveness of Trade and Specialization of Czech Industrial Sector before EU Accession. *IES Working Paper no. 19., December 2002.*
- BOČEK, P., LACHOUT, P. (1995): Linear programming approach to LMS-estimation. *Computational Statistics and Data Analysis 19, (1995) 129 – 134.*
- BOSCOVICH, R. J. (1757): De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operas, ac habentur plura eius ex exemplaria etiam sensorum impresa. *Bononiensi Scientiarum et Artium Instituto Atque Academia Commentarii 4, 353 - 396.*
- BOX, G. E. P. (1953): Non-normality and tests of variance. *Biometrika 40, 318-335.*
- ČÍŽEK, P. (2001): Robust Estimation with Discrete Explanatory Variables, [on-line] *CERGE-EI Working paper 183, Praha, URL: <http://www.cerge-ei.cz/publications/working_papers/>.*
- DURBIN, J. (1960): Estimation of parameters in time-series regression models. *Journal of Royal Statistical Society, Series B, 22, 139 – 153.*
- EGERT, B. (2002): Investigating the Balassa-Samuelson Hypothesis in Transition: Do We Understand What We See?. *Economics in Transition, vol. 10, no. 2, 2002, 1 – 36.*
- ERICKSON, J., HAR-PELED, S., MOUNT, D.M. (2004): On the Least Median Square Problem, [on-line] *Proceedings of the “20th Annual ACM Symposium on Computational Geometry”, 273-279, 2004. URL: <<http://compgeom.cs.uiuc.edu/~jeffe/pubs/pdf/halfslab.pdf>>*
- GALILEI, G. (1632): *Dialogo dei masimi sistemi.*

- HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW P. J., STAHEL W. A. (1986): *Robust Statistics -- The Approach Based on Influence Functions*. New York: J. Wiley and Sons.
- HUBER, P. J. (1964): Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73 – 101.
- HUBER, P. J. (1967): The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1*, 221 – 233.
- JUREČKOVÁ, J. (2001): *Robustní statistické metody*. Praha: Karolinum, ISBN 80-246-0259-8.
- KMENTA, J. (1986): *Elements of Econometrics*, Macmillan Publishing Company, New York.
- LAPLACE, P. S. (1793): Sur quelques points du systeme du mode. *Memoires de l'Academic Royale des Sciences de Paris*, 1 – 87.
- ROUSSEEUW, P. J., A. M. LEROY (1987): *Robust Regression and Outlier Detection*. New York: J. Wiley and Sons.
- SIEGEL, A. F. (1982): Robust regression using repeated medians. *Biometrika* 69, 242 – 244.
- TUKEY, J. W. ((1970-71) 1977): *Exploratory Data Analysis* (1970-71: preliminary edition). Addison-Wesley, Reading, Mass.
- VÍŠEK, J. Á. (1996): On high breakdown point estimation. *Computational Statistics (1996)* 11, 137-146, Berlin.
- VÍŠEK, J. Á. (1997): *Ekonomie I*. Praha: Karolinum, ISBN 80-7184-483-7.
- VÍŠEK, J. Á. (1999): The Robust Regression and the experiences from its application on estimation of parameters in a dual economy. *Proceedings of the Conference „Macromodels '99“ organized by Wroclaw University, 4 – 6 December, 1999, in Rydzyna, Poland, ISBN 83-86840-95-1*, 424 – 445.
- VÍŠEK, J. Á. (2000a): On the diversity of estimates. *Computational Statistics and Data Analysis* 34, (2000) 67 – 89.
- VÍŠEK, J. Á. (2000b): Regression with high breakdown point. In Robust 2000. (ed. Jaromír Antoch & Gejza Dohnal), published by Union of Czechoslovak Mathematicians and Physicists, ISBN 80-7015-792-5, 324 – 356.
- VÍŠEK, J. Á. (2002): Development of the Czech Export in the Nineties. (*Preprint*).
- VÍŠEK, J. Á. (2003): Modern processing data. *Proceedings of Data Analysis 2002/II Modern Statistical Methods - Modeling, Regression, Classification and data Mining*, ISBN 80-239-0204-0, 131 – 167 (invited paper).

6. Přílohy

Příloha I. Odhad globálního modelu

Tabulka I.1.a

Proměnná	Odhad koeficientu	Směrodatná odchylka	t-statistika	P-value
<i>Intercept</i>	5.4469	11.2769	0.483	[.629]
<i>log(PM)</i>	0.1190	0.0918	1.296	[.196]
<i>log(PX)</i>	-0.0519	0.0926	-0.560	[.576]
<i>log(VA)</i>	0.6870	0.0505	13.593	[.000]
<i>log(K/L)</i>	-0.6954	0.1016	-6.846	[.000]
<i>log(De/VA)</i>	0.2751	0.0647	4.249	[.000]
<i>log(FDI)</i>	0.0923	0.0414	2.230	[.026]
<i>log(Tar)</i>	-0.6197	0.0665	-9.323	[.000]
<i>log(REER)</i>	-26.6740	9.4314	-2.828	[.005]
<i>log(GDP^{EU})</i>	15.9155	7.2632	2.191	[.029]

Tabulka I.1.b

Další charakteristiky			
Mean of dep. var.	11.235	LM het. test	6.756 [.009]
Std.dev. of dep. var.	1.648	Durbin-Watson	0.181 [.000,.000]
SSR	502.900	White het. test	169.732 [.000]
Variance of residuals	1.206	Jarque-Bera test	229.443 [.000]
SE of regression	1.098	Shapiro-Wilk test	0.928 [.000]
R-squared	0.565	Ramsey's RESET2	4.200 [.041]
Adjusted R-squared	0.556	F (zero slopes)	60.216 [.000]
Log likelihood	-640.817	Schwarz B.I.C.	671.101

Tabulka I.2.a

Proměnná	Odhad koeficientu	Směrodatná odchylka	t-statistika	P-value
<i>Intercept</i>	6.3047	11.1632	0.565	[.573]
<i>log(PM)</i>	0.0732	0.0416	1.759	[.079]
<i>log(VA)</i>	0.6915	0.0499	13.861	[.000]
<i>log(K/L)</i>	-0.6915	0.1013	-6.829	[.000]
<i>log(De/VA)</i>	0.2713	0.0643	4.217	[.000]
<i>log(FDI)</i>	0.0922	0.0413	2.231	[.026]
<i>log(Tar)</i>	-0.6249	0.0658	-9.502	[.000]
<i>log(REER)</i>	-26.6916	9.4236	-2.832	[.005]
<i>log(GDP^{EU})</i>	15.6940	7.2465	2.166	[.031]

Tabulka I.2.b

Další charakteristiky			
Mean of dep. var.	11.235	LM het. test	6.224 [.013]
Std.dev. of dep. var.	1.648	Durbin-Watson	0.179 [.000,.000]
SSR	503.278	White het. test	127.632 [.000]
Variance of residuals	1.204	Jarque-Bera test	242.227 [.000]
SE of regression	1.097	Shapiro-Wilk test	0.926 [.000]
R-squared	0.565	Ramsey's RESET2	4.053 [.045]
Adjusted R-squared	0.556	F (zero slopes)	67.815 [.000]
Log likelihood	-640.977	Schwarz B.I.C.	668.233

Tabulka I.3.a

Proměnná	Odhad koeficientu	Směrodatná odchylka	t-statistika	P-value
<i>Intercept</i>	8.0386	11.1474	0.721	[.471]
<i>log(VA)</i>	0.6877	0.0500	13.763	[.000]
<i>log(K/L)</i>	-0.7973	0.0817	-9.765	[.000]
<i>log(De/VA)</i>	0.2746	0.0645	4.260	[.000]
<i>log(FDI)</i>	0.1043	0.0409	2.553	[.011]
<i>log(Tar)</i>	-0.6275	0.0659	-9.521	[.000]
<i>log(REER)</i>	-27.3998	9.4385	-2.903	[.004]
<i>log(GDP^{EU})</i>	15.9233	7.2634	2.192	[.029]

Tabulka I.3.b

Další charakteristiky			
Mean of dep. var.	11.235	LM het. test	5.833 [.016]
Std.dev. of dep. var.	1.648	Durbin-Watson	0.180 [.000,.000]
SSR	507.002	White het. test	101.104 [.000]
Variance of residuals	1.210	Jarque-Bera test	260.551 [.000]
SE of regression	1.100	Shapiro-Wilk test	0.922 [.000]
R-squared	0.562	Ramsey's RESET2	4.656 [.032]
Adjusted R-squared	0.554	F (zero slopes)	76.678 [.000]
Log likelihood	-642.551	Schwarz B.I.C.	666.778

Tabulka I.4.a

Proměnná	Odhad koeficientu	Směrodatná odchylka	t-statistika	P-value
<i>Intercept</i>	29.4824	5.3706	5.490	[.000]
<i>log(VA)</i>	0.6846	0.0502	13.645	[.000]
<i>log(K/L)</i>	-0.7935	0.0820	-9.677	[.000]
<i>log(De/VA)</i>	0.2775	0.0647	4.286	[.000]
<i>log(FDI)</i>	0.1093	0.0410	2.667	[.008]
<i>log(Tar)</i>	-0.6257	0.0662	-9.452	[.000]
<i>log(REER)</i>	-7.5083	2.6112	-2.875	[.004]

Tabulka I.4.b

Další charakteristiky			
Mean of dep. var.	11.235	LM het. test	5.421 [.020]
Std.dev. of dep. var.	1.648	Durbin-Watson	0.178 [.000,.000]
SSR	512.817	White het. test	99.649 [.000]
Variance of residuals	1.221	Jarque-Bera test	252.860 [.000]
SE of regression	1.105	Shapiro-Wilk test	0.923 [.000]
R-squared	0.557	Ramsey's RESET2	4.087 [.044]
Adjusted R-squared	0.550	F (zero slopes)	87.861 [.000]
Log likelihood	-644.986	Schwarz B.I.C.	666.185

Příloha II. Výpočet modelů metodou LTS

Rok 1993

Tabulka II.93.a Odhady modelů

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
<i>Intercept</i>	15.336	15.253	15.287	14.480	14.032	14.232	23.157	22.359	22.306	21.865
<i>log(PM)</i>	-0.572	-0.613	-0.626	-0.582	-0.549	-0.481	-0.785	-0.722	-0.699	-0.636
<i>log(PX)</i>	0.378	0.405	0.407	0.373	0.364	0.277	0.145	0.120	0.098	0.048
<i>log(VA)</i>	0.403	0.408	0.398	0.381	0.396	0.392	0.695	0.614	0.613	0.622
<i>log(KL)</i>	-0.768	-0.754	-0.741	-0.637	-0.592	-0.602	-1.430	-1.394	-1.391	-1.317
<i>log(De/VA)</i>	0.535	0.566	0.563	0.589	0.504	0.488	0.818	0.895	0.899	0.830
<i>log(FDI)</i>	0.530	0.538	0.541	0.542	0.497	0.473	0.302	0.369	0.378	0.425
<i>log(Tar)</i>	-0.673	-0.679	-0.671	-0.667	-0.500	-0.477	-1.435	-1.422	-1.428	-1.479

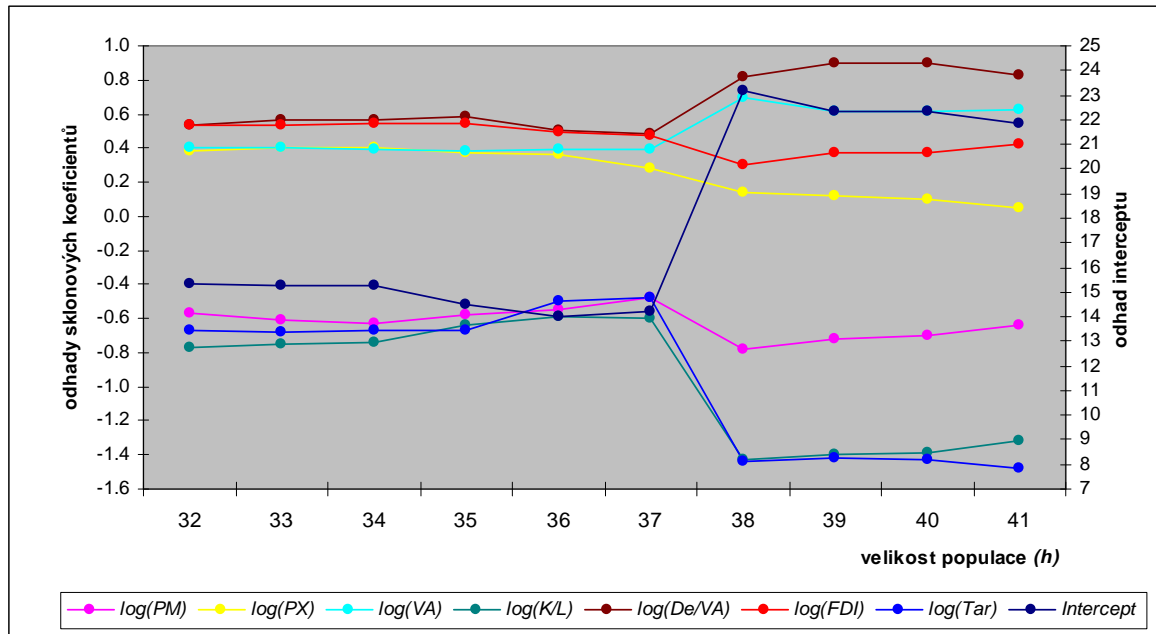
Tabulka II.93.b P - values

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.003]	[.060]	[.123]	[.214]	[.553]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(KL)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(De/VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka II.93.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
R-squared	0.983	0.980	0.975	0.972	0.964	0.959	0.972	0.970	0.968	0.963
Adjusted R-squared	0.978	0.974	0.969	0.965	0.955	0.949	0.966	0.963	0.961	0.955
$\sum r_{(i;n)}^2(\beta)$	0.711	0.900	1.113	1.320	1.725	2.084	2.505	2.806	3.096	3.569
$r_{(i)}^2(\beta)$	0.092	0.146	0.178	0.155	0.189	0.265	0.209	0.206	0.264	0.385
Variance of residuals	0.030	0.036	0.043	0.049	0.062	0.072	0.083	0.091	0.097	0.108
SE error of regression	0.172	0.190	0.207	0.221	0.248	0.268	0.289	0.301	0.311	0.329
F (zero slopes)	200.266	173.924	146.156	134.171	107.686	97.018	151.003	143.769	136.684	122.456
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.664	2.533	2.311	2.313	2.191	2.009	1.750	1.633	1.953	2.149
P-value	[<.999]	[<.997]	[<.983]	[<.983]	[<.960]	[<.883]	[<.648]	[<.497]	[<.836]	[<.945]
Shapiro-Wilk test	0.979	0.991	0.989	0.974	0.978	1.230	0.939	0.947	0.951	0.971
P-value	[.762]	[.991]	[.975]	[.556]	[.676]	[.541]	[.039]	[.066]	[.084]	[.361]
Jarque-Bera test	1.019	0.379	0.430	1.292	1.211	0.966	2.531	2.239	2.219	1.493
P-value	[.601]	[.827]	[.806]	[.524]	[.546]	[.301]	[.282]	[.326]	[.330]	[.474]
LM het. Test	0.038	0.297	0.436	0.016	0.830	3.248	0.520	3.207	2.689	0.041
P-value	[.846]	[.586]	[.509]	[.899]	[.362]	[.072]	[.471]	[.073]	[.101]	[.840]
White het. Test	n/a	n/a	n/a	n/a	n/a	37.000	35.659	36.340	35.250	36.937
P-value	n/a	n/a	n/a	n/a	n/a	[.377]	[.437]	[.406]	[.456]	[.379]

