

Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

## DIPLOMOVÁ PRÁCE



Vojtěch Šaroch

## Numerická studie simultánních rovnic

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Petr Lachout, CSc.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2013

## **Poděkování**

Děkuji tímto doc. RNDr. Petru Lachoutovi, CSc. za odborné vedení diplomové práce, cenné rady, připomínky a především čas, který mi věnoval. Dále bych poděkoval rodině za materiální a morální podporu a za pracovnu, kde jsem mohl nerušeně kompletovat svojí práci.

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Název práce: Numerická studie simultánních rovnic

Autor: Vojtěch Šaroch

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Petr Lachout, CSc.

Abstrakt: V této práci se zabýváme ekonometrickými soustavami simultánních rovnic.

V první kapitole se seznámíme s teoretickými aspekty uvedené problematiky, zejména odhadovým procedurám a jejich vlastnostem. Zmíníme také otázky identifikace a nekonzistence OLS-odhadů při simultánním modelování. Ve druhé kapitole se věnujeme základní teorii odhadu. Zejména se budeme soustředit na intervalové odhady a přesnost odhadu. Zmíníme se také o empirickém přístupu v dané problematice. Ve třetí kapitole potom provedeme numerickou studii na jednoduchém makroekonomickém modelu na uměle vy- tvořených datech. Zajímat nás budou mimo jiné vlastnosti intervalových odhadů parametrů, rychlosť konvergence či rozdíl mezi empirickým a teoretickým odhadem.

Klíčová slova: Soustava simultánních rovnic, intervalové odhady, empirické odhady

Title: Numerical study on simultaneous equations

Author: Vojtěch Šaroch

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Petr Lachout, CSc.

Abstract: In this thesis we deal with simultaneous equation model. In first chapter we introduce to theoretical aspect of this problem, especially estimation procedures and their properties. We mention issues of identification and an inconsistency of OLS-estimates for the simultaneous modeling. In second chapter we introduce theory of estimation, especially we will focus on interval estimation and precision. We mention empirical approach too. In the third chapter we perform a numerical study on simple macroeconomic model on generated dates. We are interested in properties interval estimations of parameters, the convergence rate, difference between empirical and theoretical estimation etc.

Keywords: simultaneous equations model, interval estimation, empirical estimation

# Obsah

<b>1 Soustava simultánních rovnic</b>	<b>6</b>
1.1 Nevhodnost OLS-odhadu v rámci simultánního modelování . . . . .	9
1.2 Problematika identifikace . . . . .	11
1.3 Nepřímý odhad metodou nejmenších čtverců . . . . .	14
1.4 Dvoustupňový odhad metodou nejmenších čtverců . . . . .	15
1.5 Třístupňový odhad metodou nejmenších čtverců . . . . .	16
1.6 Dynamické modely . . . . .	18
<b>2 Teorie odhadu</b>	<b>20</b>
2.1 Bodový odhad . . . . .	20
2.2 Intervalový odhad . . . . .	21
2.3 Konstrukce intervalového odhadu . . . . .	22
2.4 Intervalový odhad parametru . . . . .	22
2.5 Přesnost a chyba odhadu . . . . .	23
2.6 Testování hypotéz . . . . .	24
2.7 Empirický přístup . . . . .	25
<b>3 Numerická studie simultánních rovnic</b>	<b>27</b>
3.1 Základní model . . . . .	27
3.2 Rozšíření modelu . . . . .	30
3.3 Odhady parametrů modelu . . . . .	34

# 1. Soustava simultánních rovnic

V této práci se seznámíme s problémy, které vznikají z povahy, interpretace a odhadování vícerovnicových soustav. Většina z nich se týká odhadů, avšak například musíme zvážit i to, zdali jsou parametry v našich modelech vůbec odhadnutelné. Úvodní teoretická část práce je zpracována na základě knih [Cipra: Finační Ekonometrie] a [Greene: Econometric Analysis].

Soustava simultánních rovnic (dále jen SEM z anglické zkratky Simultaneous Equation Models) je v současnosti hojně využívaný ekonometrický koncept. Vyznačuje se tím, že mezi regresory mohou být i proměnné, které jsou v jiných rovnicích dané soustavy vysvětlující proměnné. Tyto proměnné nazýváme *endogenní* a nejčastěji značíme  $y$ . Proměnné, které ve všech rovnicích soustavy vystupují jen jako regresory, nazýváme *exogenní* a nejčastěji značíme  $x$ .

Soustavu můžeme obecně zapsat například takto:

$$y_{jt} = \sum_{i=1, i \neq j}^m \gamma_{ji} y_{it} + \sum_{i=1}^k \beta_{ji} x_{it} + u_{jt}, \quad j = 1, 2, \dots, m, \quad t = 1, 2, \dots, T, \quad (1.1)$$

kde  $x_{jt}, y_{jt}$  jsou proměnné,  $\gamma_{ji}$  a  $\beta_{ji}$  jsou parametry a  $u_{jt}$  je reziduální složka modelu. Takto definovaná soustava tedy obsahuje  $m$  endogenních a  $k$  exogenních proměnných.

**Příklad 1.** Typickým příkladem je situace, kdy mezi dvěma proměnnými  $x$  a  $y$  je obousměrný kauzální vztah, neboli že  $x$  je potřeba k vysvětlení  $y$  a naopak. V praxi používáme například tzv. *Keynesův makroekonomický model*, který má tvar

$$C_t = \gamma D_t + \beta + u_t,$$

$$D_t = C_t + G_t + v_t,$$

kde

- $C_t$  je spotřeba v čase  $t$ .
- $D_t$  je disponibilní příjem v čase  $t$ .
- $G_t$  jsou vládní výdaje v čase  $t$ .

Spotřeba a příjem jsou tedy endogenní proměnné. Pro další operace je výhodnější používat maticový zápis než vzorec (1.1). Můžeme zapsat jednotlivé rovnice každou zvlášť

$$\mathbf{y}_j = \mathbf{Y}_j \boldsymbol{\gamma}_j + \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{u}_j = \mathbf{Z}_j \boldsymbol{\delta}_j + \mathbf{u}_j, \quad j = 1, \dots, m, \quad (1.2)$$

kde  $\mathbf{y}_j$  je vektor  $(T \times 1)$  pozorovaných hodnot  $j$ -té proměnné,  $\mathbf{Y}_j$  je matice  $(T \times (m_j - 1))$  pozorování endogenních proměnných na pravé straně odpovídajícími parametry  $\boldsymbol{\gamma}_j$  rozměru  $((m_j - 1) \times 1)$ .  $\mathbf{X}_j$  je matice  $(T \times k_j)$  pozorovaných hodnot exogenních proměnných s odpovídajícími parametry  $\boldsymbol{\beta}_j$  rozměru  $(k_j \times 1)$  a  $\mathbf{u}_j$  je reziduální složka  $(T \times 1)$ . Dále  $\mathbf{Z}_j = (\mathbf{Y}_j, \mathbf{X}_j)$  a  $\boldsymbol{\delta}_j = (\boldsymbol{\gamma}_j^T, \boldsymbol{\beta}_j^T)^T$ .

Proměnné s nulovými parametry ignorujeme, proto počty proměnných  $j$ -té rovnice značíme  $m_j \leq m$  a  $k_j \leq k$ . Soustavu lze napsat také jediným maticovým zápisem pro všech  $m$  rovnic najednou

$$\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B} + \mathbf{U} = \mathbf{Z}\Delta + \mathbf{U} = \mathbf{0}, \quad (1.3)$$

kde  $\mathbf{Y}$  je matice  $(T \times m)$  endogenních proměnných soustavy s maticí parametrů  $\Gamma$  rozměru  $(m \times m)$ ,  $\mathbf{X}$  je matice  $(T \times k)$  exogenních proměnných soustavy s maticí parametrů  $\mathbf{B}$  rozměru  $(k \times m)$ ,  $\mathbf{Z} = (\mathbf{Y}, \mathbf{X})$  a  $\Delta = (\Gamma^T, \mathbf{B}^T)^T$ . Matice reziduí  $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_m)$  je rozměru  $(T \times m)$ .

Zanalyzujme matici  $\Gamma$ . Můžeme předpokládat na diagonále hodnoty  $-1$ , aby byl možný jednoduchý převod do tvaru (1.2) (*normalizační pravidlo*). Mnohem důležitější předpoklad ovšem je regularita matice  $\Gamma$ , aby existovala matice  $\Gamma^{-1}$  a my mohli přejít k tzv. *redukovánu tvaru SEM*

$$\mathbf{Y} = \mathbf{B}\Pi + \mathbf{V}, \quad (1.4)$$

kde

$$\Pi = -\mathbf{X}\Gamma^{-1}, \quad \mathbf{V} = -\mathbf{U}\Gamma^{-1}. \quad (1.5)$$

Redukovaný tvar je potvrzením toho, že na endogenní proměnné můžeme pohlížet jako na výstupy dané soustavy, protože  $\mathbf{y}_j = \mathbf{X}\boldsymbol{\pi}_j + \mathbf{v}_j$ , kde  $\mathbf{v}_j$  je  $j$ -tý sloupec matice  $\mathbf{V}$  a  $\boldsymbol{\pi}_j$  je  $j$ -tý sloupec matice  $\Pi$ .

Uved'me nejprve jednu definici

**Definice 1.** Nechť  $b_t$  je posloupnost a  $\beta$  je reálné číslo. Pokud  $\forall \delta > 0$  platí

$$\lim_{t \rightarrow \infty} P(|b_t - \beta| > \delta) = 0, \quad (1.6)$$

potom řekneme, že posloupnost  $b_t$  konverguje v pravděpodobnosti k parametru  $\beta$  a značíme  $p\lim_{t \rightarrow \infty} b_t = \beta$ .  $\square$

Nyní uved'me obecné předpoklady, které se v teorii simultánních rovnic obvykle uvažují:

- (1) Reziduální složky jsou nekorelované v čase  $E(u_{it}) = 0$ ,  $E(u_{it}, u_{is}) = 0$ ,  $\forall t \neq s$ , ale jsou současně korelované  $Cov(u_{it}, u_{jt}) = E(u_{it}u_{jt}) = \sigma_{ij}$ ,  $\forall i, j, t$ . Souhrnně v maticovém tvaru jí můžeme zapsat jako  $Var(u_{1t}, \dots, u_{Jt}) = \Sigma$ , kde  $\Sigma$  je pozitivně definitní matice.
- (2) Pro exogenní proměnné platí  $E(\mathbf{x}_{t.}^T \mathbf{x}_{t.}) = \mathbf{M}$  pro všechna  $t$ , kde  $\mathbf{M}$  je regulární a pozitivně definitní rozměru  $(k \times k)$ .
- (3) Předpokládáme nekorelovanost reziduální složky na exogenních proměnných, tedy  $E(\mathbf{x}_{t.}^T \mathbf{u}_{t.}) = E(\mathbf{x}_{t.})E(\mathbf{u}_{t.}) = \mathbf{0}$  pro všechna  $t$ .
- (4) Matice  $\Gamma$  je regulární.

