

# LINEÁRNÍ REGRESNÍ MODELY

15.10.2012 A 22.10.2012

## NÁPLŇ CVIČENÍ

- Nalezení nejlepšího (lineárního) regresního modelu pro zadaná data.
- Testování hypotéz o parametrech modelu.
- Bodová a intervalová předpověď na základě modelu.

DATA: Soubor `gdp.xls` obsahuje čtvrtletní pozorování různých makroekonomických ukazatelů z let 1995–2000 v USA. Proměnné zahrnuté v datech jsou uvedeny v následující tabulce:

<code>date</code>	datum
<code>inv</code>	investice v privátním sektoru (v mld. USD)
<code>gdp</code>	hrubý domácí produkt (v mld. USD)
<code>cons</code>	spotřeba (mld. USD)
<code>m1</code>	nominální peněžní zásoba (v mld. USD)
<code>tbill</code>	výnosy 90denních T-bills v %
<code>infl</code>	míra inflace

1. Načtěte si data do EViews, např. pomocí `File → Open → Foreign Data as Workfile`.
2. Hlavním cílem bude sestavit model pro investice na základě pozorování z let 1995–1999. Nastavte si proto `Sample` na `1995Q1 1999Q4`.
3. Otevřete si všechny proměnné z dat jako skupinu a nějak si ji pojmenujte. Nejprve si prohlédneme data, a to zejména:
  - základní popisné statistiky,
  - grafy jednotlivých proměnných,
  - grafy zobrazující vztah mezi `INV` a ostatními proměnnými,
  - korelační matici (resp. parciální korelační koeficienty).

Které proměnné budou asi pro predikci investic zásadní a které naopak asi nebudou mít vliv?

4. Odhadněte obecný model

$$E(INV) = \beta_0 + \beta_1 GDP + \beta_2 CONS + \beta_3 M1 + \beta_4 TBILL + \beta_5 INFL + \beta_6 t, \quad (M1)$$

kde  $t$  je čas (v EViews např. pomocí `@trend`). Prohlédněte si jednotlivé části výstupu. Podívejte se také na graf skutečných a vyrovnaných hodnot. Co lze soudit o kvalitě modelu?

5. Proveďte kontrolu předpokladů lineárního modelu:
  - Otestujte předpoklad normality.
  - Pomocí několika vhodných obrázků si udělejte představu o homoskedasticitě/heteroskedasticitě.
  - Pomocí vhodného obrázku si udělejte představu o možné `AR(1)` korelaci reziduí.

Připomeňte si, k čemu jsou jednotlivé předpoklady důležité a jaké jsou následky jejich porušení.

6. Postupně vyřadíte z modelu „nepotřebné“ proměnné, až dospějete k nejlepšímu možnému modelu pro investice.
7. Jelikož jsme prováděli vícenásobné testy, proveďte ještě pro kontrolu test, zda lze skutečně model (M1) zjednodušit na Váš výsledný model.  
V obecném modelu View → Coefficient diagnostics → Wald test - coefficient restrictions
8. Ve výsledném modelu otestujte, zda by bylo možné namísto GDP a CONS uvažovat pouze jejich rozdíl nebo jejich kombinaci  $GDP - 2/3 \cdot CONS$ . Lze výsledek testu zohlednit v modelu?
9. Na základě výsledného modelu proveďte predikci pro rok 2000. Spočtete bodovou i intervalovou předpověď na Q1 až Q4.  
Sestrojte obrázek zobrazující predikce společně se skutečnými hodnotami z roku 2000.
10. U makroekonomických dat se často u veličin, které udávají objem financí, pracuje s jejich logaritmy. Uvažujte proto model

$$E[\log(INV)] = \beta_0 + \beta_1 \log(GDP) + \beta_2 \log(CONS) + \beta_3 \log(M1) + \beta_4 TBILL + \beta_5 INFL + \beta_6 t. \quad (M2)$$

Pro tento model proveďte stejný postup jako pro model (M1). Tj.

- Ověřte splnění předpokladů.
- Najděte nejlepší model.
- Proveďte predikce na rok 2000.

ULOŽENÍ DNEŠNÍ PRÁCE. Dnešní data společně s výslednými modely si uložte do společného workfile. Za čas se k těmto datům i modelům vrátíte.