

Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

Bakalářská práce

2004

Jakub Jeřábek

**Univerzita Karlova v Praze
Fakulta sociálních věd**

Institut ekonomických studií

BAKALÁŘSKÁ PRÁCE

**Ekonometrický model českého zahraničního obchodu
se zeměmi Evropské unie**

(Aplikace robustní metody LTS)

Vypracoval:

Vedoucí:

Akademický rok:

Jakub Jeřábek

Doc. RNDr. Jan Ámos Víšek, CSc.

2003/2004

Prohlášení

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

V Praze dne 20. 7. 2004

Jakub Jeřábek

Obsah:

1. Úvod.....	2
2. Teorie mezinárodního obchodu	3
3. Robustní metody	4
3.1. Metoda nejmenších čtverců	4
3.2. Vlastnosti robustních metod	5
3.3. Metody LMS, LTS a SLTS.....	7
3.4. Algoritmy pro metodu LTS	9
3.5. Implementace algoritmu	10
4. Použitá data	11
4.1. Klasifikace	11
4.2. Proměnné	12
5. Sestavení ekonometrického modelu.....	14
5.1. Tvar „exportní funkce“	14
5.2. Hypotézy.....	15
6. Průběh výpočtů – diagnostika	18
6.1. Multikolinearita	18
6.2. Normalita	20
6.3. Homoskedasticita.....	21
6.4. Autokorelace	21
7. Výsledky analýzy – export.....	23
7.1. Souhrnný model pro export	23
7.2. Rovnice pro export pro každý rok zvlášť	26
7.3. Porovnání obou přístupů.....	27
8. Výsledky analýzy – import	30
3.1. Souhrnný model pro import.....	30
8.2. Rovnice pro import pro každý rok zvlášť	31
8.3. Porovnání obou přístupů.....	32
9. Závěr	34
Seznam literatury	35
Abstrakt / Abstract	37
Příloha A: Komentovaný zdrojový kód programu.....	38
Příloha B: Tabulky výsledků	41
Příloha C: Projekt bakalářské práce	43

1. Úvod

Tato bakalářská práce se zabývá ekonometrickou analýzou zahraničního obchodu České republiky se zeměmi Evropské unie. Protože Českou republiku je možné považovat za otevřenou ekonomiku, má zahraniční obchod pro české hospodářství mimořádný význam.

V období po pádu komunistického režimu prožívala česká ekonomika nelehké období. Hospodářský blok bývalé RVHP se rozpadl a dosud samozřejmí obchodní partneři už nebyli k dispozici. Otevíraly se dveře do Evropské unie; bylo nutné se přeorientovat na nové trhy a nové obchodní příležitosti. A české podniky se přeorientovaly opravdu rychle. V období od roku 1993 do roku 2001 vzrostl český export do zemí Evropské unie z 6,3 na 25,6 miliard EUR.

Nyní, v době, kdy již začínají být k dispozici kvalitnější ekonomická data o české ekonomice, je úkolem ekonomů, aby se pokusili osvětlit průběh tohoto přerodu a zachytit jeden ze základních procesů české transformace.

Cílem této práce je vybudovat vysvětlující ekonometrický model českého zahraničního obchodu (exportu i importu) se zeměmi Evropské unie v letech 1993 až 1999. K tomuto účelu budu používat robustní regresní metodu LTS, která zahradí některé nepříjemné vlastnosti dostupných statistických dat. Navíc bych chtěl prostřednictvím této metody identifikovat průběh transformace v jednotlivých odvětvích, protože odvětví, která metoda LTS do modelu vybere, se dají v daném roce považovat za transformovaná.

V druhé kapitole nastíním základní makroekonomické teorie zahraničního obchodu, které používám při sestavení modelu. Ve třetí představím použité ekonometrické metody, zejména metodu LTS. Ve čtvrté části přiblížím zdroje dat, se kterými jsem pracoval a problémy, které při práci s nimi vyvstaly. Potom sestavím vlastní model ve tvaru „exportní a importní funkce“.

Následně provedu na datech základní regresní diagnostiku a pokusím se ověřit nejzákladnější předpoklady, které by měla data použitá pro regresní analýzu splňovat. Sedmá a osmá kapitola jsou věnovány výsledkům mé regresní analýzy a vlastnostem vytvořených modelů. V závěru shrnu svá pozorování a nastíním další možná pokračování výzkumu v této oblasti.

2. Teorie mezinárodního obchodu

V této kapitole bych chtěl jen stručně shrnout základní teoretické koncepce, na kterých je postaven ekonomický výzkum zahraničních obchodu. Zaměřím se převážně na nejdůležitější dvě – Ricardovu teorii komparativní výhody a Heckscher-Ohlinův model s více výrobními faktory. Pro bližší seznámení doporučuji následující publikace: [Benáček, 1999 a Krugman, Obstfeld, 2003].

Ricardova teorie je postavena na základní myšlence komparativní výhody. To je hypotéza, že i když jsou státy, které nevyrábějí nic efektivněji než ostatní a naopak některé státy vyrábějí všechny druhy výrobků s vyšší efektivitou, mohou si všichni polepšit, když se budou specializovat. Když se bude každý stát specializovat na výrobu těch statků, které vyrábí relativně nejefektivněji (celková produktivita faktorů je pro dané odvětví nejvyšší), může ostatní statky získat obchodem a celkově vydělat. Jeho teorie byla empiricky ověřena tak, že v těch odvětvích, ve kterých byl poměr produktivity mezi dvěma státy vyšší, byl vyšší i poměr vývozu a naopak (v logaritmickém modelu) [Krugman, Obstfeld, 2003:33]. Proto by měla být produktivita jednotlivých odvětví nějak zahrnuta i do našeho modelu zahraničních obchodu s Evropskou unií.

Heckscher-Ohlinova teorie vychází z odlišného konceptu rozdílné vybavenosti států základními faktory produkce. Každý stát se podle ní má specializovat na takový druh produkce, pro nějž má dostatek odpovídajících vstupů – kapitálu nebo práce. Ve státech s přebytkem pracovní síly je výhodnější zaměřit výrobu do pracovně náročných odvětví, naopak ve státech s dostatečným množstvím fyzického kapitálu bude lepší soustředit síly do výroby v kapitálově náročných odvětvích. Tak budou výrobní kapacity obou států optimálně využity a požadované množství spotřebních statků se získá zahraničním obchodem.

Pro ověření Heckscher-Ohlinovy teorie zařadíme do modelu proměnnou K/L , tedy množství fyzického kapitálu připadajícího na jednoho pracovníka v daném odvětví. Pro odvětví s vysokou kapitálovou vybaveností bude mít tato proměnná vysokou hodnotu, pro odvětví zaměřená na využití lidské práce bude nižší (dobrým příkladem přímo z dat je textilní výroba s výjimečně nízkou vybaveností fyzickým kapitálem, naopak kapitálově nejnáročnější je v datech o české ekonomice chemický průmysl). Pokud Heckscher-Ohlinova hypotéza platí i pro český zahraniční obchod, měl by být její vliv na export obrácený než na import. S rostoucím poměrem kapitálu a práce v odvětvích by tak mohl export klesat a import stoupat, nebo naopak.

Třetí důležitou hypotézou je Krugmanův model úspor z rozsahu a vnitroodvětvového obchodu. Ten do modelu zahrneme jednoduchým způsobem – objem importu i exportu bude záviset i na velikosti produkce daného odvětví.

3. Robustní metody

3.1. Metoda nejmenších čtverců

Odedávna nejpoužívanější ekonometrickou metodou je bezpochyby metoda nejmenších čtverců (dále ji budeme označovat OLS – Ordinary Least Squares). Je založena na jednoduchém a pochopitelném principu, že malá odchylka není příliš důležitá a proto ji můžeme potlačit tím, že ji umocníme na druhou. Velká odchylka naproti tomu vadí, proto ji umocněním ještě zvýrazníme. Při určování koeficientů regresní nadroviny pak minimalizujeme součet čtverců odchylek všech pozorování od této nadroviny. Matematické vyjádření se dá zapsat v následujícím tvaru:

$$\mathbf{b}^{OLS} = \arg \min_{\mathbf{b} \in \mathcal{R}^{p+1}} \sum_{i=1}^n r_i^2(\mathbf{b})$$

kde p určuje počet vysvětlujících proměnných, n je celkový počet pozorování a $r_i^2(\mathbf{b})$ je vzdálenost i -tého pozorování od regresní nadroviny, definovaná jako $r_i^2(\mathbf{b}) = (\mathbf{y}_i - \mathbf{x}_i^T \mathbf{b})^2$.

Metoda OLS má řadu nepopíratelných výhod. Nejdůležitější z nich je numerická snadnost výpočtu, protože vektor regresních koeficientů \mathbf{b} se dá explicitně vyjádřit ve tvaru

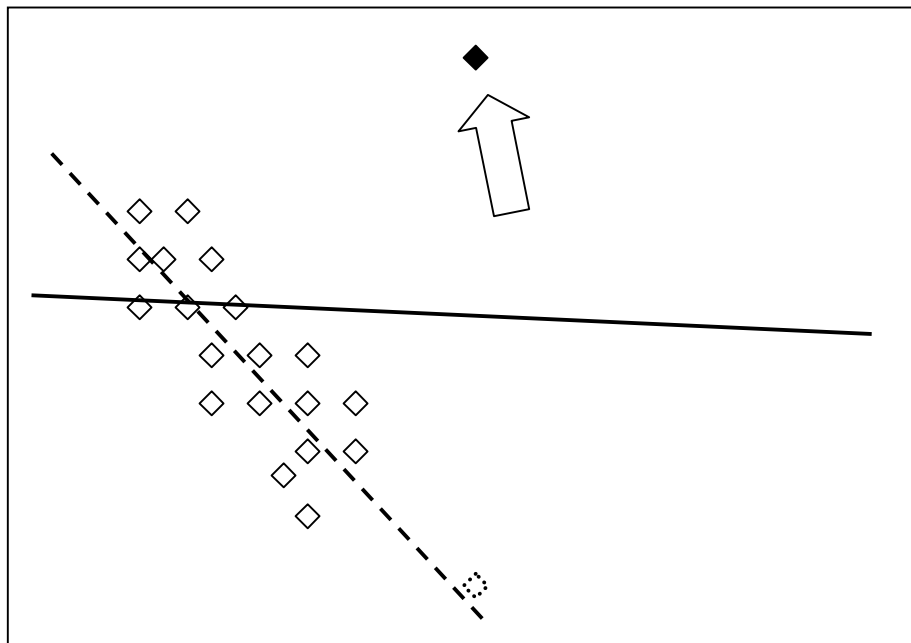
$$\mathbf{b}^{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Pro regresní rovnice o několika málo proměnných se dá jednoduše vypočítat „na papíře“. Tato vlastnost byla v minulosti klíčovým předpokladem pro využití matematických metod v praxi a je i jedním z důvodů, proč je metoda OLS tak oblíbená.

Když v roce 1820 navíc Carl Friedrich Gauss odvodil normální rozdělení, potvrdil tak teoreticky její výjimečnost mezi ostatními nestrannými odhady regresní nadroviny. Díky centrální limitní větě totiž za určitých předpokladů konverguje součet nezávislých náhodných veličin právě k veličině s normálním rozdělením. A právě pro normálně rozdělená rezidua (e_i) ve vzorci $\mathbf{y}_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i$ má odhad metodou OLS nejmenší rozptyl mezi všemi nestrannými odhady.

Pro platnost této silné věty však musí být splněny všechny její předpoklady, například to, že jednotlivá rezidua jsou mezi sebou nezávislá, mají stejný rozptyl a přibližně normální rozdělení. Stačí, aby alespoň jeden z předpokladů nebyl splněn a metoda OLS může dávat silně zkreslené, někdy i přímo nepoužitelné odhady. Na příkladu si můžeme ukázat, co se stane s odhadem metodou nejmenších čtverců, když se v datech vyskytnou odlehlá pozorování – to jsou takové body, které se v jedné nebo několika souřadnicích nacházejí mimo rozložení většiny dat. Odlehlá pozorování snadno vzniknou třeba při chybném zpracování dat nebo překlepem.

Na obrázku vidíme situaci, kdy byl čárkovaný bod dole uprostřed omylem zaznamenán jako černý bod nahoře označený šipkou. Přerušovaná čára naznačuje, kudy by vedla regresní přímka v případě správného umístění bodu. Plná čára ukazuje přibližnou polohu přímky odhadnuté na poškozených datech (s černým bodem místo čárkovaného). Zatímco původní přímka udržuje směr velké skupiny bodů, může jediné odlehlé pozorování vést ke zhroucení odhadu metodou OLS.



Graf 1: Ilustrace nevýhod metody OLS

3.2. Vlastnosti robustních metod

Nevýhody metody OLS při nesplnění předpokladů si ekonometrové a statistici uvědomovali už dávno. Francouzský matematik Pierre-Simon Laplace zavedl L_1 -regresi, která minimalizuje nečtverce, ale absolutní hodnoty reziduí. Ta už je odolná vůči odlehlým pozorováním v x-ové souřadnici, ale ne v y-ové. Navíc tento odhad zdaleka není tak vydatný jako OLS.