Graf II.93.a Stabilita koeficientů vzhledem k velikosti populace (h)



Tabulka II.93.d Vyloučená pozorování

Velikost populace (h)	32	33	34	35	36	37	38	39	40	41
1	1	1	1	1	1	1				
2							2	2	2	2
3							3	3	3	
4	4	4	4	4	4	4	4	4	4	4
6	6	6	6	6	6	6				
7	7	7	7	7	7	7	7	7	7	7
9	9	9	9	9	9	9	9	9	9	9
10	10	10	10	10	10	10				
11	11	11	11	11	11	11				
12	12	12	12	12	12	12	12	12	12	12
15							15	15	15	15
16	16	16	16	16	16	16	16	16	16	16
18							18	18	18	18
19	19	19	19	19	19	19	19	19	19	19
21	21	21	21	21	21	21	21			
22							22	22	22	22
23	23	23	23	23	23	23	23	23	23	23
24	24	24	24	24	24	24				
25	25	25	25	25	25	25				
26	26	26	26	26	26	26				
27	27	27	27	27	27	27	27	27	27	27
30	30	30	30	30	30	30				
32	32									
33	33	33	33	33	33	33	33	33	33	33
34	34	34	34	34	34	34	34	34	34	34
36	36	36	36	36	36	36	36	36	36	36

Vyloučená pozorování - pokračování

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
							37	37	37	
	38	38	38	38	38	38				
	42	42	42	42	42	42				
	44	44	44	44	44	44				
	49	49	49	49	49	49	49	49	49	49
	50									
	51	51	51	51	51					
							52	52	52	52
							53	53	53	53
							54	54		
	57	57	57				57	57	57	57
	58	58	58	58		58				58

Rok 1994

Tabulka II.94.a Odhady modelů

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	18.018	17.541	17.436	16.903	17.391	17.195	15.131	15.565	15.338	15.071
<i>log(PM)</i>	-0.676	-0.658	-0.749	-0.729	-0.765	-0.659	-0.272	-0.304	-0.144	-0.215
<i>log(PX)</i>	0.343	0.339	0.434	0.448	0.480	0.396	0.095	0.122	-0.023	0.060
<i>log(VA)</i>	0.632	0.634	0.652	0.659	0.597	0.607	0.655	0.595	0.588	0.611
<i>log(K/L)</i>	-0.940	-0.870	-0.851	-0.800	-0.899	-0.906	-0.671	-0.759	-0.746	-0.714
<i>log(De/VA)</i>	0.539	0.533	0.560	0.557	0.601	0.546	0.393	0.441	0.368	0.385
<i>log(FDI)</i>	0.408	0.410	0.411	0.398	0.445	0.403	0.233	0.277	0.278	0.268
<i>log(Tar)</i>	-1.180	-1.198	-1.262	-1.290	-1.273	-1.143	-0.838	-0.830	-0.760	-0.801

Tabulka II.94.b P - values

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.001]	[.140]	[.112]	[.423]	[.249]
<i>log(PX)</i>	[.018]	[.026]	[.006]	[.008]	[.007]	[.028]	[.595]	[.508]	[.895]	[.740]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.002]	[.001]	[.001]	[.002]
<i>log(De/VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.004]	[.002]	[.008]	[.008]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.004]	[.001]	[.001]	[.002]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

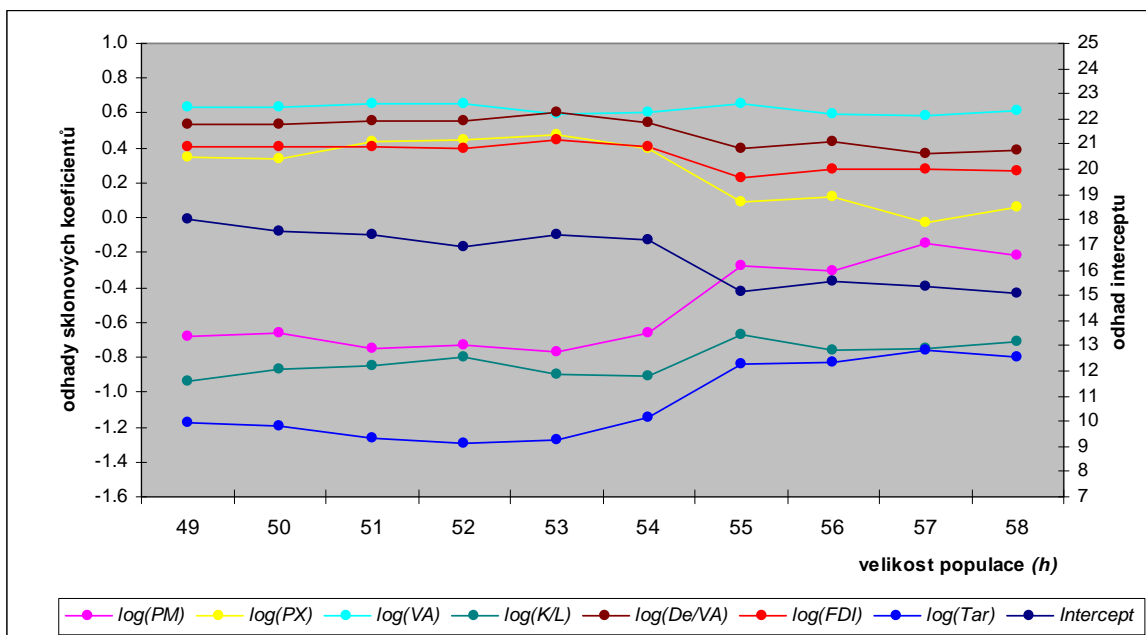
Tabulka II.94.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
R-squared	0.895	0.881	0.869	0.853	0.831	0.809	0.779	0.759	0.739	0.7213
Adjusted R-squared	0.877	0.861	0.848	0.830	0.805	0.780	0.746	0.724	0.702	0.6823
$\sum r_{(i:n)}^2(\beta)$	10.297	11.858	13.721	15.704	18.233	20.609	22.926	25.228	27.609	30.899
$r_{(h)}^2(\beta)$	1.270	1.439	1.640	1.816	2.006	1.991	2.073	2.162	2.112	2.9999
Variance of residuals	0.251	0.282	0.319	0.357	0.405	0.448	0.488	0.526	0.563	0.618
SE error of regression	0.501	0.531	0.565	0.597	0.637	0.669	0.698	0.725	0.751	0.7861
F (zero slopes)	50.116	44.235	40.755	36.610	31.699	27.912	23.697	21.647	19.817	18.488
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.454	2.327	2.298	2.323	2.229	2.107	2.029	1.985	1.972	1.9305
P-value	[<.996]	[<.986]	[<.982]	[<.985]	[<.967]	[<.918]	[<.864]	[<.824]	[<.808]	[<.760]

Další charakteristiky modelů – pokračování

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
Shapiro-Wilk test	0.985	0.988	0.988	0.984	0.977	0.970	0.983	0.986	0.979	0.97798
P-value	[.793]	[.892]	[.874]	[.720]	[.388]	[.183]	[.625]	[.741]	[.411]	[.371]
Jarque-Bera test	0.717	0.061	0.377	0.684	1.497	1.695	1.043	0.884	1.447	1.63036
P-value	[.699]	[.970]	[.828]	[.710]	[.473]	[.428]	[.593]	[.643]	[.485]	[.443]
LM het. Test	0.178	0.381	0.664	0.669	0.213	0.051	0.281	0.235	0.920	0.5371
P-value	[.673]	[.537]	[.415]	[.413]	[.644]	[.821]	[.596]	[.628]	[.337]	[.464]
White het. Test	38.822	38.431	34.443	35.889	37.681	36.640	35.988	37.790	33.928	35.5661
P-value	[.301]	[.317]	[.495]	[.427]	[.348]	[.393]	[.422]	[.343]	[.520]	[.442]

Graf II.94.a Stabilita koeficientů vzhledem k velikosti populace (*h*)



Tabulka II.94.d Vyloučená pozorování

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
	4	4	4							
	7	7	7	7	7	7	7	7	7	7
	9	9					9	9	9	
	12	12	12	12	12	12	12	12		
	16	16	16	16	16	19				
	19	19	19	19	19					
	22	22	22	22			22			
	23	23	23	23	23	23	23	23	23	23
	33									
	34	34	34	34	34	34				
	36	36	36	36	36	36				
	49	49	49	49	49	49	49	49	49	49

Tabulka II.95.a Odhady modelů

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	13.356	13.248	14.273	14.726	14.589	14.507	13.403	15.924	14.820	14.547
<i>log(PM)</i>	-0.688	-0.654	-0.692	-0.729	-0.826	-0.784	-0.703	-0.552	-0.461	-0.276
<i>log(PX)</i>	0.673	0.637	0.611	0.642	0.743	0.692	0.663	0.400	0.360	0.191
<i>log(VA)</i>	0.733	0.738	0.717	0.667	0.690	0.712	0.709	0.668	0.664	0.667
<i>log(K/L)</i>	-0.479	-0.462	-0.575	-0.660	-0.645	-0.665	-0.552	-0.828	-0.715	-0.700
<i>log(De/VA)</i>	0.179	0.175	0.222	0.255	0.280	0.359	0.394	0.403	0.445	0.366
<i>log(FDI)</i>	0.255	0.240	0.254	0.296	0.292	0.292	0.229	0.311	0.245	0.249
<i>log(Tar)</i>	-0.645	-0.630	-0.630	-0.625	-0.675	-0.722	-0.650	-0.915	-0.846	-0.759

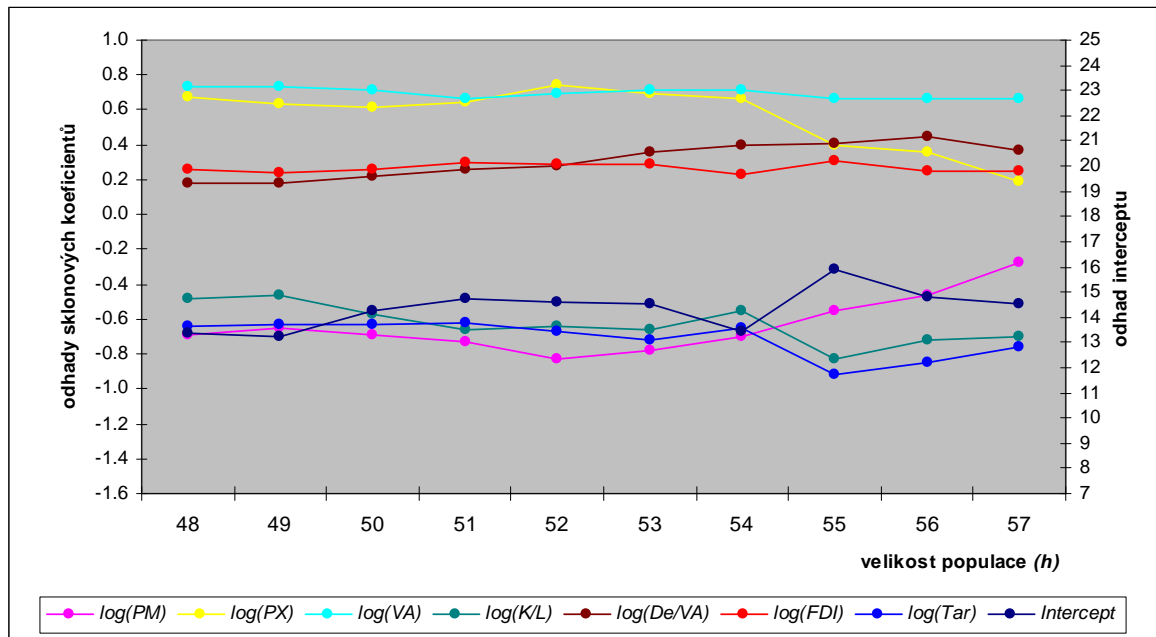
Tabulka II.95.b P - values

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.005]	[.021]	[.175]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.001]	[.031]	[.064]	[.340]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.001]	[.003]	[.000]	[.000]	[.000]	[.001]	[.005]	[.000]	[.001]	[.002]
<i>log(De/VA)</i>	[.047]	[.063]	[.023]	[.014]	[.013]	[.003]	[.002]	[.002]	[.002]	[.012]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.002]	[.000]	[.002]	[.004]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka II.95.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
R-squared	0.900	0.888	0.874	0.856	0.839	0.829	0.798	0.799	0.768	0.727
Adjusted R-squared	0.883	0.869	0.853	0.832	0.814	0.803	0.767	0.769	0.734	0.688
$\sum r_{(h,m)}^2(\beta)$	7.158	8.035	9.163	10.748	12.932	15.708	18.583	21.335	24.596	29.701
$r_{(h)}^2(\beta)$	0.732	0.832	0.882	1.261	1.920	2.404	2.386	2.104	2.564	3.969
Variance of residuals	0.179	0.196	0.218	0.250	0.294	0.349	0.404	0.454	0.512	0.606
SE error of regression	0.423	0.443	0.467	0.500	0.542	0.591	0.636	0.674	0.716	0.779
F (zero slopes)	51.653	46.594	41.640	36.466	32.828	31.252	25.989	26.654	22.704	18.679
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.143	2.002	2.124	1.942	1.905	1.664	1.550	1.677	1.621	1.444
P-value	[<.938]	[<.852]	[<.929]	[<.793]	[<.749]	[<.414]	[<.256]	[<.419]	[<.332]	[<.131]
Shapiro-Wilk test	0.986	0.984	0.980	0.985	0.981	0.974	0.982	0.965	0.980	0.977
P-value	[.814]	[.742]	[.535]	[.766]	[.551]	[.303]	[.572]	[.114]	[.455]	[.361]
Jarque-Bera test	0.419	0.591	0.666	0.637	2.072	3.273	0.885	2.226	0.378	1.993
P-value	[.811]	[.744]	[.717]	[.727]	[.355]	[.195]	[.643]	[.329]	[.828]	[.369]
LM het. Test	2.514	1.520	1.567	1.510	1.935	4.456	6.148	2.323	3.982	4.315
P-value	[.113]	[.218]	[.211]	[.219]	[.164]	[.035]	[.013]	[.128]	[.046]	[.038]
White het. Test	36.420	30.105	34.426	37.870	38.618	37.814	40.321	40.460	43.252	45.705
P-value	[.402]	[.703]	[.496]	[.340]	[.309]	[.342]	[.247]	[.242]	[.160]	[.106]