Uved'me ještě asymptotické verze těchto předpokladů

- (A1) Pro matici reziduí platí

$$p\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{U}^T \mathbf{U} = \Sigma,$$

- (A2) Pro matici exogenních proměnných  $\mathbf{X}$  platí

$$p\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}^T \mathbf{X} = \mathbf{M},$$

která je regulární a pozitivně definitní rozměru  $k \times k$ .

- (A3)

$$p\lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}^T \mathbf{U} = \mathbf{0}.$$

Mezi exogenní proměnné můžeme například zařadit i zpožděně endogenní proměnné (mluvíme pak o *Dynamické SEM*), protože jsou nekorelované s reziduální složkou. Za zmínku také stojí, že tyto proměnné jsou vytvořeny přímo v soustavě v dřívějších časech.

Někdy exogenní proměnné mohou být nenáhodné. V takových případech pak můžeme vynechat asymptotickou podmíinku (A3).

Pro úplnost dodejme, že rozptylová matice vektoru  $\mathbf{u} = (\mathbf{u}_1^T, \dots, \mathbf{u}_m^T)^T$  je rozměru  $(mT \times mT)$  a má tvar

$$Var(\mathbf{u}) = \begin{pmatrix} \sigma_{11}\mathbf{I}_T & \cdots & \sigma_{1m}\mathbf{I}_T \\ \vdots & \ddots & \vdots \\ \sigma_{m1}\mathbf{I}_T & \cdots & \sigma_{mm}\mathbf{I}_T \end{pmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I}_T. \quad (1.7)$$

**Definice 2.** Nechť matice  $\mathbf{A}$  je rozměru  $(k \times l)$  a matice  $\mathbf{B}$  je rozměru  $(p \times q)$ . Potom definujme  $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$  jako matici o rozměrech  $(kp \times lq)$  složenou z  $kl$  bloků a platí  $c_{ij} = a_{ij}\mathbf{B}$ ,  $i = 1, \dots, k$   $j = 1, \dots, l$ . Tuto operaci nazýváme jako *Kronecknerův součin matic* a používáme symbol  $\otimes$ .

## 1.1 Nevhodnost OLS-odhadu v rámci simultánního modelování

Nyní si ukážeme, že klasická metoda nejmenších čtverců (OLS-metoda) není dostatečná. Pro nějaké  $j$  platí

$$\mathbf{y}_j = \mathbf{Z}_j\boldsymbol{\delta}_j + \mathbf{u}_j. \quad (1.8)$$

OLS-odhad je dán

$$\tilde{\boldsymbol{\delta}}_j = (\mathbf{Z}_j^T \mathbf{Z}_j)^{-1} \mathbf{Z}_j^T \mathbf{y}_j. \quad (1.9)$$

V odhadu (1.9) dosadíme  $\mathbf{y}_j$  z rovnice (1.8) a dostaneme

$$\tilde{\boldsymbol{\delta}}_j = \boldsymbol{\delta}_j + (\mathbf{Z}_j^T \mathbf{Z}_j)^{-1} \mathbf{Z}_j^T \mathbf{u}_j. \quad (1.10)$$

Odtud je zřejmé, že odhad  $\tilde{\boldsymbol{\delta}}_j$  nemusí být nutně roven  $\boldsymbol{\delta}_j$ . Tento odhad nemusí dokonce ani existovat, proto OLS-odhad v SEM není obecně nestranný. Popsaný efekt označujeme jako *vychýlení v důsledku simultánního modelování*.

Ověření konzistence je náročnější, proto si zde uvedeme jenom náhledové řešení. Úplný důkaz najdeme například v knize (Dhrymes: Topics in Advanced Econometrics). Potřebujeme zjistit pravděpodobnostní limity

$$\operatorname{plim}_{T \rightarrow \infty} \frac{\mathbf{Z}_j^T \mathbf{Z}_j}{T}, \quad \operatorname{plim}_{T \rightarrow \infty} \frac{\mathbf{Z}_j^T \mathbf{u}_j}{T}. \quad (1.11)$$

Výraz  $\mathbf{Z}_j^T \mathbf{Z}_j$  není nic jiného, než

$$\begin{pmatrix} \mathbf{Y}_j^T \mathbf{Y}_j & \mathbf{Y}_j^T \mathbf{X}_j \\ \mathbf{X}_j^T \mathbf{Y}_j & \mathbf{X}_j^T \mathbf{X}_j \end{pmatrix}$$

Řešíme matici po jednotlivých buňkách

$$\begin{aligned} \mathbf{Y}_j^T \mathbf{Y}_j &= \mathbf{\Pi}_j^T \mathbf{X}_j^T \mathbf{X}_j \mathbf{\Pi}_j + \mathbf{\Pi}_j^T \mathbf{X}_j^T \mathbf{V} + \mathbf{V}^T \mathbf{X}_j \mathbf{\Pi}_j + \mathbf{V}^T \mathbf{V} \\ \mathbf{Y}_j^T \mathbf{X}_j &= \mathbf{\Pi}_j \mathbf{X}_j^T \mathbf{X}_j + \mathbf{V}^T \mathbf{X}_j \end{aligned}$$

Výraz  $\operatorname{plim}_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}_j^T \mathbf{V} = 0$  díky předpokladu (4) o reziduích. Rozptylová matice má tvar

$$E(\mathbf{v}_{t.}^T \mathbf{v}_{t.}) = (\mathbf{\Gamma}^{-1})^T \mathbf{\Sigma} \mathbf{\Gamma}^{-1} = \mathbf{\Omega}, \quad (1.12)$$

kde  $\mathbf{v}_{t.}$  je  $t$ -tý řádkový vektor matice  $\mathbf{V}$  a  $\mathbf{\Sigma}$  je rozptylová matice reziduální složky  $u$ .

Nyní můžeme dopočítat limitu z (1.11) a za pomoci předpokladu (2) dostaneme

$$\begin{aligned} \operatorname{plim}_{T \rightarrow \infty} \frac{\mathbf{Y}_j^T \mathbf{Y}_j}{T} &= \mathbf{\Pi}_j^T \mathbf{M} \mathbf{\Pi}_j + \mathbf{\Omega}_j \\ \operatorname{plim}_{T \rightarrow \infty} \frac{\mathbf{Y}_j^T \mathbf{X}_j}{T} &= \mathbf{\Pi}_j^T \mathbf{M} \\ \operatorname{plim}_{T \rightarrow \infty} \frac{\mathbf{X}_j^T \mathbf{X}_j}{T} &= \mathbf{M} \end{aligned} \quad (1.13)$$

Druhá limita z (1.11) se vypočte o poznání jednodušeji

$$\begin{aligned} \operatorname{plim}_{T \rightarrow \infty} \frac{\mathbf{Y}_j^T \mathbf{u}_j}{T} &= \mathbf{\Pi}_j^T \mathbf{X}_j^T \mathbf{u}_j + \mathbf{u}_j^T \mathbf{u}_j = \mathbf{\Sigma} \\ \operatorname{plim}_{T \rightarrow \infty} \frac{\mathbf{X}_j^T \mathbf{u}_j}{T} &= \mathbf{X}_j^T \mathbf{u}_j = \mathbf{0} \end{aligned} \quad (1.14)$$

Za platnosti našich předpokladů na model, které jsme definovali před tímto odstavcem, ověříme konzistenci odhadu parametrů  $\boldsymbol{\delta}_j$ .

$$\operatorname{plim}_{T \rightarrow \infty} \tilde{\boldsymbol{\delta}}_j = \boldsymbol{\delta}_j + \begin{pmatrix} \mathbf{\Pi}_j^T \mathbf{M} \mathbf{\Pi}_j + \mathbf{\Omega}_j & \mathbf{\Pi}_j^T \mathbf{M} \\ \mathbf{M}^T \mathbf{\Pi}_j & \mathbf{M} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{\Sigma} \\ \mathbf{0} \end{pmatrix} \quad (1.15)$$

Nekonzistence tedy vzniká vzhledem k přítomnosti druhé matice na pravé straně rovnice (1.15), která může být nenulová. Důvodem nekonzistence odhadu je tedy porušení asymptotických podmínek ortogonality ze strany endogenních vysvětlujících proměnných.

## 1.2 Problematika identifikace

V předcházejícím odstavci jsme si ukázali, že použití klasického OLS-odhadu je nedostačující. Mohou za to endogenní proměnné na pravé straně rovnice. Ovšem pokud se nám povede převést model do redukovaného tvaru rovnice (1.4), tj.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Pi} + \mathbf{V},$$

zůstanou na pravé straně už pouze exogenní proměnné. Tento model už patří do rodiny SUR soustav (*seemingly unrelated regression*). SUR soustava už má nestranné a konzistentní OLS-odhady parametrů matice  $\boldsymbol{\Pi}$ , viz např. kniha (Greene: Econometric Analysis). Matice  $\boldsymbol{\Pi}$  tedy můžeme odhadnout pomocí klasického OLS-odhadu. Označme ho např.  $\mathbf{P}$  a

$$\mathbf{P} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (1.16)$$

Pokud se nám tedy matice  $\boldsymbol{\Pi}$  povede transformovat jednoznačně zpět na matice  $\boldsymbol{\Gamma}$  a  $\mathbf{B}$ , můžeme takto transformovat i matici  $\mathbf{P}$ . Transformace mezi původním a redukovaným tvarem soustavy je spojena s tzv. *problémem identifikace*.

Problém spočívá ve specifikaci podmínek, za kterých jsme schopni nalézt vztahy mezi parametry redukované a strukturální soustavy, přičemž využíváme apriorní informace o přítomnosti či nepřítomnosti některých parametrů v jednotlivých rovnicích původního tvaru soustavy.

Strukturální tvar (1.2) obsahuje matice parametrů  $\boldsymbol{\Gamma}$ ,  $\mathbf{B}$  a  $\mathbf{U}$ . Pokud uvažujeme normalizovaný tvar, můžeme mít maximálně  $m(m-1) + km + m^2$  parametrů, zatímco redukovaný tvar obsahuje jenom matice  $\boldsymbol{\Pi}$  a  $\mathbf{V}$ , ve kterých může být maximálně  $km + m^2$  parametrů. Rozdíl v počtu parametrů mezi jednotlivými tvary soustavy činí tedy  $m(m-1)$ .