Mnozí další matematici přišli s takzvanými neparametrickými regresními metodami [např. Jurečková, 1971], které nepředpokládaly žádné konkrétní rozdělení pravděpodobnosti a počítaly na základě pořadí jednotlivých pozorování. Takovou statistikou je například medián, kde jediným předpokladem nestrannosti výběrového mediánu je symetrické rozdělení.

Často zmiňovanou možností, která už se trochu blíží robustním metodám, je pouhé odstranění odlehlých pozorování ze souboru. To se může provádět buď na základě subjektivního úsudku, nebo na základě nějakého pravidla, třeba že určitý druh vzdálenosti takového bodu je větší než k -násobek průměrné vzdálenosti všech bodů od „středu“ dat. V datech o nižší dimenzi je

pravděpodobné, že se takovým způsobem dají odlehlá pozorování opravdu odhalit. Tento přístup ale s sebou přináší celkem vysokou ztrátu efektivnosti, protože často přinášejí odlehlá pozorování důležitou informaci, třeba o tom, jak se model chová v extrémních situacích. Ve vyšších dimenzích k tomu ještě mohou být odlehlá pozorování v datech důmyslně zamaskována, takže je není možné najít ani v grafickém editoru, který umožňuje otáčení dat. Přitom většina zkoumaných datových souborů, aniž by si toho byl výzkumník vědom, obsahuje často i více než 1% významných chyb (*gross errors*) [Hampel, Ronchetti, Rousseeuw, Stahel, 1986:25].

Při odvozování robustních statistických metod se vychází z odlišného principu. Autor sice předpokládá určité pravděpodobnostní rozdělení, ale snaží se, aby jeho metody fungovaly i pro „podobná“ rozdělení. To znamená, třeba když je do souboru přimícháno určité procento pozorování s jiným rozdělením. Patří sem i metody, které zvládají práci s odlehlými pozorováními nebo takové, které vytvoří model pro většinu dat a označí ta, která se „nehodí“.

Základním pojmem při práci s daty, o kterých si myslíme, že jsou nějakým způsobem kontaminována, je „bod zlomu“ (*breakdown point*). Uveďme nyní alespoň jeho základní definici [Rousseeuw, Leroy, 1987]:

Mějme náhodný výběr o n pozorováních:

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

T buď regresní odhad. To znamená, že aplikací T na náhodný výběr Z dostaneme vektor regresních koeficientů. Mějme nyní výběr Z' , který z původního výběru dostaneme nahrazením m pozorování libovolnými hodnotami. Potom bude veličina $\text{bias}(m; T, Z)$ označovat maximální zkreslení, kterého se můžeme dopustit, když provedeme nahrazení m bodů v původním výběru:

$$\text{bias}(m; T, Z) = \sup_{Z'} \| T(Z') - T(Z) \|.$$

Pokud je toho zkreslení nekonečné, znamená to, že m odlehlých pozorování může mít libovolně velký vliv na odhad regresních koeficientů $T(Z)$. Z toho vyplývá, že odhad T se „zhroutí“ (breaks down) a nepodává žádnou informaci o regresních koeficientech původního výběru. Proto „bod zlomu“ (breakdown point) odhadu T na náhodném výběru Z definujeme jako:

$$\varepsilon_n^*(T, Z) = \min \{m/n ; \text{bias}(m; T, Z) = \infty\},$$

tedy podíl „kontaminace“ náhodného výběru Z , při kterém se odhad T „zhroutí“.

Pokud nyní tuto definici „bodu zlomu“ aplikujeme na odhad metodou nejmenších čtverců, okamžitě vidíme, že jediné odlehlé pozorování způsobí „zhroutení“ odhadu. Proto je „bod zlomu“

pro metodu OLS roven $1/n$. S rostoucím rozsahem výběru se $1/n$ blíží nule, proto můžeme říct, že odhad metodou nejmenších čtverců má *breakdown point* rovný 0%.

Naopak, když vezmeme medián jako odhad střední hodnoty z „kontaminovaného“ normálního rozdělení, snadno odvodíme, že „zhroucení“ odhadu by způsobila až 50% kontaminace vzorku.

3.3. Metody LMS, LTS a SLTS

Při práci s kontaminovanými daty je naším cílem, aby používané metody měly co nejvyšší „bod zlomu“ při zachování vydatnosti odhadu. Mezi takovéto „robustní“ metody patří LMS (Least Median of Squares), LTS (Least Trimmed Squares) a SLTS (Smoothed LTS).

První zmíněný robustní odhad regresních koeficientů, LMS, používá místo součtu čtverců reziduí jejich medián. Tento odhad, navržený P. Rousseeuwem [Rousseeuw, 1984], je definován tak, že minimalizuje medián z druhých mocnin reziduí. Jak jsme si ukázali, je medián velice robustní odhad. Při LMS tak vlastně záleží jenom na „lepší“ polovině pozorování ve výběru. Pokud je hladina „kontaminace“ dat menší než 50% (a to je, protože jinak se menšina stává většinou), pak by měl LMS odhad poskytovat „správné“ řešení. Na LMS navazují různé varianty, například, že se minimalizuje nikoliv medián z reziduí, ale h -tá pořadová statistika [Rousseeuw, Leroy, 1987:132]. Další metodou založenou na LMS jsou RLS (reweighted least squares), kdy se vypočítá jak odhad OLS tak LMS a při novém odhadu pomocí OLS se vyřadí ta pozorování, která mají rezidua pro OLS k -krát větší než pro LMS (např. pro $k=2,5$, viz [Rousseeuw, Leroy, 1987:18]).

Zůstávají ale pořád dva problémy. První je, jak výsledný LMS odhad spočítat, protože na rozdíl od OLS pro něj neexistuje explicitní vyjádření. Rousseeuw navrhl algoritmus, který náhodně vybírá skupiny pozorování, pro ně počítá zmíněný medián a porovnává, jestli se zlepšila dosavadní minimální hodnota. Tento postup je však celkem zdoluhavý, neefektivní a navíc nespolehlivý [Višek, 2000]. V roce 1995 byl publikován lepší algoritmus založený na simplexové metodě lineárního programování [Boček, Lachout, 1995].

Druhým problémem, který znevýhodňuje metodu LMS oproti jiným, je pomalá konvergence. Standardní odhad metodou OLS je \sqrt{n} -konzistentní [Višek, 1997], to znamená, že konverguje ke „správnému“ řešení „jako \sqrt{n} “. Naproti tomu odhad metodou LMS konverguje mnohem pomaleji, je totiž jenom $\sqrt[3]{n}$ -konzistentní. Dá se říci, že pokud bychom pro danou přesnost odhadu metodou OLS potřebovali 100 pozorování, tak pro stejnou přesnost při metodě LMS jich už potřebujeme 1000.

Další zmíněná robustní metoda je LTS [Rousseeuw, 1984] (Least Trimmed Squares - doslovný překlad by byl „metoda ořezaných nejmenších čtverců“, protože při odhadu doslova

„odřezáváme“ data, která nám do modelu „nesedí“). Tato metoda je bližší původnímu odhadu OLS, protože rovněž počítá součet druhých mocnin reziduí (a ne medián). Na rozdíl od OLS ale LTS minimalizuje součet čtverců reziduí pouze u některých prvků. Vlastně by se mělo psát LTS_h , protože při výpočtu musíme zadat h - počet pozorování, který chceme zahrnout do modelu. LTS

$$\mathbf{b}^{LTS} = \arg \min_{\mathbf{b} \in \mathfrak{R}^{p+1}} \sum_{i=1}^h r_{[i]}^2(\mathbf{b})$$

potom sestavuje model pro celá data, nýbrž vybere ze souboru pouze taková pozorování, aby byl součet čtverců jejich reziduí minimální. Matematická definice by se dala vyjádřit asi takhle:

[i] u reziduí naznačuje, že sčítáme přes seřazená rezidua, tedy vybereme jenom h nejmenších a ostatní na výsledek odhadu vůbec nemají vliv. „Bod zlomu“ metody LTS závisí na zvoleném h . Když položíme h rovné celé části z $n/2$ (n je celkový počet pozorování), dostáváme *breakdown point* 50%. Navíc má metoda LTS asymptoticky normální rozdělení a je \sqrt{n} -konzistentní [Víšek, 1999a]. Optimální počet pozorování, které je třeba zařadit do modelu, aby měla metoda LTS maximální *breakdown point* je dosažen při hodnotě $h = [n/2] + [(p+1)/2]$ [Rousseeuw, Leroy, 1987:132].

Další možné rozšíření metody LTS naznačil Pavel Čížek ve své disertaci [Čížek, 2002]. Svoji metodu nazval SLTS (Smoothed Least Trimmed Squares). Úprava metody LTS spočívá v použití vah, které změni vliv jednotlivých reziduí na celkovou sumu. Potom, co jsou rezidua pro danou regresní přímku seřazena, přidělí se jim váhy podle vhodné klesající funkce (např. logistické), přičemž větší rezidua dostanou nižší ohodnocení než malá. Součet čtverců se potom počítá pro všechna rezidua, ale každé se započítává s příslušnou vahou. Takto „vyhlazenou“ metodu LTS lze potom použít i pro diskrétní vysvětlující proměnné.

3.4. Algoritmy pro metodu LTS

Mějme nyní data složená ze dvou populací. Jedna z nich se bude řídit navrženým modelem, zatímco druhá část odpovídá nějakému jinému modelu, nebo se vůbec žádným modelem neřídí. Když budeme chtít tyto dvě skupiny při naší analýze rozdělit a identifikovat, bude výhodné použít na naše data nějakou z robustních metod. Z předchozí kapitoly vyplývá, že nejvhodnější pro takovou analýzu bude metoda LTS – má vysoký „bod zlomu“, poměrně rychle konverguje a navíc, když budeme postupně zvyšovat „ořezávací“ konstantu h , můžeme se pokusit najít datové body příslušející první subpopulaci.

Nejjednodušší cestou, jak počítat metodu LTS by bylo vyzkoušet všechny kombinace h prvků z výběru. To by ale bylo se současnými výpočetními možnostmi technicky neproveditelné;

například pro 50 pozorování a pro h rovnající se 30 bychom museli porovnat sumu čtverců asi pro 10^{13} kombinací. Proto byly vyvinuty různé přibližné algoritmy, které postup výpočtu urychlí.

Původní schéma, jak ho navrhl objevitel metody [Rousseeuw, Leroy, 1987], spočívá v náhodných výběrech vždy $p+1$ pozorování z celkového souboru. Pro ně se vypočítají koeficienty regresní nadroviny a pro zjištěnou nadrovinu určíme pro každé pozorování ze souboru reziduum a jeho druhou mocninu. Ty potom seřadíme a vypočteme součet prvních h . Tento postup opakujeme pro zvolený počet opakování a zjistíme nejmenší hodnotu tohoto součtu. Odhad metodou LTS pak bude vypočítán jako odhad obyčejnou metodou OLS na vybraných h pozorování, pro které byl nejmenší součet. Zmíněný postup je velmi náročný na počet opakování a při nízkém počtu opakování zase hrozí, že nebude nalezena optimální kombinace h bodů.

Proto byl postup výpočtu zlepšen a zrychlen. Pro první iteraci se vybere buď $p+1$ nebo h pozorování. Z nich se vypočte regresní nadrovina. Pro takto zjištěnou nadrovinu se spočtou rezidua pro všechna pozorování a h nejmenších z nich určí prvky pro druhou iteraci. Tento postup se opakuje, dokud při výběru prvků pro další iteraci dochází k snižování součtu čtverců reziduí vybraných h prvků. Tím se získá odhad pro první opakování. Nyní se zmíněný postup opakuje vždy pro jinou náhodou první iteraci dokud se buď několikrát nezíská stejný výsledek, nebo dokud počet opakování nedosáhne zvoleného maxima [Víšek, 1994; Rousseeuw, Van Driessen, 1998].

Otázkou zůstává, jak vybrat regresní nadrovinu pro první iteraci. Jednodušší se zdá vybrat náhodně h pozorování a vypočítat koeficienty nadroviny metodou nejmenších čtverců. Tento postup ale skrývá následující úskalí. Předpokládejme, že máme v souboru 30 „správných“ a 20 „odlehlých“ pozorování. Potom při h ve výši 30 s největší pravděpodobností budou v našem prvním výběru i nějaká odlehlá pozorování a to může celý odhad výrazně vychýlit. Proto je výhodnější zvolit na začátku pouze $p+1$ pozorování, z nich vypočítat koeficienty regresní nadroviny jako lineární rovnici a teprve potom iterovat vždy výběrem h prvků.

3.5. Implementace algoritmu

Při implementaci algoritmu LTS jsem zvolil výpočetní prostředí statistického programu R. Má několik zásadních výhod – je volně šiřitelný, snadno se v něm programuje a obsahuje širokou knihovnu nejrůznějších statistických testů. Komentovaný zdrojový kód algoritmu je v Příloze A.