Graf II.95.a Stabilita koeficientů vzhledem k velikosti populace (h)



Tabulka II.95.d Vyloučená pozorování

Velikost populace (h)	48	49	50	51	52	53	54	55	56	57
1	1	1	1	1	1	1	1			
7	7	7	7	7	7	7	7	7	7	7
9	9	9	9	9						
11	11	11	11	11	11	11	11			
12	12	12	12	12	12	12	12	12	12	
22	22	22	22							
23	23	23	23	23	23	23	23	23	23	23
24	24	24	24	24	24	24	24	24	24	24
25	25	25	25	25	25					
30	30									
34										
36	36	36	36	36	36	36		36		
49	49	49	49	49	49	49	49	49	49	49

Rok 1996

Tabulka II.96.a Odhady modelů

Velikost populace (h)	44	45	46	47	48	49	50	51	52	53
Intercept	17.332	16.929	16.573	17.205	16.546	16.718	16.528	16.583	16.422	13.384
$\log(PM)$	-0.556	-0.745	-0.743	-0.726	-0.632	-0.546	-0.420	-0.271	-0.086	0.026
$\log(PX)$	0.382	0.556	0.564	0.530	0.466	0.385	0.258	0.111	-0.056	-0.030
$\log(VA)$	0.658	0.627	0.628	0.577	0.569	0.585	0.587	0.586	0.581	0.763
$\log(K/L)$	-1.020	-0.910	-0.861	-0.898	-0.834	-0.868	-0.847	-0.863	-0.851	-0.666
$\log(De/VA)$	0.207	0.220	0.218	0.154	0.185	0.152	0.144	0.090	0.067	0.267
$\log(FDI)$	0.307	0.349	0.350	0.358	0.319	0.290	0.272	0.278	0.272	0.219
$\log(Tar)$	-0.605	-0.692	-0.703	-0.685	-0.637	-0.558	-0.529	-0.467	-0.538	-0.687

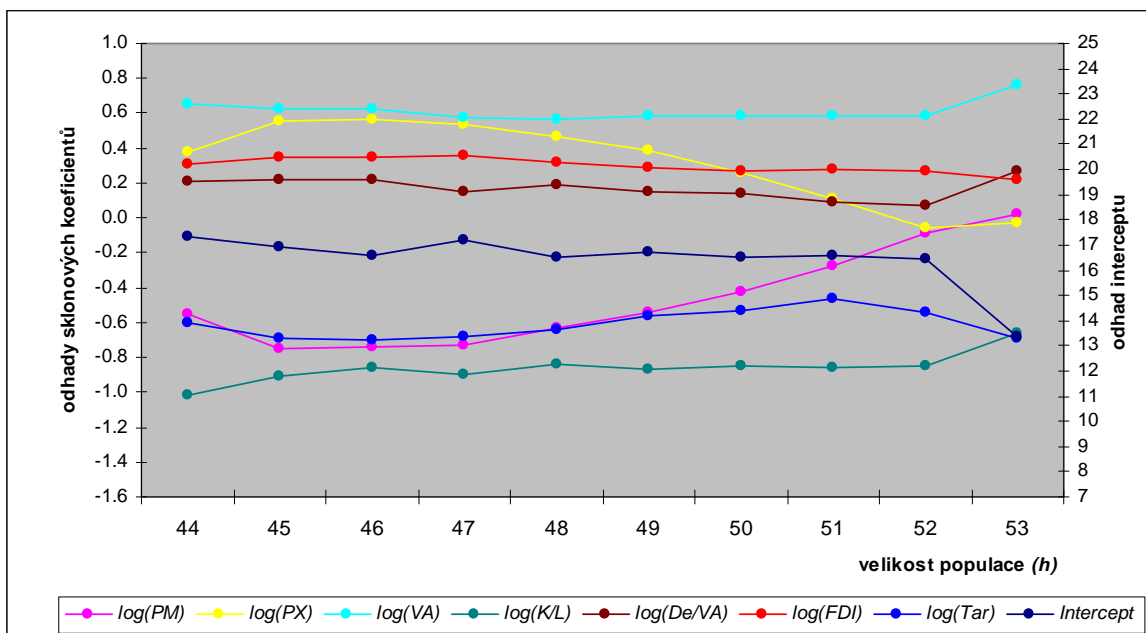
Tabulka II.96.b P - values

Velikost populace (<i>h</i>)	44	45	46	47	48	49	50	51	52	53
Intercept	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
log(PM)	[.000]	[.000]	[.000]	[.000]	[.000]	[.002]	[.018]	[.115]	[.611]	[.882]
log(PX)	[.004]	[.000]	[.000]	[.001]	[.005]	[.023]	[.134]	[.509]	[.742]	[.866]
log(VA)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
log(K/L)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
log(De/VA)	[.009]	[.009]	[.014]	[.082]	[.050]	[.121]	[.171]	[.404]	[.561]	[.031]
log(FDI)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.002]
log(Tar)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka II.96.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	44	45	46	47	48	49	50	51	52	53
R-squared	0.933	0.920	0.910	0.897	0.884	0.867	0.842	0.829	0.803	0.818
Adjusted R-squared	0.920	0.905	0.893	0.879	0.863	0.845	0.816	0.801	0.772	0.790
$\sum r_{(h,m)}^2(\beta)$	4.414	5.232	5.933	6.966	8.128	9.367	11.203	12.796	15.210	17.085
$r_{(h)}^2(\beta)$	0.703	0.649	0.642	0.797	0.889	0.984	1.638	1.303	1.788	1.394
Variance of residuals	0.123	0.141	0.156	0.179	0.203	0.228	0.267	0.298	0.346	0.380
SE error of regression	0.350	0.376	0.395	0.423	0.451	0.478	0.516	0.546	0.588	0.616
F (zero slopes)	71.353	60.898	54.628	48.555	43.449	38.289	31.980	29.726	25.651	28.950
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.231	2.690	2.435	2.306	2.405	2.287	2.212	2.170	2.062	2.338
P-value	[<.968]	[<1.00]	[<.994]	[<.982]	[<.993]	[<.980]	[<.963]	[<.949]	[<.892]	[<.987]
Shapiro-Wilk test	0.983	0.975	0.976	0.979	0.971	0.973	0.970	0.956	0.974	0.974
P-value	[.772]	[.436]	[.439]	[.548]	[.274]	[.315]	[.224]	[.058]	[.324]	[.292]
Jarque-Bera test	0.858	1.340	1.032	1.026	1.833	0.752	0.068	0.075	0.186	1.967
P-value	[.651]	[.512]	[.597]	[.599]	[.400]	[.686]	[.967]	[.963]	[.911]	[.374]
LM het. Test	0.988	1.138	1.306	3.720	9.502	7.803	3.047	5.814	5.179	6.086
P-value	[.320]	[.286]	[.253]	[.054]	[.002]	[.005]	[.081]	[.016]	[.023]	[.014]
White het. Test	39.850	36.890	39.085	39.802	36.654	40.972	29.084	32.686	36.381	37.681
P-value	[.263]	[.382]	[.291]	[.265]	[.392]	[.225]	[.749]	[.580]	[.404]	[.348]

Graf II.96.a Stabilita koeficientů vzhledem k velikosti populace (*h*)



Tabulka II.96.d Vyloučená pozorování

Velikost populace (<i>h</i>)	44	45	46	47	48	49	50	51	52	53
	1	1	1	1	1	1	1	1		
	4	4	4	4	4	4	4	4	4	
	6	6	6	6	6	6	6	6	6	
	7	7	7	7	7	7	7	7	7	7
	9	9	9	9	9	9	9	9	9	9
	12	12	12	12	12	12	12			
	13	13	13							13
		16	16	16	16					
	23	23	23	23	23	23	23	23	23	23
	24	24	24	24	24	24	24	24	24	24
	25	25	25	25	25	25	25	25	25	25
	26	26	26	26	26	26	26	26	26	
	27									
	32									32
	33	33	34	34	34	34				
	34	34	36	36						
	36	36								
	49	49	49	49	49	49	49	49	49	49

Rok 1997

Tabulka II.97.a Odhady modelů

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	14.224	13.943	13.697	13.480	13.507	13.032	11.842	11.619	10.502	13.723
<i>log(PM)</i>	-0.965	-0.895	-0.921	-0.981	-0.897	-0.690	-0.516	-0.374	-0.239	-0.365
<i>log(PX)</i>	0.893	0.830	0.857	0.922	0.852	0.671	0.551	0.431	0.347	0.427
<i>log(VA)</i>	0.801	0.805	0.830	0.845	0.868	0.888	0.861	0.850	0.871	0.808
<i>log(K/L)</i>	-0.686	-0.656	-0.647	-0.620	-0.641	-0.613	-0.498	-0.482	-0.382	-0.664
<i>log(De/VA)</i>	0.266	0.272	0.320	0.336	0.315	0.284	0.350	0.331	0.363	0.090
<i>log(FDI)</i>	0.283	0.266	0.261	0.254	0.220	0.214	0.164	0.164	0.138	0.124
<i>log(Tar)</i>	-0.825	-0.798	-0.805	-0.848	-0.773	-0.707	-0.643	-0.712	-0.730	-0.710

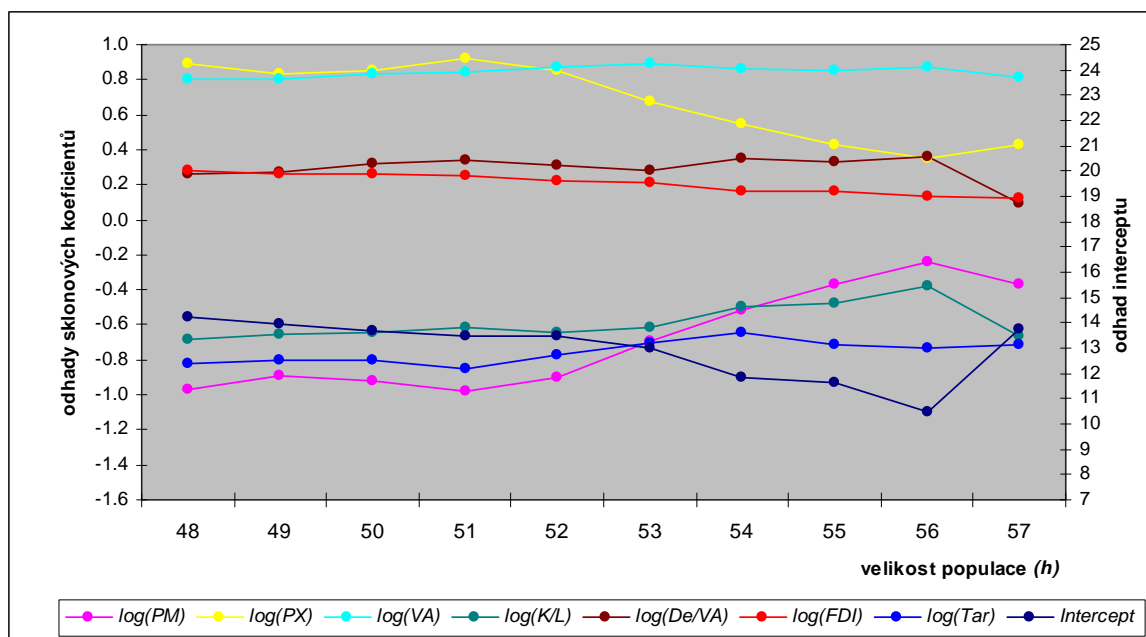
Tabulka II.97.b P - values

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.004]	[.041]	[.185]	[.071]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.001]	[.012]	[.046]	[.025]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.004]	[.009]	[.042]	[.001]
<i>log(De/VA)</i>	[.005]	[.006]	[.002]	[.002]	[.004]	[.014]	[.004]	[.011]	[.008]	[.448]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.000]	[.001]	[.002]	[.018]	[.028]	[.074]	[.120]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka II.97.c **Další charakteristiky modelů**