Abychom předešli problému, vyplatí se zformulovati identifikační kritéria. Vezměme nyní  $j$ -tou rovnici strukturálního tvaru soustavy (1.2) i s vynechanými proměnnými

$$\mathbf{y}_j = \mathbf{Y}_j \boldsymbol{\gamma}_j + \mathbf{Y}_j^* \boldsymbol{\gamma}_j^* + \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{X}_j^* \boldsymbol{\beta}_j^* + u_j, \quad (1.17)$$

kde  $\mathbf{Y}_j^*$  je matice endogenních proměnných nevyskytujících se v rovnici  $j$  a  $\mathbf{X}_j^*$  je matice exogenních proměnných nevyskytujících se v rovnici  $j$ . Počet vynechaných proměnných označme  $m_j^*$  resp.  $k_j^*$ . Odtud vyplývá, že  $\boldsymbol{\gamma}_j^* = \mathbf{0}$  a  $\boldsymbol{\beta}_j^* = \mathbf{0}$ . Rozepišme si podrobně redukovaný tvar soustavy, přičemž si nově rozčleňme a označme matici  $\boldsymbol{\Pi}$

$$(\mathbf{y}_j, \mathbf{Y}_j, \mathbf{Y}_j^*) = (\mathbf{X}_j, \mathbf{X}_j^*) \begin{pmatrix} \boldsymbol{\pi}_j, \cdot\boldsymbol{\Pi}_j, \overline{\boldsymbol{\Pi}}_j \\ \boldsymbol{\pi}_j^*, \underline{\boldsymbol{\Pi}}_j^*, \overline{\boldsymbol{\Pi}}_j^* \end{pmatrix} + \mathbf{V}_j . \quad (1.18)$$

Nyní aplikujeme převodní vztah  $\boldsymbol{\Pi}\boldsymbol{\Gamma} = -\mathbf{B}$  na matici  $\boldsymbol{\Pi}$  z (1.18) a  $j$ -tý sloupec matic  $\boldsymbol{\Gamma}$  a  $\mathbf{B}$ . Dostaneme

$$\begin{pmatrix} \boldsymbol{\pi}_j, \cdot\boldsymbol{\Pi}_j, \overline{\boldsymbol{\Pi}}_j \\ \boldsymbol{\pi}_j^*, \underline{\boldsymbol{\Pi}}_j^*, \overline{\boldsymbol{\Pi}}_j^* \end{pmatrix} \begin{pmatrix} 1 \\ -\gamma_j \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_j \\ \mathbf{0} \end{pmatrix} ,$$

ze kterého můžeme získat 2 soustavy rovnic,

$$\boldsymbol{\pi}_j - \cdot\boldsymbol{\Pi}_j \boldsymbol{\gamma}_j = \boldsymbol{\beta}_j , \quad (k_j \text{ rovnice}) \quad (1.19)$$

$$\boldsymbol{\pi}_j^* - \boldsymbol{\Pi}_j^* \boldsymbol{\gamma}_j = \mathbf{0} . \quad (k_j^* \text{ rovnice}) \quad (1.20)$$

Řešení soustavy rovnic (1.19) je plně určeno řešením soustavy (1.20). Aby měl systém rovnic (1.20) řešení pro parametry vektoru  $\boldsymbol{\gamma}_j$ , musíme mít v soustavě alespoň stejně rovnic jako je jeho délka vektoru  $\boldsymbol{\gamma}_j$ . Na tomto poznání je postavena *rozměrová podmínka identifikace*, která se aplikuje na každou rovnici zvlášť. Počet exogenních proměnných vynechaných v rovnici  $j$  musí být alespoň stejně velký jako je počet endogenních proměnných obsažených v rovnici  $j$

$$k_j^* = k - k_j \geq m_j - 1 .$$

Jedná se však pouze o nutnou podmínu, která nám zajišťuje alespoň jedno řešení soustavy (1.20). Nezajišťuje nám však jednoznačnost. Pro tu potřebujeme znát přímo hodnotu matice  $h(\boldsymbol{\pi}_j^*, \boldsymbol{\Pi}_j^*)$ . *Hodnotní podmínka identifikace* rovnice  $j$  je splněna, pokud

$$h(\boldsymbol{\pi}_j^*, \boldsymbol{\Pi}_j^*) = m_j - 1 .$$

Splnění hodnostní podmínky už zaručuje jednoznačné řešení strukturálních parametrů při znalosti redukovaných parametrů. V praxi se používá hlavně *rozměrová podmínka*, protože její aplikace je na rozdíl od hodnostní podmínky rychlá a jednoduchá. Na základech předchozí diskuse definujme pojmy

**Definice 3.** Rovnice Soustavy simultánních rovnic se nazývá

- *neidentifikovaná*, jestliže z redukovaných parametrů nejde získat ani jeden soubor strukturálních parametrů. Situace nastává, pokud rovnice nevyhovuje hodnostní podmínce nebo

$$k_j* = k - k_j < m_j - 1 .$$

V opačném případě hovoříme o *identifikované* rovnici.

- *přesně identifikovaná*, jestliže z redukovaných parametrů najdeme právě jeden soubor strukturálních parametrů. Situace nastává za splnění hodnostní podmínky a zároveň

$$k - k_j = m_j - 1 .$$

- *přeidentifikovaná*, jestliže z redukovaných parametrů lze získat alespoň dva odlišné soubory strukturálních parametrů. Situace nastává za splnění hodnostní podmínky a zároveň

$$k - k_j > m_j - 1 .$$

Soustavu ve strukturálním tvaru nazveme *identifikovanou*, pokud je každá její rovnice je identifikovaná.  $\square$

**Příklad 2.** Mějme soustavu rovnic tvaru

$$\begin{aligned} y_{1t} &= \gamma_{12}y_{2t} + \gamma_{13}y_{3t} + \beta_{12}x_{2t} + u_{1t}, \\ y_{2t} &= \gamma_{21}y_{1t} + \beta_{21}x_{1t} + u_{2t}, \\ y_{3t} &= \gamma_{31}y_{1t} + \beta_{31}x_{1t} + \beta_{32}x_{2t} + \beta_{33}x_{3t} + u_{3t}. \end{aligned} \tag{1.21}$$

Tvar matic  $\Gamma$  a  $\mathbf{B}$  je

$$\Gamma = \begin{pmatrix} -1 & \gamma_{21} & \gamma_{31} \\ \gamma_{12} & -1 & 0 \\ \gamma_{13} & 0 & -1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & \beta_{21} & \beta_{31} \\ \beta_{12} & 0 & \beta_{32} \\ 0 & 0 & \beta_{33} \end{pmatrix}.$$

Aplikujeme rozměrovou podmínu na první rovnici a dostaneme  $k_1^* = 2 = m_1 - 1$ , tedy rovnice je přesně identifikovaná. Při aplikaci na druhou rovnici dostaneme  $3 - 1 > 1$ , tedy rovnice je přeidentifikovaná. Třetí rovnice je neidentifikovaná, neboť není vyneschána žádná exogenní proměnná ( $k_3^* = 0$ ) a obsahuje endogenní proměnnou.

Vyzkoušejme také hodnotní podmínu. Mějme obecnou matici redukoveného tvaru

$$\boldsymbol{\Pi} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix}.$$

Pro první rovnici dostaneme

$$(\boldsymbol{\pi}_1^*, \boldsymbol{\Pi}_1^*) = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix}. \quad (1.22)$$

Můžeme předpokládat maximální hodnost naší vybrané matice. Tedy hodnotní podmína  $h(\boldsymbol{\pi}_1^*, \boldsymbol{\Pi}_1^*) = 2 = m_1 - 1$  je splněna. Pro druhou rovnici platí

$$h(\boldsymbol{\pi}_2^*, \boldsymbol{\Pi}_2^*) = h\left(\begin{array}{cc} \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{array}\right) = 2 > m_2 - 1$$

Rovnici tedy máme při maximální hodnosti matice (což v podstatě je lineární nezávislost parametrů  $\pi_{21}$ ,  $\pi_{22}$  s parametry  $\pi_{31}$  a  $\pi_{32}$ ) přeidentifikovanou. Jelikož ve třetí rovnici není vyneschána žádná exogenní proměnná, je hodnost matice  $(\boldsymbol{\pi}_3^*, \boldsymbol{\Pi}_3^*) = 0$  a tedy rovnice je neidentifikovaná

### 1.3 Nepřímý odhad metodou nejmenších čtverců

Nyní se začneme zabývat některými nejpoužívanějšími metodami odhadu parametrů v rámci teorie soustavy simultánních rovnic. Mezi nejvíce intuitivní patří nepřímá metoda nejmenších čtverců (Indirect Least Squares, ILS-odhad). Idea je odhadnout parametry matice  $\boldsymbol{\Pi}$  pomocí OLS-odhadu a potom zpětně transformovat na matice  $\boldsymbol{\Gamma}$  a  $\mathbf{B}$ , tak jak jsme uvažovali na začátku předcházejícím odstavci. To že je matice  $\boldsymbol{\Pi}$  asymptoticky odhadnutelná, můžeme ukázat např. za pomocí pravděpodobnostních limit (1.13)

$$\boldsymbol{\Pi} = \mathbf{M}^{-1} \mathbf{M} \boldsymbol{\Pi} = \left[ \operatorname{plim}_{T \rightarrow \infty} \left( \frac{\mathbf{X}_j^T \mathbf{X}_j}{T} \right) \right]^{-1} \operatorname{plim}_{T \rightarrow \infty} \left( \frac{\mathbf{X}_j^T \mathbf{Y}_j}{T} \right). \quad (1.23)$$

Podobně můžeme odhadnout i Rozptylovou matici  $\Omega$ .

$$\Omega = \operatorname{plim}_{T \rightarrow \infty} \left( \frac{\mathbf{Y}_j^T \mathbf{Y}_j}{T} \right) - \operatorname{plim}_{T \rightarrow \infty} \left( \frac{\mathbf{Y}_j^T \mathbf{X}_j}{T} \right) \left( \frac{\mathbf{X}_j^T \mathbf{X}_j}{T} \right)^{-1} \left( \frac{\mathbf{X}_j^T \mathbf{Y}_j}{T} \right) \quad (1.24)$$

Pro praktickou aplikaci je většinou potřebné znát i odhady parametrů strukturálního tvaru. Zde narázíme na problém identifikace, o kterém jsme mluvili v kapitole (1.2). Pro ILS-odhad  $j$ -té rovnice odtud vyplývá následující závěr

- Pro neidentifikovanou rovnici neexistuje ILS-odhad.
- Pro přesně identifikovanou rovnici existuje právě jeden ILS-odhad, který konzistentní a asymptoticky eficientní.
- Pro přeidentifikovanou rovnici existují alespoň 2 odlišné odhady, které sice jsou konzistentní, ale žádný není asymptoticky eficientní, neboť žádný odhad plně nevyužije veškerou dostupnou informaci.

Tento odhad se většinou používá u jednoduchých modelů obsahujících málo proměnných, které jsou přesně identifikované. V ekonometrické a finanční praxi se však častěji používají obsáhlé modely s řídkými maticemi  $\Gamma$  a  $\mathbf{B}$ , které většinou zaručují přeidentifikovanost soustavy. Proto je lepší se poohlédnout po jiném odhadu.