Výpočet probíhá podle tohoto schématu [Víšek, 2000:17]:

- 1) *Referenční součet čtverců* polož rovný $+\infty$
- 2) Vyber náhodně $p+1$ pozorování ze základního souboru

- 3) Zjisti, jestli matice vytvořená těmito $p+1$ řádky není singulární, v případě potřeby opakuj krok 1)
- 4) Najdi koeficienty regresní nadroviny a vypočítej rezidua pro všechna pozorování, indexy bodů, pro které najdeš h nejmenších reziduí, ulož do i
- 5) Dokud se snižuje *součet čtverců* nebo počet opakování dosáhne 1000 opakuj:
 - a. Vypočti regresní nadrovinu pro všechna pozorování v i metodou OLS
 - b. Najdi h pozorování, která mají nejmenší rezidua vzhledem k této nadrovině
 - c. Pokud je součet reziduí pro tato pozorování menší než současný *součet čtverců*, změň ho a vlož indexy nejlepších h pozorování do i
 - d. Vrať se zpět do „a.“
- 6) Porovnej *součet čtverců* s *referenčním součtem čtverců*; pokud je *součet čtverců* menší, nahraď *referenční součet čtverců*, zapamatuj si indexy pozorování, pro které byl tento součet dosažen (-> *nejlepší výběr*) a polož *počítadlo* rovné 0
- 7) Pokud je *součet čtverců* ve zvolené toleranci od *referenčního součtu čtverců*, zvyš *počítadlo* o 1
- 8) Pokud není *počítadlo* větší nebo rovné číslu SHODA nebo počet opakování nedosáhl MAXOPAK, vrať se na krok 2)
- 9) Vypiš výsledek - odhad lineárního regresního modelu metodou OLS pro pozorování uložená v poli *nejlepší výběr*
- 10) Jako výsledek činnosti programu vrať indexy pozorování v poli *nejlepší výběr*

4. Použitá data

4.1. Klasifikace

Kromě použité metody jsou nejdůležitější součástí pro odhadování ekonometrického modelu data. V případě odvětvového modelu pro export a import bylo třeba dát dohromady data nejméně ze dvou zdrojů. Jedním z nich byla data o vývozu a dovozu členěná podle ekonomických odvětví. Druhým zdrojem byla data o českých podnicích, která se odvětvově člení podle převažujícího výrobního zaměření podniku. Tady se vynořuje první problém při zpracování dat, kterému se není možné vyhnout. Budu citovat jednoho kolegu z Ministerstva průmyslu a obchodu, když mi objasňoval principy klasifikace dat o produkci: „Kdybychom tady na ministerstvu na střeše chovali krávu, tak by to mléko byla produkce nikoliv odvětví 155 – *výroba a zpracování mléka*, ale odvětví 751 – *vláda a ústřední orgány*“.

Je vidět že pokud 20% výroby daného podniku spadá do jednoho odvětví a zbytek je rozdělen po 10% do dalších, pak je podnik zařazen do odvětví, ve kterém vyrábí 20% a jeho statistické výkazy jsou zpracovávány v rámci tohoto odvětví. To pochopitelně způsobuje podstatné zkreslení, pokud se snažíme spojit statistiky po hospodářských výsledcích jednotlivých odvětví se statistikami o vývozu a dovozu zboží, které příslušná odvětví produkují.

Dalším problémem zpracování dat jsou odlišné klasifikace. Množství a ceny zboží vyvezeného a dovezeného se standardně zpracovávají v celní klasifikaci a také v klasifikaci SITC (Standard International Trade Classification). Dvojmístné kódy podle klasifikace SITC rozdělují průmyslové a zemědělské výrobky do 61 kategorií.

Oproti tomu statistiky o výrobě a podnicích se zpracovávají podle klasifikace SKP (Standardní klasifikace produkce) a OKEČ (Odvětvová klasifikace ekonomických činností), které odpovídá evropský standard NACE (Nomenclature général des Activités économiques dans les Communautés Européens) a mezinárodní klasifikace ISIC. OKEČ se člení na 9 sekcí, 27 subsekcí, 93 oddílů a 217 pododdílů.

Hlavní nevýhodou této „dichotomie“ klasifikací je, že neexistuje možnost jednoznačného převodu z jedné klasifikace do druhé. Je možné provést nepřesný a zdlouhavý převod výběrem jednotlivých odpovídajících si odvětví. Stejně by ale bylo nutné některá odvětví rozdělovat (k transformacím mezi různými odvětvovými klasifikacemi více viz. [Papageorgiu, 2002]).

Díky laskavosti doc. Benáčka jsem mohl při své analýze využít již transformovaná data za roky 1993-1999. Bylo by jistě zajímavé doplnit data alespoň o nejnovější dva až tři roky, ale bohužel to nebylo v mých silách. Sehnal jsem sice údaje o tržbách a cenách vývozu a dovozu

v klasifikace SITC, ale údaje o podnicích žádný subjekt v klasifikaci SITC opravdu neshromažďuje (navíc by to ani nemělo smysl, protože SITC je klasifikace výrobků, ne podniků). Data o podnicích v trojmístné klasifikaci OKEČ rozdělená na 217 kategorií se mi sice podařilo opatřit na Ministerstvu průmyslu a obchodu, ale protože cíl této práce se soustřeďuje spíše na použití sofistikovaných ekonometrických metod než na složité transformování, rozhodl jsem se omezit analýzu na dostupná data za roky 1993-1999.

4.2. Proměnné

Celkem jsem měl k dispozici 427 pozorování – pro každý ze sedmi roků data ze 61 odvětví podle klasifikace SITC. Ústředními proměnnými byla hodnota exportu do Evropské unie a importu rovněž z EU. Je vyjádřen v běžných cenách v tisících dolarů. Zde je nutné podotknout, že se rozhodně nemusí rovnat součet exportu z České republiky a importu do zemí, kam ČR vyváží. Důvodem jsou různé statistické nepřesnosti a také fakt, že pro účely vývozu se zboží registruje netto (bez dopravy) a pro dovoz se započítává brutto. Další chybu ve statistikách můžou vyvolat změny směnného kurzu během roku, protože pro účely modelu započítáváme pouze meziroční změny reálného směnného kurzu [Benáček, Prokop, Víšek, 2003a:10-11]. V modelu jsou tyto veličiny označeny VX (values of export) a VM (values of import).

Další proměnnou jsou kilogramové ceny vývozu a dovozu. Ty slouží jako pomocná proměnná k odhadnutí změn v kvalitě obchodovaného zboží. Vysoké ceny znamenají konkurenci kvalitou, nízké konkurenci cenou. Tento ukazatel ale může být dost ambivalentní, protože pokud hypotéza jednotného trhu platí jen omezeně, pak změny v jednotkových cenách můžou naznačovat změnu ceny a nikoliv kvality. Kromě toho je v naprosté většině odvětví produkt silně diferencován a změna v „kilogramových“ cenách může znamenat přechod k jinému typu produkce. Nicméně tyto jednotkové ceny jsou uvedeny v běžných cenách v Kč za tunu a v modelu je značím PX (exportní) a PM (importní).

Pro účely zkoumání absorpční schopnosti českého a evropského trhu jsou do modelu zařazeny také veličiny GDPEU a GDPCZ. Český hrubý domácí produkt je měřen v konstantních cenách roku 1995 v Kč, evropský hrubý domácí produkt je součtem pro všechny státy EU-15 a je počítán v Eurech. Do modelu jsem je zařadil pod názvem GEU a GCZ.

Následující proměnná má v ekonometrickém modelu pro import a export význačné místo; je totiž základním článkem při ověřování Heckscher-Ohlinova modelu zahraničního obchodu na základě rozdílné vybavenosti výrobními faktory. Je to veličina K/L, tedy množství fyzického kapitálu připadajícího na jednoho pracovníka. S jejím zjišťováním se zároveň pojí největší problémy. Jako počet pracovníků v jednotlivých odvětvích je brán přepočtený počet pracovníků (ve

„full-time“ ekvivalentech. Ale pro množství fyzického kapitálu existují dvě časové řady, z nichž žádná nepokrývá celé sledované období. První je pro roky 1993-1997 a zahrnuje veškerý pracovní kapitál v pořizovacích cenách bez zahrnutí odpisů. Druhá je od roku 1994 do 2000 a obsahuje hodnotu fyzického kapitálu v tržních cenách očištěnou o odpisy. Převod druhé časové řady na první je jen nepřesný a může v modelu způsobovat problémy [Benáček, Prokop, Víšek, 2003b:13]. Kvůli neexistenci OECD statistik pro fyzický kapitál nemůžeme do modelu zařadit také obdobný poměr pro země Evropské unie. Proměnnou K/L označuji v modelu zkratkou KL.

Další proměnná v modelu je objem přímých zahraničních investic, které v daném roce směřovaly do příslušného průmyslového odvětví. Jsou uváděny v miliónech Kč v běžných cenách. V modelu slouží jako pomocná proměnná, protože jsou korelovány s „celkovou produkcí faktorů“, parametrem důležitým pro ricardiánský model komparativní výhody [Benáček, Prokop, Víšek, 2003a:19]. V modelu jsou označeny FDI a spolu s proměnnou KL poskytují dostatečnou základnu pro zkoumání nabídkové strany zahraničního obchodu.

Nesmíme ovšem zapomenout na tak samozřejmou věc, že čím je odvětví ve větší míře zastoupené v české ekonomice, tím větší budou vývozy a dovozy. Pro ilustraci této skutečnosti použijeme proměnnou závislou na přidané hodnotě každého odvětví. Proměnná VA (value added) v modelu je tisícinásobkem poměru přidané hodnoty daného odvětví na přidané hodnotě celé ekonomiky. V několik prvních zkoumaných letech možná nebyla přidaná hodnota nejlepším ukazatelem velikosti odvětví, protože spousta významných odvětví byla v krizi a jejich přidaná hodnota byla spíš záporná, ale v poslední době lze VA s úspěchem použít.

Poslední veličinou v modelu je reálný směnný kurz. Jeho význam je jasný – apreciacie (zhodnocení) české koruny zdraží exporty a zlevní importované zboží. Depreciacie (znehodnocení) působí opačně. Reálný směnný kurz označíme zkratkou RER a vyšší RER pak znamená apreciaci české koruny (k čemuž v 90. letech pravidelně docházelo).

5. Sestavení ekonometrického modelu

5.1. Tvar „exportní funkce“

Při vytváření ekonometrického modelu pro vývoz z České republiky do Evropské unie vycházíme z teoretických předpokladů jednotlivých makroekonomických teorií. Náš model se nebude snažit o predikci, ale především o vysvětlení závislosti jednotlivých proměnných a ve spojení s metodou LTS také o rozdělení odvětví na dvě skupiny – tu, která se modelem řídí a druhou, kterou LTS vyřadí jako odlehlá pozorování.

Základní funkční předpis pro export nebo import sestavíme v exponenciálním tvaru jako zobecněnou Cobb-Douglasovu funkci. Jako vysvětlující proměnné pro export použijeme PX (exportní ceny), VA (přidanou hodnotu), KL (poměr kapitálu a práce), FDI (objemy přímých zahraničních investic), GEU (GDP Evropské unie) a RER (reálný směnný kurs Kč a Eura). Do rovnice pro import místo PX dosadíme PM (importní ceny), a místo GEU dáme GCZ (GDP České republiky). Funkční předpisy pak budou v následujícím tvaru:

$$X_{it} = e^{\beta_0} (PX_{it})^{\beta_1} (VA_{it})^{\beta_2} (KL_{it})^{\beta_3} (FDI_{it})^{\beta_4} (GEU_t)^{\beta_5} (RER_t)^{\beta_6} \quad i = 1 \dots 61; t = 93, \dots$$

$$M_{it} = e^{\beta_0} (PM_{it})^{\beta_1} (VA_{it})^{\beta_2} (KL_{it})^{\beta_3} (FDI_{it})^{\beta_4} (GCZ_t)^{\beta_5} (RER_t)^{\beta_6} \quad i = 1 \dots 61; t = 93, \dots$$

Pokud z této rovnice vynecháme poslední dvě proměnné, dostáváme tvar „exportní“ a „importní“ funkce pro jednotlivé roky:

$$X_{it} = e^{\beta_{0t}} (PX_{it})^{\beta_{1t}} (VA_{it})^{\beta_{2t}} (KL_{it})^{\beta_{3t}} (FDI_{it})^{\beta_{4t}} \quad i = 1 \dots 61; \text{ pro každé } t = 93, \dots$$

$$M_{it} = e^{\beta_{0t}} (PM_{it})^{\beta_{1t}} (VA_{it})^{\beta_{2t}} (KL_{it})^{\beta_{3t}} (FDI_{it})^{\beta_{4t}} \quad i = 1 \dots 61; \text{ pro každé } t = 93, \dots$$

Nyní obě strany rovnice zlogaritmujeme, abychom dostali lineární funkční závislost. Vypočtené koeficienty z lineární rovnice pak budeme moci přímo interpretovat jako elasticity substituce pro exporty na domácím trhu a pro import na trhu zahraničním – nárůst příslušné proměnné o 1% tak vyvolá nárůst exportu, případně importu o $\beta_i\%$. Dostaneme tak v konečné podobě rovnice pro celý model (všechny roky dohromady):

$$X_{it} = \beta_0 + \beta_1 \ln(PX_{it}) + \beta_2 \ln(VA_{it}) + \beta_3 \ln(KL_{it}) + \beta_4 \ln(FDI_{it}) + \beta_5 \ln(GEU_t) + \beta_6 \ln(RER_t)$$

$$i = 1 \dots 61; t = 93, \dots$$

$$M_{it} = \beta_0 + \beta_1 \ln(PM_{it}) + \beta_2 \ln(VA_{it}) + \beta_3 \ln(KL_{it}) + \beta_4 \ln(FDI_{it}) + \beta_5 \ln(GCZ_t) + \beta_6 \ln(RER_t)$$

$$i = 1 \dots 61; t = 93, \dots$$

Pro druhý model, ve kterém budeme odhadovat koeficienty pro každý rok zvlášť, provedeme stejnou úpravu:

$$X_{it} = \beta_0 + \beta_{1t} \ln(PX_{it}) + \beta_{2t} \ln(VA_{it}) + \beta_{3t} \ln(KL_{it}) + \beta_{4t} \ln(FDI_{it})$$

$$i = 1 \dots 61; \text{ pro každé } t = 93, \dots$$

$$M_{it} = \beta_{0t} + \beta_{1t} \ln(PM_{it}) + \beta_{2t} \ln(VA_{it}) + \beta_{3t} \ln(KL_{it}) + \beta_{4t} \ln(FDI_{it})$$

$$i = 1 \dots 61; \text{ pro každé } t = 93, \dots$$

Když už máme regresní rovnice sestaveny, můžeme přistoupit k odhadu parametrů. Vzhledem k tomu, že česká ekonomika procházela v devadesátých letech restrukturalizací, nemůžeme očekávat, že se nám podaří vytvořit model, který bude dokonale vystihovat chování všech průmyslových a zemědělských odvětví. Některá odvětví budou ve zkoumaných letech již transformovaná a budou se svým chováním blížit zjištěnému modelu. V jiných odvětvích ale ještě nemusí být proměna dokončena a budou se chovat rozdílně. Proto budeme koeficienty ve všech modelech odhadovat pomocí popsané metody LTS, která je schopná vytržít zadaný počet odlehlých pozorování.