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
R-squared	0.921	0.910	0.910	0.902	0.890	0.874	0.856	0.831	0.817	0.782
Adjusted R-squared	0.907	0.895	0.895	0.887	0.872	0.855	0.834	0.806	0.790	0.751
$\sum r_{(h)}^2(\beta)$	7.191	8.160	9.144	10.356	11.793	13.724	15.958	19.109	21.648	25.021
$r_{(h)}^2(\beta)$	0.617	0.898	0.894	1.085	1.227	1.565	1.676	2.581	2.104	2.300
Variance of residuals	0.180	0.199	0.218	0.241	0.268	0.305	0.347	0.407	0.451	0.511
SE error of regression	0.424	0.446	0.467	0.491	0.518	0.552	0.589	0.638	0.672	0.715
F (zero slopes)	66.611	59.471	60.663	56.814	50.631	44.709	39.108	33.043	30.547	25.064
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.060	2.032	2.048	2.006	2.025	2.102	1.976	1.811	1.962	1.872
P-value	[<.895]	[<.875]	[<.885]	[<.851]	[<.865]	[<.916]	[<.819]	[<.614]	[<.800]	[<.689]
Shapiro-Wilk test	0.969	0.978	0.975	0.980	0.968	0.976	0.987	0.984	0.976	0.976
P-value	[.239]	[.501]	[.379]	[.524]	[.180]	[.355]	[.810]	[.664]	[.341]	[.322]
Jarque-Bera test	1.899	1.295	1.179	1.161	1.508	1.632	0.704	1.554	1.870	0.345
P-value	[.387]	[.523]	[.555]	[.560]	[.471]	[.442]	[.703]	[.460]	[.393]	[.842]
LM het. Test	5.282	2.733	4.570	4.435	4.170	5.924	9.139	5.028	5.171	9.788
P-value	[.022]	[.098]	[.033]	[.035]	[.041]	[.015]	[.003]	[.025]	[.023]	[.002]
White het. Test	42.270	33.666	27.585	33.639	38.395	36.700	39.125	42.393	45.916	41.580
P-value	[.186]	[.532]	[.810]	[.534]	[.318]	[.390]	[.290]	[.182]	[.103]	[.206]

Graf II.97.a **Stabilita koeficientů vzhledem k velikosti populace (*h*)**



Tabulka II.97.d Vyloučená pozorování

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
1	1	1	1	1	1	1	1			
4	4	4	4	4	4	4	4	4		4
7	7	7	7	7	7	7	7	7	7	7
9	9	9								
12	12	12	12	12	12					
13	13	13	13	13	13	13	13	13	13	
16	16	16	16	16						
23	23	23	23	23	23	23	23	23	23	23
25	25	25	25	25	25	25	25	25	25	25
26	26									
34										
36	36	36	36	36	36	36				
49	49	49	49	49	49	49	49	49	49	

Rok 1998

Tabulka II.98.a Odhady modelů

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
<i>Intercept</i>	12.941	12.701	12.847	12.807	12.390	12.771	11.471	10.134	11.562	9.246
<i>log(PM)</i>	-0.398	-0.452	-0.508	-0.526	-0.520	-0.588	-0.282	-0.265	-0.272	-0.151
<i>log(PX)</i>	0.548	0.606	0.653	0.670	0.675	0.707	0.501	0.466	0.466	0.382
<i>log(VA)</i>	0.802	0.817	0.831	0.849	0.823	0.818	0.852	0.891	0.845	0.853
<i>log(KL)</i>	-0.662	-0.648	-0.670	-0.685	-0.642	-0.663	-0.576	-0.400	-0.563	-0.273
<i>log(De/VA)</i>	-0.018	0.003	0.027	0.060	0.101	0.101	0.146	0.274	0.151	0.305
<i>log(FDI)</i>	0.243	0.262	0.260	0.260	0.267	0.275	0.192	0.142	0.179	0.097
<i>log(Tar)</i>	-0.849	-0.851	-0.848	-0.855	-0.853	-0.840	-0.777	-0.720	-0.745	-0.663

Tabulka II.98.b P - values

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.002]	[.001]	[.001]	[.001]	[.001]	[.001]	[.066]	[.089]	[.096]	[.331]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.001]	[.001]	[.007]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(KL)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.006]	[.000]	[.067]
<i>log(De/VA)</i>	[.751]	[.965]	[.691]	[.380]	[.155]	[.182]	[.057]	[.006]	[.075]	[.005]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.002]	[.028]	[.007]	[.158]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

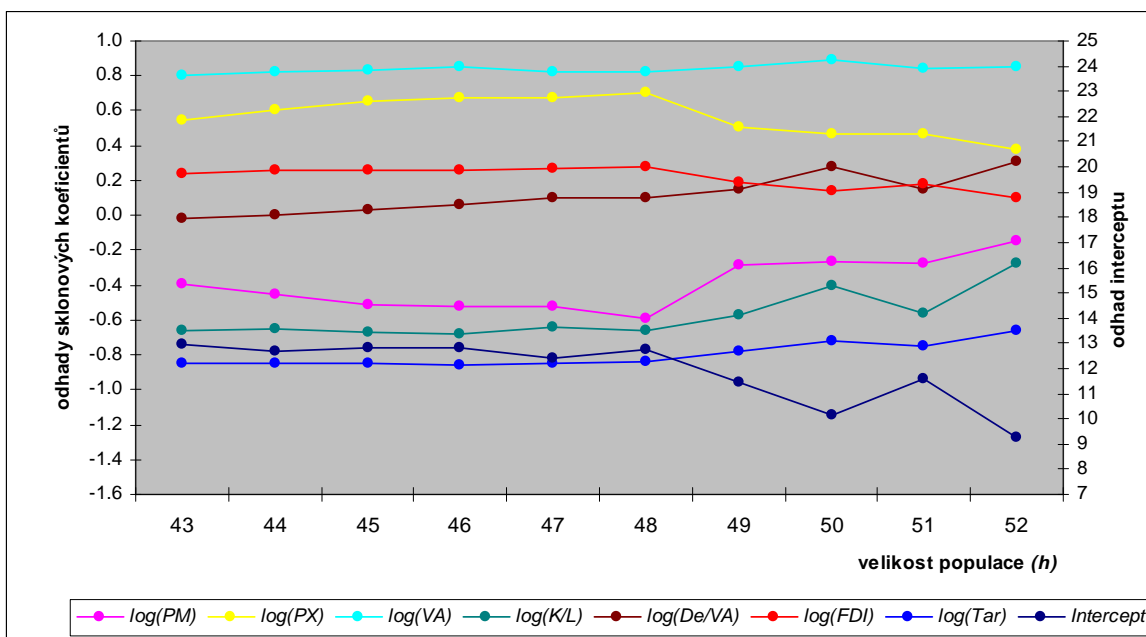
Tabulka II.98.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
R-squared	0.960	0.955	0.949	0.947	0.939	0.929	0.921	0.909	0.898	0.887
Adjusted R-squared	0.952	0.946	0.939	0.937	0.928	0.917	0.907	0.894	0.881	0.869
$\sum r_{(h,m)}^2(\beta)$	3.030	3.742	4.393	4.970	5.753	6.652	7.616	8.643	9.825	11.196
$r_{(h)}^2(\beta)$	0.414	0.588	0.581	0.512	0.675	0.795	0.700	0.836	1.014	1.120
Variance of residuals	0.087	0.104	0.119	0.131	0.148	0.166	0.186	0.206	0.228	0.254
SE error of regression	0.294	0.322	0.345	0.362	0.384	0.408	0.431	0.454	0.478	0.504
F (zero slopes)	119.076	109.649	98.188	96.816	85.301	74.966	67.863	60.239	53.852	49.437
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.450	2.346	2.223	2.006	1.962	1.858	2.056	2.081	2.032	1.966
P-value	[<.995]	[<.987]	[<.966]	[<.861]	[<.823]	[<.711]	[<.892]	[<.906]	[<.872]	[<.814]

Další charakteristiky modelů – pokračování

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
Shapiro-Wilk test	0.945	0.975	0.982	0.989	0.988	0.994	0.982	0.978	0.979	0.986
P-value	[.039]	[.446]	[.699]	[.932]	[.906]	[.997]	[.640]	[.453]	[.477]	[.782]
Jarque-Bera test	2.640	0.127	0.115	0.512	0.650	0.137	1.079	1.402	0.286	0.660
P-value	[.267]	[.938]	[.944]	[.774]	[.723]	[.934]	[.583]	[.496]	[.867]	[.719]
LM het. Test	1.895	0.261	0.740	0.063	0.099	0.003	0.042	0.168	0.195	1.217
P-value	[.169]	[.610]	[.390]	[.801]	[.753]	[.987]	[.838]	[.682]	[.659]	[.270]
White het. Test	38.824	34.817	38.551	37.576	37.831	21.883	39.970	22.585	23.246	28.778
P-value	[.301]	[.477]	[.312]	[.352]	[.341]	[.959]	[.259]	[.948]	[.936]	[.762]

Graf II.98.a Stabilita koeficientů vzhledem k velikosti populace (*h*)



Tabulka II.98.d

Vyloučená pozorování

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
	1	1	1	1	1	1	1	1	1	1
	7	7	7	7	7	7	7	7	7	7
	9	9	9	9	9	9	9	9	9	9
	13	13	13	13	13	13	13	13	13	13
	16	16	16	16	16	16				
	23	23	23	23	23	23	23	23	23	23
	24	24	24	24	24	24	24	24	24	24
	25	25	25	25	25	25	25	25	25	25
	26	26	26							
	29	29	29	29						
	30	30	30	30	30	30	30	30	30	30
	32	32	32	32	32	32	32	32	32	
	34	34	34	34	34	34	34			
	36	36	36	36	36	36	36	36	36	
								49		49
	52									
	53	53	53	53	53	53				
	55	55								
	59	59	59	59	59		59			

Rok 1999

Tabulka II.99.a Odhady modelů

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	13.756	13.637	13.403	12.970	12.593	12.268	11.380	11.184	10.740	11.146
<i>log(PM)</i>	-0.525	-0.542	-0.581	-0.495	-0.478	-0.518	-0.341	-0.384	-0.202	-0.182
<i>log(PX)</i>	0.534	0.554	0.613	0.532	0.524	0.588	0.468	0.520	0.362	0.332
<i>log(VA)</i>	0.909	0.935	0.927	0.898	0.868	0.857	0.817	0.839	0.821	0.775
<i>log(K/L)</i>	-0.793	-0.808	-0.795	-0.740	-0.703	-0.682	-0.585	-0.566	-0.542	-0.518
<i>log(De/VA)</i>	0.088	0.120	0.093	0.232	0.275	0.252	0.302	0.311	0.300	0.207
<i>log(FDI)</i>	0.207	0.215	0.238	0.179	0.187	0.210	0.155	0.135	0.147	0.126
<i>log(Tar)</i>	-0.700	-0.727	-0.743	-0.729	-0.723	-0.741	-0.680	-0.725	-0.807	-0.785

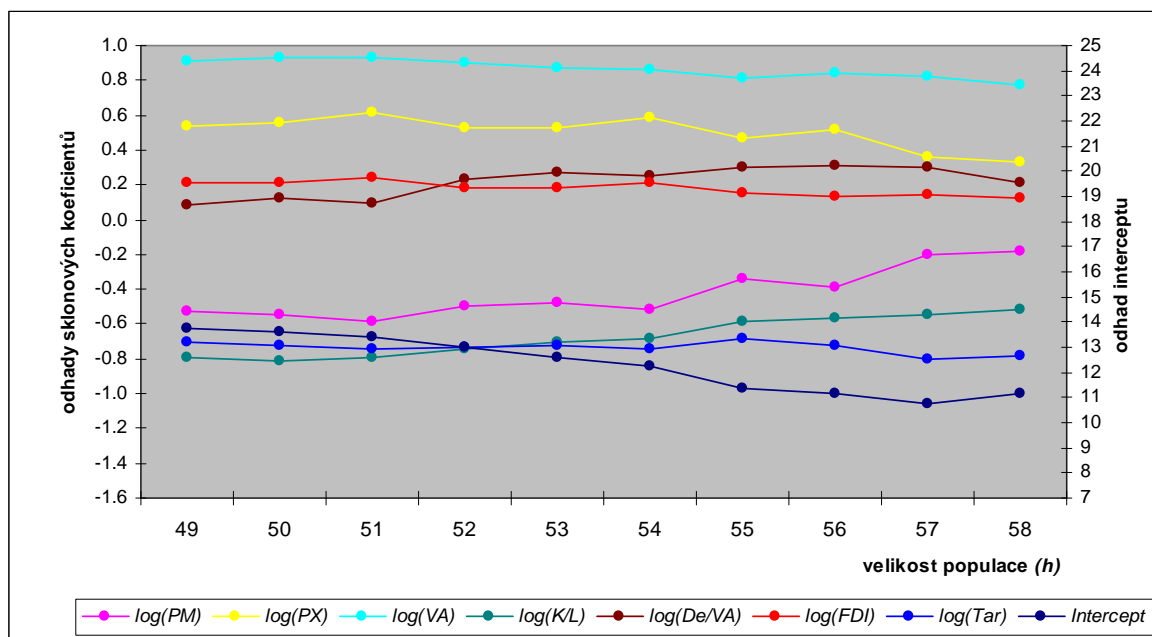
Tabulka II.99.b P - values

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.002]	[.003]	[.002]	[.041]	[.032]	[.257]	[.366]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.003]	[.002]	[.030]	[.076]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.003]
<i>log(De/VA)</i>	[.309]	[.176]	[.312]	[.012]	[.003]	[.009]	[.004]	[.005]	[.011]	[.107]
<i>log(FDI)</i>	[.003]	[.003]	[.001]	[.017]	[.016]	[.009]	[.063]	[.129]	[.127]	[.243]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka II.99.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
R-squared	0.913	0.920	0.913	0.903	0.895	0.886	0.867	0.849	0.825	0.771
Adjusted R-squared	0.898	0.906	0.898	0.888	0.878	0.868	0.847	0.827	0.800	0.739
$\sum r_{(h)}^2(\beta)$	8.150	9.014	10.153	11.379	12.423	13.909	16.645	19.816	23.490	30.779
$r_{(h)}^2(\beta)$	0.837	0.724	0.983	0.846	0.918	1.291	1.847	2.942	2.867	6.339
Variance of residuals	0.199	0.215	0.236	0.259	0.276	0.302	0.354	0.413	0.479	0.616
SE error of regression	0.446	0.463	0.486	0.509	0.525	0.550	0.595	0.643	0.692	0.785
F (zero slopes)	61.087	68.853	64.118	58.779	54.607	50.874	43.731	38.522	33.028	24.048
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.031	1.901	1.818	1.611	1.700	1.700	1.617	1.499	1.463	1.562
P-value	[<.874]	[<.752]	[<.645]	[<.349]	[<.466]	[<.458]	[<.334]	[<.187]	[<.147]	[<.243]
Shapiro-Wilk test	0.949	0.951	0.967	0.957	0.958	0.962	0.977	0.974	0.964	0.957
P-value	[.032]	[.037]	[.171]	[.056]	[.060]	[.084]	[.382]	[.255]	[.084]	[.039]
Jarque-Bera test	1.157	1.667	1.775	0.921	1.097	1.322	0.552	2.244	4.070	6.965
P-value	[.561]	[.435]	[.412]	[.631]	[.578]	[.516]	[.759]	[.326]	[.131]	[.031]
LM het. Test	0.084	1.633	3.273	1.431	1.573	2.168	3.228	2.710	1.090	4.416
P-value	[.772]	[.201]	[.070]	[.232]	[.210]	[.141]	[.072]	[.100]	[.296]	[.036]
White het. Test	36.635	37.102	39.152	40.686	41.929	45.465	48.728	37.179	39.989	36.829
P-value	[.393]	[.372]	[.289]	[.234]	[.196]	[.111]	[.061]	[.369]	[.258]	[.384]