## 1.4 Dvoustupňový odhad metodou nejmenších čtverců

Dvoustupňový odhad metodou nejmenších čtverců v současnosti patří k nejpoužívanějším odhadovým procedurám v teorii SEM, protože poskytuje v praxi přijatelné výsledky. Jak název napovídá, konstrukce 2SLS-odhadu probíhá ve dvou stupních. Pokud vyjdeme z rovnice (1.2)

$$\mathbf{y}_j = \mathbf{Y}_j \boldsymbol{\gamma}_j + \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{u}_j = \mathbf{Z}_j \boldsymbol{\delta}_j + \mathbf{u}_j, \quad j = 1, \dots, m, \quad (1.25)$$

odhadneme v prvním stupni všechny endogenní proměnné  $\mathbf{Y}_j$  pomocí klasického OLS-odhadu pomocí exogenních proměnných

$$\widehat{\mathbf{Y}}_j = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_j, \quad (1.26)$$

ve druhém stupni nahradíme původní matici  $\mathbf{Y}_j$  maticí odhadnutou z prvního stupně  $\widehat{\mathbf{Y}}_j$  a zkonstruujeme OLS-odhad parametrů  $\gamma_j$  a  $\beta_j$ . Odhady parametrů označme  $\mathbf{c}_{2SLS\ j}$  a  $\mathbf{b}_{2SLS\ j}$  a ty můžeme odhadnout jednotným vzorcem

$$\begin{pmatrix} \mathbf{c}_{2SLS\ j} \\ \mathbf{b}_{2SLS\ j} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_j^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_j & \mathbf{Y}_j^T \mathbf{X}_j \\ \mathbf{X}_j^T \mathbf{Y}_j & \mathbf{X}_j^T \mathbf{X}_j \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}_j^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_j \\ \mathbf{X}_j^T \mathbf{y}_j \end{pmatrix}. \quad (1.27)$$

Za konzistentní odhad jeho rozptylové matice můžeme vzít

$$\mathbf{S}_{jj}^{2SLS} = \widehat{\sigma}_{jj}^{2SLS} \begin{pmatrix} \mathbf{Y}_j^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_j & \mathbf{Y}_j^T \mathbf{X}_j \\ \mathbf{X}_j^T \mathbf{Y}_j & \mathbf{X}_j^T \mathbf{X}_j \end{pmatrix}^{-1}, \quad (1.28)$$

kde

$$\widehat{\sigma}_{jj}^{2SLS} = \frac{1}{T} (\mathbf{y}_j - \mathbf{Y}_j \mathbf{c}_{2SLS\ j} - \mathbf{X}_j \mathbf{b}_{2SLS\ j})^T (\mathbf{y}_j - \mathbf{Y}_j \mathbf{c}_{2SLS\ j} - \mathbf{X}_j \mathbf{b}_{2SLS\ j}). \quad (1.29)$$

2SLS-odhad  $j$ -té rovnice má (za platnosti asymptotických předpokladů **(A1)**–**(A3)**) následující vlastnosti:

- je konzistentní,
- je asymptoticky normální,
- je asymptoticky eficientní v rámci uvažované  $j$ -té rovnice.
- 2SLS-odhad je totožný s ILS-odhadem, pokud je  $j$ -tá rovnice přesně identifikovaná.

## 1.5 Třístupňový odhad metodou nejmenších čtverců

Doposud jsme měli pouze odhady pouze pro jednotlivé rovnice zvlášť. Takové odhady nejsou asymptoticky eficientní v rámci celé soustavy, protože nezohledňují současnou korelovanost mezi reziduálními složkami jednotlivých rovnic, konkrétně tedy je nutné konzistentně odhadnout všechny kovariance  $\sigma_{ij}$  v (1.7). To můžeme realizovat přidáním třetího stupně k předchozí dvoustupňové metodě.

V prvních dvou stupních získáme 2SLS-rezidua pro všechny rovnice soustavy a s jejich pomocí odhadneme kovariance  $\sigma_{ij}$  jako

$$\hat{\sigma}_{ij} = \frac{(\mathbf{y}_i - \mathbf{Y}_i \mathbf{c}_{2SLS\ i} - \mathbf{X}_i \mathbf{b}_{2SLS\ i})^T (\mathbf{y}_j - \mathbf{Y}_j \mathbf{c}_{2SLS\ j} - \mathbf{X}_j \mathbf{b}_{2SLS\ j})}{T}, \quad i, j = 1, \dots, m. \quad (1.30)$$

Ve třetím stupni dostaneme 3SLS-odhad. Ovšem vyskytuje se tu problém, který potřebujeme nejdříve ošetřit. Naši původní rovnici

$$\mathbf{y} = \mathbf{Z}\delta + \mathbf{u}$$

musíme vynásobit zleva maticí  $\mathbf{H}^T = (\mathbf{I}_m \otimes \mathbf{X})^T$  a to kvůli zachování konzistence odhadu. Transformovaný model má tvar

$$\mathbf{H}^T \mathbf{y} = \mathbf{H}^T \mathbf{Z}\delta + \mathbf{H}^T \mathbf{u}. \quad (1.31)$$

Podle knihy (Cipra: Ekonometrie) přibližně platí

$$Var(\mathbf{H}^T \mathbf{u}) \sim \mathbf{H}^T E(\mathbf{u}\mathbf{u}^T) \mathbf{H} = (\mathbf{I}_m \otimes \mathbf{X}^T) (\Sigma \otimes \mathbf{I}_m) (\mathbf{I}_m \otimes \mathbf{X}) = \Sigma \otimes \mathbf{X}^T \mathbf{X}. \quad (1.32)$$

Rovnost zde platí v případě úplné nezávislosti procesů  $\mathbf{X}$  a  $\mathbf{U}$ , avšak naše předpoklady (A1)-(A3) vyžadují jenom jejich současnou nekorelovanost.

Odhad  $\mathbf{d}_{3SLS}$  parametrů  $\boldsymbol{\delta} = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T)^T$  dostaneme jako přípustný Aitkenův odhad v transformovaném modelu (1.31), který rozepsaný po jednotlivých rovnicích má explicitní tvar

$$\begin{aligned} \begin{pmatrix} \mathbf{d}_{3SLS1} \\ \vdots \\ \mathbf{d}_{3SLSm} \end{pmatrix} &= \begin{pmatrix} \hat{\sigma}_{11} \mathbf{Z}_1^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_1 & \cdots & \hat{\sigma}_{1m} \mathbf{Z}_1^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_m \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{m1} \mathbf{Z}_m^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_1 & \cdots & \hat{\sigma}_{mm} \mathbf{Z}_m^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_m \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \hat{\sigma}_{11} \mathbf{Z}_1^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_1 + \cdots + \hat{\sigma}_{1m} \mathbf{Z}_1^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_m \\ \vdots \\ \hat{\sigma}_{m1} \mathbf{Z}_m^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_1 + \cdots + \hat{\sigma}_{mm} \mathbf{Z}_m^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_m \end{pmatrix}. \end{aligned} \quad (1.33)$$

Za konzistentní rozptylový odhad matice odhadu  $\mathbf{d}_{3SLS}$  můžeme vzít

$$\begin{pmatrix} \hat{\sigma}_{11} \mathbf{Z}_1^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_1 & \cdots & \hat{\sigma}_{1m} \mathbf{Z}_1^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_m \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{m1} \mathbf{Z}_m^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_1 & \cdots & \hat{\sigma}_{mm} \mathbf{Z}_m^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}_m \end{pmatrix}^{-1}. \quad (1.34)$$

Opět shrňme vlastnosti 3SLS-odhadu za platnosti asymptotických předpokladů

- je konzistentní
- je asymptoticky normální
- je asymptoticky eficientní v rámci celé soustavy
- pokud jsou současně všechny rovnice nekorelované nebo přesně identifikované, je totožný s 2SLS-odhadem.

## 1.6 Dynamické modely

Pod tímto označením máme na mysli soustavu simultánních rovnic, která obsahuje zpožděné endogenní proměnné typu  $y_{j,t-k}$ , ( $k \geq 1$ ). Pro  $k = 1$  má dynamický model tvar

$$\mathbf{y}_t^T \boldsymbol{\Gamma} + \mathbf{x}_t^T \mathbf{B} + \mathbf{y}_{t-1}^T \boldsymbol{\Phi} = \mathbf{u}_t^T, \quad t = 2, \dots, T, \quad (1.35)$$

kde  $\boldsymbol{\Phi}$  je matice o rozměrech  $(m \times m)$ . Pokud model obsahuje zpožděné endogenní proměnné až do výše  $p$ , jeho *strukturální tvar*

$$\mathbf{y}_t^T \boldsymbol{\Gamma} + \sum_{k=1}^p L^k (\mathbf{y}_t^T) \boldsymbol{\Phi}_k + \mathbf{x}_t^T \mathbf{B} = \mathbf{u}_t^T, \quad t = p, \dots, T, \quad (1.36)$$

kde  $L$  je lineární operátor zpětného časového posunutí a  $\boldsymbol{\Phi}_k$  jsou matice o rozměrech  $(m \times m)$ . Chceme-li maticový zápis bez sumy a operátorů posunutí, musíme k (1.36) přidat následující identity  $\mathbf{y}_{t-j} = \mathbf{y}_{t-j}$  pro  $j = 1, \dots, p-1$ . Dostaneme

$$\mathbf{y}_t^{*T} \boldsymbol{\Gamma}^* + \mathbf{x}_t^{*T} \mathbf{B}^* + \mathbf{y}_{t-1}^{*T} \boldsymbol{\Phi}^* = \mathbf{u}_t^{*T}, \quad t = p+1, \dots, T, \quad (1.37)$$

kde platí

$$\mathbf{y}_t^* = (\mathbf{y}_t, \dots, \mathbf{y}_{t-p+1}) , \quad \mathbf{u}_t^* = (\mathbf{u}_t, \mathbf{0}, \dots, \mathbf{0}) , \quad \mathbf{B}^* = (\mathbf{B}, \mathbf{0}, \dots, \mathbf{0}) ,$$

$$\boldsymbol{\Gamma}^* = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_m \end{pmatrix} , \quad \boldsymbol{\Phi}^* = \begin{pmatrix} \boldsymbol{\Phi}_1 & \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\Phi}_2 & \mathbf{0} & \mathbf{I}_m & & \mathbf{0} \\ \vdots & \vdots & & \ddots & \vdots \\ \boldsymbol{\Phi}_{p-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_m \\ \boldsymbol{\Phi}_p & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} .$$

Tyto posunuté proměnné, jak bylo již zmíněno v úvodu této kapitoly, jsou nekorelované se současnou reziduální složkou (splňují podmínu ortogonality), takže je můžeme řadit pod exogenní proměnné. Dynamický model můžeme také převést do redukovaného tvaru rovnice

$$\mathbf{y}_t^* = \mathbf{x}_t^* \boldsymbol{\Pi}^* + \mathbf{y}_{t-1}^* \boldsymbol{\Delta}^* + \mathbf{v}_t^* , \quad t = p+1, \dots, T , \quad (1.38)$$

kde  $\boldsymbol{\Pi}^* = -\mathbf{B}^*(\boldsymbol{\Gamma})^{*-1}$ ,  $\boldsymbol{\Delta}^* = -\boldsymbol{\Phi}^*(\boldsymbol{\Gamma})^{*-1}$  a  $\mathbf{v}_t^* = -\mathbf{u}_t^*(\boldsymbol{\Gamma})^{*-1}$ , za předpokladu existence inverzní matice  $\boldsymbol{\Gamma}^*$ .