5.2. Hypotézy

Při výpočtu bude naším cílem rozhodnout o uměřenosti některých předpokladů o chování českého exportu a importu v průběhu devadesátých let. Zdrojem naší analýzy nebudou jen výsledné koeficienty u příslušných proměnných v daných modelech. Důležité informace přinese i průběh výpočtu, zejména rozdílnost výsledků pro rozdílné hodnoty dělící konstanty h , tedy počet pozorování, která budeme skutečně chtít zařadit do modelu.

Obraťme nejprve pozornost k předpokladům o konečných výsledcích regresních odhadů. Ve druhé kapitole je prezentováno několik teorií, které se snaží popsat příčiny a důsledky mezinárodního obchodu. Pokusme se vyvodit z nich důsledky i pro naši situaci.

První, Ricardův model komparativní výhody, se zakládá na schopnosti ekonomiky vyrábět některé statky efektivněji než jiné, samozřejmě v porovnání s hospodářskými podmínkami v zahraničí [Benáček, 1999]. V našich modelech by tento fakt mohla ilustrovat proměnná FDI. Je totiž velmi pravděpodobné, že právě odvětví, ve kterých je celková produktivita faktorů (total factor productivity, TFP) relativně nejvyšší, budou přitahovat i největší objemy zahraničních investic. Za předpokladu volného pohybu kapitálu, a ten můžeme v české republice v 90. letech s výhradami

považovat za splněný (pomiňme situace kolem privatizace strategických podniků a podobně), budou investice proudit do těch oblastí ekonomiky, které jsou konkurenceschopné i na zahraničních trzích. Naopak v těch odvětvích, které konkurenceschopné nejsou, můžeme očekávat nízké nebo žádné přímé zahraniční investice. V případě FDI je tedy hypotéza jasná, pokud má pro české hospodářství platit Ricardova teorie komparativní výhody, bude koeficient u této proměnné v modelu pro vývoz jednoznačně kladný – vyšší investice budou znamenat vyšší vývoz.

Pro případ dovozu není předem vůbec jasné, jaké znaménko bude mít koeficient u FDI. Na jedné straně znamenají investice zvýšení lidského kapitálu a produktivity, to znamená, že by český trh měl být v daném odvětví soběstačnější a dovoz by měl poklesnout. Naproti tomu budou vysoké investice doprovázet také vstup strategického partnera do některého z významných podniků v daném odvětví. V tom případě by byl efekt opačný, zvýšené investice by zvyšovaly dovoz ať už nových technologií nebo meziproductů pro dané odvětví (Například když do mladoboleslavské Škodovky vstoupila skupin VW Group, zvýšil se sice prodej Škodovek na českém trhu, ale většina dílů pro jejich výrobu se dovážela z jiných továren Volkswagenu, převážně z Německa.)

Základní proměnnou pro ověření Heckscher-Ohlinova modelu je poměr nákladů na kapitál a práci v příslušném odvětví a časovém období. [Krugman, Obstfeld, 2003] Česká republika je pro tento účel tradičně považována za zemi spíše lépe vybavenou pracovní silou než kapitálem (určitě alespoň vzhledem k zemím Evropské unie). Pokud by tato hypotéza měla být udržitelná, znamenalo by to, že koeficient u proměnné KL v modelu pro vývoz by měl být záporný a v modelu pro dovoz ze zemí Evropské unie kladný. Tím by se prokázal vyšší vývoz v odvětvích náročných na pracovní sílu a vyšší dovoz v těch částech ekonomiky, které potřebují velké množství fyzického kapitálu.

Další vysvětlující proměnné v modelech pro import i export jsou importní a exportní ceny. Jak jsem již naznačil v kapitole o analyzovaných datech, není jednoznačné vysvětlení významu cen úplně jednoduché. Dále budeme předpokládat platnost hypotézy jedné ceny, tedy že zvýšení ceny znamená zároveň zvýšení kvality a naopak. Potom by kladná cenová elasticita vývozu naznačovala, že české zboží konkuruje na evropských trzích kvalitou, záporná, že konkuruje cenou. V případě dovozů je situace analogická: kladná elasticita = evropské zboží konkuruje na českém trhu kvalitou, záporná elasticita = konkuruje cenou. U téhle proměnné bude navíc zajímavé sledovat vývoj v čase, jestli se původní cenová konkurence změní na kvalitativní nebo naopak.

Poměr přidané hodnoty daného odvětví vzhledem k přidané hodnotě celé ekonomiky je do modelu zařazen, aby bylo možné měřit relativní význam odvětví pro celé národní hospodářství. Znaménko koeficientu u této veličiny by mělo být kladné jak pro vývoz, tak pro dovoz, potvrzující interpretaci, že v průměru větší odvětví víc dováží i vyváží.

Reálný směnný kurs je makroekonomický ukazatel, který má velký vliv na vývoj exportu i importu. Jeho růst zdražuje domácí měnu vůči cizí, což za jinak stejných podmínek zdraží vývoz a zlevní dovoz. Při zvýšení RER české koruny vzhledem k Euru by tedy měl poklesnout český export do zemí Evropské unie a naopak import z těchto zemí kvůli relativnímu zlevnění vzrůst. Pro vývoz tedy očekáváme zápornou „kursovou“ elasticitu a pro dovoz kladnou.

Posledními proměnnými v modelu je hrubý domácí produkt České republiky a zemí Evropské unie. Ty zde slouží jako ukazatel velikosti cílového trhu pro export a import. Měly by tedy mít v obou modelech koeficienty s kladným znaménkem, protože větší trh je schopen absorbovat větší množství dovezeného zboží.

Na závěr této kapitoly můžeme ještě učinit předpovědi ohledně chování modelů při použití robustní regresní metody LTS. Pokud předpokládáme, že v průběhu devadesátých let v České republice postupovala ekonomická transformace, bude naší hypotézou, že odhadované modely pro jednotlivé roky budou stále lépe a lépe vysvětlovat vývoz a dovoz.

Navíc se někdy při použití metody LTS může podařit přesně odhadnout počet odvětví které do modelu skutečně patří. A to v tom případě, že když postupně zvyšujeme počet požadovaných odvětví v modelu, dojde v jednom místě k výraznému zlomu v koeficientech α (nebo) k prudkému poklesu vysvětlovací schopnosti modelu vyjádřené koeficientem determinace. Potom můžeme prohlásit, že pozorování, která jsou do modelu zařazena těsně předtím, než ke zmíněnému „skoku“ dojde, jsou skutečně transformovaná odvětví a ostatní odvětví ještě neprošla dostatečně hlubokou restrukturalizací, která by je přiblížila tržním poměrům.

6. Průběh výpočtů - diagnostika

6.1. Multikolinearita

Je nepříjemné, když se v modelu vyskytnou dvě nebo více vysvětlujících proměnných, které jsou navzájem silně korelované. Tento fakt může mít vliv na celou řadu získaných výsledků, hlavně může ovlivnit získané koeficienty u takto korelovaných faktorů a také jejich signifikanci. Když budou v modelu dvě proměnné, které jsou velmi silně korelovány, můžou se „přetahovat“ o vliv a to sníží jejich významnost. Nebo dokonce můžou mít koeficienty protichůdné znaménka, když jedna proměnná působí „příliš silně“ a druhá se tento vliv snaží korigovat.

Proto je třeba předem ošetřit, jestli nejsou některé proměnné v modelu navzájem korelovány. Z korelační matice pro data o exportu pro celkový model vidíme, že RER – reálný směnný kurs je korelován s hrubým domácím produktem zemí Evropské unie (GEU).

	px	va	kl	fdi	geu	rer
px	1	-0.02	-0.61	-0.01	0.01	0.01
va	-0.02	1	0.04	0.57	-0.01	-0.02
kl	-0.61	0.04	1	0.2	0.06	0.06
fdi	-0.01	0.57	0.2	1	0.34	0.33
geu	0.01	-0.01	0.06	0.34	1	0.98
rer	0.01	-0.02	0.06	0.33	0.98	1

Tabulka 1: Korelační matice vysvětlujících proměnných v modelu pro export

Koeficient 0.98 ukazuje na silnou závislost. Když ověříme ještě druhé přibližné kritérium pro kolinearitu – číslo podmíněnosti matice vysvětlujících proměnných, dospějeme k hodnotě 733. Oba tyto důvody nás vedou k jedinému možnému řešení – měli bychom z modelu vynechat jednu z dvojice proměnných RER-GEU, nebo si alespoň být vědomi toho, že kolinearita v modelu zhoršuje vlastnosti našich odhadů.

Provedeme proto předběžnou diagnostiku – metodou LTS odhadneme model ve dvou variantách – bez koeficientu RER a s ním. Pokud se odhady budou výrazně lišit nebo budou nestabilní, budeme se muset rozhodnout pro model bez koeficientu RER. Výpočty jsem provedl pro obě varianty a dospěl jsem k závěru, že model s koeficientem RER je sice mnohem méně stabilní co se týče změn koeficientů při přidávání dalších pozorování a signifikance proměnných je v něm výrazně nižší, ale rozdíly mezi oběma modely jsou natolik zajímavé, že uvedu výsledky pro oba dva. V následující tabulce (*tabulka 2*) vidíme odhady obou modelů pro 314 pozorování (později

dojdeme k závěru, že to je přibližně optimální počet pozorování, které je možné zařadit do modelu pro „lepší skupinu odvětví“):

Model s proměnnou RER

Koeficienty:		
	Odhad	Pr(> t)
()	-5.22824	0.36313
px	0.08585	0.00107 **
va	0.59184	< 2e-16 ***
kl	-0.24724	1.73e-05 ***
fdi	0.15399	3.20e-13 ***
geu	2.39640	0.15339
rer	-1.09552	0.60602

R-Squared: 0.8318

Model bez RER

Koeficienty:		
	Odhad	Pr(> t)
()	-2.81078	0.39743
px	0.08606	0.00102 **
va	0.59212	< 2e-16 ***
kl	-0.24663	1.77e-05 ***
fdi	0.15394	3.04e-13 ***
geu	1.55330	3.46e-05 ***

R-Squared: 0.8322

Tabulka 2: Porovnání koeficientů pro model s a bez proměnné RER

Z porovnání obou modelů můžeme odvodit několik důležitých faktů. Zaprvé proměnná RER ani GEU není v prvním modelu (se směnným kursem) signifikantní. To je typický průvodní jev multikolinearity. Tyto dvě proměnné jsou navzájem silně korelovány a tak působí proti sobě. Nicméně toto působení je v souladu s ekonomickou teorií, protože směnný kurs vyrovnává vysokou důchodovou elasticitu českých vývozu. Můžeme říci, že kdyby v devadesátých letech koruna tak silně neposilovala, rostl by český export do Evropské unie ještě mnohem rychleji.

Pokud bychom použili pouze druhý model, získali bychom důchodovou elasticitu českých vývozu do EU na mnohem nižší úrovni a o působení směnného kurzu bychom se nedozvěděli nic. To je dobrý příklad toho, že když si uvědomíme, jaké problémy při odhadu nastávají, můžeme využít i jinak nekvalitní odhad, pokud ho ověříme ještě jiným způsobem a správně interpretujeme.