Graf II.99.a Stabilita koeficientů vzhledem k velikosti populace (h)



Tabulka II.99.d

Vyloučená pozorování

Velikost populace (h)	49	50	51	52	53	54	55	56	57	58
1	1	1	1	1	1	1	1	1		
7	7	7	7	7	7	7	7	7	7	7
8	8	8		8	8					
9	9	9	9	9	9	9	9			
12	12	12	12							
13	13	13	13	13	13	13	13	13	13	
23	23	23	23	23	23	23	23	23	23	23
24										
25	25	25	25	25	25	25	25	25	25	25
29	29	29	29	29						
36	36	36	36	36	36	36				
53	53	53	53							

Příloha III. Výpočet modelů metodou RLS

Rok 1993

Tabulka III.93.a Odhady modelů

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
<i>Intercept</i>	14.055	14.325	14.461	14.480	14.263	14.232	23.157	22.359	23.008	22.598
<i>log(PM)</i>	-0.606	-0.603	-0.596	-0.582	-0.441	-0.481	-0.785	-0.722	-0.706	-0.647
<i>log(PX)</i>	0.396	0.383	0.377	0.373	0.251	0.277	0.145	0.120	0.094	0.043
<i>log(VA)</i>	0.385	0.374	0.372	0.381	0.384	0.392	0.695	0.614	0.598	0.595
<i>log(K/L)</i>	-0.582	-0.617	-0.636	-0.637	-0.609	-0.602	-1.430	-1.394	-1.436	-1.381
<i>log(De/VA)</i>	0.621	0.628	0.614	0.589	0.464	0.488	0.818	0.895	0.799	0.778
<i>log(FDI)</i>	0.552	0.563	0.550	0.542	0.468	0.473	0.302	0.369	0.484	0.458
<i>log(Tar)</i>	-0.669	-0.678	-0.648	-0.667	-0.480	-0.477	-1.435	-1.422	-1.566	-1.523

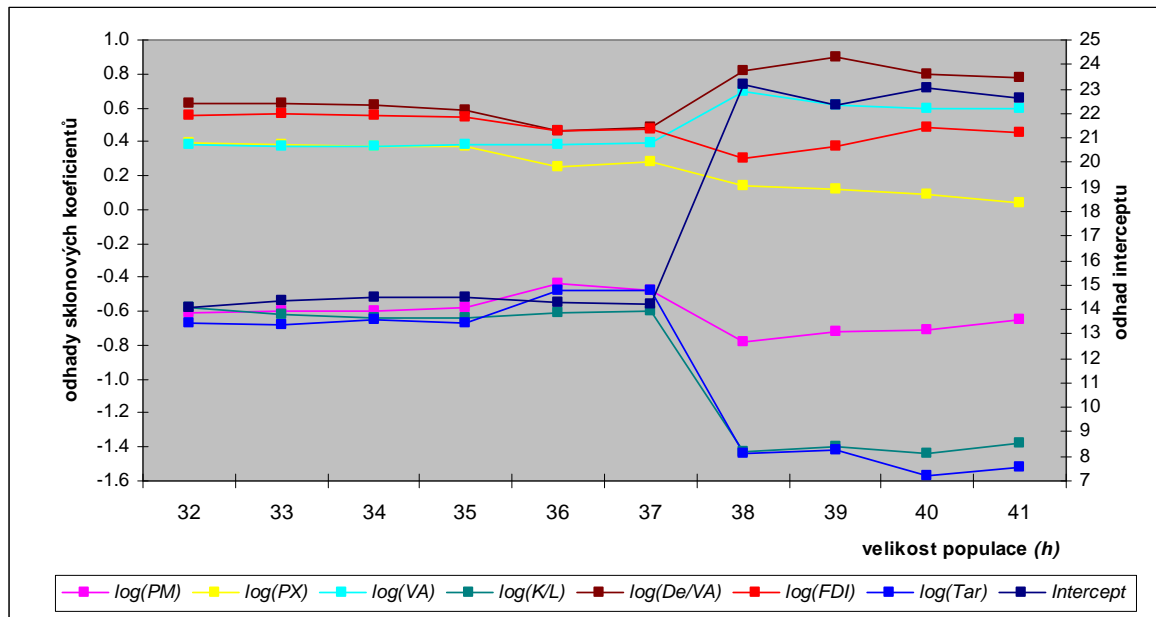
Tabulka III.93.b P - values

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.006]	[.003]	[.060]	[.123]	[.277]	[.599]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(De/VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka III.93.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
R-squared	0.977	0.976	0.975	0.972	0.961	0.959	0.972	0.970	0.962	0.960
Adjusted R-squared	0.971	0.970	0.968	0.965	0.951	0.949	0.966	0.963	0.953	0.952
$\sum r_{(lm)}^2(\beta)$	1.034	1.098	1.181	1.320	1.877	2.084	2.505	2.806	3.381	3.649
$r_{(h)}^2(\beta)$	0.121	0.125	0.124	0.155	0.240	0.265	0.209	0.206	0.332	0.359
Variance of residuals	0.043	0.044	0.045	0.049	0.067	0.072	0.083	0.091	0.106	0.111
SE error of regression	0.208	0.210	0.213	0.221	0.259	0.268	0.289	0.301	0.325	0.333
F (zero slopes)	147.877	147.820	144.755	134.171	98.413	97.018	151.003	143.769	114.520	114.477
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.171	2.407	2.358	2.313	2.198	2.009	1.650	1.497	2.225	2.033
P-value	[<.957]	[<.992]	[<.988]	[<.983]	[<.962]	[<.883]	[<.528]	[<.333]	[<.967]	[<.889]
Shapiro-Wilk test	0.946	0.937	0.958	0.974	0.953	1.230	0.939	0.947	0.956	0.960
P-value	[.111]	[.057]	[.219]	[.556]	[.129]	[.541]	[.039]	[.066]	[.123]	[.162]
Jarque-Bera test	2.055	2.291	1.867	1.292	1.432	0.966	2.531	2.239	1.951	1.775
P-value	[.358]	[.318]	[.393]	[.524]	[.489]	[.301]	[.282]	[.326]	[.377]	[.412]
LM het. Test	0.549	0.438	0.042	0.016	3.859	3.248	0.520	3.207	0.582	1.196
P-value	[.459]	[.508]	[.838]	[.899]	[.049]	[.072]	[.471]	[.073]	[.446]	[.274]
White het. Test	n/a	n/a	n/a	n/a	n/a	37.000	35.659	36.340	35.729	36.635
P-value	n/a	n/a	n/a	n/a	n/a	[.377]	[.437]	[.406]	[.434]	[.393]

Graf III.93.a Stabilita koeficientů vzhledem k velikosti populace (h)



Tabulka III.93.d Vyloučená pozorování

Velikost populace (h)	32	33	34	35	36	37	38	39	40	41
1	1	1	1	1	1	1				
2							2	2	2	2
3							3	3		
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5							
6	6	6	6	6	6	6				
7	7	7	7	7	7	7	7	7	7	7
9	9	9	9	9	9	9	9	9	9	9
10	10	10	10	10	10	10				
11	11	11	11	11	11	11				
12	12	12	12	12	12	12	12	12	12	12
15							15	15	15	15
16	16	16	16	16	16	16	16	16	16	16
18							18	18	18	18
19	19	19	19	19			19	19	19	19
21	21	21	21	21	21	21	21			
22							22	22	22	22
23	23	23	23	23	23	23	23	23	23	23
24	24	24	24	24	24	24				
25	25	25	25	25	25	25				
26	26	26	26	26	26	26			26	26
27	27	27	27	27	27	27	27	27	27	27
30	30	30	30	30	30	30				
33	33	33	33	33	33	33	33	33	33	33
34	34	34	34	34	34	34	34	34	34	34
35										
36	36	36	36	36	36	36	36	36	36	36
37							37	37		
38	38	38	38	38	38	38				

Vyloučená pozorování – pokračování

Velikost populace (<i>h</i>)	32	33	34	35	36	37	38	39	40	41
	42	42	42	42	42	42				
	44	44	44	44	44	44				
	49	49	49	49	49	49	49	49	49	49
					50					
	51	51	51	51					51	
							52	52	52	52
							53	53		
							54	54		
	56	56								
							57	57	57	57
	58	58	58	58	58	58			58	58

Rok 1994

Tabulka III.94.a Odhady modelů

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	17.314	16.437	16.271	14.853	16.123	15.618	14.885	14.905	15.438	15.301
<i>log(PM)</i>	-0.502	-0.063	-0.091	-0.074	-0.161	-0.139	-0.108	-0.112	0.042	0.048
<i>log(PX)</i>	0.262	-0.135	-0.092	-0.089	-0.056	-0.063	-0.035	-0.051	-0.141	-0.148
<i>log(VA)</i>	0.519	0.437	0.458	0.663	0.628	0.631	0.453	0.648	0.591	0.598
<i>log(K/L)</i>	-1.037	-0.980	-0.962	-0.663	-0.806	-0.734	-0.758	-0.657	-0.755	-0.735
<i>log(De/VA)</i>	0.574	0.504	0.509	0.334	0.348	0.337	0.498	0.320	0.201	0.194
<i>log(FDI)</i>	0.423	0.384	0.383	0.235	0.263	0.264	0.391	0.234	0.198	0.183
<i>log(Tar)</i>	-0.925	-0.682	-0.748	-0.749	-0.746	-0.756	-0.804	-0.768	-0.722	-0.703

Tabulka III.94.b P - values

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.007]	[.669]	[.544]	[.645]	[.310]	[.393]	[.533]	[.521]	[.805]	[.783]
<i>log(PX)</i>	[.113]	[.349]	[.527]	[.572]	[.716]	[.696]	[.837]	[.765]	[.432]	[.420]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.000]	[.000]	[.002]	[.000]	[.001]	[.000]	[.003]	[.001]	[.002]
<i>log(De/VA)</i>	[.000]	[.000]	[.000]	[.007]	[.005]	[.008]	[.001]	[.018]	[.108]	[.127]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.003]	[.001]	[.001]	[.000]	[.005]	[.011]	[.021]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

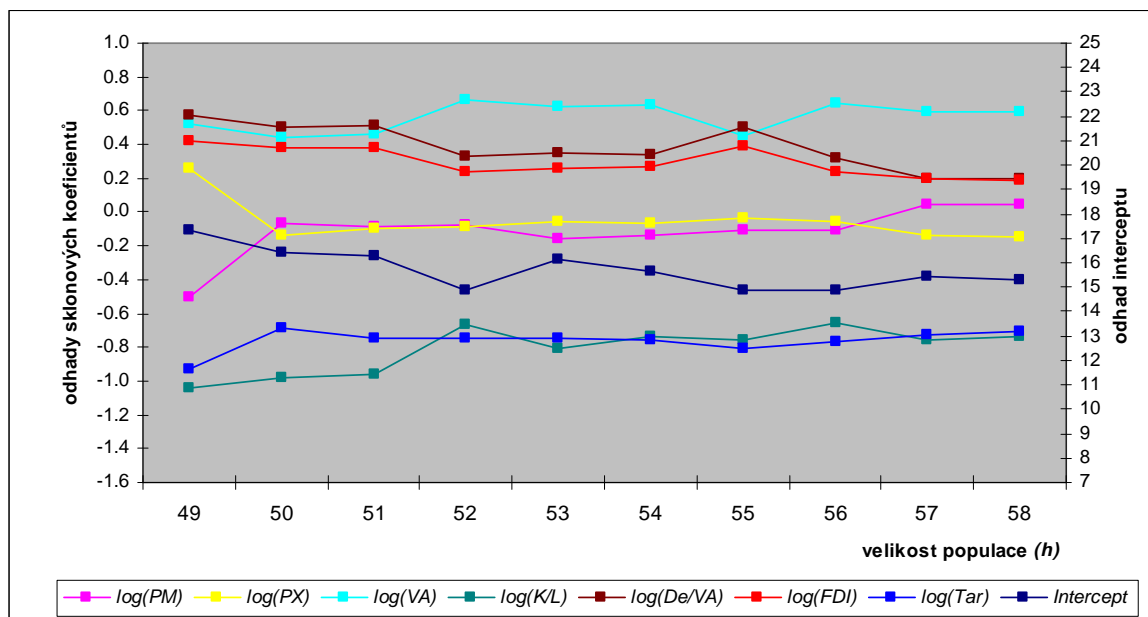
Tabulka III.94.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
R-squared	0.851	0.841	0.837	0.809	0.809	0.793	0.768	0.758	0.714	0.698
Adjusted R-squared	0.825	0.814	0.810	0.779	0.779	0.762	0.733	0.723	0.673	0.656
$\sum r_{(lm)}^2(\beta)$	15.211	14.848	16.032	19.083	19.182	20.973	23.876	25.324	30.036	31.954
$r_{(h)}^2(\beta)$	2.580	1.193	1.293	1.608	1.810	1.757	2.031	2.078	2.572	2.474
Variance of residuals	0.371	0.354	0.373	0.434	0.426	0.456	0.508	0.528	0.613	0.639
SE error of regression	0.609	0.595	0.611	0.659	0.653	0.675	0.713	0.726	0.783	0.799
F (zero slopes)	33.377	31.636	31.451	26.670	27.261	25.212	22.220	21.514	17.479	16.515
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	1.909	1.542	1.689	1.714	1.977	1.858	1.746	1.997	2.036	1.999
P-value	[<.765]	[<.278]	[<.465]	[<.493]	[<.822]	[<.683]	[<.520]	[<.834]	[<.867]	[<.833]