## 2. Teorie odhadu

V této kapitole si uvedeme základy teorie odhadu potřebné pro naší numerickou studii. Začneme u bodového odhadu, který nám poslouží hlavně k ilustraci problému. Dále si definujeme intervalový odhad a jeho konstrukci. Provedeme krátkou diskusi o přesnosti a chybě odhadu a kapitolu zakončíme testováním hypotéz.

### 2.1 Bodový odhad

Máme k dispozici nějaká data (například jednotlivé endogenní a exogenní proměnné u soustavy simultánních rovnic), která považujeme za realizaci náhodného výběru  $X_1, \dots, X_n$  z nějakého neznámého rozdělení. Náhodný výběr pochází z rozdělení, které můžeme popsat pomocí pravděpodobnostní funkce  $f(x, \theta)$ . Toto rozdělení je funkcí neznámého parametru  $\theta$ . Potřebujeme bližší informace o charakteristikách tohoto rozdělení (hustotu, střední hodnotu, rozptyl,...), proto budeme konstruovat jejich odhady.

**Definice 4.** Odhadem neznámé charakteristiky  $\theta$  rozumíme jakoukoli funkci  $\hat{\theta}_n$  pozorování  $X_1, \dots, X_n$ .

Podle definice může odhad být jakákoli funkce, proto od "správného" odhadu vyžadujeme některé vlastnosti. Zcela intuitivně vyžadujeme nejčastější následující dvě:

- **nestrannost** - požadavek, aby použitý bodový odhad skutečnou hodnotu charakteristiky nadhodnocoval ani nepodhodnocoval. Matematicky  $E[\hat{\theta}_n] = \theta$ .
- **konzistence** - požadavek, aby s přibývajícím rozsahem výběru se použitý odhad blížil skutečnému parametru. Matematicky  $\lim_{t \rightarrow \infty} \hat{\theta}_n = \theta$ .

Bodový odhad parametru  $\theta$  je jedno číslo, které by mělo nejlépe vystihnout naši hledanou charakteristiku rozdělení. Pokud například chceme zjistit střední hodnotu našeho rozdělení, mělo by číslo našeho bodového odhadu mít jistou spojitost se střední hodnotou.

**Příklad 3.** Výběrový průměr je konzistentní a nestranný odhad střední hodnoty. Mějme  $X_1, \dots, X_n$  iid náhodný výběr se střední hodnotou  $\mu$ , potom

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu ,$$

$$\lim_{n \rightarrow \infty} \bar{X} = E(X_i) = \mu .$$

□

V některých příkladech lze najít více statistik, které jsou nestranné a konzistentní. V tomto případě použijeme ten, který má nejmenší rozptyl. Tu statistiku, která má nejmenší rozptyl, nazveme *nejlepším nestranným odhadem* parametru  $\theta$ .

## 2.2 Intervalový odhad

Musíme si uvědomit, že bodový odhad parametru se téměř vždy liší od jeho skutečné hodnoty. Pokud chceme zvýšit kvalitu odhadu, musíme jedno číslo bodového odhadu nahradit intervalom.

**Definice 5.** Mějme náhodný výběr  $X_1, \dots, X_n$  z rozdělení, jež je prvkem rodiny  $\{F_\theta, \theta \in \Theta\}$ . Jsou-li  $T_d(x_1, \dots, x_n)$  a  $T_h(x_1, \dots, x_n)$  statistiky, pro které platí

$$P(T_d < \theta < T_h) = 1 - \alpha , \quad (2.1)$$

potom  $(T_d, T_h)$  nazveme  $(1 - \alpha)100\%$  *interval spolehlivosti* pro parametr  $\theta$ . Číslo  $0 < \alpha < 1$  nazveme *riziko odhadu* a  $(1 - \alpha)$  je *koeficient spolehlivosti*.

Jinak řečeno intervalový odhad parametru  $\theta$  je interval s náhodnýmimezemi, který překryje  $\theta$  s předepsanou pravděpodobností  $1 - \alpha$ .

Konkrétní data  $x_1, \dots, x_n$ , která experimentálně získáme, nejsou náhodné funkce z pravděpodobnostního prostoru  $\Omega \rightarrow R^1$ , ale jsou to výsledky pokusu pro určitě  $\omega \subset \Omega$ . Proto sestrojíme-li na jejich základě intervalový odhad parametru  $\theta$ , potom nemá smysl mluvit o pravděpodobnosti  $P(T_d < \theta < T_h)$ , protože všechny tři hodnoty jsou čísla a nerovnost bud' platí nebo neplatí. Proto číslo  $(1 - \alpha)$  nazýváme místo toho raději spolehlivost a udává nám četnost případů, kdy náš intervalový odhad bude správný. Číslo  $\alpha$  většinou volíme jako 0,01 nebo 0,05. Poznamenejme, že interval spolehlivosti můžeme zadat nerovností  $\theta > T_d$  nebo  $\theta < T_h$  a tím dostaneme *jednostranné intervaly spolehlivosti*.

## 2.3 Konstrukce intervalového odhadu

Ukážeme si konstrukci oboustranného intervalového odhadu, která je použitelná ve většině případů. Najdeme nějakou funkci  $h$  náhodného výběru  $\mathbf{X}$  a parametrické funkce  $f(\theta)$  tak, aby její rozdělení nezáviselo na parametru  $\theta$ . Nechť  $h_{\alpha/2}$  a  $h_{1-\alpha/2}$  jsou kvantily toho rozdělení. Pro všechna  $\theta$  platí

$$P(h_{\alpha/2} < h[\mathbf{X}, f(\theta)] < h_{1-\alpha/2}) = 1 - \alpha. \quad (2.2)$$

Pokud se nám povede nerovnosti v závorce převést ekvivalentními úpravami na tvar, kde prostřední člen na levé straně výrazu (2.2) stojí jen  $f(\theta)$ , potom jsme sestrojili intervalový odhad o spolehlivosti  $1 - \alpha$ . V následujícím příkladu si ukážeme jak dostat intervalový odhad střední hodnoty.

**Příklad 4.** Nechť  $X_1, \dots, X_n$  náhodný výběr z normálního rozdělení  $N(\mu, \sigma^2)$  s neznámým rozptylem. Potom lze dokázat (kniha Dupač, Hušková: Pravděpodobnost a matematická statistika), že výběrový průměr  $\bar{X}_n$  a výběrový rozptyl  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{X}_n - X_i)^2$  jsou nezávislé náhodné veličiny a náhodná veličina

$$T = \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \quad (2.3)$$

má Studentovo rozdělení o  $(n - 1)$  stupních volnosti. Označíme-li  $t_{\alpha/2,n-1}$  jako  $\alpha$  kvantil  $t$ -rozdělení o  $(n - 1)$  stupních volnosti. Postupnými úpravami dostaneme

$$\left( t_{\alpha/2,n-1} < \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} < t_{1-\alpha,n-1} \right), \quad (2.4)$$

$$\left( t_{\alpha/2,n-1} \frac{S_n}{\sqrt{n}} < \bar{X}_n - \mu < t_{1-\alpha/2,n-1} \frac{S_n}{\sqrt{n}} \right), \quad (2.5)$$

$$\left( \bar{X}_n - t_{1-\alpha/2,n-1} \frac{S_n}{\sqrt{n}} < \mu < \bar{X}_n + t_{1-\alpha/2,n-1} \frac{S_n}{\sqrt{n}} \right) \quad (2.6)$$

a oboustranný intervalový odhad je hotov.  $\square$

## 2.4 Intervalový odhad parametru

Následující vzorce lze nalézt například v knize (Mrkvíčka, Petrášková: Úvod do statistiky). Nechť máme model lineární regrese  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ , kde  $\mathbf{Y}$  a  $\mathbf{u}$  jsou rozměru  $(n \times 1)$ ,

$\mathbf{X}$  je rozměru  $(n \times k)$  a  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^T$ . Chceme zjistit intervalové odhady jednotlivých parametrů  $\beta_j$ ,  $j = 1, \dots, k$ . Bodový odhad  $\boldsymbol{\beta}$  je

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

Označme

$$s^2 = \frac{SSe}{n - k - 1} = \frac{\sum_{i=1}^n (Y_i - \sum_{j=1}^k x_{ij} b_j)^2}{n - k - 1} = \frac{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}{n - k - 1}. \quad (2.7)$$

Intervalový odhad parametru  $\beta_j$  o spolehlivosti  $(1 - \alpha)$  je

$$\left( b_j - t_{n-k-1, \alpha/2} s \sqrt{(\mathbf{X}^T \mathbf{X})_{jj}^{-1}}, b_j + t_{n-k-1, \alpha/2} s \sqrt{(\mathbf{X}^T \mathbf{X})_{jj}^{-1}} \right), \quad (2.8)$$

kde  $(\mathbf{X}^T \mathbf{X})_{jj}^{-1}$  je  $j$ -tý diagonální prvek matice  $(\mathbf{X}^T \mathbf{X})^{-1}$ .

## 2.5 Přesnost a chyba odhadu

Za přesnost odhadu rozumíme maximální chybu, které se při odhadu s danou spolehlivostí dopustíme. S rostoucí šírkou intervalu spolehlivosti klesá přesnost odhadu. Čím bude daný interval širší, tím sice bude odhad spolehlivější, ale odhad bude méně přesný a jeho praktická využitelnost bude malá. Mezi přesností a spolehlivostí odhadu existuje nepřímá úměrnost.

