Posudek na práci Thi Yen Nguyenové: Ekonometrické modelování cen nemovitostí v České republice

Předložená práce si vytýčila jako cíl vytvořit ekonometrický model analyzující ceny bytů na českém trhu, s tím že čerpá data z ČSÚ, ČNB a ARAD. Práce využívá postupy pro analýzu časových řad, konkrétně běžný lineární regresní model a dále metodologii stochatistických časových řad ARIMA.

Na počátku práce diplomantka vhodně popisuje klíčové determinanty poptávky a nabídky, které následně využívá ve svých modelech. V této části bych vytkl užívání měřítek u některých grafů, které (výrazně) zkreslují interpretaci (nejvíce je to patrné u grafu 5, kde autorka hovoří o strmém růstu, ač jsou rozdíly věcně nepatrné). Ačkoli jsou uvedeny grafické či tabulkové popisy jednotlivých prediktorů, prediktorem pak není absolutní hodnota, ale řetězový index (viz str. 23). Bylo by možná proto dobré, aby práce obsahovala i základní popisné statistiky těchto hodnot.

V obecném výkladu časových řad se nevhodně míchají pojmy z oblasti deterministického modelování (trend, cykličnost, sezónnost) a stochastických přístupů (bílý šum) a nepoučený čtenář může být náležitě zmaten. Jasné vysvětlení odlišností obou přístupů chybí.

Ostatně díky tomu se nedozvíme, proč autorka užívá pro stejná data regresní i ARIMA modely a co je výhodou této kombinace, resp. který přístup by byl vhodnější. Autorka nejdříve konstruuje několik regresních modelů, komentáře jsou adekvátní. Bohužel se občas chytila do pasti rozdílu mezi statistickou a věcnou významností. Ač jsou vlivy mnoha prediktorů statisticky průkazné, věcně jsou zcela nezajímavé, což autorka nedokáže postihnout. Ostatně dle míry statistické významnosti srovnává i jednotlivé prediktory a to není správně (např. s. 30-31).

Lze ocenit, že autorka provedla i testy splnění předpokladů pro užití regresního modelu.

U ARIMA modelů mi není jasné, jak autorka dospěla k určení konkrétního modelu, tj. řádu AR či MA procesu, protože nejdříve je aplikován model a teprve následně jsou prezentovány grafy ACF a PACF funkcí.

Mám-li celkově shrnout své hodnocení, pak oceňuji, že se autorka pustila do poměrně komplexního modelování časových řad fenoménu, který je více než aktuální. Vyrovnala se poměrně obstojně s technickými složitostmi, mj. i s výpočty v Gretl. Trošku mi ale schází analytický nadhled a výraznější pokus o formulaci věcných závěrů. Samozřejmě lze vytýkat nuance v popisech statistických procedur a jejich algoritmů, ale to je nad síly studentů nestatistických oborů.

Celkově hodnotím práci jako velmi dobrou, skrze stupnici užívanou na ČVUT jako **C** a navrhuji tyto otázky k obhajobě:

1. Které prediktory jsou v OLS (regrese) modelu opravdu ty nejdůležitější a jak to lze určit?
2. Dle čeho vybíráme řád AR a MA procesu?

V Praze 11. 6. 2021

Petr Soukup

FSV UK

s. 30 Výsledná analýza pomocí jednoduché regrese popisuje 73,22 % celého modelu Zcela jistě ne celého modelu, ale variability závisle proměnné. Paradoxně na str. 41 uvádí autorka tuto skutečnost správně.