Další charakteristiky modelů – pokračování

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
Shapiro-Wilk test	0.957	0.971	0.970	0.976	0.985	0.982	0.976	0.976	0.981	0.980
P-value	[.071]	[.251]	[.231]	[.359]	[.724]	[.581]	[.327]	[.317]	[.505]	[.444]
Jarque-Bera test	4.450	2.013	1.921	1.939	1.328	1.617	1.474	1.700	1.303	1.361
P-value	[.108]	[.366]	[.383]	[.379]	[.515]	[.445]	[.479]	[.427]	[.521]	[.506]
LM het. Test	0.000	2.665	2.127	2.187	2.829	2.930	1.464	1.297	4.716	3.691
P-value	[.987]	[.103]	[.145]	[.139]	[.093]	[.087]	[.226]	[.255]	[.030]	[.055]
White het. Test	31.740	37.578	37.977	37.572	38.164	38.807	33.975	32.823	42.178	40.051
P-value	[.626]	[.352]	[.335]	[.352]	[.328]	[.302]	[.517]	[.574]	[.188]	[.256]

Graf III.94.a Stabilita koeficientů vzhledem k velikosti populace (*h*)



Tabulka III.94.d Vyloučená pozorování

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
	4	4	4	4	4	4				
		6								
	7	7	7	7	7	7	7	7	7	7
	9	9	9	9	9	9	9	9	9	9
	12									
	21	21	21				21			
				22	22	22		22		
	23	23	23	23	23	23	23	23	23	23
	27	27	27							
	32	32	32	32						
	33	33	33		33					
	34	34	34	34	34	34	34		34	
				35						
	36									
	49	49	49	49	49	49	49	49		

Tabulka III.95.a Odhady modelů

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	14.591	15.396	14.273	14.726	13.636	15.825	14.682	14.813	14.820	15.484
<i>log(PM)</i>	-0.749	-0.770	-0.692	-0.729	-0.648	-0.539	-0.445	0.069	-0.461	-0.246
<i>log(PX)</i>	0.646	0.662	0.611	0.642	0.611	0.355	0.314	-0.088	0.360	0.209
<i>log(VA)</i>	0.713	0.659	0.717	0.667	0.664	0.709	0.707	0.725	0.664	0.675
<i>log(K/L)</i>	-0.627	-0.750	-0.575	-0.660	-0.548	-0.758	-0.640	-0.716	-0.715	-0.778
<i>log(De/VA)</i>	0.278	0.272	0.222	0.255	0.290	0.438	0.476	0.152	0.445	0.232
<i>log(FDI)</i>	0.256	0.311	0.254	0.296	0.234	0.297	0.228	0.122	0.245	0.142
<i>log(Tar)</i>	-0.635	-0.644	-0.630	-0.625	-0.553	-1.056	-0.983	-0.686	-0.846	-0.788

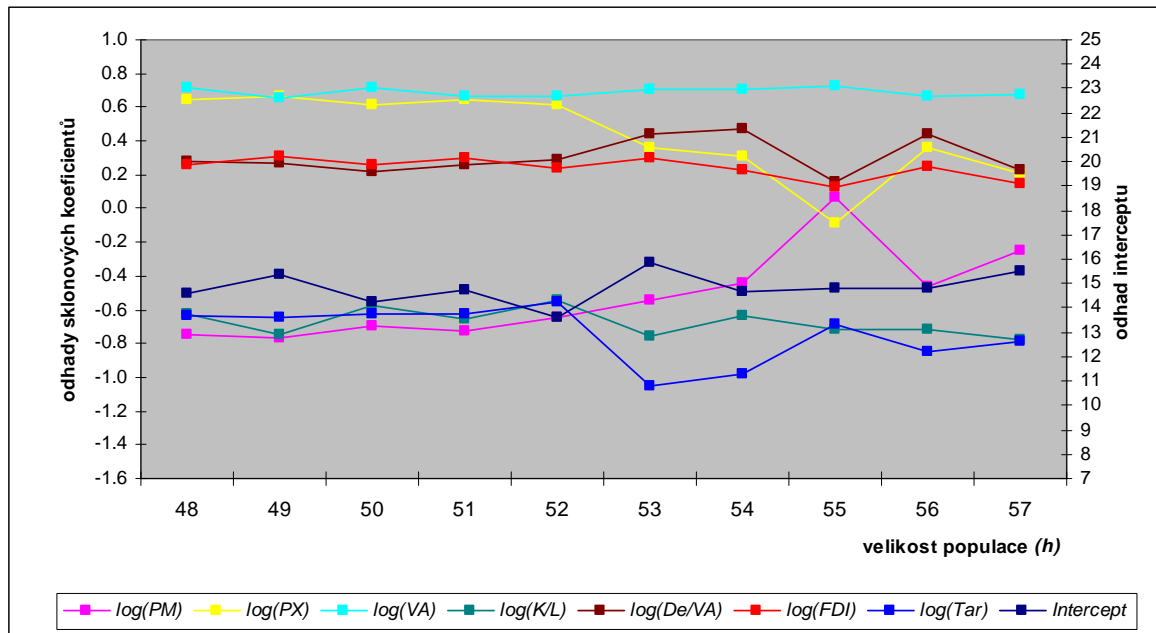
Tabulka III.95.b P - values

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.003]	[.017]	[.721]	[.021]	[.232]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.001]	[.036]	[.082]	[.656]	[.064]	[.314]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.000]	[.000]	[.000]	[.002]	[.000]	[.002]	[.002]	[.001]	[.001]
<i>log(De/VA)</i>	[.004]	[.005]	[.023]	[.014]	[.012]	[.001]	[.000]	[.226]	[.002]	[.080]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.004]	[.125]	[.002]	[.075]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka III.95.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
R-squared	0.895	0.883	0.874	0.856	0.818	0.838	0.807	0.732	0.768	0.709
Adjusted R-squared	0.877	0.863	0.853	0.832	0.789	0.813	0.777	0.692	0.734	0.668
$\sum r_{(h,m)}^2(\beta)$	7.564	8.690	9.163	10.748	13.566	16.940	20.226	27.501	24.596	30.848
$r_{(h)}^2(\beta)$	0.755	1.030	0.882	1.261	2.209	1.469	2.577	2.494	2.564	3.676
Variance of residuals	0.189	0.212	0.218	0.250	0.308	0.376	0.440	0.585	0.512	0.630
SE error of regression	0.435	0.460	0.467	0.500	0.555	0.614	0.663	0.765	0.716	0.793
F (zero slopes)	48.770	44.044	41.640	36.466	28.267	33.281	27.425	18.333	22.704	17.068
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.278	2.020	2.124	1.942	1.779	1.871	1.802	1.699	1.621	1.632
P-value	[<.978]	[<.866]	[<.929]	[<.793]	[<.586]	[<.705]	[<.607]	[<.451]	[<.332]	[<.341]
Shapiro-Wilk test	0.974	0.977	0.980	0.985	0.985	0.976	0.988	0.987	0.980	0.986
P-value	[.367]	[.453]	[.535]	[.766]	[.774]	[.361]	[.847]	[.806]	[.455]	[.743]
Jarque-Bera test	1.521	1.001	0.666	0.637	0.557	1.174	0.055	0.350	0.378	0.125
P-value	[.467]	[.606]	[.717]	[.727]	[.757]	[.556]	[.973]	[.840]	[.828]	[.939]
LM het. Test	1.930	3.047	1.567	1.510	3.691	3.971	5.716	8.609	3.982	7.994
P-value	[.165]	[.081]	[.211]	[.219]	[.055]	[.046]	[.017]	[.003]	[.046]	[.005]
White het. Test	28.138	43.064	34.426	37.870	38.681	39.205	44.147	47.098	43.252	46.952
P-value	[.788]	[.164]	[.496]	[.340]	[.307]	[.287]	[.138]	[.083]	[.160]	[.085]

Graf III.95.a Stabilita koeficientů vzhledem k velikosti populace (h)



Tabulka III.95.d Vyloučená pozorování

Velikost populace (h)	48	49	50	51	52	53	54	55	56	57
	1	1	1	1	1					
	7	7	7	7	7	7	7	7	7	7
	9	9	9	9	9			9		
	11	11	11	11	11					
	12	12	12	12	12	12	12		12	12
						19	19			
	22		22			22	22	22		
	23	23	23	23	23	23	23	23	23	23
	24	24	24	24	24	24	24	24	24	24
	25	25	25	25	25					
								32		
	33	33								
	36	34								
		36	36	36		36				
	37									
	49	49	49	49	49	49	49		49	

Rok 1996

Tabulka III.96.a Odhady modelů

Velikost populace (h)	44	45	46	47	48	49	50	51	52	53
Intercept	15.405	16.725	15.387	16.031	16.546	13.706	12.293	13.088	12.728	12.928
$\log(PM)$	-0.712	-0.473	-0.604	-0.602	-0.632	-0.032	-0.080	-0.076	-0.119	-0.151
$\log(PX)$	0.589	0.327	0.494	0.453	0.466	0.018	0.122	0.070	0.125	0.159
$\log(VA)$	0.560	0.649	0.499	0.491	0.569	0.641	0.811	0.795	0.816	0.825
$\log(K/L)$	-0.772	-0.961	-0.746	-0.813	-0.834	-0.764	-0.506	-0.592	-0.554	-0.655
$\log(De/VA)$	0.277	0.236	0.242	0.251	0.185	0.404	0.270	0.291	0.315	0.404
$\log(FDI)$	0.405	0.274	0.377	0.378	0.319	0.292	0.176	0.183	0.172	0.201
$\log(Tar)$	-0.733	-0.567	-0.670	-0.641	-0.637	-0.717	-0.774	-0.750	-0.787	-0.802

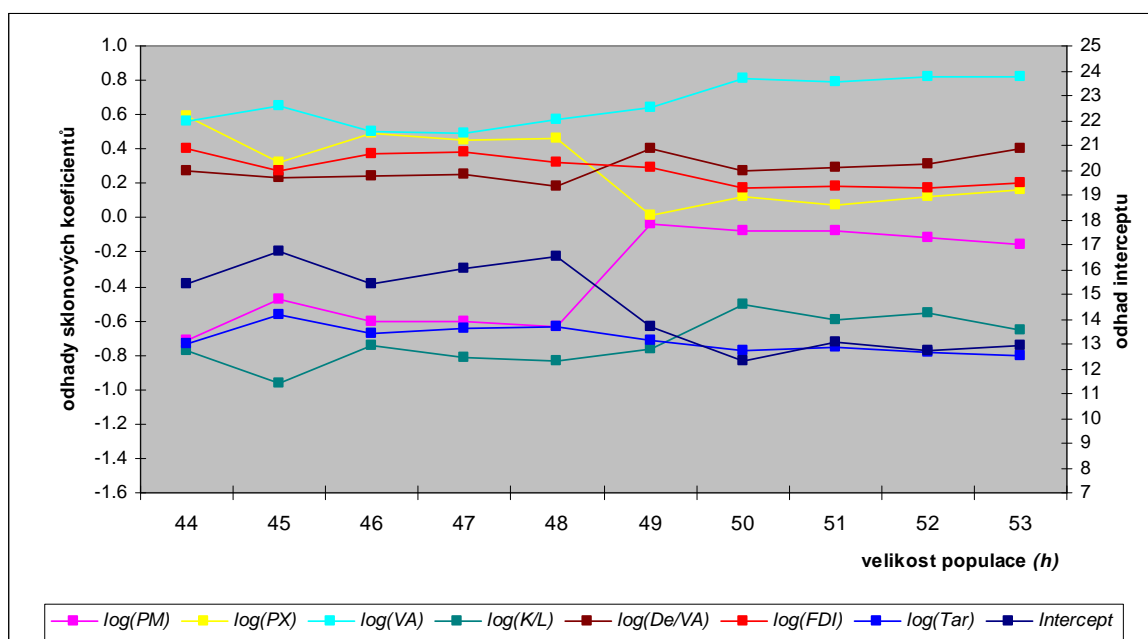
Tabulka III.96.b P - values

Velikost populace (<i>h</i>)	44	45	46	47	48	49	50	51	52	53
Intercept	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
log(PM)	[.000]	[.001]	[.000]	[.000]	[.000]	[.854]	[.647]	[.668]	[.516]	[.433]
log(PX)	[.000]	[.018]	[.003]	[.005]	[.005]	[.914]	[.496]	[.696]	[.499]	[.411]
log(VA)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
log(K/L)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.006]	[.001]	[.003]	[.001]
log(De/VA)	[.003]	[.006]	[.017]	[.015]	[.050]	[.002]	[.024]	[.016]	[.012]	[.002]
log(FDI)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.009]	[.007]	[.014]	[.005]
log(Tar)	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka III.96.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	44	45	46	47	48	49	50	51	52	53
R-squared	0.921	0.921	0.895	0.890	0.884	0.855	0.844	0.835	0.826	0.843
Adjusted R-squared	0.906	0.906	0.875	0.870	0.863	0.831	0.819	0.809	0.798	0.819
$\sum r_{(km)}^2(\beta)$	5.060	5.369	7.201	7.611	8.128	12.742	13.881	14.737	16.250	18.259
$r_{(h)}^2(\beta)$	0.664	0.734	0.814	0.854	0.889	1.437	1.172	1.391	1.381	1.534
Variance of residuals	0.141	0.145	0.190	0.195	0.203	0.311	0.331	0.343	0.369	0.406
SE error of regression	0.375	0.381	0.435	0.442	0.451	0.557	0.575	0.585	0.608	0.637
F (zero slopes)	60.107	61.566	46.182	44.924	43.449	34.672	32.573	31.196	29.824	34.543
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.189	2.181	2.263	2.124	2.405	1.977	2.496	2.407	2.297	1.854
P-value	[<.957]	[<.954]	[<.975]	[<.931]	[<.993]	[<.831]	[<.997]	[<.993]	[<.981]	[<.683]
Shapiro-Wilk test	0.966	0.973	0.950	0.947	0.971	0.969	0.953	0.960	0.949	0.953
P-value	[.213]	[.371]	[.047]	[.033]	[.274]	[.217]	[.047]	[.084]	[.028]	[.036]
Jarque-Bera test	1.934	1.399	2.903	3.055	1.833	2.014	2.733	2.555	3.201	2.970
P-value	[.380]	[.497]	[.234]	[.217]	[.400]	[.365]	[.255]	[.279]	[.202]	[.227]
LM het. Test	0.537	3.828	9.453	8.886	9.502	3.112	2.732	2.912	2.435	6.154
P-value	[.464]	[.050]	[.002]	[.003]	[.002]	[.078]	[.098]	[.088]	[.119]	[.013]
White het. Test	36.663	36.190	39.154	39.884	36.654	37.637	33.011	36.031	35.265	30.603
P-value	[.392]	[.413]	[.289]	[.262]	[.392]	[.349]	[.564]	[.420]	[.456]	[.680]