Chybou oboustraného intervalového odhadu rozumíme polovinu délky intervalu spolehlivosti. Někdy potřebujeme kvantifikovat, abychom jí mohli podle našeho uvážení reguloval. Zaved'me  $\Delta$  jako přípustnou chybu tak, aby mohli zapsat intervalový odhad jako  $(\bar{X} - \Delta, \bar{X} + \Delta)$ . Velice zajímavou otázkou v této oblasti je stanovení rozsahu souboru, aneb jak velký musí být rozsah výběrového souboru, abychom ze zjištěných dat mohli provést intervalový odhad průměru  $\mu$  s předem danou přesností  $\Delta$ . Pokud neznáme rozptyl výběru (což většinou platí), musíme provést napřed takzvaný dvoufázový výběr. Nejprve provedeme předvýběr, kde vezmeme  $m$  pozorování, ze kterého vypočítáme výběrový rozptyl  $S_m$ . Máme tedy

$$\Delta = t_{1-\alpha/2, m-1} \frac{S_m}{\sqrt{n}}$$

a postupnými úpravami převedeme na konečný tvar

$$n = \frac{(t_{1-\alpha/2,m-1})^2 S_m^2}{\Delta^2}. \quad (2.9)$$

Ve druhé fázi porovnáme  $m$  a  $n$ . Je-li  $m > n$ , nemusíme již dále dělat žádná další šetření, protože předvýběr dává již dostatečně přesný odhad. Je-li  $m < n$ , musíme doplnit do výběru ještě  $n - m$  pozorování na požadovaný rozsah. Výběrový soubor o  $n$  jednotkách pak již dává dostatečně přesný odhad.

## 2.6 Testování hypotéz

Testování hypotéz je ověřování platnosti výroku na základě náhodného výběru. Mějme nějaký výrok, o jeho pravdivosti chceme rozhodnout. Definujeme si nulovou hypotézu  $H_0$ , což je tvrzení o jehož pravdivosti rozhodujeme (např. zda se výběry liší, je větší než,...). Naproti tomu postavíme alternativní hypotézu  $H_1$ , která obsahuje doplněk  $H_0$  k prostoru parametrů  $\theta$ .

K rozhodnutí, která z výše uvedených hypotéz platí používáme *statistický test*, na jehož základě rozhodneme o platnosti  $H_0$ . Definujme náhodnou veličinu  $T_n = T_n(X_1, \dots, X_n)$  jako testovou statistiku a kritický obor  $C$  jako množinu výsledků pokusu, kdy  $H_0$  zamítáme.

Zdůrazněme, že hypotézy  $H_0$  a  $H_1$  nemají symetrické postavení. Pokud  $T_n \in C$ , potom nulovou hypotézu zamítáme ve prospěch hypotézy  $H_1$  (naše data svědčí proti  $H_0$ ). Pokud však  $T_n \notin C$ , potom na základě našich dat nemůžeme nulovou hypotézu zamítnout, avšak netvrdíme, že  $H_0$  platí. Při našem rozhodnutí nemusí být správné, můžeme dojít k následujícím závěrům:

- (1) Platí  $H_0$  a my jsme se rozhodli nezamítnout  $H_0$ .
- (2) Platí  $H_0$  a my jsme se rozhodli zamítнуть  $H_0$ .
- (3)  $H_0$  neplatí a my jsme se rozhodli nezamítnout  $H_0$ .
- (4)  $H_0$  neplatí a my jsme se rozhodli zamítнуть  $H_0$ .

V případě (1) a (4) jsme se rozhodli správně. V případě (2) se dopouštíme chyby prvního druhu a u (3) chyby druhého druhu. Za nulovou hypotézu volíme tu, jejíž neoprávněné zamítnutí (chyba prvního druhu) je závažnější. Konstrukci hypotézy o střední hodnotě normálního rozdělení založené na intervalovém odhadu si ukážeme na příkladě.

**Příklad 5.** Nechť  $X_1, \dots, X_n$  je náhodný výběr z normálního rozdělení  $N(\mu, \sigma^2)$ , kde  $\mu$  ani  $\sigma^2$  neznáme a  $g(\theta)$  je parametrická funkce  $\mu$ . Zajímá nás střední hodnota. Testujeme hypotézu  $H_0 : g(\theta) = \mu_0$  proti alternativě  $H_1 : g(\theta) \neq \mu_0$ . Nechť  $(T_d(\mathbf{X}), T_h(\mathbf{X}))$  je intervalový odhad parametrické funkce  $g(\theta)$  o spolehlivosti  $(1 - \alpha)$ , potom kritický obor  $C$  testu  $H_0$  na hladině  $(1 - \alpha)$  má tvar

$$C = \{\mathbf{X} \in R^n; \mu_0 \notin (T_d(\mathbf{X}), T_h(\mathbf{X}))\} . \quad (2.10)$$

Odtud za našich předpokladů můžeme dojít postupnými úpravami ke tvaru

$$C = \left\{ \mathbf{X} \in R^n; \mu_0 \notin \left( \bar{X}_n - t_{1-\alpha/2, n-1} \frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{1-\alpha/2, n-1} \frac{S_n}{\sqrt{n}} \right) \right\} , \quad (2.11)$$

což lze přepsat do testové statistiky

$$\frac{|X_n - \mu_0| \sqrt{n}}{S_n} \geq t_{1-\alpha/2, n-1} . \quad (2.12)$$

Upozorněme, že levá strana rovnice (2.12) má Studentovo rozdělení pouze za platnosti hypotézy  $H_0$ . Rovnice (2.12) udává kritický obor, proto při pravdivosti nerovnosti hypotézu  $H_0$  zamítáme.

Pro jednostranné hypotézy typu  $H_0 : g(\theta) \geq \mu_0$  proti  $H_1 : g(\theta) < \mu_0$  postupujeme obdobně.

## 2.7 Empirický přístup

Nečtě  $X_1, \dots, X_n$  je náhodný výběr. Potom definujme empirickou distribuční funkci jako

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{I}_{[X_j \leq x]} , \quad (2.13)$$

kde  $\mathbf{I}_A$  je *indikátor množiny*  $A$ , který je definován jako

$$\mathbf{I}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases} .$$

Empirická distribuční funkce je tedy nespojitá a po částech lineární. Při velkém  $n$  konverguje k distribuční funkci, ze které byl proveden náhodný výběr.

V kapitole (2.3) jsme zmínili, že při konstrukci intervalového odhadu hledáme nějakou distribuční funkci  $h$  náhodného výběru  $X$ . Většinou používáme normovaná rozdělení (jako normální rozdělení, t-rozdělení,  $\chi^2$ -rozdělení), na která transformujeme náhodný výběr  $X$  při užití Centrální limitní věty. Avšak za funkci  $h$  můžeme vzít empirickou distribuční funkci vytvořenou z výběru  $X$  a následně při intervalovém odhadu vzít její kvantily  $h_{\alpha/2}$  a  $h_{1-\alpha/2}$ .

# 3. Numerická studie simultánních rovnic

V první kapitole jsme uvedli několik metod odhadu simultánních rovnic. V této kapitole provedeme jejich numerickou studii a porovnáme jednotlivé metody odhadu. Jako výpočtový software jsme použili program Mathematica 8.0 od firmy Wolfram, který má kvalitní generátor náhodných čísel, nemá však implementovány odhadové procedury simultánních rovnic. Ty jsme si museli vytvořit sami pomocí existujících vzorců. Vhodná alternativa může být například program EViews od firmy Quantitative Micro Software QMS, který je v současnosti hojně rozšířený díky jeho uživatelskému rozhraní a širokou nabídkou ekonometrických vzorců.

## 3.1 Základní model

Zvolíme si základní Keynesův makroekonomický model tvaru

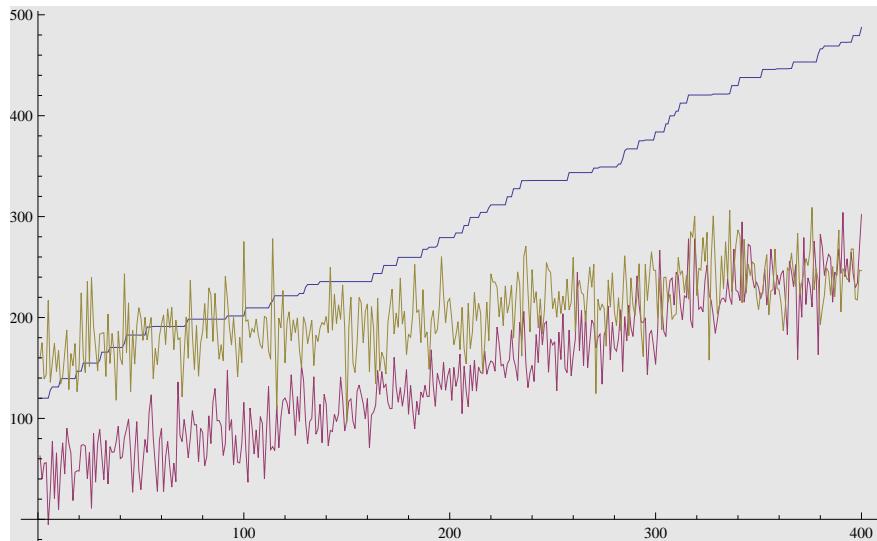
$$\begin{aligned} C_t &= \gamma_1 D_t + \beta_1 + e_{1t} , \\ D_t &= \gamma_2 C_t + \beta_2 G_t + e_{2t} , \end{aligned} \tag{3.1}$$

kde  $C_t$  je spotřeba v čase  $t$ ,  $D_t$  je disponibilní příjem v čase  $t$  a  $G_t$  jsou vládní výdaje v čase  $t$ . Připomeňme, že obdobný model máme v příkladu 1 v první kapitole, tady však ve druhé rovnici jsme za známé parametry o hodnotě 1 ve druhé rovnici nahradili neznámými parametry  $\gamma_2$  a  $\beta_2$ . Obě rovnice modelu jsou přesně identifikované, proto jsou ILS-odhad, 2SLS-odhad i 3SLS-odhad totožný. Proto u tohoto modelu použijeme jenom 2SLS-odhad.

Nejprve zvolme hodnoty parametrů modelu

$$\gamma_1 = 0,4 , \quad \gamma_2 = -0,3 , \quad \beta_1 = 150 , \quad \beta_2 = 0,7$$

a nyní můžeme generovat naše data. Začneme naší exogenní proměnnou. Jedná se o vládní výdaje, tj. složku, kterou můžeme v případě nutnosti upravovat (pomocí různých zásahů či dotací). K této proměnné můžeme přistupovat různě a podle toho nám bude ovlivňovat zbytek modelu.



Obrázek 3.1: Generovaná data pro počet pozorování 400. Modré jsou vládní výdaje, spotřeba žlutě a příjem červeně

- Necháme jí fluktuovat kolem jedné hodnoty. To nám ve výsledku dá poměrně úzký rozsah hodnot dat i při velkém počtu pozorování.
- Dáme jí nějaký drift, tudíž hodnoty dat se nám rozprostřou po intervalu, který se s rostoucím počtem pozorování bude rozšiřovat.
- Vytvoříme skokovou funkci. Tak se nám hodnoty dat rozdělí do několika shluků. S rostoucím počtem dat bude přibývat i počet shluků.

Vytvořili jsme skokovou funkci pomocí lognormálního rozdělení. Rovnice  $C_t$  a  $D_t$  jsme vygenerovali pomocí modelu (3.1), přičemž rezidua jsme vzali z výběru  $e_1 \sim N(0, 625)$  a  $e_2 \sim N(0, 900)$ . Následující tabulka udává odhad parametrů pro různý počet pozorování  $n$ .