Nyní obraťme pozornost k designové matici pro import. I zde by se mohl objevit problém multikolinearity. Vypočítáme korelační matici:

	pm	va	kl	fdi	gcz	rer
pm	1	0.04	-0.61	0	-0.04	-0.06
va	0.04	1	0.04	0.57	0	-0.02
kl	-0.61	0.04	1	0.20	0.06	0.06
fdi	0	0.57	0.20	1	0.29	0.33
gcz	-0.04	0	0.06	0.29	1	0.76
rer	-0.06	-0.02	0.06	0.33	0.76	1

Tabulka 3: Korelační matice vysvětlujících proměnných v modelu pro import

Vidíme, že český HDP je také svázán s reálným směnným kursem, ale ne tak silně jako v předchozím případě HDP zemí Evropské unie. Číslo podmíněnosti je rovno 340. Znamená to, že hrozí stejné nebezpečí jako v předchozím případě, sice že multikolinearita zhorší signifikanci těchto dvou závislých proměnných. Porovnání obou odhadů necháme až do kapitoly o výsledcích analýzy modelu pro import.

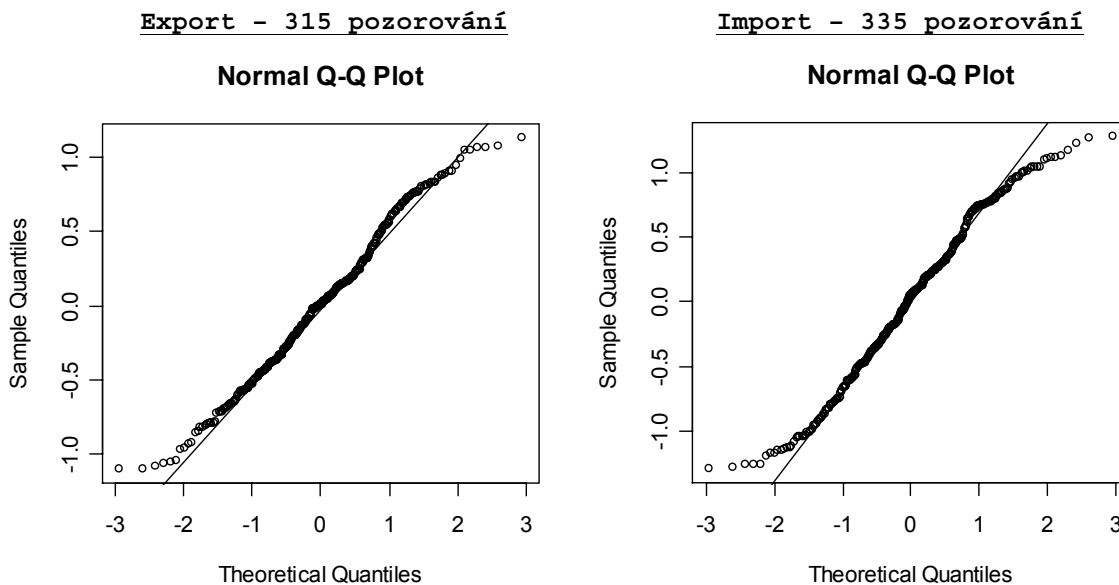
Rovněž je potřeba otestovat kolinearitu v modelech pro jednotlivé roky. Korelační matice pro vysvětlující proměnné exportu i importu jsou pro všechny roky 1993 – 1999 velmi podobné maticím pro celkový model a nejvyšší korelační koeficienty dosahují v absolutní hodnotě 0.75 (model pro jednotlivé roky neobsahuje proměnné RER a GEU/GCZ, protože by byly pro všechna pozorování stejné – GDP a směnný kurs se měří pro celou ekonomiku). Korelační koeficienty mezi 0.70 a 0.75 v těchto maticích jsou dosaženy pro FDI a VA – to je celkem logické, že do významnějších odvětví bude směřovat víc zahraničních investic. Protože jsou čísla podmíněnosti těchto matic většinou kolem hodnoty 10, nemusíme se kolinearitou obávat.

6.2. Normalita

Aby byl odhad metodou LTS pro vybranou skupinu pozorování nejlepším nestranným odhadem, měla by mít rezidua alespoň přibližně normální rozdělení. Pro ověření normality reziduí použijeme Shapiro-Wilkovu statistiku [Víšek 1999b:88].

Výsledky testování ukazují nepříjemný fakt. Protože p-hodnota pro Shapiro-Wilkovu v modelu pro import pro žádnou dělicí konstantu h nedosahuje ani 0,01, musíme vycházet z toho, že předpoklad normality není splněn. Naopak v modelu pro export se pro určité hodnoty dělicí konstanty h nedá normalita vyvrátit – je to jednak pro hodnoty kolem 309 až 314 a potom 340 až 350. Je zajímavé, že, jak později zjistíme, hranice normality na hodnotě h rovné 314 koliduje s optimálním množstvím pozorování v modelu pro export – rovněž 314.

Pro kontrolu ještě použijeme normálový graf. Na grafu kvantilů reziduí z modelu pro export v závislosti na kvantilech normálního rozdělení vidíme, že body grafu přibližně vystihují přímkou příslušného normálního rozdělení. Graf pro rezidua z modelu pro import je již trochu odlišný. Vidíme, že v levé části je graf konvexní a v pravé konkávní. To by naznačovalo, že rozdělení reziduí je spíše *long tailed*, že má „těžké chvosty“ a je spíše bližší t-rozdělení pro nízký počet stupňů volnosti. Odlišnosti vůči normalitě ale nejsou příliš výrazné, proto nebudeme podnikat žádná opatření k zajištění normálního rozdělení reziduí (třeba Box-Coxovu transformaci [Greene, 1993:329]).



Grafy 2 a 3: Normalita v celkovém modelu pro export a pro import

Kromě testování normality pro celkové modely exportu a importu bylo nutné ještě otestovat normalitu reziduí pro modely jednotlivých let. Také zde jsem použil Shapiro-Wilkovu statistiku a nyní s mnohem přívětivějším výsledkem. Až na několik výjimek se ukázalo, že rezidua v obou modelech pro všechny roky mají normální rozdělení.

6.3. Homoskedasticita

Dalším z předpokladů, které je třeba otestovat předtím, než můžeme začít rozebírat výsledky regresní analýzy, je homoskedasticita, tedy rovnost rozptylů pro všechna pozorování v modelu. Pro testování homoskedasticity jsem použil Brausch-Paganův test homoskedasticity [Višek 1999b:84].

Zde už se projeví výhody metody LTS: Zatímco pro vysoké hodnoty dělicí konstanty h byly oba celkové modely zřetelně heteroskedastické, pro nižší hodnoty tomu bylo právě naopak. V modelu pro export byl předpoklad homoskedasticity narušen až potom, co jsem do modelu zařadil více než 288 pozorování, v modelu pro import už při 265 pozorováních ze 427. Modely sestavené po jednotlivých letech v použitém testu nevykazovaly žádné známky heteroskedasticity.

6.4. Autokorelace

Obvyklým testem na autokorelaci reziduí v případě časové řady je Durbin-Watsonova statistika [např.: Johnston, DiNardo 1997:429]. Pro panelová data tento test striktně vzato není použitelný, protože část sousedících pozorování jsou různá odvětví a nikoli stejné odvětví v různých letech. Navíc pro výsledná pozorování vybraná metodou LTS je mnohem větší pravděpodobnost, že

budou vedle sebe různá odvětví, protože je možné, že některé odvětví nebude do modelu zařazeno pro všechny roky. Nicméně i takto „zeslabená“ Durbin-Watsonova statistika (nejméně 1/7 dvojic pozorování by měla působit proti určení správné autokorelace) dává znepokojivé výsledky. V odhadu celkového modelu pro export i import jsou hodnoty DW-statistiky velmi nízké. Pro minimální dělicí konstantu 214 je to zhruba 1,2 a se zvyšováním h hodnota statistiky klesá až na 0,6. Příslušné P-hodnoty tak pro všechny možnosti „potvrzují“ hypotézu, že rezidua jsou kladně korelovaná.

Náprava sériové korelace reziduí pro celkový model by byla velmi obtížná, ne-li nemožná, protože zkoumám panelová data a navíc používám metodu, která vyřazuje z modelu jednotlivá pozorování (bez použití metody LTS by se dal aplikovat některý z postupů navržených v [Greene, 1993:456]). I z tohoto důvodu jsem se nespolehal jenom na odhad modelu pro všechny roky dohromady, ale odhadoval jsem i rovnici pro export a import za každý rok zvlášť – pro tyto jednotlivé modely samozřejmě autokorelace nehrozí, protože se v nich vůbec nestřetávají hodnoty z různých časových období.

7. Výsledky analýzy – export

7.1. Souhrnný model pro export

Model pro export z České republiky do Evropské unie sestavený v kapitole 5 jsem testoval metodou LTS pro všechny hodnoty dělicí konstanty h od 214 (polovina počtu pozorování) až do 427 (počet všech pozorování – 61 odvětví pro roky 1993 až 1999). Výpočet jsem prováděl tak, aby program vždy našel 40-krát stejnou hodnotu nebo provedl 100 iterací. Navíc jsem ještě celý postup desetkrát zopakoval, přičemž se vždy nahrazovaly horší výsledky pro dané h lepšími (měřeno podle hodnoty reziduálního součtu čtverců – čím nižší, tím lepší).

Asi nejdůležitějším výsledkem, kterého se mi pro celkový model podařilo dosáhnout, je určení optimálního počtu pozorování, které by bylo vhodné do modelu zahrnout. Ve výsledných koeficientech se opravdu objevil skok, který se na stejném místě vyskytl i v hodnotě koeficientu determinace (byť ne tak výrazně) a rovněž v signifikanci některých koeficientů. Zlom se vyskytl na téměř stejném místě pro model se zahrnutou vysvětlující proměnnou RER i pro model bez ní. Došlo k tomu při hodnotě dělicí konstanty h 315 (314 v modelu se zahrnutým reálným směnným kurzem RER). Dále budeme v souladu s hypotézou předpokládat, že do „skoku“ model obsahuje pouze odvětví již plně transformovaná (popřípadě za roky, ve kterých již byla transformovaná), a po skoku již mají v modelu významný vliv i odvětví, která ještě neprošla restrukturalizací (Jak je vidět se zařazování, po skoku se změni zařazení asi 10 pozorování do modelu, především se odeberou některá, která tam již byla, a místo toho se objeví asi dvě nová odvětví pro všech 7 let.).

Nejvýraznější je skok pro hodnotu koeficientu u proměnné K/L . Do bodu zlomu se hodnota koeficientu pohybuje mezi 0,20 až 0,25, při zlomu náhle vyskočí na hodnotu 0,60, kolem které se pohybuje až do maximální výše dělicí konstanty h . Obě hodnoty naznačují, že vyšší export lze očekávat v odvětvích, která spoléhají spíše na využití české pracovní síly než na nákladný fyzický kapitál. Protože je ale hodnota před skokem mnohem nižší, vyvozují z toho, že u transformovaných odvětví není tento efekt tak silný jako u odvětví, která ještě neprošla restrukturalizací.

K další změně dochází v hodnotách koeficientu u exportních cen (PX). Do bodu zlomu je kladný, naznačující, že odvětví, která nemají s exportem problém, soupeří na Evropském trhu vyšší kvalitou. Potom dojde k jeho změně na záporný, to znamená, že ostatní odvětví, která dosud neukončila transformaci, konkurují evropským firmám spíše nižší cenou.

Hodnota koeficientu u přidané hodnoty je podle očekávání kladná. Také koeficient u FDI je kladný, což nám ověřuje platnost hypotézy, že odvětví, do kterých přitéká více přímých zahraničních investic, také více vyvážejí. Protože jsou přímé zahraniční investice indikátorem

„lidského kapitálu“ v odvětví, jsou korelovány s ricardiánským ukazatelem „celkové produktivity faktorů“ a kladným koeficientem se tak zároveň potvrzuje platnost Ricardovy teorie zahraničního obchodu i pro obchod České republiky s Evropskou unií.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-2.22047	3.32951	-0.667	0.50533	
px	0.08228	0.02607	3.156	0.00176	**
va	0.59187	0.02521	23.474	< 2.00E-16	***
kl	-0.24967	0.05692	-4.386	1.58E-05	***
fdi	0.15345	0.02032	7.55	4.93E-13	***
geu	1.49293	0.37115	4.022	7.25E-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.5045 on 309 degrees of freedom
Multiple R-Squared: 0.8321, Adjusted R-squared: 0.8293
F-statistic: 306.2 on 5 and 309 DF, p-value: < 2.2e-16

Tabulka 4: Koeficienty v modelu pro export, 315 pozorování (před skokem)

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-2.07332	3.34237	-0.62	0.536	
px	-0.13372	0.02459	-5.439	1.09E-07	***
va	0.54115	0.02563	21.118	< 2.00E-16	***
kl	-0.62002	0.05168	-11.997	< 2.00E-16	***
fdi	0.18356	0.02098	8.751	< 2.00E-16	***
geu	1.9134	0.37194	5.144	4.77E-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.5082 on 310 degrees of freedom
Multiple R-Squared: 0.8222, Adjusted R-squared: 0.8194
F-statistic: 286.8 on 5 and 310 DF, p-value: < 2.2e-16

Tabulka 5: Koeficienty v modelu pro export, 316 pozorování (po skoku)

Uvedené hodnoty byly zhruba stejné pro oba modely, ten bez RER i ten s tímto ukazatelem. Zajímavější je, jak se rozložil koeficient u GEU při přidání proměnné RER do modelu. V modelu „bez“ je hodnota koeficientu „před skokem“ asi 1,5 a „po skoku“ asi 2. To znamená, že méně „přizpůsobená“ odvětví více spoléhají na důchodovou elasticitu při vývozu do Evropské unie a jsou závislejší na případných negativních výkyvech ekonomické situace v EU.