Graf III.96.a Stabilita koeficientů vzhledem k velikosti populace (*h*)



Tabulka III.96.d Vyloučená pozorování

Velikost populace (<i>h</i>)	44	45	46	47	48	49	50	51	52	53
	1	1	1	1	1					
	4	4	4	4	4					
	6	6	6	6	6					
	7	7	7	7	7	7	7	7	7	7
	9	9	9	9	9	9	9	9		
	12	12	12	12	12	12	12	12	12	12
	13					13	13	13	13	13
	16	16	16	16	16					
	21	21	21	21		21				
							22	22	22	
	23	23	23	23	23	23	23	23	23	23
	24	24	24	24	24	24	24	24	24	
	25	25	25	25	25	25	25	25	25	25
	26	26	26	26	26					
						27				
	30	30	30				30			
						32	32	32	32	32
						33				33
	34	34	34	34	34					
	36									
	49	49	49	49	49	49	49	49	49	49

Rok 1997

Tabulka III.97.a Odhady modelů

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	14.377	13.943	12.450	12.183	12.033	12.189	9.290	10.411	10.502	12.097
<i>log(PM)</i>	-0.575	-0.895	-0.637	-0.663	-0.410	-0.475	-0.116	-0.176	-0.239	-0.197
<i>log(PX)</i>	0.545	0.830	0.658	0.684	0.456	0.498	0.317	0.307	0.347	0.328
<i>log(VA)</i>	0.870	0.805	0.822	0.847	0.850	0.839	0.897	0.883	0.871	0.878
<i>log(KL)</i>	-0.780	-0.656	-0.507	-0.496	-0.549	-0.536	-0.296	-0.409	-0.382	-0.545
<i>log(De/VA)</i>	0.170	0.272	0.300	0.352	0.340	0.333	0.347	0.369	0.363	0.174
<i>log(FDI)</i>	0.231	0.266	0.142	0.138	0.174	0.177	0.127	0.136	0.138	0.083
<i>log(Tar)</i>	-0.670	-0.798	-0.638	-0.646	-0.610	-0.603	-0.758	-0.732	-0.730	-0.729

Tabulka III.97.b P - values

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.000]	[.000]	[.000]	[.015]	[.006]	[.498]	[.321]	[.185]	[.304]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.004]	[.002]	[.052]	[.070]	[.046]	[.076]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(KL)</i>	[.000]	[.000]	[.001]	[.002]	[.001]	[.001]	[.103]	[.026]	[.042]	[.005]
<i>log(De/VA)</i>	[.086]	[.006]	[.005]	[.001]	[.003]	[.004]	[.007]	[.005]	[.008]	[.149]
<i>log(FDI)</i>	[.000]	[.000]	[.023]	[.032]	[.008]	[.008]	[.078]	[.070]	[.074]	[.293]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

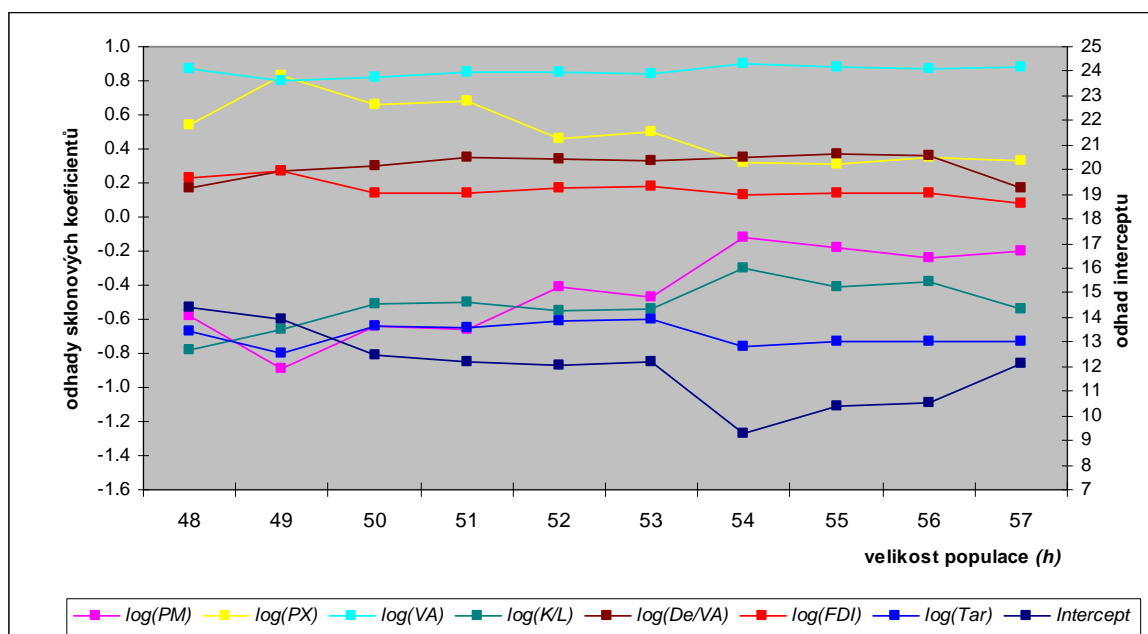
Tabulka III.97.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
R-squared	0.913	0.910	0.882	0.884	0.878	0.868	0.847	0.831	0.817	0.787
Adjusted R-squared	0.897	0.895	0.862	0.865	0.858	0.847	0.824	0.806	0.790	0.757
$\sum r_{(h,m)}^2(\beta)$	8.204	8.160	10.468	11.536	13.034	14.115	17.954	19.879	21.648	25.168
$r_{(h)}^2(\beta)$	0.530	0.898	1.255	1.309	1.379	1.493	1.557	1.775	2.104	2.135
Variance of residuals	0.205	0.199	0.249	0.268	0.296	0.314	0.390	0.423	0.451	0.514
SE error of regression	0.453	0.446	0.499	0.518	0.544	0.560	0.625	0.650	0.672	0.717

Další charakteristiky modelů – pokračování

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
F (zero slopes)	59.674	59.471	44.793	46.673	45.045	42.157	36.474	33.120	30.547	25.891
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.119	2.032	2.141	2.079	2.166	2.068	2.120	2.012	1.962	1.943
P-value	[<.928]	[<.875]	[<.937]	[<.904]	[<.946]	[<.895]	[<.925]	[<.850]	[<.800]	[<.777]
Shapiro-Wilk test	0.956	0.978	0.981	0.983	0.981	0.983	0.966	0.966	0.976	0.975
P-value	[.072]	[.501]	[.613]	[.656]	[.548]	[.631]	[.125]	[.123]	[.341]	[.293]
Jarque-Bera test	2.623	1.295	0.669	0.908	1.254	0.921	2.071	2.018	1.870	0.556
P-value	[.269]	[.523]	[.716]	[.635]	[.534]	[.631]	[.355]	[.365]	[.393]	[.757]
LM het. Test	4.940	2.733	6.689	8.744	10.262	12.167	5.145	4.014	5.171	6.145
P-value	[.026]	[.098]	[.010]	[.003]	[.001]	[.000]	[.023]	[.045]	[.023]	[.013]
White het. Test	36.515	33.666	42.989	40.078	41.315	44.094	35.603	42.194	45.916	43.602
P-value	[.398]	[.532]	[.166]	[.255]	[.214]	[.139]	[.440]	[.188]	[.103]	[.151]

Graf III.97.a Stabilita koeficientů vzhledem k velikosti populace (*h*)



Tabulka III.97.d Vyloučená pozorování

Velikost populace (<i>h</i>)	48	49	50	51	52	53	54	55	56	57
1	1	1	1	1	1	1				
4	4	4	4	4	4	4				
7	7	7	7	7	7	7	7	7	7	7
9	9	9	9	9	9	9				
12		12	12	12						
13	13	13	13	13	13	13	13	13	13	13
16										
23	23	23	23	23	23	23	23	23	23	23
25	25	25	25	25	25	25	25	25	25	25
26	26	26	26							
29										
30							30			
31			31	31						
32					32		32	32		
33										
36	36									
49	49	49	49	49	49	49	49	49	49	49

Rok 1998

Tabulka III.98.a Odhady modelů

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
<i>Intercept</i>	12.119	9.599	9.457	9.998	10.341	11.539	11.471	9.032	8.875	9.246
<i>log(PM)</i>	-0.412	-0.423	-0.384	-0.398	-0.375	-0.378	-0.282	-0.161	-0.189	-0.151
<i>log(PX)</i>	0.567	0.658	0.621	0.618	0.578	0.578	0.501	0.393	0.439	0.382
<i>log(VA)</i>	0.846	0.775	0.759	0.904	0.878	0.839	0.852	0.896	0.890	0.853
<i>log(K/L)</i>	-0.594	-0.419	-0.402	-0.374	-0.432	-0.572	-0.576	-0.215	-0.208	-0.273
<i>log(De/VA)</i>	0.089	0.346	0.357	0.253	0.260	0.155	0.146	0.314	0.294	0.305
<i>log(FDI)</i>	0.223	0.305	0.302	0.151	0.193	0.224	0.192	0.032	0.046	0.097
<i>log(Tar)</i>	-0.841	-0.832	-0.818	-0.791	-0.807	-0.829	-0.777	-0.633	-0.647	-0.663

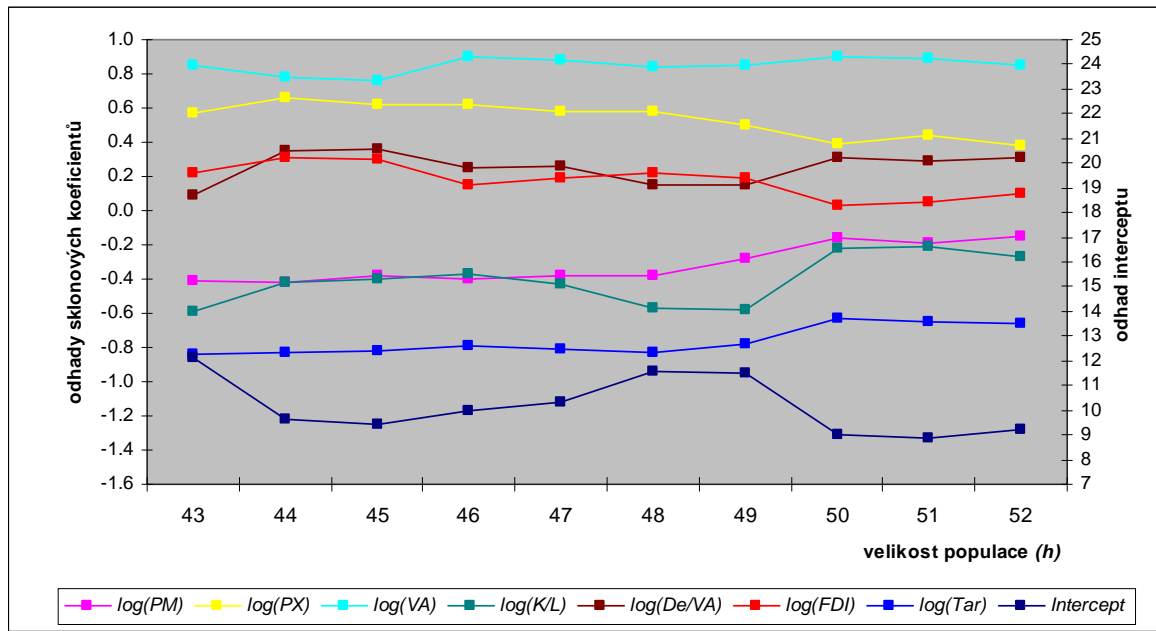
Tabulka III.98.b P - values

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.003]	[.004]	[.008]	[.006]	[.011]	[.015]	[.066]	[.285]	[.218]	[.331]
<i>log(PX)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.006]	[.002]	[.007]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.001]	[.001]	[.003]	[.001]	[.000]	[.000]	[.142]	[.163]	[.067]
<i>log(De/VA)</i>	[.250]	[.000]	[.000]	[.003]	[.003]	[.035]	[.057]	[.003]	[.006]	[.005]
<i>log(FDI)</i>	[.000]	[.000]	[.000]	[.015]	[.001]	[.000]	[.002]	[.652]	[.528]	[.158]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka III.98.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	43	44	45	46	47	48	49	50	51	52
R-squared	0.957	0.947	0.945	0.941	0.938	0.929	0.921	0.896	0.893	0.887
Adjusted R-squared	0.948	0.937	0.935	0.930	0.927	0.917	0.907	0.879	0.875	0.869
$\sum r_{(h)n}^2(\beta)$	3.546	4.822	5.132	5.427	5.869	6.713	7.616	9.777	10.446	11.196
$r_{(h)}^2(\beta)$	0.501	0.472	0.434	0.402	0.437	0.701	0.700	0.899	0.978	1.120
Variance of residuals	0.101	0.134	0.139	0.143	0.150	0.168	0.186	0.233	0.243	0.254
SE error of regression	0.318	0.366	0.372	0.378	0.388	0.410	0.431	0.482	0.493	0.504
F (zero slopes)	110.003	92.215	91.674	86.950	84.215	75.242	67.863	51.872	51.022	49.437
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.108	1.992	1.934	1.527	1.813	1.843	2.056	1.621	1.608	1.966
P-value	[<.926]	[<.855]	[<.803]	[<.295]	[<.662]	[<.692]	[<.892]	[<.378]	[<.353]	[<.814]
Shapiro-Wilk test	0.941	0.930	0.936	0.953	0.948	0.968	0.982	0.985	0.988	0.986
P-value	[.029]	[.010]	[.015]	[.059]	[.036]	[.219]	[.640]	[.757]	[.898]	[.782]
Jarque-Bera test	2.732	3.103	3.082	2.694	2.674	1.690	1.079	1.019	0.804	0.660
P-value	[.255]	[.212]	[.214]	[.260]	[.263]	[.430]	[.583]	[.601]	[.669]	[.719]
LM het. Test	3.594	0.113	0.467	0.059	0.002	0.000	0.042	0.399	0.704	1.217
P-value	[.058]	[.737]	[.494]	[.807]	[.962]	[.990]	[.838]	[.528]	[.402]	[.270]
White het. Test	37.975	37.001	37.747	35.942	36.435	40.080	39.970	25.175	27.482	28.778
P-value	[.335]	[.377]	[.345]	[.424]	[.402]	[.255]	[.259]	[.890]	[.814]	[.762]

Graf III.98.a Stabilita koeficientů vzhledem k velikosti populace (h)



Tabulka III.98.d Vyloučená pozorování

Velikost populace (h)	43	44	45	46	47	48	49	50	51	52
	1	1	1	1	1	1	1	1	1	1
	7	7	7	7	7	7	7	7	7	7
								8		
	9	9	9	9	9	9	9	9	9	9
	13	13	13	13	13	13	13	13	13	13
	16	16	16	16	16	16				
		21	21							
	23	23	23	23	23	23	23	23	23	23
	24	24	24	24	24	24	24	24	24	24
	25	25	25	25	25	25	25	25	25	25
	29									
	30	30	30	30	30	30	30	30	30	30
				31				31	31	
	32	32	32	32	32	32	32			
	34	34	34	34	34	34	34			
	36	36	36	36	36	36	36			
	49	49	49	49	49			49	49	49
	52									
	53									
	55									
	59	59	59	59	59	59	59			
		60	60							
		61								

Rok 1999

Tabulka III.99.a Vyloučená pozorování

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	13.570	8.561	10.063	12.970	12.593	12.268	11.380	11.573	10.740	11.146
<i>log(PM)</i>	-0.526	-0.059	-0.369	-0.495	-0.478	-0.518	-0.341	-0.361	-0.202	-0.182
<i>log(PX)</i>	0.551	0.368	0.493	0.532	0.524	0.588	0.468	0.468	0.362	0.332
<i>log(VA)</i>	0.939	0.778	0.938	0.898	0.868	0.857	0.817	0.858	0.821	0.775
<i>log(K/L)</i>	-0.827	-0.394	-0.424	-0.740	-0.703	-0.682	-0.585	-0.635	-0.542	-0.518
<i>log(De/VA)</i>	0.124	0.338	0.389	0.232	0.275	0.252	0.302	0.252	0.300	0.207
<i>log(FDI)</i>	0.225	0.200	0.101	0.179	0.187	0.210	0.155	0.201	0.147	0.126
<i>log(Tar)</i>	-0.736	-0.725	-0.751	-0.729	-0.723	-0.741	-0.680	-0.869	-0.807	-0.785

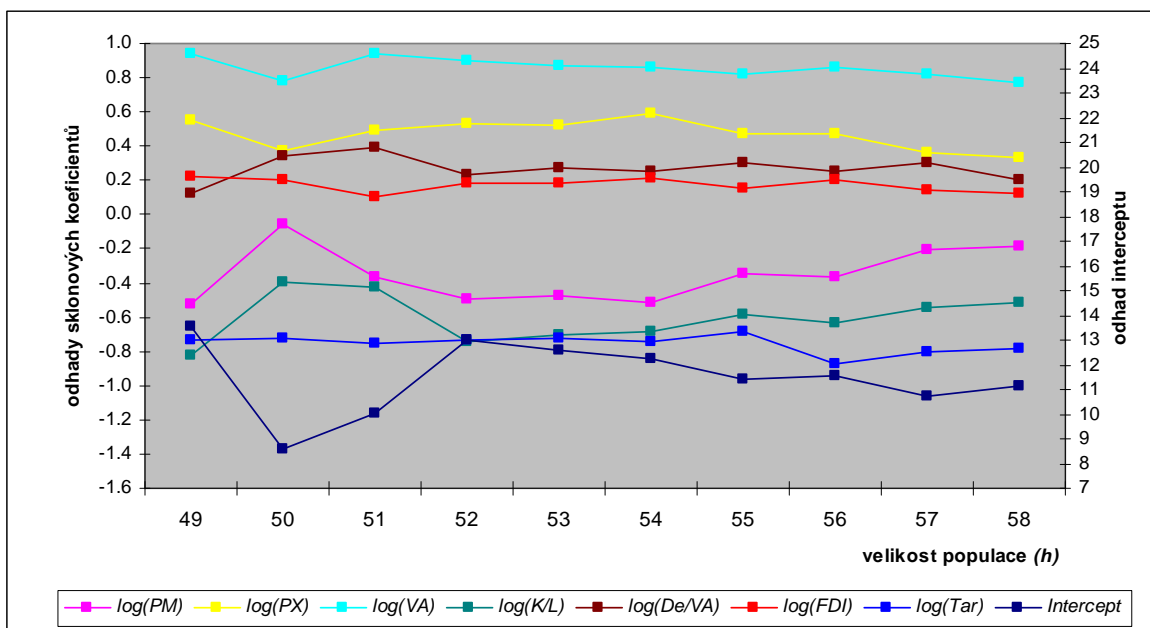
Tabulka III.99.b P - values

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
<i>Intercept</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(PM)</i>	[.000]	[.716]	[.047]	[.002]	[.003]	[.002]	[.041]	[.052]	[.257]	[.366]
<i>log(PX)</i>	[.000]	[.010]	[.002]	[.000]	[.000]	[.000]	[.003]	[.005]	[.030]	[.076]
<i>log(VA)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
<i>log(K/L)</i>	[.000]	[.002]	[.014]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.003]
<i>log(De/VA)</i>	[.150]	[.000]	[.002]	[.012]	[.003]	[.009]	[.004]	[.027]	[.011]	[.107]
<i>log(FDI)</i>	[.001]	[.008]	[.246]	[.017]	[.016]	[.009]	[.063]	[.036]	[.127]	[.243]
<i>log(Tar)</i>	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]

Tabulka III.99.c Další charakteristiky modelů

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
R-squared	0.926	0.913	0.888	0.903	0.895	0.886	0.867	0.840	0.825	0.771
Adjusted R-squared	0.913	0.898	0.871	0.888	0.878	0.868	0.847	0.817	0.800	0.739
$\sum r_{(km)}^2(\beta)$	8.239	10.626	13.757	11.379	12.423	13.909	16.645	20.967	23.490	30.779
$r_{(n)}^2(\beta)$	0.696	0.801	2.639	0.846	0.918	1.291	1.847	3.175	2.867	6.339
Variance of residuals	0.201	0.253	0.313	0.259	0.276	0.302	0.354	0.437	0.479	0.616
SE error of regression	0.448	0.503	0.559	0.509	0.525	0.550	0.595	0.661	0.692	0.785
F (zero slopes)	73.088	62.673	50.019	58.779	54.607	50.874	43.731	36.030	33.028	24.048
P-value	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
Durbin-Watson	2.093	1.768	1.834	1.611	1.700	1.700	1.617	1.486	1.463	1.562
P-value	[<.914]	[<.584]	[<.661]	[<.349]	[<.466]	[<.458]	[<.334]	[<.174]	[<.147]	[<.243]
Shapiro-Wilk test	0.949	0.973	0.954	0.957	0.958	0.962	0.977	0.951	0.964	0.957
P-value	[.033]	[.299]	[.043]	[.056]	[.060]	[.084]	[.382]	[.023]	[.084]	[.039]
Jarque-Bera test	2.290	1.831	4.644	0.921	1.097	1.322	0.552	6.207	4.070	6.965
P-value	[.318]	[.400]	[.098]	[.631]	[.578]	[.516]	[.759]	[.045]	[.131]	[.031]
LM het. Test	0.579	2.727	0.828	1.431	1.573	2.168	3.228	0.713	1.090	4.416
P-value	[.447]	[.099]	[.363]	[.232]	[.210]	[.141]	[.072]	[.399]	[.296]	[.036]
White het. Test	33.027	40.525	34.785	40.686	41.929	45.465	48.728	39.712	39.989	36.829
P-value	[.564]	[.240]	[.478]	[.234]	[.196]	[.111]	[.061]	[.268]	[.258]	[.384]

Graf III.99.a Stabilita koeficientů vzhledem k velikosti populace (*h*)



Tabulka III.99.d Vyloučená pozorování

Velikost populace (<i>h</i>)	49	50	51	52	53	54	55	56	57	58
	1	1	1	1	1	1	1			
	7	7	7	7	7	7	7	7	7	7
	8		8	8	8					
	9	9	9	9	9	9	9			
	12									
	13	13	13	13	13	13	13	13	13	
		15								
	23	23	23	23	23	23	23	23	23	23
	25	25	25	25	25	25	25	25	25	25
	29			29						
		30	30							
	32	32								
	36		36	36	36	36		36		
			49							
	53									
		59								
		60		60						

Příloha IV. Vyloučená pozorování - porovnání metod

Vyloučená pozorování v jednotlivých letech

Rok	1993		1994		1995		1996		1997		1998		1999	
Metoda	LTS	RLS	LTS	RLS	LTS	RLS	LTS	RLS	LTS	RLS	LTS	RLS	LTS	RLS
Velikost souboru (h)	37	37	54	54	54	52	49	48	54	53	48	49	54	54
	1	1			1	1	1	1	1	1	1	1	1	1
	4	4		4			4	4	4	4				
	6	6					6	6						
	7	7	7	7	7	7	7	7	7	7	7	7	7	7
	9	9		9		9	9	9		9	9	9	9	9
	10	10												
	11	11			11	11								
	12	12	12		12	12	12	12						
	16	16						16			13	13	13	13
			19								16			
	21	21		22										
	23	23	23	23	23	23	23	23	23	23	23	23	23	23
	24	24			24	24	24	24			24	24		
	25	25				25	25	25	25	25	25	25	25	25
	26	26					26	26						
	27	27												
	30	30									30	30		
											32	32		
	33	33												
	34	34	34	34			34	34			34	34		
	36	36	36								36	36	36	36
	38	38												
	42	42												
	44	44												
	49	49	49	49	49	49	49	49	49	49				
											53			
	58	58												
												59		

Příloha V. Klasifikace SITC 02

Kategorie	SITC 02 kód	Číslo pozorování
Meat and meat preparations	01	1
Dairy products and birds' eggs	02	2
Fish (not marine mammals), crustaceans, molluscs and aquatic invertebrates, and preparations thereof	03	3
Cereals and cereal preparations	04	4
Vegetables and fruit	05	5
Sugars, sugar preparations and honey	06	6
Coffee, tea, cocoa, spices, and manufactures thereof	07	7
Feeding stuff for animals (not including unmilled cereals)	08	8
Miscellaneous edible products and preparations	09	9
Beverages	11	10
Tobacco and tobacco manufactures	12	11
Hides, skins and furskins, raw	21	12
Oil-seeds and oleaginous fruits	22	13
Crude rubber (including synthetic and reclaimed)	23	14
Cork and wood	24	15
Pulp and waste paper	25	16
Textile fibres (other than wool tops and other combed wool) and their wastes (not manufactured into yarn or	26	17
Crude fertilizers, other than those of division 56, and crude minerals (excluding coal, petroleum and	27	18
Metalliferous ores and metal scrap	28	19
Crude animal and vegetable materials, n.e.s.	29	20
Coal, coke and briquettes	32	21
Petroleum, petroleum products and related materials	33	22
Gas, natural and manufactured	34	23
Animal oils and fats	41	24
Fixed vegetable fats and oils, crude, refined or fractionated	42	25
Animal or vegetable fats and oils, processed; waxes of animal or vegetable origin; inedible mixtures or preparations of animal or vegetable fats or oils, n.e.s.	43	26
Organic chemicals	51	27
Inorganic chemicals	52	28
Dyeing, tanning and colouring materials	53	29
Medicinal and pharmaceutical products	54	30
Essential oils and resinoids and perfume materials; toilet, polishing and cleansing preparations	55	31
Fertilizers (other than those of group 272)	56	32
Plastics in primary forms	57	33
Plastics in non-primary forms	58	34
Chemical materials and products, n.e.s.	59	35
Leather, leather manufactures, n.e.s., and dressed furskins	61	36
Rubber manufactures, n.e.s.	62	37
Cork and wood manufactures (excluding furniture)	63	38
Paper, paperboard and articles of paper pulp, of paper or of paperboard	64	39
Textile yarn, fabrics, made-up articles, n.e.s., and related products	65	40
Non-metallic mineral manufactures, n.e.s.	66	41
Iron and steel	67	42
Non-ferrous metals	68	43
Manufactures of metals, n.e.s.	69	44
Power-generating machinery and equipment	71	45
Machinery specialized for particular industries	72	46
Metalworking machinery	73	47
General industrial machinery and equipment, n.e.s., and machine parts, n.e.s.	74	48
Office machines and automatic data-processing machines	75	49
Telecommunications and sound-recording and reproducing apparatus and equipment	76	50
Electrical machinery, apparatus and appliances, n.e.s., and electrical parts thereof (including non-electrical counterparts, n.e.s., of electrical household-type equipment)	77	51
Road vehicles (including air-cushion vehicles)	78	52
Other transport equipment	79	53
Prefabricated buildings; sanitary, plumbing, heating and lighting fixtures and fittings, n.e.s.	81	54
Furniture, and parts thereof; bedding, mattresses, mattress supports, cushions and similar stuffed	82	55
Travel goods, handbags and similar containers	83	56
Articles of apparel and clothing accessories	84	57
Footwear	85	58
Professional, scientific and controlling instruments and apparatus, n.e.s.	87	59
Photographic apparatus, equipment and supplies and optical goods, n.e.s.; watches and clocks	88	60
Miscellaneous manufactured articles, n.e.s.	89	61