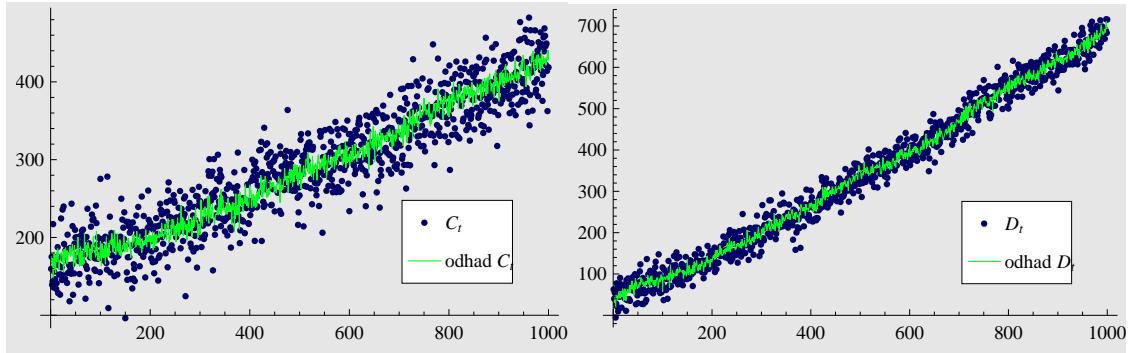
2SLS-odhad parametrů

<b>počet pozorování</b>	<b>c<sub>1</sub></b>	<b>c<sub>2</sub></b>	<b>b<sub>1</sub></b>	<b>b<sub>2</sub></b>
50	0,87681	-0,06397	119,26	0,44959
100	0,73136	-0,17123	127,32	0,56995
200	0,40023	-0,29589	148,68	0,69721
400	0,39174	-0,28199	149,01	0,68511
1000	0,40127	-0,30536	148,49	0,70193

Zběžným pohledem na tabulkou vidíme, že odhad dává poměrně přesné výsledky od zhruba 200 pozorování. Zvolíme si hladinu významnosti  $\alpha = 5\%$ . Vytvoříme intervalový odhad první rovnice na hladině  $\alpha = 5\%$ . To můžeme interpretovat tak, že interval obsahuje řešení  $C_t - c_1 D_t - b_1$  s 95% pravděpodobností. Obdobně vytvoříme intervalový odhad druhé rovnice

<b>počet pozorování</b>	první rovnice		druhá rovnice	
	$T_d^1$	$T_h^1$	$T_d^2$	$T_h^2$
50	-11,3559	11,3559	-5,9410	5,9410
100	-6,8176	6,8176	-4,7895	4,7895
200	-4,2838	4,2838	-3,3031	3,3031
400	-2,9766	2,9766	-2,4045	2,4045
1000	-1,9167	1,9167	-1,5753	1,5753

Dvojice  $(T_d^1, T_h^1)$  je intervalový odhad první rovnice a  $(T_d^2, T_h^2)$  je intervalový odhad druhé rovnice. S rostoucím počtem pozorování se chyba odhadu snižuje, což je významný argument k tvrzení, že naše metoda odhadu je konzistentní (což jsme očekávali viz. kapitola 1.4). Naše hodnoty proměnných jsou řádu stovek, jak je vidět na obrázku (3.1). Pokud bychom chtěli chybu odhadu snížit například na maximálně 1% průměru regresoru, použijeme rovnici na stanovení rozsahu souboru zmíněnou v kapitole (2.5). Zvolme za  $m = 1000$  a při volbě  $\Delta_1 = 0,01 * E[C_t] = 2,91$  resp.  $\Delta_2 = 0,01 * E[D_t] = 3,55$  pro druhou rovnici dostaneme  $n_1 = 391$  a  $n_2 = 196$ . Pro druhou rovnici potřebuje znatelně méně pozorování, protože přípustná šířka intervalu je větší než u prvního.



Obrázek 3.2: Průběh vysvětlovaných proměnných a jejich vypočtených hodnot na základě 2SLS-odhadu ekonometrického modelu pro 1000 pozorování.

Na obrázku (3.2) máme znázorněný průběh endogenních proměnných pro 1000 pozorování srovnáný s jejich odhadem.

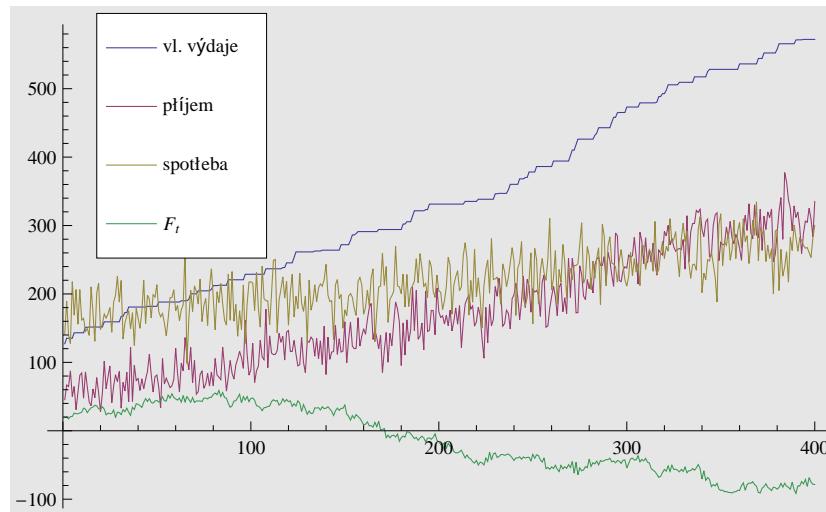
## 3.2 Rozšíření modelu

K první rovnici modelu (3.1) přidejme ještě jednu exogenní proměnnou, kterou označme například  $F_t$ . Půjde o jakousi fiktivní proměnnou, kterou si vytvoříme nezávisle na ostatních a při generování dat jí nebudeme v modelu uvažovat. Tvar soustavy tedy máme

$$\begin{aligned} C_t &= \gamma_1 D_t + \beta_1 + \beta_3 F_t + e_{1t}, \\ D_t &= \gamma_2 C_t + \beta_2 G_t + e_{2t} \end{aligned} \quad (3.2)$$

a proměnné  $C_t$ ,  $D_t$ ,  $G_t$  vygenerujeme stejně jako v předchozím odstavci. V této situaci je druhá rovnice přeidentifikovaná a proto zde už budou rozdíly mezi jednotlivými odhady.

Nejprve aplikujeme dvojstupňový odhad. Následující tabulka udává 2SLS-odhady parametrů při různém rozsahu výběru.



Obrázek 3.3: Vygenerovaná data rozšířeného modelu pro 400 pozorování

2SLS-odhad parametrů

<b>počet pozorování</b>	<b>c<sub>1</sub></b>	<b>c<sub>2</sub></b>	<b>b<sub>1</sub></b>	<b>b<sub>2</sub></b>	<b>b<sub>3</sub></b>
50	0,93796	0,03302	131,65	0,38454	-1,32415
100	0,34852	-0,01351	153,06	0,43425	-0,16793
200	0,41828	-0,19326	143,28	0,61781	0,08566
400	0,33102	-0,26009	159,23	0,67313	-0,13965
1000	0,40521	-0,29876	147,44	0,69884	0,01810

U odhadnutých parametrů **c<sub>2</sub>** a **b<sub>3</sub>** provedeme t-test o jejich nulovosti. U parametru **c<sub>2</sub>** u nízkého počtu pozorování nám vycházejí hodnoty blízké 0 a tedy ve druhé rovnici se příjem stává závislým jenom na vládních výdajích.

<b>n</b>	<b>c<sub>2</sub></b>	<i>t</i> -stat	<i>p</i> -hodnota	<b>b<sub>3</sub></b>	<i>t</i> -stat	<i>p</i> -hodnota
50	0,03302	2,05004	0,04573	-1,32415	-4,36324	0,00006
100	-0,01351	-1,13863	0,25760	-0,16793	-0,48397	0,62947
200	-0,19326	-22,3134	0,00000	0,08566	1,37298	0,17130
400	-0,26009	-45,1013	0,00000	-0,13965	1,25012	0,21198
1000	-0,298764	-86,7972	0,00000	0,01810	0,07557	0,93978

Zde je vidět, že předpokládaný tvar modelu významně ovlivňuje celkové výsledky. Při 100 pozorování na 95% hladině významnosti dokonce můžeme zamítnout nulovost parametru  $\mathbf{c}_2$  a  $\mathbf{b}_3$ . U rozsahu výběru 200, 400 a 1000 nám  $t$ -test zamítá hypotézu o nulovosti parametru  $\mathbf{c}_2$  a zároveň nemůže zamítnout nulovost parametru  $\mathbf{b}_3$ , což jsme v podstatě chtěli. V tomto rozšířeném tvaru potřebujeme mnohonásobně více pozorování, protože naše odhady konvergují mnohem pomaleji ke skutečnosti.

Na stejná data aplikujeme třístupňový odhad, který je asymptoticky eficientní v rámci celé soustavy, odtud je také zřejmé, že ho budeme apriori preferovat oproti 2SLS-odhadu.

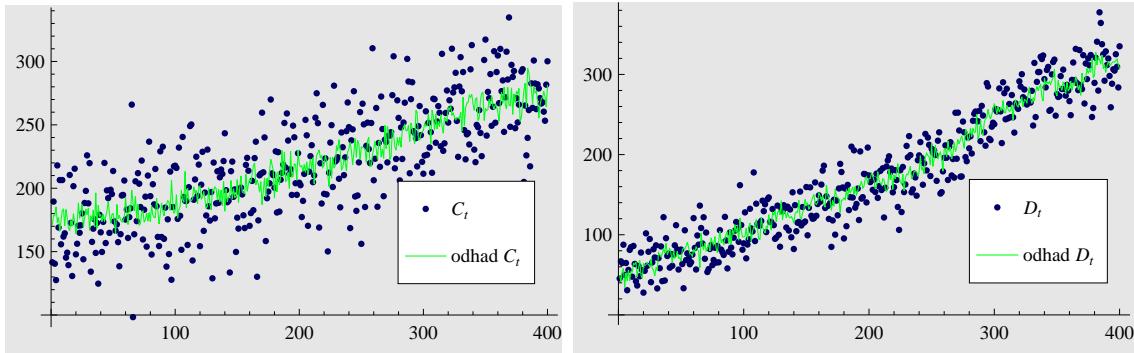
3SLS-odhady parametrů

<b>počet pozorování</b>	<b>c<sub>1</sub></b>	<b>c<sub>2</sub></b>	<b>b<sub>1</sub></b>	<b>b<sub>2</sub></b>	<b>b<sub>3</sub></b>
50	1,01086	-0,01590	128,15	0,43686	-1,41550
100	0,35986	-0,01135	152,06	-0,15337	0,43218
200	0,41576	-0,19355	143,04	0,61804	0,10034
400	0,33103	-0,26007	159,23	0,67312	-0,13963
1000	0,40518	-0,29876	147,45	0,69884	0,01782

Odhadnuté hodnoty jsou pro více než 200 pozorování blízké dvojstupňovému odhadu. T-testy parametrů  $\mathbf{c}_2$  a  $\mathbf{d}_3$  dopadnou obdobně. Raději se zaměříme na intervalové odhady. S jejich pomocí zjistíme, který z odhadů je přesnější nebo který konverguje rychleji ke správnému výsledku. Pro porovnání bude stačit první rovnice.