V modelu s proměnnou RER je situace trochu složitější, důchodová elasticita (koeficient u GEU) je zde před skokem asi 2,4 a po skoku náhle vzroste na 5,4. Proti ní působí koeficient u RER, který je před skokem asi -1,1 a po skoku - 4,1. Opět zde vidíme, že „netransformovaná odvětví“ jsou (ještě silněji) závislá na výkyvech agregátní poptávky v EU a směnného kurzu. A rozdíl mezi

1,5 a 2,4 pro „transformovaná“ odvětví znamená, že kdyby býval reálný směnný kurs české koruny nerostl v devadesátých letech tak rychle, mohl by být růst vývozu ještě silnější.

Zbývá jen dodat, že v obou modelech jsou všechny vysvětlující proměnné (s jedinou výjimkou) statisticky významné. V průběhu výpočtu docházelo ke změnám signifikance pro různou hodnotu konstanty h , většinou se jednalo o marginální efekty (např. pro hodnoty h 276-286 byla proměnná K/L nesignifikantní). Tou výjimkou v signifikanci jsou proměnné GEU a RER v druhém modelu, které se díky kolinearitě navzájem přetlačují.

Poslední zajímavost, které si můžeme pro tyto dva modely všimnout, je relativně významný pokles koeficientu determinace při přechodu přes „skok“. Zatímco standardně při změně dělicí konstanty h docházelo k jeho změně asi o 0,002, zde se v obou modelech najednou posune o 0,01. V modelech pro import, jak uvidíme, je tento zlom ještě výraznější.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-5.22824	5.74045	-0.911	0.36313	
px	0.08585	0.02598	3.304	0.00107	**
va	0.59184	0.02508	23.599	< 2.00E-16	***
kl	-0.24724	0.05663	-4.366	1.73E-05	***
fdi	0.15399	0.02021	7.619	3.20E-13	***
geu	2.3964	1.67438	1.431	0.15339	
rer	-1.09552	2.12189	-0.516	0.60602	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Residual standard error: 0.5017 on 307 degrees of freedom
 Multiple R-Squared: 0.835, Adjusted R-squared: 0.8318
 F-statistic: 258.9 on 6 and 307 DF, p-value: < 2.2e-16

Tabulka 6: Koeficienty v modelu pro export se zahrnutím směnného kursu, 314 pozorování (před skokem)

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-13.3779	5.7366	-2.332	0.02034	*
px	-0.16344	0.02419	-6.757	7.04E-11	***
va	0.53737	0.02578	20.841	< 2.00E-16	***
kl	-0.6394	0.05128	-12.469	< 2.00E-16	***
fdi	0.17499	0.02113	8.28	3.79E-15	***
geu	5.37477	1.68503	3.19	0.00157	**
rer	-4.12314	2.1425	-1.924	0.05522	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Residual standard error: 0.5052 on 308 degrees of freedom
 Multiple R-Squared: 0.8197, Adjusted R-squared: 0.8161
 F-statistic: 233.3 on 6 and 308 DF, p-value: < 2.2e-16

Tabulka 7: Koeficienty v modelu pro export se zahrnutím směnného kursu, 315 pozorování (po skoku)

Koeficient determinace je pro oba modely na úrovni přijatelných 0,83, hodnotě, které by bez metody LTS bylo jen velmi obtížné dosáhnout (koeficient determinace pro metodu OLS je 0,45). V příloze B je ještě uvedena tabulka, jak se vyvíjely koeficienty pro jednotlivé hodnoty konstanty h , jsou tam průměrné hodnoty vždy pro 20 hodnot h pro oba dva modely.

7.2. Rovnice pro export pro každý rok zvlášť

Data pro jednotlivé roky obsahovala čtyři vysvětlující proměnné a 61 pozorování. Výsledky pro každý rok zvlášť jsem počítal podobně jako pro celkový model, opět metodou LTS a opět desetkrát pro každé nastavení h . Zajímavé bylo sledovat, jak se vyvíjela hodnota koeficientu determinace pro dané množství pozorování v modelu v jednotlivých letech. Ale ještě zajímavější je vidět, jak se mění počet odvětví, zařazených do modelu, když chceme zachovat danou hodnotu koeficientu determinace:

R²	1993	1994	1995	1996	1997	1998	1999
0,90	39	38	40	42	43	43	44
0,85	43	42	46	46	46	47	47
ZLOM	39	41	42	43	46	47	51

Tabulka 9: Počet odvětví v modelu pro export pro daný rok, aby koeficient determinace byl vyšší než zadaná hodnota

Vidíme, že počet odvětví, která v modelu dosáhnou daného koeficientu determinace, s postupem času narůstá. To znamená, že se jednotlivá odvětví s postupem času transformují a přibližují našemu modelu. V posledním řádku označeném zlom vidíme počet pozorování v modelu, při kterém pro daný rok došlo ke „skoku“ v koeficientech, stejně jako v celkovém modelu pro export. Tady je vidět průběh transformace ještě výrazněji.

Nyní se podívejme, jak se v průběhu devadesátých let měnily koeficienty v našem modelu pro jednotlivé roky. Uvedené hodnoty koeficientů jsou pro hodnotu h těsně před skokem. (Kromě roku 1995, který je zjevně atypický. Rok 1995 má až do hodnoty $h = 50$ koeficient u K/L významně kladný a do hodnoty $h = 42$ má záporný koeficient u FDI. V koeficientech je pro tento rok hned několik skoků, jeden pro hodnotu 44 a druhý pro $h = 52$. Použijeme proto hodnoty pro $h = 52$.)

V tabulce 10 vidíme, že vývoj většiny koeficientů byl převážně příznivý. Strategie prosadit se nižší cenou se v průběhu let změnila na konkurenci v kvalitě (koeficient u PX změnil znaménko ze záporného na kladné). Dále také klesl koeficient u FDI, což se dá interpretovat tak, že klesl význam přímých zahraničních investic pro export a vyvázejí už i odvětví, do kterých tolik zahraničního kapitálu nepřichází. (V souvislosti s konkurováním kvalitou můžeme zamítnout

hypotézu, že bychom začali vyvážet „nekvalifikovanou práci“ v odvětvích s nízkou úrovní lidského kapitálu.)

rok	(Intercept)	px	va	kl	fdi
1993	10.71	-0.012	0.619	-0.119	0.139
1994	12.106	-0.124	0.592	-0.178	0.134
1995	10.479	0.126	0.658	-0.203	0.089
1996	10.903	0.035	0.647	-0.167	0.132
1997	10.002	0.179	0.666	-0.169	0.073
1998	10.339	0.212	0.591	-0.267	0.175
1999	8.709	0.332	0.707	-0.134	0.045

Tabulka 10: Koeficienty modelu pro export, roky 1993 - 1999

Koeficient u K/L je záporný a v průběhu let mírně narůstá, což naznačuje, že Česká republika pořád ještě není zemí relativně bohatou na fyzický kapitál. Je ovšem třeba poznamenat důležitý fakt, že proměnná KL v modelech uvedených v tabulce není statisticky významná, tedy že nemůžeme z jejích hodnot dělat nijak zvlášť silné závěry. Ostatní proměnné v modelech už významné jsou, kromě koeficientu u PM v letech 1994-1996 a FDI v letech 1995, 1997 a 1999.

7.3. Porovnání obou přístupů

Pro porovnání obou modelů zahraničního obchodu – po jednotlivých letech a celkově – použijeme tabulku 11: zařazování jednotlivých odvětví do modelu. Zároveň tím přiblížíme teoretické výsledky blíž praktické rovině, tedy kdo vyvážá a kdo ne. Sloupec roky v tabulce znamená model pro daný rok s „přirozeným počtem pozorování“ – tedy před „bodem zlomu“, sloupec celk. označuje celkový model pro 315 pozorování (bez RER). Písmeno A znamená, že dané odvětví je v příslušném roce do modelu zařazeno. V levém sloupci jsou anglické názvy odvětví a jejich číselné označení podle SITC.

Vidíme, že zatímco odvětví v horní části tabulky jsou často do modelu nezařazena (je to především potravinářský průmysl, některé suroviny nebo farmaceutický průmysl), jsou téměř všechna odvětví ve spodní části tabulky zařazena stále (to je strojírenství, elektrotechnika, spotřební průmysl). U některých odvětví dochází k tomu, že zatímco v prvních letech nejsou do modelu zařazena, tak v pozdější době k jejich zařazení dochází – můžeme to brát jako signál, že tato odvětví se stala konkurenceschopnými a začala vyvážet do Evropské unie podle našeho modelu. Jsou to třeba odvětví kancelářské stroje, telekomunikace, přesné strojírenství a transportní zařízení, z chemického průmyslu parfémů a plasty.

Některá odvětví jsou jako výjimky ve svém oboru zařazena: z potravinářského průmyslu jsou to ryby, zelenina a ovoce a průmysl krmiv. Naopak ze surovin není zařazen vývoz dřeva a

zemního plynu. Takovéto odchylky jsou většinou způsobeny tím, že dané subodvětví nějakým způsobem není standardní – třeba vývoz ryb může znamenat jihočeské kapry nebo do zemního plynu může být zahrnuta tranzitní doprava přes naše území.

Asi nejdůležitější na této tabulce je fakt, že jak do celkového modelu tak do modelů po jednotlivých letech byla zařazena zhruba stejná odvětví. To do jisté míry potvrzuje vhodný výběr metody pro daná data: Přestože byla data nezávisle na sobě zpracována dvěma odlišnými způsoby, výsledky se téměř shodují. K přibližné shodě dochází i pro koeficienty (viz tabulky 4, 6 a 10).

rok	typ odhadu	93		94		95		96		97		98		99	
		roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.
Meat	01														
Milk, eggs	02	A	A	A	A	A	A	A	A					A	
Fish	03	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Cereals, flour	04					A	A							A	
Fruits, vegetables	05	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Suggar	06													A	
Coffee, tea	07														
Animal fodder	08	A	A		A		A	A	A	A	A	A		A	
Other food	09														
Drinks	11														A
Tobacco	12														
Leather	21	A	A	A	A		A	A	A	A	A	A	A	A	A
Oil seeds	22	A	A	A	A	A	A	A	A	A	A	A	A		
Raw rubber	23	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Raw wood	24														
Pulp, cellulosis	25	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Textile fibres	26	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Minerals, fertilizers	27	A	A		A	A	A	A	A	A	A	A	A	A	
Iron ore	28		A	A	A	A	A	A	A	A	A	A	A	A	A
Organic raw inputs	29	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Coal, coke	32	A	A	A	A		A	A	A	A	A	A	A	A	A
Crude oil, petrol	33		A					A	A	A	A	A	A	A	A
Natur.+ indust. gas	34														
Animal oils, fats	41														
Raw vegetable oils	42														
Processed oils	43														A
Organic chemistry	51	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Anorganic chem.	52	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Paints, pigments	53	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Pharmaceuticals	54														
Perfumes	55						A					A	A	A	A
Artif. Fertilizers	56	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Raw plastic mat.	57	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Processed plastic	58					A	A	A	A	A	A	A	A	A	A
Other chemicals	59		A				A	A	A	A	A	A	A	A	A
Processed leather	61	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Rubber products	62	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Processed wood	63	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Paper	64	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Textile products	65	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Non-metallic prod.	66	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Iron, steel	67	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Non-ferrous metals	68	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Metallic products	69	A	A	A	A	A	A	A	A	A	A	A	A	A	A

rok typ odhadu	93		94		95		96		97		98		99	
	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.
Machines-energy 71	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Industrial machines 72	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Metalworking mach. 73	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Mach.eq.-heat,etc. 74	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Business machines 75		A		A		A		A		A		A		A
Telecom, recorders 76			A	A	A	A	A	A	A	A	A	A	A	A
Electr. appl., equip. 77	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Road vehicles 78	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Transport. equip. 79		A		A		A		A		A		A		A
Prefab.build., equip. 81	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Furniture 82	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Haberdashery 83	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Clothing, apparel 84	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Footware 85	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Precision instrum. 87			A		A		A		A		A		A	
Optical instrum. 88	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Other products 89	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Počet (roky/celk.)	39	44	41	43	42	47	43	46	46	45	47	46	51	44

Tabulka 11: Zařazování odvětví do modelů pro export v závislosti na čase, roky 1993 - 1999

8. Výsledky analýzy – import

8.1. Souhrnný model pro import

Celkový model pro import jsem odhadoval stejně jako model pro export. Zmíněný skok v koeficientech, který se objevil v modelu pro export, se dá nalézt i v modelu pro import. Zde je ještě výraznější. Koeficient determinace s přidáním jednoho dalšího pozorování poklesl z 0,83 na 0,77, tedy o 0,06. Bylo to při hodnotě konstanty $h = 335$, proto jsem usoudil, že optimální počet pozorování v modelu je roven 335, tedy asi třem čtvrtinám ze všech importních odvětví. Výrazný skok byl ovšem pouze v modelu se zařazenou proměnnou RER, v modelu bez RER byl skok mnohem méně výrazný a došlo k němu až při $h = 348$.

Dále budu uvádět pouze hodnoty pro model s proměnnou RER: V následujících tabulkách (tabulky 12 a 13) vidíme výsledky před a po „skoku“.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-16.4307	2.75929	-5.955	6.71E-09	***
pm	0.45343	0.02738	16.563	2.00E-16	***
va	0.89446	0.04087	21.884	2.00E-16	***
kl	0.36472	0.06438	5.665	3.22E-08	***
fdi	-0.05408	0.03202	-1.689	0.0921	.
gcz	1.38648	0.30201	4.591	6.30E-06	***
rer	1.20241	0.80768	1.489	0.1375	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Residual standard error: 0.6278 on 328 degrees of freedom
 Multiple R-Squared: 0.8316, Adjusted R-squared: 0.8286
 F-statistic: 270 on 6 and 328 DF, p-value: < 2.2e-16

Tabulka 12: Koeficienty v modelu pro import, 335 pozorování (před skokem)

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-18.9348	2.72172	-6.957	1.89E-11	***
pm	0.46694	0.02695	17.328	2.00E-16	***
va	0.70808	0.0408	17.355	2.00E-16	***
kl	0.22178	0.06678	3.321	0.000998	***
fdi	-0.14097	0.03164	-4.456	1.15E-05	***
gcz	1.24755	0.29992	4.16	4.07E-05	***
rer	2.41722	0.81466	2.967	0.003226	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Residual standard error: 0.6306 on 329 degrees of freedom
 Multiple R-Squared: 0.7728, Adjusted R-squared: 0.7686
 F-statistic: 186.5 on 6 and 329 DF, p-value: < 2.2e-16

Tabulka 13: Koeficienty v modelu pro import, 336 pozorování (po skoku)

Opět se pokusím interpretovat získané výsledky: Největší změna při „skoku“ je ve výši koeficientu u reálného směnného kurzu. To se dá vysvětlit tak, že v méně konkurenceschopných odvětvích závisí import ve větší míře na směnném kurzu než v konkurenceschopnějších odvětvích. Tedy že české firmy v transformovaných odvětvích si „drží“ svůj trh, zatímco netransformovaná odvětví jsou vydána na pospas makroekonomickým vlivům jako je posilování koruny. Naproti tomu velikost odvětví a proměnná KL má v „transformovaných“ odvětvích větší váhu než v „netransformovaných“.

Zatímco pro transformovaná odvětví nemá výše zahraničních investic vliv na velikost importu, pro odvětví, která ještě neprošla restrukturalizací, snižují přímé zahraniční investice import z Evropské unie. Důležitý je ještě koeficient u proměnné PM, tedy importní ceny. Ten je kladný v obou případech, v zásadě tedy dovážíme více zboží s vyššími jednotkovými cenami.

8.2. Rovnice pro import pro každý rok zvlášť

Analýza importu po jednotlivých letech je velmi podobná stejné analýze pro export. Opět uvedu tabulku počtu odvětví v modelu pro minimální dané hodnoty koeficientu determinace a počtu zařazených odvětví, při němž dochází ke „skoku“ (tedy optimální počet odvětví v modelu). U importu po jednotlivých letech nicméně tento skok není tak výrazný a mnohdy je sporné, zda k němu vůbec došlo. Z tabulky 14 vidíme, že opět dochází ke zvyšování počtu odvětví zařazených do modelu s postupujícími roky.

R²	1993	1994	1995	1996	1997	1998	1999
0,90	35	43	42	41	43	46	46
0,85	41	47	47	46	48	49	50
ZLOM	45	45	43	49	49	48	54

Tabulka 14: Počet odvětví v modelu pro import pro daný rok, aby koeficient determinace byl vyšší než zadaná hodnota

Nyní se podívejme, jak vypadá vývoj koeficientů pro import v čase. Jediné dvě proměnné, které zaznamenávají výraznější změnu, jsou FDI a KL. KL se v průběhu času snižuje ze zhruba 0,5 na asi 0,2 – to se dá interpretovat jako snižování vlivu kapitálové náročnosti na importu do ČR z Evropské unie. FDI se přesouvá z kladných čísel do záporných, to znamená, že zpočátku znamenaly přímé zahraniční investice zvýšený import do odvětví, nyní už si (možná) zahraniční firmy nacházejí české subdodavatele a naopak jejich investice český import snižují. Ještě je samozřejmě třeba zmínit signifikanci proměnných – s výjimkou roku 1996 je proměnná KL

nesignifikantní a proměnná FDI je nesignifikantní v letech 1997 a 1999. V tabulce 15 jsou uvedeny hodnoty všech koeficientů, vždy pro optimální počet pozorování v modelu v daném roce

rok	(Intercept)	pm	va	kl	fdi
1993	2.801	0.51	0.569	0.497	0.057
1994	0.27	0.625	0.549	0.729	0.2
1995	1.542	0.515	0.983	0.75	-0.158
1996	4.616	0.435	0.983	0.371	-0.171
1997	5.031	0.482	1.23	0.285	-0.389
1998	6.323	0.351	1.194	0.195	-0.229
1999	5.719	0.505	0.916	0.168	-0.19

Tabulka 15: Koeficienty modelu pro import, roky 1993 - 1999

8.3. Porovnání obou přístupů

Pro import opět uvádím tabulku zařazování odvětví do modelu pro všechny roky 1993-1999. Tentokrát jsou modely pro jednotlivé roky vybrány tak, aby obsahovaly stejně pozorování a dohromady dávaly zhruba počet pozorování celkového modelu, to je model pro import těsně před „skokem“ – tedy 335 pozorování. Opět vidíme, že celkový model se s dílčími velmi dobře shoduje. Kdybychom tuto tabulku srovnali s tabulkou k modelům pro export, zjistíme, že se příliš neliší – rovněž často nejsou zařazena odvětví potravinářského průmyslu nebo energetických surovin. Z toho můžeme dovodit, že existuje souvislost mezi tím, jestli bude dané odvětví zařazeno do modelu pro export a do modelu pro import. Znamená to, že „stupeň transformovanosti“ odvětví je obdobný v oblasti importní i oblasti exportní.

rok	typ odhadu	93	94	95	96	97	98	99
		roky celk.	roky celk.	roky celk.	roky celk.	roky celk.	roky celk.	roky celk.
	Meat 01		A A				A	
	Milk, eggs 02	A A	A A	A A	A A			
	Fish 03	A A		A A	A A	A A	A A	A A
	Cereals, flour 04	A A	A A	A A	A A		A A	A A
	Fruits, vegetables 05							
	Suggar 06	A A	A A	A A	A A	A A	A A	A A
	Coffee, tea 07	A A	A A	A A	A A	A A	A A	A A
	Animal fodder 08	A A	A A	A A	A A	A A	A A	A A
	Other food 09	A A	A A	A A	A A	A A	A A	A A
	Drinks 11	A A				A		A
	Tobacco 12							
	Leather 21		A A	A A	A A	A A	A A	A A
	Oil seeds 22	A A	A A	A A	A A	A A	A A	A A
	Raw rubber 23	A A	A A	A A	A A	A A	A A	A A
	Raw wood 24	A A	A A	A A	A A	A A	A A	A A
	Pulp, cellusosis 25	A A	A A	A A	A A	A A	A A	A A
	Textile fibres 26	A A	A A	A A	A A	A A	A A	A A
	Minerals, fertilizers 27	A A	A A	A A	A A	A A	A A	A A

rok typ odhadu	93		94		95		96		97		98		99	
	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.	roky	celk.
Iron ore	28	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Organic raw inputs	29	A	A				A							
Coal, coke	32													
Crude oil, petrol	33		A A								A A		A	
Natur.+ indust. gas	34								A A				A	A
Animal oils, fats	41		A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A	
Raw vegetable oils	42	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Processed oils	43	A A	A		A A	A A	A A	A A	A		A A	A A	A A	A A
Organic chemistry	51	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Anorganic chem.	52			A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Paints, pigments	53	A												
Pharmaceuticals	54	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Perfumes	55				A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Artif. Fertilizers	56	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Raw plastic mat.	57	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Processed plastic	58	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Other chemicals	59	A A	A A	A A	A A	A A	A A	A A	A					
Processed leather	61			A A	A A	A A	A A	A A	A A	A A				
Rubber products	62	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Processed wood	63	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Paper	64	A A	A A	A A	A A	A A	A A	A A	A		A A	A A	A A	A A
Textile products	65	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Non-metallic prod.	66	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Iron, steel	67	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Non-ferrous metals	68	A A	A A	A A	A A	A A	A A	A A	A		A A	A A	A A	A A
Metallic products	69	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Machines-energy	71	A A	A		A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Industrial machines	72	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Metalworking mach.	73	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Mach.eq.-heat,etc.	74	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Business machines	75	A							A					
Telecom, recorders	76	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Electr. appl., equip.	77	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Road vehicles	78	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Transport. equip.	79										A A	A A	A A	A A
Prefab.build., equip.	81	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Furniture	82	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Haberdashery	83	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Clothing, apparel	84	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Footware	85		A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Precision instrum.	87	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Optical instrum.	88	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Other products	89	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A	A A
Součet (roky / celk.)		48 45	48 48	48 48	48 48	48 50	48 47	48 49	48 48	48 48	48 48	48 48	48 48	48 48

Tabulka 16: Zařazování odvětví do modelů pro import v závislosti na čase, roky 1993 - 1999

9. Závěr

Výsledky analýz ukázaly, že základní makroekonomické teorie o zahraničním obchodu lze použít i na obchod České republiky s Evropskou unií. Podařilo se identifikovat průmyslová odvětví, která se chovají velmi standardně podle ekonometrického modelu založeného na Cobb-Douglasově produkční funkci a navíc počet těchto odvětví mezi lety 1993 a 1999 významně stoupl z 39 na 51.

Hodnoty regresních koeficientů naznačují, že v české produkci pro export stále převažují odvětví s vyššími nároky na počet pracovníků než na množství fyzického kapitálu. Nicméně původně cenová konkurence na evropských trzích se v průběhu let změnila na kvalitativní. České exporty pozitivně ovlivňuje růst HDP v Evropské unii, důchodová elasticita dosahuje hodnot kolem 2,5. Naopak negativní vliv má téměř permanentní posilování české koruny, které okamžitě tento růst snižuje zápornou elasticitou o velikosti zhruba -1. Pozitivní vliv mají rovněž přímé zahraniční investice, které zvyšují úroveň lidského kapitálu v příslušných odvětvích.

Charakteristiky českého vývozu jsou obdobné, závisí pozitivně na výši HDP České republiky a na růstu reálného směnného kursu. Rovněž pozitivní vliv má vyšší podíl fyzického kapitálu na jednotku „práce“ v příslušném odvětví, naopak lehce negativně ovlivňuje import příliv zahraničních investic. Evropské zboží na českých trzích konkuruje kvalitou, stejně jako české zboží na trzích Evropské unie.

Hodnoty koeficientů potvrzují v tomto případě platnost jak Ricardovy, tak Heckscher-Ohlinovy teorie zahraničního obchodu. Odvětví s vyšší celkovou produktivitou faktorů (vyšší úroveň lidského kapitálu indikovaná přílivem zahraničních investic) mají komparativní výhodu a více vyvážejí. Odvětví s nižší produktivitou mají naopak vyšší objemy dovozů. V odvětvích s nízkým poměrem kapitálu na jednoho pracovníka Česká republika (v souladu s Heckscher-Ohlinovým modelem) více vyváží a v odvětvích s vysokým poměr kapitálu se orientujeme na dovoz.

Navíc analýza ukázala, jak je v tomto případě výhodné použití metody LTS. Podařilo se nám určit nejen hodnoty koeficientů pro větší skupinu „transformovaných“ odvětví, ale dokonce jsme přesně zjistili i velikost této skupiny a identifikovali průmyslová odvětví, která do ní patří.

Pokračováním výzkumu v této oblasti by měla být transformace navazujících časových řad, aby bylo možné předpovídat základní trendy českého vývozu a dovozu i do budoucna, nejen vysvětlovat minulost.