<b>počet pozorování</b>	2SLS odhad		3SLS odhad	
	$T_d^{2SLS}$	$T_h^{2SLS}$	$T_d^{3SLS}$	$T_h^{3SLS}$
50	-9,45072	9,45072	-9,66342	9,71483
100	-6,08884	6,08884	-6,10741	6,10447
200	-4,20347	4,20347	-4,20064	4,20099
400	-3,00449	3,00449	-3,00450	3,00450
1000	-1,82478	1,82478	-1,82472	1,82472

Za povšimnutí stojí, že třístupňový odhad nemá interval symetrický kolem nuly při malém počtu pozorování. Může za to fakt, že do vzorce zasahují nepřesně odhadnuté



Obrázek 3.4: Průběh vysvětlovaných proměnných a jejich vypočtených hodnot na základě 3SLS-odhadu ekonometrického modelu pro 400 pozorování.

parametry z dvojstupňového odhadu. U malého vzorku má užší intervalový odhad dvoustupňový odhad, u vzorku velikosti 200 má už však užší interval třístupňový odhad. Odtud můžeme usuzovat, že 3SLS-odhad konverguje ke správnému výsledku rychleji. Na obrázku (3.4) máme ukázaný průběh spotřeby  $C_t$  a příjmu  $D_t$  pro vzorek 400. Použitý je pouze 3SLS-odhad, neboť v našem měřítku se oba odhady výrazně překrývají.

Při vysokém počtu pozorování je tedy lepší použít třístupňový odhad. Výhodou také je, že je oproti dvoustupňovému odhadu asymptoticky eficientní v rámci celé soustavy, což je důležitá vlastnost zejména u silně přeidentifikovaných soustav simultánních rovnic. Pokud máme jenom malý počet pozorování, je důležité věnovat zvýšenou pozornost správné identifikaci modelu, neboť pak model může nově přidané proměnné přisoudit mnohem větší důležitost než jí náleží. Jak se například stalo v našem příkladě u 50 pozorování, kdy  $c_2 = -0,016$  a  $b_3 = -1,42$ . Tento typ chyby lze například identifikovat následovně. Soustředíme se na proměnnou, u níž s jistotou víme, že ovlivní náš model. Spotřeba v našem modelu určitě má ovlivňovat rovnici příjmu. Avšak vidíme, že hodnota parametru  $c_2$  je poměrně nevýznamná (velikost pouze v řádu setin), což může vypovídat o špatném tvaru soustavy.

### 3.3 Odhad parametrů modelu

V následujícím odstavci provedeme a porovnáme mezi sebou teoretické a empirické intervaly spolehlivosti jednotlivých parametrů. V odstavci (3.2) jsme numericky zjistili, že 3SLS-odhad a 2SLS-odhad parametrů v našem jednoduchém modelu vycházejí přibližně stejně, proto pro následující šetření použijeme pouze 2SLS-odhad.

Nejprve vezměme výběr o velikosti 100. Intervaly jsou uvedeny v následující tabulce.

intervalové odhady při rozsahu výběru 100

Parametr $k$	teoretický odhad		empirický odhad	
	$T_d^{r,k}$	$T_h^{r,k}$	$T_d^{e,k}$	$T_h^{e,k}$
$\gamma_1$	0,0729	0,6241	0,0832	0,5876
$\gamma_2$	-0,1307	0,1080	-0,1139	0,1177
$\beta_1$	128,98	177,14	129,87	173,95
$\beta_2$	0,3156	0,5488	0,3320	0,5583
$\beta_3$	-0,6752	0,3393	-0,6563	0,2721

Zajímavý je absolutní člen první rovnice, neboli parametr  $\beta_1$ . Šířka téměř 50 je veliká při porovnání s hodnotami příjmu a spotřeby, které jsou v řádu několika stovek. Z toho lze usuzovat, že model je špatně podmíněný, tedy malá změna ve vstupních data vyvolá velkou změnu výstupních dat.

Intervalový odhad  $(T_d^{r,\gamma^2}, T_h^{r,\gamma^2})$ , resp.  $(T_d^{r,\beta^2}, T_h^{r,\beta^2})$  nepokrývá na hladině spolehlivosti 95% parametr  $\gamma_2$ , resp.  $\beta_2$ , který jsme zvolili v kapitole (3.1). Jak již bylo zmíněno dříve, pro tento počet pozorování model upřednostňuje informaci z proměnné  $F_t$ . Empirické intervalové odhady jsou pro všechny parametry užší než teoretické založené na centrální limitní větě, předpokládají menší chybu. Nyní zvýšíme počet pozorování na 400.

intervalové odhady při rozsahu výběru 400

<b>Parametr <math>k</math></b>	teoretický odhad		empirický odhad	
	$T_d^{r,k}$	$T_h^{r,k}$	$T_d^{e,k}$	$T_h^{e,k}$
$\gamma_1$	0,2499	0,4121	0,2439	0,4072
$\gamma_2$	-0,3028	-0,2173	-0,3035	-0,2179
$\beta_1$	146,60	171,87	145,67	171,10
$\beta_2$	0,6471	0,6991	0,6467	0,6987
$\beta_3$	-0,2857	0,0065	-0,2965	-0,0024

Opět se podíváme na absolutní člen první rovnice, tentokrát jeho intervalový odhad je poloviční, tj. má velikost přibližně 25. Intervalové odhady pokrývají naše hledané parametry s výjimkou  $\beta_2$  a empirického odhadu  $\beta_3$ . U  $\beta_2$  je však rozdíl parametru použitého při generování dat a  $T_h^{r,\beta_2}$  jen 0,009, proto při i malém zvýšení koeficientu významnosti by již interval spolehlivosti parametr obsahoval. Ostatně všechny parametry je vyskytuje blízko své jedné hranice intervalu, tedy jsme sice na dobré cestě, ale výběr velikosti 400 ještě není dostatečný u našeho modelu. Rozdíl empirického a teoretického odhadu je řádu setin (u parametru  $\beta_1$  třetí platná číslice).

Na závěr proved'me ještě jedno šetření s rozsahem výběru 1000.

intervalové odhady při rozsahu výběru 1000

<b>Parametr <math>k</math></b>	teoretický odhad		empirický odhad	
	$T_d^{r,k}$	$T_h^{r,k}$	$T_d^{e,k}$	$T_h^{e,k}$
$\gamma_1$	0,3930	0,4174	0,3923	0,4166
$\gamma_2$	-0,3250	-0,2725	-0,3261	-0,2735
$\beta_1$	142,63	152,24	142,32	151,94
$\beta_2$	0,6874	0,7103	0,6868	0,7098
$\beta_3$	-0,0385	0,0746	-0,0421	0,0710

Interval u absolutního členu má velikost přibližně 10. I ostatní odhady jsou již velice přesné a všechny pokrývají zvolené parametry. Odhad  $\beta_3$  je správně předpokládán jako nulový, avšak na zjištění tohoto faktu jsme potřebovali výběr takhle velkého rozsahu. Naproti tomu, empirický intervalový odhad konverguje k teoretickému rychleji, už i pro 400 pozorování byl výsledek velice blízký.

Nevhodně zvolený model se ukázal jako velký problém, kdy konvergence parametrů ke skutečné hodnotě je poměrně pomalá, neboť bylo potřeba kolem 1000 pozorování. Naproti tomu u základního modelu (3.1) stačil výběr o rozsahu 200. Je proto důležité při konstrukci simultánního modelu dbát vysoké opatrnosti, neboť v případě nesplnění správnosti modelu nám naše dostupná data nemusí stačit k objevení správných závislostí.

Empirický odhad v našich podmínkách konverguje k teoretickému odhadu a je použitelný už při vzorku velikosti 400, kdy se liší na úrovni třetí platné číslice.

# Závěr

Cílem práce byla prezentace teoretických základů ekonometrických soustav simultánních rovnic a jejich následná numerická studie. Cílem studie bylo ověřit přesnost a chybu odhadu, zejména rychlosť konvergence, rozdíl mezi empirickou a asymptotickou hodnotou a to i v případě nevhodně zvoleného modelu.

V první kapitole jsme se seznámili s teorií soustav simultánních rovnic a některými jejími problémy, jako například nevhodnost klasického odhadu nejmenších čtverců či problematiku identifikace. Ve druhé části jsme uvedli některé používané metody odhadu.

Ve druhé kapitole jsme se seznámili se základní statistickou teorií odhadu, u které jsme hlavně kladli důraz na intervalové odhady a jejich konstrukce. Na konci kapitoly byl také zmíněn empirický přístup k dané problematice.

Ve třetí kapitole textu byla prezentována vlastní implementace předchozích teoretických poznatků a výsledků v rámci Keynesova makroekonomického modelu, který byl rozšířen o proměnnou, která je nezávislá na ostatních proměnných. Tento model jsme pak podrobili následné numerické studii. K odhadu parametrů příslušné soustavy simultánních rovnic jsme zvolili dva nejběžnější přístupy, konkrétně 2SLS-odhad a 3SLS-odhad. Nevhodně zvolený model se ukázal jako velký problém při konvergenci odhadů parametrů ke skutečné hodnotě. Na závěr jsme porovnali empirický a teoretický (asymptotický) intervalový odhad.

# Literatura

- [1] *T. Cipra: Finační Ekonometrie*  
Ekopress, 2008
- [2] *P. J. Dhrymes: Topics in Advanced Econometrics: Linear and non-linear simultaneous equations*  
Springer, 1994
- [3] *W. H. Greene: Econometric Analysis*  
Prentice Hall, 2003
- [4] *R. Hendrych: Ekonometrické soustavy simultánních rovnic v životním pojištění*  
Diplomová práce, 2010
- [5] *T. Cipra: Ekonometrie*  
SPN Praha, 1984
- [6] *V. Dupač, M. Hušková: Pravděpodobnost a matematická statistika*  
Karolinum, 2009
- [7] *T. Mrkvička, V. Petrášková: Úvod do statistiky*  
Jihočeská univerzita, 2006
- [8] *L. Čechura a spol.: Cvičení z ekonometrie*  
České zemědělská univerzita v Praze, 2008