Seznam literatury:

- Benáček, V.** (1999): Teorie mezinárodního obchodu (in Kubista et al. - Mezinárodní ekonomické vztahy. Praha, Nakl. Editio HZ, 1999)
- Benáček, V., Prokop, L., Víšek, J. Á.** (2003a): Determining Factors of the Czech Foreign Trade Balance: Structural Issues in Trade Creation. Working Paper of Czech National Bank, no. 3, s.1-55
- Benáček, V., Prokop, L., Víšek, J. Á.** (2003b): Structure and Dynamics of Trade in a Small Economy in Transition before the EU Accession: The Case of Czech Exports and Imports. Interim Report IR-03-034, IIASA Laxenburg, Austria
- Berndt, E.** (1991): The Practice of Econometrics. Pearson Education, Boston
- Boček, P., Lachout, P.** (1995): Linear programming approach to LMS-estimation I. In: Computational Statistics and Data Analysis, 1995, 19, s. 129-134
- Čížek, P.** (2002): Essays on Robust Estimation in Econometrics. Disertační práce na CERGE-EI, Praha, 2002
- Greene, W. H.** (1993): Econometric Analysis. (Second Edition) Macmillan Press, New York
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A.** (1986): Robust Statistics. The Approach based on Influence Functions. Wiley, New York
- Johnston, J., DiNardo, J.** (1997): Econometric methods 4th ed. McGraw-Hill, New York
- Jurečková, J.** (1971): Nonparametric estimate of regression coefficients, The Annals of Mathematical Statistics, 42, 1328-1338
- Krugman, P. R., Obstfeld, M.** (2003): International Economics: Theory and Policy. Pearson Education, Boston
- Papageorgiu, H.** (2002): Harmonisation of Economic Statistical Classification and Related Transformation. Department of Mathematics, University of Athens
- Rousseeuw, P. J.** (1984): Least Median of Squares Regression. Journal of the American Statistical Association, 79, 871-880
- Rousseeuw, P. J., Van Driessen, K.** (1998): Computing LTS regression on large data sets. Technical Report, University of Antwerp, submitted
- Rousseeuw, P. J., Leroy, A. M.** (1987): Robust regression and outlier detection. Wiley, New York
- Víšek, J. Á.** (1994): A cautionary note on the method of Least Median of Square reconsidered. Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Prague, 1994, 254-259
- Víšek, J. Á.** (1999a): The least trimmed squares – random carriers. Bulletin of the Czech econometric society, Vol. 10/1999, s.1-30
- Víšek, J. Á.** (1999b): Ekonometrie I. (skripta). Karolinum 1999
- Víšek, J. Á.** (2000): Regression with high breakdown point. In Robust '2000: Proceedings of the 11th Conference on Robust Statistics, 324 - 356 (invited paper)

Ekonometrický model českého zahraničního obchodu se zeměmi Evropské unie

(Aplikace robustní metody LTS)

Abstrakt:

Cílem této práce je vytvořit ekonometrický model zahraničního obchodu České republiky s Evropskou unií pro roky 1993 – 1999 pomocí robustní regresní metody LTS. V práci je nalezen logaritmický model pro export a import a jsou zjištěny hodnoty koeficientů pro základní ekonomické proměnné z Ricardova a Heckscher-Ohlinova modelu zahraničního obchodu. Navíc práce identifikuje průmyslová odvětví, která se modelem řídí, a zjistí vzestupný trend jejich počtu v závislosti na čase. Z toho lze usuzovat na postupné dokončování transformace české ekonomiky.

Econometric model of the Czech foreign trade with countries of the European union

(An application of the robust LTS method)

Abstract:

The main goal of the thesis is to build an econometric model of the Czech foreign trade with European Union for the years 1993 – 1999 using the LTS robust regression method. We compose a logarithmical model for export and import and find out values of the coefficients for main economic variables contained in Ricardian and Heckscher-Ohlin model. Moreover we identify the economic branches that stick to the model and find out a growing trend of their number. Based on this, we can show the progress of the Czech transition in that time.

Příloha A: Komentovaný zdrojový kód programu v jazyce R

```
LTS<-function(y,x,h,s,maxopak,eps)
  # význam promenných je následující:
  # y - vektor vysvětlovaných promenných
  # x - matice vysvětlujících promenných
  # h - počet pozorování, která chceme zahrnout do modelu
  # s - počet, kolikrát se musí zopakovat správný výsledek,
  # aby program mohl skončit
  # maxopak - počet opakování, po kterých program skončí,
  # i když nebyl nalezen správný výsledek
  # eps - povolená tolerance při hledání shody výsledku
{
  #inicializace

n<-length(y)
p<-length(x)/n
xx<-cbind(seq(1,1,length=n),x)
  # p je počet sloupců matice x - tedy počet vysvětlujících promenných
  # minus 1 za intercept
  # xx bude rozšířená matice x i s vektorem jedniček (pro intercept)

opak<-1
  # opak zachycuje, kolikrát jsme náhodně vybrali počátečních p+1 pozorování

shoda<-0
ssmin<-1e100
  # shoda je počet, kolikrát jsme našli při opakování stejný model
  # ssmin je zatím nejmenší dosažený součet čtverců pro h pozorování

  # ZACÁTEK VNEJSÍHO CYKLU
  # vnější cyklus probíhá pokud jsme nezískali nejlepší model s-krát
  # nebo dokud počet opakování nedosáhne čísla maxopak

uprava<-1
while((shoda<s)&(opak<maxopak))
{

kappa<-1e20
while(kappa>1e6)
  {
    i<-sample(1:n,p+1)
    xprvni<-xx[i,]
    kappa<-kappa(xprvni)
  }
yprvni<-y[i]
  # i je náhodný výběr o rozsahu p+1, z něj vypočteme počáteční hodnoty
koeficientu,
  # dále už potom porovnáme vždy součet čtverců pro nejlepších h pozorování
  # vytvoříme designovou matici z náhodných p+1 pozorování a příslušný vektor y
  # pokud je matice špatně podmíněná, nemůžeme ji invertovat
  # zkoušíme to, dokud není v pořádku

b<-solve(xprvni,yprvni)
  # b je aktuální vektor odhadnutých parametrů včetně interceptu
  # solve znamená, že se první parametry (matice) invertuje
  # vyřešíme

r2<-(y-xx%*%b)^2
  # r2 je vektor čtverců reziduí pro všechna pozorování 1 až n
```

```

i<-sort(order(r2)[1:h])
# i je serazeny vektor indexu tech h pozorovani, ktere maji nejmensi rezidua

ssstare<-1e100
ss<-sum(r2[i])
# ss bude obsahovat soucet rezidui vybranych h pozorovani
# promenne ss a ssstare pouzijeme k porovnani, zda doslo ke snizeni
# souctu ctvercu pri iteracich

# ZACATEK VNITRNIHO CYKLU
# vnitri cyklus se bude provadet, dokud bude klesat soucet ctvercu,
# nebo dokud pocet iteraci neprekroci maxiter=1000

iter<-1
while((ss<ssstare)&(iter<1000))
{
ssstare<-ss
istare<-i
xv<-xx[i,]
yv<-y[i]
# xv je rozsirena matice regresoru,
# bude obsahovat aktualnich h pozorovani
# yv bude obsahovat vysvetlovanou promennou pro
# aktualnich h pozorovani
b<-solve(crossprod(xv,xv),t(xv)%*%yv)
r2<-(y-xx%*%b)^2
i<-sort(order(r2)[1:h])
ss<-sum(r2[i])
iter<-iter+1
}
# KONEC VNITRNIHO CYKLU
# ulozeni vysledku

if(abs(ssstare-ssmin)<eps)
{
shoda<-shoda+1
print(c('opak c.',opak,'shoda c.',shoda,'iteraci',iter))
}
# kdyz je shoda, vypiseme a zaznamename do indikatoru

if(ssstare<ssmin)
{
print(c(uprava,'uprava pri',opak,'opakovani'))
uprava<-uprava+1
ssmin<-ssstare
imin<-istare
shoda<-0
}

# pokud je soucet ctvercu zatim nejlepsi, zaznamename to do poctu zmen,
# a pamatujeme si, pro jaky vektor indexu situace nastala,
# pocitadlo shodnosti vynulujeme

opak<-opak+1
}
# KONEC VNEJSIHO CYKLU

print('Vysledek:')
print(summary(lm(y[imin]~x[imin,])))
print(c('pocet iteraci',opak))

```

```
print(c('vysledna suma ctvercu rezidui',sum(resid(lm(y[imin]~x[imin,]))^2)))  
  
  # vystupem funkce je vektor indexu vybranych  
  #      do vysledneho najlepsiho modelu  
return(imin);  
}
```

Příloha B: Tabulky výsledků

h	(intercept)	px	va	kl	fdi	geu
217-237	-4.079	0.099	0.717	-0.197	0.117	1.622
237-257	-3.04	0.08	0.699	-0.231	0.125	1.548
257-277	-2.793	0.072	0.678	-0.268	0.142	1.552
277-297	-2.834	0.119	0.663	-0.129	0.119	1.441
297-317	-2.323	0.069	0.615	-0.23	0.141	1.502
317-337	-2.703	-0.104	0.553	-0.584	0.174	1.934
337-357	-5.689	-0.084	0.569	-0.602	0.157	2.263
357-377	-10.306	-0.021	0.608	-0.565	0.097	2.705
377-397	-13.943	0.037	0.668	-0.487	0.039	3.005
397-417	-18.394	0.071	0.722	-0.417	-0.033	3.427

h	(intercept)	px	va	kl	fdi	geu	rer
217-237	-15.265	0.07	0.706	-0.215	0.111	5.39	-4.732
237-257	-8.798	0.07	0.695	-0.246	0.127	3.669	-2.792
257-277	-7.12	0.07	0.676	-0.264	0.142	3.104	-2.043
277-297	-4.724	0.112	0.659	-0.138	0.122	2.245	-1.113
297-317	-5.55	0.057	0.611	-0.249	0.141	2.554	-1.279
317-337	-7.118	-0.114	0.551	-0.597	0.171	3.403	-1.83
337-357	-11.807	-0.084	0.568	-0.601	0.157	4.365	-2.715
357-377	-20.306	-0.023	0.608	-0.566	0.099	6.184	-4.518
377-397	-22.878	0.039	0.665	-0.483	0.04	6.206	-4.223
397-417	-27.084	0.069	0.718	-0.422	-0.031	6.54	-4.09

Tabulka B1: Průměry koeficientů pro celkový model pro export

h	(intercept)	pm	va	kl	fdi	gcz
217-237	-8.833	0.331	1.219	0.476	-0.294	1.236
237-257	-7.111	0.34	1.149	0.452	-0.245	1.083
257-277	-10.566	0.355	1.091	0.438	-0.22	1.405
277-297	-15.8	0.401	1.02	0.424	-0.179	1.865
297-317	-15.796	0.416	0.964	0.386	-0.122	1.872
317-337	-15.447	0.438	0.908	0.352	-0.08	1.84
337-357	-13.853	0.469	0.796	0.294	-0.081	1.73
357-377	-13.099	0.497	0.774	0.292	-0.141	1.664
377-397	-12.7	0.502	0.737	0.303	-0.098	1.607
397-417	-9.731	0.463	0.596	0.108	0.006	1.469

h	(intercept)	pm	va	kl	fdi	gcz	rer
217-237	-13.749	0.333	1.23	0.486	-0.317	0.381	3.036
237-257	-11.662	0.341	1.197	0.472	-0.282	0.605	2.062
257-277	-14.936	0.379	1.071	0.446	-0.203	0.742	2.428
277-297	-19.75	0.411	1.015	0.43	-0.177	1.366	1.987
297-317	-18.057	0.418	0.983	0.397	-0.138	1.513	1.304
317-337	-17.921	0.438	0.905	0.358	-0.087	1.449	1.44
337-357	-18.444	0.477	0.74	0.23	-0.151	1.167	2.457
357-377	-16.377	0.496	0.8	0.308	-0.165	1.156	1.87
377-397	-15.454	0.501	0.743	0.298	-0.11	1.068	1.859
397-417	-13.61	0.468	0.61	0.125	-0.016	0.917	2.093

Tabulka B2: Průměry koeficientů pro celkový model pro import

Příloha C: Projekt bakalářské práce

Termín bakalářské zkoušky: letní semestr 2003/2004
Vedoucí bakalářské práce: Doc. RNDr. Jan Ámos Víšek
Autor bakalářské práce: Jakub Jeřábek

Téma: Model českého zahraničního obchodu se zeměmi Evropské unie

Cíl práce:

Cílem práce je popsat strukturu panelových dat o importu a exportu pro jednotlivá odvětví české ekonomiky za období 1993-2002. Ve velmi pravděpodobném případě heteroskedastických nebo jinak nekvalitních dat bude třeba pro vytvoření modelu použít robustní metody. Hypotézou je, že se podaří rozdělit odvětví do dvou skupin, kde u jedné bude ekonometrický model odpovídat nějaké standardní produkční funkci (např. Cobb-Douglasova nebo CES funkce), což by se dalo interpretovat jako tržní orientace převážné většiny podniků v daném odvětví. Pokud navíc bude počet odvětví v první skupině s časem narůstat, zdokumentovalo by to postupné dokončování transformace české ekonomiky.

Osnova:

1. Nalezení modelu pro export tradičními i netradičními metodami odhadu regresního modelu.
2. Zpracování dat jako panelových dat pro celé období 1993-2002.
3. Zpracování dat robustními metodami po jednotlivých letech s cílem detekovat případnou heterogenost celého systému.
4. Vyhodnocení výsledků z hlediska různých makroekonomických modelů.

Literatura:

- Benáček, V.: Teorie mezinárodního obchodu (in Kubista et al. - Mezinárodní ekonomické vztahy. Praha, Nakl. Editio HZ, 1999)
- Berndt, E.: The Practice of Econometrics. Reading, Mass., Addison-Wesley, 1990.
- Greene, W. H.: Econometric Analysis. Macmillan Press, New York, 1993.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A.: Robust Statistics. The Approach based on Influence Functions. Wiley, New York, 1986.
- Johnston J., DiNardo, J.: Econometric methods. 4th ed. New York: McGraw-Hill, 1997.
- Víšek, J. Á.: Ekonometrie I. Karolinum, Praha, 1997.

V Praze dne 23. října 2003

Podpis vedoucího bakalářské práce:

Podpis autora: