

Regresní model pro predikování vývoje cen na nemovitostním trhu[#]

*Darina Tauberová**

Úvod

Při oceňování nemovitých věcí je znalec, příp. odhadce, stále častěji konfrontován s potřebou stanovit tržní hodnotu nemovitosti v rámci perspektivy vývoje cen nemovitostí. Jedná se nejen o případy koupě za účelem investování do nemovitostí, případy výstavby, provedení analýzy trhu, ale i každé stanovení tržní hodnoty by mělo obsahovat analýzu vývoje cen oceňované nemovité věci. Tyto požadavky kladou nároky na rozvíjení oboru oceňování nemovitostí, neboť predikci vývoje nemovitostního trhu nelze stanovit pouze na základě analýzy aktuálně prodávaných (nabízených) nemovitých věcí. Je nutné přistoupit k dalším metodikám, kdy jako vhodné se jeví využít matematicko – statistické metody. Výběr vhodného matematicko – statistického přístupu byl předmětem disertační práce autorky. Závěry této disertační práce jsou prezentovány v tomto článku. Novost práce spočívá nejen v nalezení vhodného metodologického přístupu, který je v oboru oceňování nemovitostí de facto nepoužívaný, ale i v podrobném popsání tohoto postupu a srozumitelném vysvětlení použitého modelu a prováděných testů. Článek předpokládá znalost obecné problematiky oceňování nemovitostí.

Zkoumání je zaměřeno na rezidenční nemovitosti z důvodu, že pro korporátní nemovité věci typu komerčních objektů větších rozsahů, průmyslových objektů či hal a zemědělských pozemků, platí jiná pravidla a zákonitosti než pro rezidenční nemovité věci. Autorku ke zkoumání vede snaha obor oceňování nemovitostí rozvíjet, neboť tento obor je jedním z velmi důležitých oborů nezbytných pro ekonomiku státu jako celku, jeho stabilitu a vliv na globální ekonomiku. Při kvalitně prováděných posudcích je možné minimalizovat rizika trhu, kterým je např. tzv. „realitní bublina“.

Autorka zkoumala možné přístupy jevící se jako vhodné pro vytvoření systémového přístupu pro predikování vývoje cen nemovitostí a pro predikci vývoje nemovitostního trhu bylo zvoleno jako nejvhodnější vytvoření regresního modelu, konkrétně vytvoření zpožděného vícenásobného regresního modelu za podmínky splnění předpokladů regresních modelů. Článek je rozdělen na teoretickou a praktickou část. V teoretické části jsou vysvětleny regresní modely, předpoklady regresních modelů a testování hypotéz, v praktické části je demonstrován výběr proměnných a samotná tvorba regresního modelu vč. jeho verifikace.

1. Teoretická část

V této kapitole jsou vysvětleny regresní modely, předpoklady regresních modelů a testování hypotéz. Také je vysvětleno, jak probíhá výběr relevantních vysvětlujících proměnných.

[#] Článek je zpracován jako jeden z výstupů výsledků disertační práce autorky

^{*} Mgr. Darina Tauberová, znalec a certifikovaný odhadce, doktorand na ÚSI VÚT v Brně. Email: darina.tauberova@centrum.cz

1.1. Regresní modely

Regresní modely zjišťují vzájemný vztah jednotlivých veličin. Regresní modely jsou jednoduché lineární a nelineární a dále vícenásobné modely lineární a nelineární. U jednoduchých modelů je pouze jedna vysvětlující proměnná x , u vícenásobných regresních modelů je sada vysvětlujících proměnných x_1, \dots, x_n . Další typy modelů – logistické, multinomické, Poissonovy apod. – nebudou v této práci popsány, neboť nebudou využívány (Meloun, Militký 2002).

Základem regresního modelování je metoda nejmenších čtverců. Tato metoda byla prvně publikována již v roce 1805 francouzským matematikem Adrienem Mariem Legendrem a byla využívána pro určení dráhy nebeských těles (Hebák, 1998). Pojem regrese byl prvně využíván antropologem a meteorologem Francisem Galtonem v roce 1877 v oboru genetika při řešení otázky vztahu mezi výškou otců a synů (Rybář, 2014).

Lineární regrese není běžně užívána v oboru oceňování, ale na vědecké úrovni je i v tomto oboru využívána, např. Zbyněk Zazvonil o ní hovoří ve své publikaci „*Porovnávání hodnota nemovitostí*“.

Regresní analýza jako statistická metoda umožňuje odhadovat hodnotu závisle proměnné na základě znalosti nezávisle proměnných. V případě lineární regrese se jedná o nalezení vzorce závislosti, kde x je nezávislá proměnná a y proměnná závislá. Vztah mezi dvěma proměnnými je sumarizován přímkou, z níž lze vypočítat její parametry a určit její rovnici.

$$y = a + b * x, \quad (1)$$

kde	a, b	–	určují svojí hodnotou vlastnosti dané přímky,
	y	–	je hodnota závisle proměnné,
	x	–	je hodnota nezávisle proměnné,

Data jsou zakreslena do bodového grafu, tzv. korelačního pole a je nutné ověřit, zda mezi veličinami existuje závislost, tzv. regrese. Smyslem regresní analýzy je určit koeficienty " a " a " b ", neboť hodnoty x (x -ové souřadnice jsou přesné) a y (y -ové souřadnice mohou být zatíženy chybou) jsou známy. Koeficient a určuje sklon přímky, koeficient b určuje posun přímky a určuje průsečík přímky s osou y , jedná se o regresní koeficient, který určuje svoji hodnotou vlastnosti dané přímky. Korelační koeficient určuje mezi dvěma proměnnými jejich souvislost, zda existuje či nikoliv a zda je možno vyslovit nějakou predikci. Regresní analýza je navíc schopna popsat velikost vlivu nezávisle proměnné x na závisle proměnnou y . Platí, že závisle proměnná y je měřena na intervalové úrovni a nezávisle proměnná x je intervalová nebo dichotomická (tzn. proměnná, nabývající pouze dvou hodnot). Rozložení obou proměnných by mělo být normální, při větším souboru toto není podmínkou, neboť dle teorie pravděpodobnosti a dle centrální limitní věty, pak nenormální rozložení nemá na výsledky velký účinek. Lineární regrese představuje aproximaci hodnot přímkou metodou nejmenších čtverců (Řezanková, Loster, 2013). Lze shrnout, že při tvorbě modelu je tedy závisle proměnná y to, co vysvětlují, a nezávisle proměnná x podle čeho vysvětlují.

Obecně lze výpočty provádět v profesionálních statistických programech nebo v programech jako je OpenOffice či Microsoft Excel, pro přesnější a jednodušší výpočty je však vhodnější používat profesionální statistické software (GRET, Statgraphics apod.).

V případě, že je do výpočtu uvažována sada nezávisle proměnných a jedna závisle proměnná, pak se jedná o mnohonásobnou (vícenásobnou) lineární regresi (Meloun, Militký 2002).

$$y = a + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n, \quad (2)$$

kde a, b – určují svoji hodnotou vlastnosti dané přímky,
 y – je hodnota závisle proměnné,
 $x_1 \dots 2$ – je hodnota nezávisle proměnné,

Další druhy modelů, nelineární modely, se běžně užívají (kvadratické, kubické, logaritmické), nicméně platí, že pokud model lineární regrese je pouze o málo horší než model nelineární regrese, bývá lineární model regrese upřednostňován z důvodu jasnější interpretace regresních koeficientů, což v případě nelineárních modelů není možné (Rybář, 2014).

1.2. Zpoždění proměnných

Při tvorbě regresního modelu je nutno uvažovat s možným zpožděním proměnných. Důvodem je, že v ekonomice reagují jevy s určitým zpožděním na daný podnět. Tedy, hodnota vysvětlované proměnné y v určitém čase t nemusí nutně záviset na proměnné x_1 v čase t , ale závisí na ní v čase $t-j$, kde j vyjadřuje míru zpoždění. Pokud toto zpoždění není odhaleno, může nezahrnutí zpoždění vést k chybným výsledkům a nezjištění významných korelací.

Zjištění optimálních zpoždění je možné za použití statistických softwarů, které nabízejí možnost automatického zanalyzování zpoždění proměnných, popř. lze využít MS Excel, kde se zpoždění zjišťuje posunutím postupně o jedno časové období a poté vždy vypočítat korelační koeficient (cesta v MS Excel: Analýza dat - Korelace). Jak v statistickém softwaru, tak při použití programu MS Excel je nutné znát okolnosti, vývoj a vztahy proměnných a vždy zanalyzovat, zda v praxi jsou zpoždění reálná či nikoliv. Zároveň při konečné stavbě zpožděného vícenásobného regresního modelu je nutné některá zpoždění přizpůsobit tak, aby splňovala některé nezbytné předpoklady regresního modelu, zejména předpoklad autokorelace reziduí (vysvětleno níže) (Rybář, 2014).

1.3. Předpoklady regresních modelů

Regresní modely musí splňovat několik předpokladů, kterými jsou:

- Proměnné musí být v lineárním vztahu. Vícenásobná lineární regrese je založena na Pearsonově korelačním koeficientu, takže neexistence linearity může způsobit, že důležité vztahy zůstanou neobjasněny.
- Proměnné musí být normálně rozloženy, jinak je zde možnost nepřesnosti výsledků. Tedy je nutné zkoumat rozložení každé proměnné vstupující do analýzy, v rámci platnosti centrální limitní věty.
- Nezávisle proměnná musí být intervalová nebo dichotomická
- Nezávisle proměnné nesmí být mezi sebou příliš vysoce korelovány, protože by to bylo porušením požadavku na multikolinearitě
- V datech nesmí být odlehle hodnoty
- Vztahy mezi proměnnými by měly vykazovat homoskedasticitu, tj. homogenitu rozptylu, která znamená, že rozptyl v datech jedné proměnné bude přibližně odpovídat rozptylům ostatních proměnných. Heteroskedasticita naopak má za následek, že model nebude vypovídající a nestranný (Hebák, 1998).

Při výběru proměnných je vhodné upřednostnit kvalitu nad kvantitou, neexistuje přímý vztah, že čím více proměnných do výpočtu bude zahrnuto, tím více bude výsledek přesnější. V rámci

efektivitu je vhodné dosáhnout s minimálními vstupy maximálně možného efektu, tedy počet proměnných maximálně redukovat. Tedy, u regresních modelů je potřebné zajistit následné předpoklady: správný funkční tvar dané závislosti, homoskedasticitu reziduí, nepřítomnost autokorelace reziduí, nepřítomnost odlehlých hodnot a normalitu reziduí (Meloun, Militký 2002). Jak výše uvedené zajistit a ověřit je vysvětleno níže.

Koeficient determinace, Pearsonův korelační koeficient

Při tvorbě regresních modelů je sledován koeficient determinace. Vzájemný vztah mezi veličinami vyjadřuje korelace. Pokud ve vztahu veličin existuje korelace, pak na sobě veličiny pravděpodobně závisejí. Korelace však neřeší kauzalitu- vztah příčiny a následku. V případě regresních modelů se využívají koeficienty R a R^2 , R je koeficient korelace a R^2 je koeficient determinace. Hodnota R je tzv. Pearsonův korelační koeficient, který je možno vypočítat dle vzorce:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y}, \quad (3)$$

kde r – Pearsonův korelační koeficient,
 \bar{x} a \bar{y} – jsou výběrové průměry,
 s_x a s_y – jsou výběrové směrodatné odchylky,

V případě, že druhé momenty náhodných veličin x a y jsou konečné, korelační koeficient nabývá hodnot z intervalu $<-1,1>$. Při nezávislosti veličin je korelační koeficient roven 0. Platí vztah, že čím vyšší je v regresní analýze hodnota R , tím se nabývá na jistotě, že model vyhovuje vloženým datům, tedy v případě hodnot blízkých se $+1$ se jedná o přímou lineární závislost, v případě hodnot blízkých se -1 se jedná o nepřímou lineární závislost mezi veličinami. Koeficient determinace R^2 určuje, jak přesná je predikce dle zjištěné regresní rovnice, nabývá hodnot $<0,1>$, Pokud jsou data rozložena příliš daleko od regresní přímky, jedná se o velkou chybu a hodnota R^2 bude nízká. Naopak, pokud data budou v blízkosti regresní přímky, jedná se o malou chybu a R^2 bude nabývat vysokých hodnot. R^2 určuje, jaký podíl rozptylu v pozorování závisle proměnné se podařilo regresí vysvětlit. Čím vyšší je procento, tím větší je úspěšnost regresní analýzy. Při vynásobení hodnoty stem je získána hodnota v procentech, tzn. kolik procent rozptylu hledané proměnné je možné objasnit modelem a kolik zůstane neobjasněno, hodnoty nabývají $<0\%,100\%>$. Pokud existuje lineární závislost, koeficient determinace je počítán jako druhá mocnina korelačního koeficientu r . Tedy $R^2 = r^2$. Koeficienty R a R^2 lze použít pouze v případě lineární závislosti. V případě nelineární závislosti jsou užívány indexy I , index korelace, a index I^2 , index determinace (Hebák, 1998). Nicméně indexy determinace a koeficienty determinace jsou ve statistických softwarech užívány shodně pod označením R^2 .

Koeficient determinace R^2 nám říká, kolik rozptylu vysvětlované proměnné je vysvětleno, ale nezohledňuje počet vysvětlujících proměnných v modelu. Proto se používá v případě vícenásobných regresních modelů adjustovaný koeficient determinace R^2_{ADJ} , který zohledňuje také přidání nevýznamných proměnných. Tento koeficient lze vypočítat dle vzorce:

$$R^2_{adj} = 1 - (1 - R^2) \cdot \frac{n-1}{n-p}, \quad (4)$$

kde R^2_{ADJ} – je adjustovaný koeficient determinace,
 R^2 – je koeficient determinace
 p – označuje počet parametrů v regresním modelu.,

Adjustovaný koeficient determinace je vlastně očištěním koeficientu determinace R^2 a lépe ukazuje vhodnost přidání či odebrání statisticky významných či naopak nevýznamných vysvětlujících proměnných do modelu.

Statistická významnost p , p -hodnota

V rámci testování hypotéz je užívána p -hodnota (p -value). Jedná se o nejmenší hladinu významnosti testu, kdy je zamítnuta nulová hypotéza. Tato nulová hypotéza H_0 může být následující: náhodné chyby jsou homoskedastické. V případě nízkých hodnot je vysoká pravděpodobnost, že platí nulová hypotéza. Standardní je hranice 5%. Výsledek testu je srovnáván se zvolenou hladinou významnosti α . Tedy pokud výsledek testu bude méně než 5%, může být nulová hypotéza zamítnuta a jedná se pak o statisticky významný výsledek (Hebák, 1998).

Multikolinearita

Multikolinearitu, vzájemnou statistickou závislost, je nutné odstranit v případě vícerozměrných regresních modelů. Samotná multikolinearita je chápána jako nežádoucí korelace (závislost) mezi vysvětlujícími proměnnými. Pokud závislost mezi vysvětlujícími proměnnými nabývá v absolutní hodnotě 0,8, pak se jedná o vysokou multikolinearitu (Rybář, 2014). Nezávisle proměnné nemají být mezi sebou příliš vysoce korelovány, protože by to bylo porušením požadavku na multikolinearitu a porušení klasického lineárního modelu. Pokud v datech existuje multikolinearita, výsledky regrese jsou nespolehlivé. Je to z důvodu, že vysoká multikolinearita zvyšuje pravděpodobnost, že jinak vhodná nezávisle proměnná bude uvažována jako statisticky bezvýznamná a vyřazena z modelu. Multikolinearita vzniká ve chvíli, kdy některé proměnné jsou příliš vzájemně korelovány a jsou příliš závislé, multikolinearita má vliv na odhady mezních vlivů. Řešením je vyloučení některé ze silně korelované proměnné. Důvodů vzniku multikolinearity může být mnoho- např. nevhodná specifikace modelu, chybné zařazení zpoždění, nadměrný počet vysvětlujících proměnných, jejich nevhodná volba a kombinace apod. Koeficient determinace poté vychází relativně vysoký, ale t -testy jsou statisticky nevýznamné (Hebák, 1998). Zjistit multikolinearitu lze několika způsoby, např. určením párového korelačního koeficientu mezi vysvětlujícími proměnnými, kdy platí, že pokud je $|r| \geq 0,8$ je zde multikolinearita, popř. je možné jednotlivé vysvětlující proměnné postupně měnit na vysvětlované proměnné a díky ostatním vysvětlujícím proměnným modelovat lineární regresi a zjišťovat koeficient determinace (Melou, Militký, 2002).

Zpětná hřebenová regrese

Pro odstranění multikolinearity je využívána zpětná hřebenová regrese (backword regression, ridge regression), kdy jsou vyřazovány takové vysvětlující proměnné, které mají největší p -hodnoty.

Studentův t -test, F -test a χ^2 test

Testovacích statistik a přístupů je mnoho, mezi nejčastěji užívané patří Studentův t -test, F -test a χ^2 test (chí kvadrát).

a) Studentův t -test testuje rozdíly dvou středních hodnot, jedná se o nejčastěji užívaný test. Tento test slouží pro zkoumání významnosti jednotlivých odhadnutých koeficientů. Případné zamítnutí nulové hypotézy lze zjistit pomocí intervalů spolehlivosti, porovnáním testové statistiky s kritickou hodnotou a pomocí p -hodnoty (Meloun, Militký 2002).

b) F -test testuje rozdíly dvou rozptylů. Tento test je užíván pro testování jakékoliv hypotézy, kterou lze zapsat lineární kombinací regresních proměnných. Jedná se o analýzu rozptylu, nebo-li ANOVA (Analysis of variance) nebo také disperzní analýza. Analýza

rozptylu testuje rozdíly více středních hodnot, tedy testování rozdílů mezi jednotlivými soubory, kdy se vícenásobně porovnají jejich průměry. Tento test ověřuje, zda na hodnotu náhodné veličiny má statisticky významný vliv hodnota některého znaku, který je pozorovatelný u zkoumané veličiny, jedná se o rozklad celkového rozptylu dat na složky objasněné, tzv. známé zdroje variability a na složku neobjasněnou, o níž se předpokládá, že je náhodná. Jedná se o samostatnou metodu řešení, kdy je úkolem zjistit, zda soubor diskrétních nezávisle proměnných ovlivňuje hodnotu kvantitativní závisle proměnné. Jedná se o statistickou metodu, která umožňuje provádět vícenásobné porovnávání středních hodnot. Tato metoda je založena na hodnocení vztahů mezi rozptyly porovnávaných výběrových souborů (testování shody středních hodnot se převádí na testování shody dvou rozptylů). Jedná se o *f-testy*, kdy se hodnotí vztahy mezi rozptyly. Podmínkou testu je normalita dat v souboru, homoskedasticita a nezávislost měření. Měří se na hladině spolehlivosti 5% (Meloun, Milítký 2002).

c) χ^2 test (chí kvadrát, test dobré shody) testuje rozdíl četnosti souborů, testuje normalitu reziduí. Při testování se specifikují kritické hodnoty (kvantily) příslušných rozdělení dle testů- *t*-rozdělení, *f*-rozdělení, χ^2 –rozdělení (Meloun, Milítký 2002).

Rezidua

Odhady regresních parametrů jsou odvozeny metodou nejmenších čtverců (viz výše). Optimální regresní parametry mají být určeny tak, aby součet čtverců dle vzorce $S_R = \sum_{i=1}^n (y_i - y_{iTeor})^2$ byl minimální. Rozdíly hodnot reálné hodnoty y_i a teoretické hodnoty y_{iTeor} jsou označovány jako rezidua regresního modelu a reprezentují v podstatě strany čtverců, které se snažíme minimalizovat. Vztah je definovaný jako $e_i = y_i - y_{iTeor}$. Rezidua by měla být rozložena zcela náhodně kolem horizontální osy a neměly by se nacházet mimo 95-ti % interval spolehlivosti. Rezidua by měla přibližně kopírovat přímku, měla by mít normální rozdělení s nulovou střední hodnotou (Rybář, 2014).

Homoskedasticita

S rezidui přímo souvisí homoskedasticita/heteroskedasticita. Předpokladem lineárního regresního modelu je homoskedasticita, která zaručuje konstantnost rozptylu reziduí. Volně lze termín homoskedasticita přeložit jako stejnorozptylovost. Když tato podmínka není splněna, pak se jedná o heteroskedasticitu. Pokud je v modelu přítomna heteroskedasticita, jsou výsledky modelu nespolehlivé, stejně tak testy hypotéz (*t-testy*, *F testy*). Analýza může být provedena na základě testování hypotéz: Breusch – Paganův test, Whiteův test, Glejserův test, Goldfeld-Quandtův test, Bartlettův test heteroskedasticity. Testy jsou založeny na nulové hypotéze, že rozptyly náhodné složky jsou konstantní (Hebák, 1998). Analýza může být provedena i graficky, kdy se v grafu vizuálně rezidui kontrolují rozptyly. Jedná se sice o nejméně přesnou metodu, ale v mnohých případech je dostačující.

Autokorelace reziduí, Durbin-Watsonův test

Nepřítomnost autokorelace reziduí v modelu je dalším předpokladem lineárního regresního modelu. Autokorelace je definována jako porušení předpokladu o vzájemné nezávislosti náhodných složek z různých pozorování. Autokorelace může být způsobena nevhodným způsobem sběru dat, nezahrnutí zpoždění proměnných, vlivy okolí, výběr nesprávného modelu. Důsledkem je potom vychýlení odhadu rozptylu modelu a statistické testy ztrácejí na síle. Autokorelaci lze odhalit grafickou analýzou reziduí v čase- rezidua v případě nepřítomnosti autokorelace náhodně kolísají. Autokorelaci je možné odstranit například využitím vhodnějších zpoždění proměnných a využití vhodnějšího trendu. Autokorelace ve vztahu ke zpoždění mezi rezidui je rozlišována jako autokorelace prvního, druhého až *m*-tého řádu (Rybář, 2014). Hodnoty jsou v intervalu $<0,4>$; pokud je hodnota *D*

menší než 2, jedná se o pozitivní autokorelaci, hodnoty vyšší než 2 značí autokorelaci negativní (Hebák, 1998).

Odlehlé hodnoty v modelu

Kvalitu modelu mohou ovlivňovat tzv. odlehlé hodnoty. Odlehlý bod se vyskytuje vně základní konfiguraci bodů v grafu a v případě, že není k dispozici větší počet zkoumaných proměnných, může i jedna odlehlá hodnota zkresluje celý model. Odlehlý bod je nazýván vlivný, pokud zjistíme, že po jeho odstranění se významně změní poloha regresní přímky (Meloun, Militký 2002).

2. Praktická část – vytvoření modelu

Výsledná volba metody řešení problému je vytvoření zpožděného vícenásobného regresního modelu za podmínky splnění předpokladů regresních modelů, které jsou uvedeny výše. Veličiny (faktory) jsou standardizovány a parametrizovány na bazické indexy. Na základě korelace a zpožděných korelací je pak z těchto veličin možné identifikovat podstatné vysvětlující proměnné.

Popis postupu tvorby zpožděného vícenásobného regresního modelu je následující:

KROK 1) Výběr vysvětlované proměnné y

KROK 2) Výběr vysvětlujících proměnných x_1, \dots, x_{24}

KROK 3) Vytvoření korelační matice bez zpoždění

KROK 4) Zjištění zpoždění vysvětlujících proměnných

KROK 5) Zpětná hřebenová regrese

KROK 6) Vytvoření zpožděného vícenásobného regresního modelu

KROK 7) Ověření pravdivosti výsledků řešení problému

2.1. Výběr vysvětlované proměnné y

Jako vysvětlovaná proměnná byl vybrán House price index (HPI). Tento syntetický cenový index je zveřejňován Českým statistickým úřadem, měří vývoj cenové hladiny rezidenčních nemovitostí a je počítán na základě harmonizované normy Evropské unie. Jeho výhodou je mezinárodní srovnatelnost. Data jsou zveřejňována čtvrtletně v bazických indexech vztažených k roku 2015 (2015=100). Jedná se o index, který jasně ukazuje růst/pokles cenové hladiny rezidenčních nemovitostí. Nejedná se pouze o statistické údaje prodaných nemovitostí, ale index navíc obsahuje i poptávku po netržních komoditách, což je obsaženo v rámci výpočtu HPI a obsahuje i tzv. neúplná data. Index HPI je tedy vypočítáván na základě (1) metody hedonické regrese (neboli metody hedonické ceny, která měří poptávku po netržních komoditách (jako je hluk, výhled z oken, okolí), (2) klouzavého průměru (metoda založená na zjišťování vývoje časové řady) a (3) na základě metody regrese na datech opakovaných prodejů. Výpočet HPI vychází tedy jak ze statistik cen nemovitostí „Ceny nemovitostí ČSÚ“ uveřejňovaných na stránkách ČSÚ, tak i z tzv. neúplných dat a tím může být vypočítáván čtvrtletně. Z tohoto důvodu dochází v časových řadách HPI k zpětným mírným revizím již zveřejněných hodnot. (Český statistický úřad)

Hlavní metodické aspekty House price indexu (HPI) jsou dle ČSÚ následující:

1. Měří vývoj cen rezidenčních nemovitostí (byty, rodinné domy, vč. souvisejících pozemků)
2. Jedná se o celkové nákupy domácností, nákupy ostatních sektorů jsou vyloučeny

3. Zahrnuje jak nové, tak starší rezidenční nemovitosti
4. Index vychází ze skutečně realizovaných cen, ale i z tzv. neúplných dat
5. HPI reprezentuje vývoj cenové hladiny na celém území České republiky
6. HPI je každoročně řetězen, je zpětně revidován
7. Aktualizuje se vnitřní váhová struktura, netržní ceny jsou vyloučeny

2.2. Výběr vysvětlujících proměnných x_1, \dots, x_{24}

Do regresního modelu musejí být vybrány relevantní proměnné. Jsou to proměnné, které vstupují do predikce vývoje nemovitostního trhu. Jsou to faktory makroekonomické, mikroekonomické, socio-ekonomické či politické. Data faktorů byla podrobně statisticky zkoumána, bylo zjištěno, že některé faktory nelze zařadit do statistického zkoumání, neboť nejsou k dispozici relevantní data, popř. existují data pouze za krátký časový úsek. Základními zdroji dat jsou oficiální stránky Českého statistického úřadu a České národní banky. Je nutné upozornit, že není možné veškeré činitele ovlivňující zkoumaný problém vyjmenovat ani popsat, neboť lidská činnost je ovlivňována nejen objektivně, ale i subjektivně. Dále, při vytváření jakýchkoliv modelů v této práci se vychází z tzv. no-event scénáře, který předpokládá, že nedojde k žádné mimořádné události, kterou je např. vyhození dluhové a bankovní krize eurozóny, migrace, válečný stav či jiné geopolitické události s významným dopadem na českou ekonomiku. Zároveň se nepředpokládá zásadní průlom v řešení problémů v České republice, resp. Eurozóně.

Jako relevantní vysvětlující proměnné byly vybrány faktory uvedené v Tab. 1. Tato data jsou dále zkoumána. Bylo vybráno 24 vysvětlujících proměnných k dalšímu statistickému zkoumání z hlediska relevantnosti pro sestavení modelu pro predikci vývoje trhu. Na základě vědeckého zkoumání, odborných rešerší a na základě odborné literatury byly tyto proměnné určeny jako nejrelevantnější. Problémem u některých faktorů, které by v teoretické rovině mohly ovlivňovat predikci, jsou data, která není možno parametrizovat a kvantifikovat, či zdroje dat jsou nedostatečné. Jednotlivé vysvětlující proměnné popisovány v tomto článku nebudou.

Tab. 1 Výběr vysvětlujících proměnných

Označení	Popis vysvětlující proměnné
x_1	<i>Hrubý domácí produkt</i>
x_2	<i>Inflace</i>
x_3	<i>Stavebnictví a rezidenční výstavba: Bytová výstavba v ČR</i>
x_4	<i>Stavebnictví a rezidenční výstavba: Počet stavebních povolení</i>
x_5	<i>Stavebnictví a rezidenční výstavba: Nové podlahové plochy dle stav. povolení</i>
x_6	<i>Stavebnictví a rezidenční výstavba: Důvěra ve stavebnictví</i>
x_7	<i>Stavebnictví a rezidenční výstavba: Index stavební produkce celkem</i>
x_8	<i>Stavebnictví a rezidenční výstavba: Index stavební produkce pozemní stavitelství</i>
x_9	<i>Hypoteční úvěrování a dostupnost hypotečních produktů: Průměrná úroková sazba</i>
x_{10}	<i>Hypoteční úvěrování a dostupnost hypotečních produktů: Počet hypoték</i>
x_{11}	<i>Hypoteční úvěrování a dostupnost hypotečních produktů: Dostupnost úvěrů</i>
x_{12}	<i>Zaměstnanost: Obecná míra nezaměstnanosti</i>
x_{13}	<i>Zaměstnanost: Míra ekonomická aktivity 15-64letých</i>
x_{14}	<i>Zaměstnanost: Mediány hrubých měsíčních mezd</i>

x_{15}	<i>Ceny nemovitostí: Indexy nabídkových cen bytů, celá ČR</i>
x_{16}	<i>Ceny nemovitostí: Indexy realizované cen nových bytů, celá ČR</i>
x_{17}	<i>Ceny nemovitostí: Indexy realizované cen starších bytů, celá ČR</i>
x_{18}	<i>Kurz koruny vůči euru</i>
x_{19}	<i>Kurz koruny vůči dolaru</i>
x_{20}	<i>Demografický vývoj</i>
x_{21}	<i>Domáci realizovaná poptávka</i>
x_{22}	<i>Index dostupnosti bydlení</i>
x_{23}	<i>Index návratnosti bydlení</i>
x_{24}	<i>Výdaje na konečnou spotřebu domácností</i>

Zdroj: vlastní tvorba

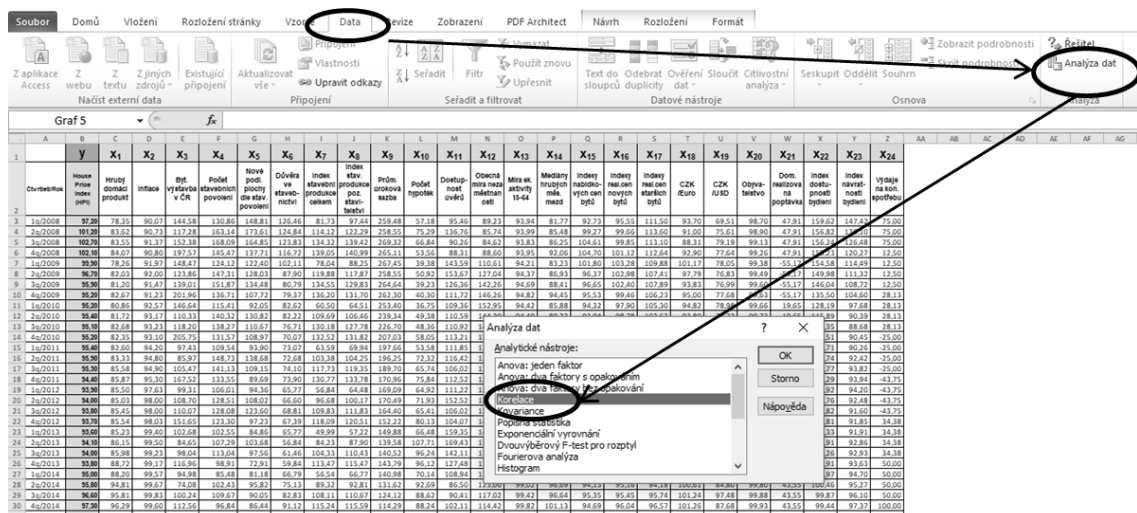
Aby bylo možné data všech faktorů dále zkoumat, je nutné je upravit do srovnatelných hodnot. Jako vhodné se v tomto případě jeví bazické indexy. Indexy jsou klíčovým nástrojem pro jakékoliv porovnání v čase a místě. Bazický index vysvětluje, jaký je rozdíl hodnot nyní a v nějakém vybraném období (např. Statistický úřad používá rok 2015 jako vybrané období a jiné roky jsou k tomuto datu srovnávány). Bazické indexy všech období jsou tedy vztaženy vždy k jednomu danému datu, ke stejné vybrané hodnotě a nejlépe zachycují vývoj v časové řadě. Bazický index vztažený k roku 2015 je počítán tak, že se zjistí průměrná hodnota v roce 2015, která se označí jako 2015=100. Pak je možné, že i v samotném roce 2015 v jednotlivých měsících či kvartálech je hodnota odlišná od hodnoty 100. Je to způsobeno právě tím, že bázi je průměr 2015=100, vychází z průměrných hodnot celého roku 2015. Z důvodu, že Český statistický úřad nejčastěji nově používá bazický index 2015=100, jsou data v této práci také vztažena k průměru roku 2015.

2.3. Vytvoření korelační matice bez zpoždění

V prvním kroku byly proměnné zkoumány bez zpoždění, byl zjišťován korelační koeficient metodou nejmenších čtverců. Korelační matice obsahuje párové korelační koeficienty jednotlivých vysvětlujících proměnných, které podávají informaci o výskytu tzv. multikolinearity (tedy závislosti) mezi dvěma či více proměnnými. Multikolinearita se v modelu vyskytuje tehdy, když hodnota párového koeficientu překročí v absolutní hodnotě 0,8. Pro výše uvedených 24 vysvětlujících proměnných x_1, \dots, x_{24} je vytvořena korelační matice Pearsonových korelačních koeficientů. Korelační matici lze vytvořit i v programu MS Excel.

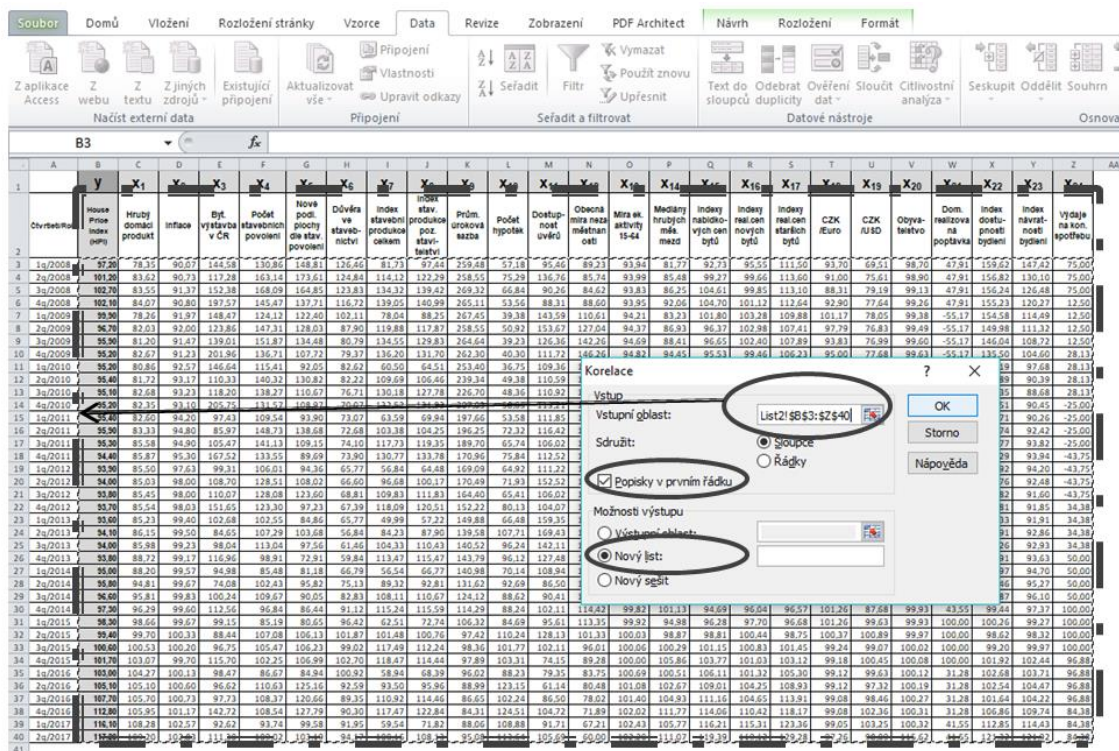
Postup v MS Excel: Data - Analýza dat – Korelace viz Obr. 1 – Označení vybraných dat (všechna vstupní data, y, x_1, \dots, x_{24}), zaškrtnutí políčka „Popisky v prvním řádku“ a políčka „Nový list“, viz Obr. 2 – Je vytvořena korelační matice na samostatném listu viz Obr. 3.

Obr. 1 Korelace, postup v MS Excel, 1. část



Zdroj: Vlastní tvorba

Obr. 2 Korelace, postup v MS Excel, 2. část



Zdroj: Vlastní tvorba

Na novém listu se zobrazí korelační matice pro všechny proměnné včetně popisu jednotlivých proměnných.

Obr. 3 Korelační matice, postup v MS Excel, 3. část

	House Price Index (HPI)	Hrubý domácí produkt	Inflace	Byt. výstavba v ČR	Počet stavebních povolení	Nové podl. plochy ve stavebnictví	Důvěra ve stavební produkci celkem	Index stavební produkce	stav. produkce stav. telství	Prům. úroková sazba	Počet hypoték	Dostupnost úvěrů	Obecná míra nezaměstnanosti	Míra ek. aktivit 15-64	Mediány hrubých měs. výd. cen bytů	Indexy nabídko- nových bytů	Indexy real.cen starších bytů	Indexy real.cen starších bytů	CZK /Euro	CZK /USD	Obyva- telstvo	Dom. realizova- ná poptávka	Index dostu- pnosti bydlení	Index návrat- nosti bydlení	Výdaje na kon. spotřebu
House Price Index (HPI)	1																								
Hrubý domácí produkt	0,743241	1																							
Inflace	0,382226	0,820575	1																						
Byt. výstavba v ČR	-0,06039	-0,40961	-0,61894	1																					
Počet stavebních povolení	-0,18141	-0,62472	-0,80693	0,486175	1																				
Počet stavebních povolení	0,213879	-0,27013	-0,5838	0,264362	0,791273	1																			
Důvěra ve stavebnictví celkem	0,509672	0,197105	-0,28204	0,211131	0,249618	0,597066	1																		
Důvěra ve stavebnictví celkem	-0,02431	-0,1378	-0,34627	0,520698	0,632894	0,437326	0,141906	1																	
stavi- telství	0,031765	-0,13738	-0,36142	0,551746	0,650421	0,483765	0,20581	0,987961	1																
Prům. úroková sazba	-0,40181	-0,86862	-0,97256	0,574267	0,796482	0,544613	0,189517	0,280596	0,29455	1															
Počet hypoték	0,552605	0,848088	0,845865	-0,48999	-0,54937	-0,20773	0,033616	-0,07415	-0,05846	-0,88107	1														
Dostupnost úvěrů	-0,41031	-0,51462	-0,19271	-0,03465	0,201506	-0,01172	-0,39583	-0,06985	-0,0838	0,303813	-0,25793	1													
místnosti	-0,88967	-0,71217	-0,32574	0,089222	0,160867	-0,31404	-0,72724	-0,00595	-0,08168	0,393977	-0,59717	0,491487	1												
Míra ek. aktivit 15-64	0,584017	0,935658	0,941817	-0,5168	-0,77518	-0,44512	-0,04174	-0,25932	-0,26277	-0,94813	0,880314	-0,33646	-0,53535	1											
Mediány hrubých měs. mezd	0,631227	0,889337	0,793402	-0,17748	-0,5487	-0,35883	-0,06292	0,103204	0,095346	-0,83718	0,813937	-0,41924	-0,52148	0,867788	1										
Indexy nabídko- vých cen bytů	0,962786	0,729658	0,335124	0,000673	-0,17864	0,223314	0,566982	0,026489	0,067397	-0,35186	0,498803	-0,43826	-0,87357	0,566433	0,598978	1									
Indexy real. cen nových bytů	0,931613	0,604706	0,240562	0,020134	-0,06731	0,221014	0,417194	0,0296	0,059875	-0,23427	0,361819	-0,26691	-0,72374	0,4452	0,531751	0,912344	1								
Indexy real. cen starších bytů	0,844219	0,328152	-0,14646	0,266105	0,250848	0,523891	0,61932	0,144244	0,205578	0,117971	0,088743	-0,30961	-0,72408	0,093692	0,232144	0,824212	0,880914	1							
CZK /Euro	0,353495	0,645147	0,581562	-0,30727	-0,69687	-0,41022	0,151408	-0,25691	-0,2708	-0,57962	0,492958	-0,22308	-0,3802	0,704325	0,502463	0,441563	0,316535	0,036946	1						
CZK /USD	0,673905	0,951383	0,778442	-0,37184	-0,62645	-0,28811	0,214415	-0,16145	-0,1734	-0,83315	0,788886	-0,46258	-0,64463	0,891469	0,814987	0,679246	0,554376	0,26297	0,683181	1					
Obyva- telstvo	0,549724	0,434229	0,35912	-0,08681	-0,16626	-0,11936	0,040976	-0,01782	0,004791	-0,29511	0,320831	-0,08166	-0,37254	0,389042	0,473707	0,492951	0,616456	0,462211	0,142263	0,319213	1				
Dom. realizovaná poptávka	0,394109	0,581834	0,372744	-0,25395	-0,37629	-0,04148	0,514541	-0,08071	-0,07051	-0,47097	0,530326	-0,48359	-0,57141	0,489233	0,396417	0,39436	0,195727	0,151374	0,406307	0,59923	0,12319	1			
Index dostu- pnosti bydlení	0,107726	-0,48715	-0,80067	0,538775	0,692964	0,716792	0,61369	0,247207	0,308898	0,81248	-0,56436	0,116127	-0,16765	-0,64299	-0,57962	0,157955	0,216329	0,547879	-0,28491	-0,48228	-0,05047	-0,22078	1		
Index návrat- nosti bydlení	0,526	0,053105	-0,33153	0,282094	0,314648	0,638411	0,834149	0,086845	0,187674	0,30351	-0,01039	-0,17588	-0,67301	-0,10673	-0,13235	0,529236	0,477876	0,725949	0,041886	0,001712	0,181112	0,202999	0,775261	1	
Výdaje na kon. spotřebu	0,5789	0,69042	0,42951	-0,24962	-0,42211	0,003595	0,554929	-0,09005	-0,07129	-0,48273	0,594357	-0,39955	-0,6816	0,649781	0,480738	0,651095	0,446321	0,337706	0,65474	0,667901	0,199164	0,720905	-0,06978	0,402467	1

Zdroj: Vlastní tvorba

Výše uvedená korelační matice je použitelná pouze pro první předvýběr relevantních proměnných, protože v sobě nezahrnuje znalost konkrétní situace na trhu a neřeší posun proměnných v čase. Kdyby byl model sestaven pouze na základě nezpožděných proměnných, hodnota adjustovaného koeficientu determinace R^2_{ADJ} by vyšla nižší, než v případě využití zpožděných vysvětlujících proměnných. Výsledek modelu se pomocí zpožděných vysvětlujících proměnných může ještě vylepšit, jak bude demonstrováno v dalších kapitolách. Předvýběrem a za pomoci statistického software bylo do užšího výběru vybráno 8 vysvětlujících proměnných. Výsledky korelace bez zpoždění jsou uvedeny na Obr. 4. Hodnocení pomocí programu MS Excel je však zdoluhavé, a proto je vhodnější využít profesionální statistický software. Bude využit software GRETL, který provede veškeré výpočty v jednom kroku.

Obr. 4 Korelace bez zpoždění

Model 1: OLS, za použití pozorování 2008:1-2017:2 (T = 38)					
Závisle proměnná: hpi					
	koeficient	směr. chyba	t-podíl	p-hodnota	

const	59,0193	27,9771	2,110	0,0436	**
hdp	-0,184255	0,111862	-1,647	0,1103	
inflace	-0,320478	0,238422	-1,344	0,1893	
urok_sazba	-0,0686474	0,0228820	-3,000	0,0055	***
nabidkove_ceny	0,555514	0,100399	5,533	5,77e-06	***
real_ceny	0,368525	0,0957712	3,848	0,0006	***
poptavka	0,0314325	0,00889432	3,534	0,0014	***
dostupnost	0,0799325	0,0277237	2,883	0,0073	***
spotreba	-0,0379612	0,0109289	-3,473	0,0016	***
Střední hodnota závisle proměnné	98,69211				
Sm. odchylka závisle proměnné	6,089707				
Součet čtverců reziduí	35,65990				
Sm. chyba regrese	1,108897				
Koeficient determinace	0,974011				
Adjustovaný koeficient determinace	0,966842				
F(8, 29)	135,8584				
P-hodnota(F)	7,16e-21				
Logaritmus věrohodnosti	-52,71204				
Akaikevo kritérium	123,4241				
Schwarzovo kritérium	138,1623				
Hannan-Quinnovo kritérium	128,6678				
rho (koeficient autokorelace)	0,464514				
Durbin-Watsonova statistika	1,033550				
zde je poznámka o zkratkách statistik modelu					
Pomíne-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 3 (inflace)					

Zdroj: SW GRETL

Neopomenutelným důvodem pro použití zpožděných vysvětlujících proměnných je také to, že pro predikci na základě zpožděných proměnných jsou používána reálná data nikoliv predikovaná. Hodnoty koeficientu determinace R^2 jsou 0,974011 a hodnoty adjustovaného koeficientu determinace R^2_{ADJ} jsou 0,966842. Vyhodnocení těchto výsledků bude provedeno v následujících kapitolách. Je vysoce pravděpodobné, že koeficient determinace R^2 a adjustovaný koeficient determinace R^2_{ADJ} se zvýší, pokud se zjistí optimální zpoždění vysvětlujících proměnných a vyselektují se statisticky nevýznamné vysvětlující proměnné.

2.3. Zjištění zpoždění vysvětlujících proměnných

V dalším kroku bylo zjišťováno optimální zpoždění vysvětlujících proměnných. Pro zjištění optimálních zpoždění podle nejvyššího zpožděného korelačního koeficientu je využit statistický software GRET. Pokud by bylo zjišťováno optimální zpoždění pomocí MS Excel, postup by byl takový, že se posunují řádky dat a vytváří se korelační matice v daném zpoždění. Jedná se o velmi pracný a časově náročný postup. Software GRET navrhne optimální zpoždění jednotlivých vysvětlujících proměnných a zároveň vyselektuje statisticky nevýznamné vysvětlující proměnné. Proměnné, které nekorelují, jsou z modelu odstraněny postupnou hřebenovou regresí, která je demonstrována v následující kapitole. Zjišťuje se, jestli se zpožděním a odstraněním vybraných vysvětlujících proměnných se model zlepšuje (dle adjustovaného koeficientu R^2_{ADJ}).

Na základě výstupu ze software GRET je vybráno 8 vysvětlujících proměnných. Optimální zpoždění je uvedeno v Tab. 2, v tabulce jsou uvedeny i zkratky jednotlivých vysvětlujících proměnných.

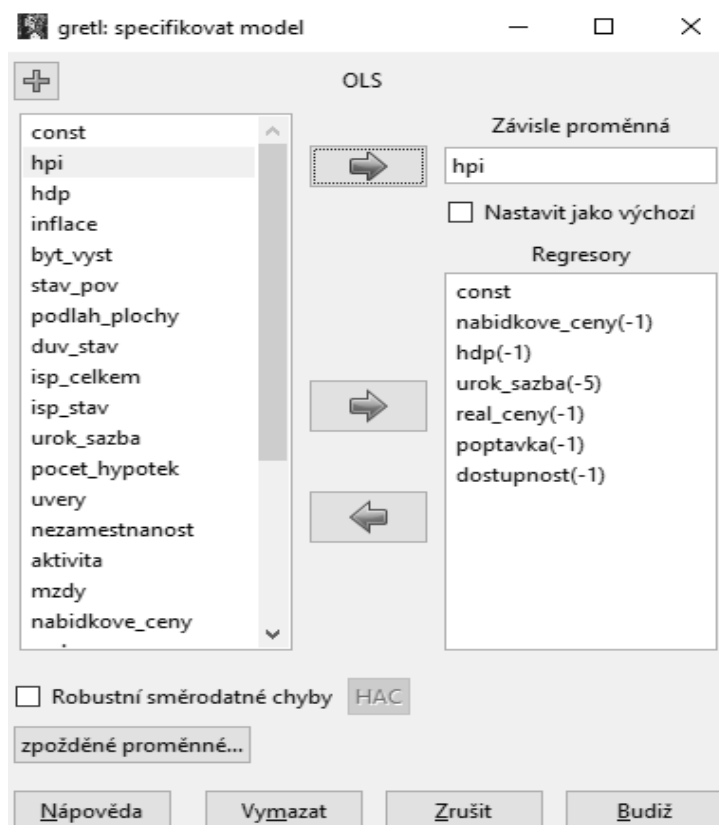
Tab. 2 Optimální vysvětlující proměnné a jejich zpoždění

Označení	Popis vysvětlující proměnné	Zkratka	Optimální zpoždění (kvartální)
y	<i>House price index</i>	HPI	x
x_1	<i>Hrubý domácí produkt</i>	HDP	1
x_2	<i>Inflace</i>	IN	1
x_9	<i>Průměrná úroková sazba</i>	US	5
x_{15}	<i>Indexy nabídkových cen bytů</i>	NC	1
x_{16}	<i>Indexy real.cen nových bytů</i>	RC	1
x_{21}	<i>Domácí realizovaná poptávka</i>	RP	1
x_{22}	<i>Index dostupnosti bydlení</i>	DB	1
x_{24}	<i>Výdaje na konečnou spotřebu</i>	VS	1

Zdroj: vlastní tvorba

Na základě vybraných vysvětlujících proměnných v jejich optimálních zpoždění je vytvořen primární regresní model pro vysvětlovanou proměnnou HPI. Do programu jsou importována data včetně zadání zjištěného zpoždění a poté vybrány určené vysvětlující proměnné, viz Obr. 5.

Obr. 5 Výběr proměnných v SW GRETL



Zdroj: SW GRETL

Program zpracuje vložená data a provede taktéž veškeré jemu zadané statistické testy. Využití statistických software zrychluje práci a výpočty. Primární navržená podoba modelu je znázorněna na Obr. 6. V modelu je použito všech 8 vybraných optimálních proměnných v jejich zpoždění.

Obr. 6 Zpožděné korelace, výsledek 1

Model 6: OLS, za použití pozorování 2008:2-2017:2 (T = 33)
Závisle proměnná: hpi

	koeficient	směr. chyba	t-podíl	p-hodnota	
const	-22,2525	47,0781	-0,4727	0,6407	
nabidkove_ceny_1	0,425910	0,136200	3,127	0,0046	***
hdp_1	0,207645	0,119247	1,741	0,0944	*
inflace_1	0,168358	0,392140	0,4293	0,6715	
urok_sazba_5	0,0249994	0,0287762	0,8688	0,3936	
real_ceny_1	0,563908	0,164472	3,429	0,0022	***
poptavka_1	-0,0165910	0,0125738	-1,319	0,1995	
dostupnost_1	-0,140115	0,0388663	-3,605	0,0014	***
spotreba_1	-0,00244279	0,0132426	-0,1845	0,8552	
Střední hodnota závisle proměnné		98,40000			
Sm. odchylka závisle proměnné		6,450920			
Součet čtverců reziduí		27,88980			
Sm. chyba regrese		1,077996			
Koeficient determinace		0,979056			
Adjustovaný koeficient determinace		0,972075			
F(8, 24)		140,2416			
P-hodnota(F)		3,06e-18			
Logaritmus věrohodnosti		-44,04891			
Akaikovo kritérium		106,0978			
Schwarzovo kritérium		119,5664			
Hannan-Quinnovo kritérium		110,6296			
rho (koeficient autokorelace)		0,150897			
Durbin-Watsonova statistika		1,691421			

zde je poznámka o zkratkách statistik modelu

Pomine-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 35 (spotreba_1)

LM test pro autokorelaci až do řádu 1 -
Nulová hypotéza: žádná autokorelace
Testovací statistika: LMF = 0,889223
s p-hodnotou = $P(F(1,23) > 0,889223) = 0,355489$

Whiteův test heteroskedasticity -
Nulová hypotéza: není zde heteroskedasticita
Testovací statistika: LM = 24,0589
s p-hodnotou = $P(\text{Chi-kvadrát}(16) > 24,0589) = 0,0882256$

Test normality reziduí -
Nulová hypotéza: chyby jsou normálně rozdělené
Testovací statistika: Chi-kvadrát(2) = 2,51364
s p-hodnotou = 0,284558

Zdroj: SW GRETL

Interpretace výsledků a výstupy ze software GRETL jsou uvedeny níže. Při výpočtech je možno zadat provedení mnohých statistických testů, pro účely této práce jsou však zvoleny tři hlavní relevantní testy: test autokorelace, test heteroskedasticity a test normality reziduí.

Obr. 7 Vysvětlení jednotlivých položek

Model 6: OLS, za použití pozorování 2008:2-2017:2 (T = 33)
Závisle proměnná: hpi

	1.	směr. chyba	t-podíl	2.	
	koeficient			p-hodnota	
const	-22,2525	47,0781	-0,4727	0,6407	
nabídkove_ceny_1	0,425910	0,136200	3,127	0,0046	***
hdp_1	0,207645	0,119247	1,741	0,0944	*
inflace_1	0,168358	0,392140	0,4293	0,6715	
urok_sazba_5	0,0249994	0,0287762	0,8688	0,3936	
real_ceny_1	0,563908	0,164472	3,429	0,0022	***
poptavka_1	-0,0165910	0,0125738	-1,319	0,1995	
dostupnost_1	-0,140115	0,0388663	-3,605	0,0014	***
spotreba_1	-0,00244279	0,0132426	-0,1845	0,8552	

Střední hodnota závisle proměnné 98,40000
 Sm. odchylka závisle proměnné 6,450920
 Součet čtverců reziduí 27,88980
 Sm. chyba regrese 1,077896
 Koeficient determinace 0,979056
 Adjustovaný koeficient determinace 0,972075
 F(8, 24) 140,2416
 P-hodnota(F) 3,06e-18
 Logaritmus věrohodnosti -44,04891
 Akaikovo kritérium 106,0978
 Schwarzovo kritérium 119,5664
 Hannan-Quinnovo kritérium 110,6296
 rho (koeficient autokorelace) 0,150897
 Durbin-Watsonova statistika 1,691421

zde je poznámka o zkratkách statistik modelu

Pomíne-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 35 (spotreba 1)

LM test pro autokorelaci až do řádu 1 -
 Nulová hypotéza: žádná autokorelace
 Testovací statistika: LMF = 0,889223
 s p-hodnotou = $P(F(1,23) > 0,889223) = 0,355489$

Whiteův test heteroskedasticity -
 Nulová hypotéza: není zde heteroskedasticita
 Testovací statistika: LM = 24,0589
 s p-hodnotou = $P(\text{Chi-kvadrát}(16) > 24,0589) = 0,0882256$

Test normality reziduí -
 Nulová hypotéza: chyby jsou normálně rozdělené
 Testovací statistika: Chi-kvadrát(2) = 2,51364
 s p-hodnotou = 0,284558

Zdroj: SW GRETL

Interpretace výsledků a vysvětlení těchto výsledků významných pro další zkoumání ze software GRETL jsou následující. Jednotlivé položky uvedené na Obr. 7 jsou označeny, očíslovány a níže vysvětleny pod tímto číslem:

1. Sloupec „Koeficient“: Ve sloupci označeném jako „Koeficient“ jsou uvedeny hodnoty, které tvoří regresní model.

2. Sloupec „P-hodnota“: P-hodnota určuje statistickou významnost jednotlivých parametrů (tedy i proměnných). Pokud je hodnota p pro daný parametr nižší než daná hladina významnosti (nejčastěji 0,05), je parametr statisticky významný. Pro přehlednější čtení

výsledků modelu program GRETl zvýrazňuje statisticky významné koeficienty hvězdičkou, kdy * je pro p menší než 0,1, ** pro p menší než 0,05, *** p menší než 0,01. V případě uvedeném v Obr. 6 (popř. Obr. 7) je jako statisticky nevýznamná označená vysvětlující proměnná x_{24} *Výdaje na konečnou spotřebu domácností*. Tato proměnná bude z modelu vyjmuta a model bude nadále zkoumán bez této proměnné.

3. „Koeficient determinace“. Koeficient determinace R^2 ukazuje, z kolika procent je uvažovaná regresní závislost schopna objasnit výchozí variability hodnot vysvětlované proměnné. Statistika R^2 (vhodnost modelu) je v tomto případě 0,979056. Tedy uvažovaná regresní závislost je schopna objasnit 97,91% výchozí variability hodnot vysvětlované proměnné.

4. „Adjustovaný koeficient determinace“. Adjustovaný koeficient determinace R^2_{ADJ} je vhodnější mírou kvality modelu než hodnoty samotného koeficientu determinace R^2 , neboť se jedná o koeficient, který vhodněji reflektuje vhodnost/nevhodnost přidání/odebrání vysvětlující proměnné. Pro celý model je hodnota adjustovaného koeficientu determinace R^2_{ADJ} 0,972075, tedy 97,21%.

5. „Durbin-Watsonova statistika“. Durbin-Watsonova statistika testuje rezidua a určuje, zda existuje statisticky významná korelace založená na pořadí, v jakém se data objevují v datovém souboru. Pokud jsou hodnoty blížíci se 2, pak náhodná složka modelu je prostá autokorelace. Hodnota Durbin-Watsonovy statistiky je v tomto případě 1,691421, což je hodnota blížíci se 2, tedy náhodná složka modelu se zdá být prostá autokorelace. Navíc p -hodnota Durbin-Watsonova testu vyšla 0,355489, tedy vyšší než 0,05. V modelu tedy není přítomna statisticky významná autokorelace na hladině významnosti 5%.

6. „Whiteův test heteroskedasticity“ Test má chí kvadrát rozdělení a spočívá v testování hypotézy o konstantním rozptylu reziduí. Testovaná hypotéza je H_0 : Není zde heteroskedasticita. Sleduje se výsledná p -hodnota na hladině významnosti 0,05. Pokud je p -hodnota větší než stanovená hladina významnosti, nulovou hypotézu nelze zamítnout. Výsledná p -hodnota testu heteroskedasticity je 0,0882256, tedy hodnota větší než 0,05. Nulovou hypotézu nelze zamítnout na hladině významnosti 0,05. Náhodná složka modelu není heteroskedastická, tedy je homoskedastická a je zaručena konstantnost rozptylu reziduí. Předpoklad nepřítomnosti heteroskedasticity v modelu tedy není porušen.

7. „Test normality reziduí“. Při testování normality reziduí se statisticky zjišťuje, zda jsou rezidua náhodně rozložena kolem horizontální osy. Je testována nulová hypotéza H_0 : Chyby jsou normálně rozdělené. Sleduje se výsledná p -hodnota na hladině významnosti 0,05. Pokud je p -hodnota větší než stanovená hladina významnosti, nulovou hypotézu nelze zamítnout. Výsledná p -hodnota testu normality reziduí je 0,284558, tedy hodnota větší než 0,05. Nulovou hypotézu nelze zamítnout na hladině významnosti 0,05. Rezidua jsou v modelu normálně rozložena.

2.4. Zpětná hřebenová regrese

Zpětnou hřebenovou regresí byl zjišťován optimální počet proměnných (je zjišťován systém podstatných veličin), odstraněny byly ty vysvětlující proměnné, které měly největší p -hodnotu. V modelu označeném jako Výsledek 1 (viz Obr. 6), byla p -hodnota u vysvětlující proměnné x_{24} *Výdaje na konečnou spotřebu domácností* ve výši 0,8552, což je hodnota nejvíce se blížíci hodnotě 1 a je tedy tato proměnná nejvíce statisticky nevýznamná. Z tohoto důvodu byla vysvětlující proměnná x_{24} *Výdaje na konečnou spotřebu domácností* z modelu odstraněna. Následně byl proveden výpočet bez této proměnné.

Obr. 8 Zpožděné korelace, výsledek 2

```

Model 7: OLS, za použití pozorování 2008:2-2017:2 (T = 33)
Závisle proměnná: hpi

      koeficient   směr. chyba   t-podíl   p-hodnota
-----
const      -22,7160      46,0938      -0,4928      0,6264
nabidkove_ceny_1    0,411150      0,108066      3,805      0,0008 ***
hdp_1       0,211599      0,115016      1,840      0,0777 *
inflace_1    0,172966      0,383708      0,4508      0,6560
urok_sazba_5    0,0258235      0,0278727      0,9265      0,3631
real_ceny_1    0,575463      0,149109      3,859      0,0007 ***
poptavka_1    -0,0185039      0,00697246     -2,654      0,0136 **
dostupnost_1    -0,142546      0,0358516     -3,976      0,0005 ***

Střední hodnota závisle proměnné      98,40000
Sm. odchylka závisle proměnné      6,450920
Součet čtverců reziduí      27,92934
Sm. chyba regrese      1,056964
Koeficient determinace      0,979027
Adjustovaný koeficient determinace      0,973154
F(7, 25)      166,7129
P-hodnota(F)      2,30e-19
Logaritmus věrohodnosti      -44,07228
Akaikovo kritérium      104,1446
Schwarzovo kritérium      116,1166
Hannan-Quinnovo kritérium      108,1728
rho (koeficient autokorelace)      0,148015
Durbin-Watsonova statistika      1,697771
zde je poznámka o zkratkách statistik modelu

Pomíne-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 31 (inflace_1)

Whiteův test heteroskedasticity -
Nulová hypotéza: není zde heteroskedasticita
Testovací statistika: LM = 20,9496
s p-hodnotou = P(Chí-kvadrát(14) > 20,9496) = 0,10293

Test normality reziduí -
Nulová hypotéza: chyby jsou normálně rozdělené
Testovací statistika: Chí-kvadrát(2) = 2,75796
s p-hodnotou = 0,251835

```

Zdroj: SW GRETL

Zhodnocení modelu výsledku 2: Po odstranění vysvětlující proměnné x_{24} *Výdaje na konečnou spotřebu domácností* se zvýšila hodnota adjustovaného koeficientu determinace R^2_{ADJ} z 0,972075 na 0,973154, tedy z 97,21% na 97,32%, byla potvrzena vhodnost odstranění proměnné. Hodnoty Durbin-Watsonovy statistiky jsou 1,697771, což je hodnota blíží se 2, tedy náhodná složka modelu je prostá autokorelace. Whiteův test heteroskedasticity: *p-hodnota* testu heteroskedasticity je 0,10293, tedy větší než 0,05 a nulovou hypotézu „ H_0 : Není zde heteroskedasticita“ nelze zamítnout na hladině významnosti 0,05, náhodná složka modelu je homoskedastická, předpoklad nepřítomnosti heteroskedasticity v modelu tedy není porušen. Test normality reziduí, chí kvadrát test, ukazuje normální rozdělení chyb, *p-hodnota* na hladině významnosti 0,05 je 0,251835. Nejvyšší *p-hodnotu* ve výši 0,6560 v modelu má vysvětlující proměnná x_2 *Inflace*.

V následujícím výpočtu byla odstraněna další statisticky nejméně významná vysvětlující proměnná x_2 *Inflace* s *p*-hodnotou 0,6560 a bylo zkoumáno, zda se model opět zlepšil, tj. zda se zvýšila hodnota R^2_{ADJ} .

Obr. 9 Zpožděné korelace, výsledek 3

Model 8: OLS, za použití pozorování 2008:2-2017:2 (T = 33)

Závisle proměnná: hpi

	koeficient	směr. chyba	t-podíl	p-hodnota	
const	-2,17210	6,79210	-0,3198	0,7517	
nabidkove_ceny_1	0,395358	0,100651	3,928	0,0006	***
hdp_1	0,183845	0,0956486	1,922	0,0656	*
urok_sazba_5	0,0142312	0,0105839	1,345	0,1904	
real_ceny_1	0,611055	0,124534	4,907	4,30e-05	***
poptavka_1	-0,0181625	0,00682416	-2,661	0,0132	**
dostupnost_1	-0,152422	0,0279395	-5,455	1,01e-05	***

Střední hodnota závisle proměnné	98,40000
Sm. odchylka závisle proměnné	6,450920
Součet čtverců reziduí	28,15635
Sm. chyba regrese	1,040642
Koeficient determinace	0,978856
Adjustovaný koeficient determinace	0,973977
F(6, 26)	200,6125
P-hodnota(F)	1,70e-20
Logaritmus věrohodnosti	-44,20585
Akaikovo kritérium	102,4117
Schwarzovo kritérium	112,8873
Hannan-Quinnovo kritérium	105,9364
rho (koeficient autokorelace)	0,127082
Durbin-Watsonova statistika	1,742615

zde je poznámka o zkratkách statistik modelu

Pomine-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 27 (urok_sazba_5)

Whiteův test heteroskedasticity -

Nulová hypotéza: není zde heteroskedasticita

Testovací statistika: LM = 32,5688

s p-hodnotou = $P(\text{Chí-kvadrát}(27) > 32,5688) = 0,211673$

Test normality reziduí -

Nulová hypotéza: chyby jsou normálně rozdělené

Testovací statistika: Chí-kvadrát(2) = 1,60816

s p-hodnotou = 0,4475

LM test pro autokorelaci až do řádu 4 -

Nulová hypotéza: žádná autokorelace

Testovací statistika: LMF = 0,805669

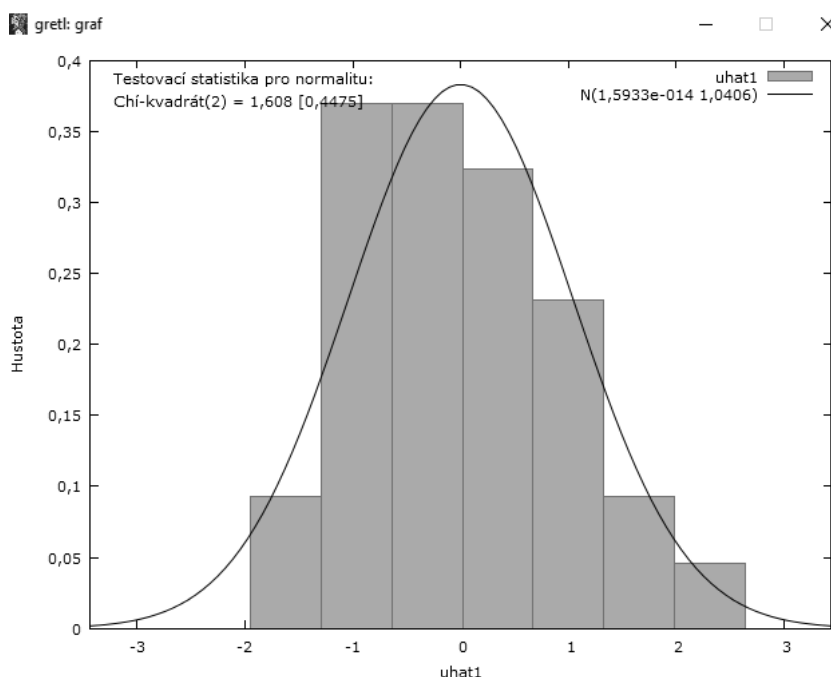
s p-hodnotou = $P(F(4,22) > 0,805669) = 0,534743$

Zdroj: SW GRETL

Zhodnocení modelu výsledku 3: Po odstranění další vysvětlující proměnné x_2 *Inflace* se opět zvýšila hodnota adjustovaného koeficientu determinace R^2_{ADJ} z 0,973154 na 0,973977, tedy z 97,32% na 97,40%, byla potvrzena vhodnost odstranění proměnné. Hodnoty Durbin-

Watsonovy statistiky jsou 1,742615, což je hodnota blíží se 2, tedy náhodná složka modelu je prostá autokorelace. P-hodnota testu pro autokorelaci má navíc hodnotu 0,534743, tedy hodnota je větší než 0,05 a je zde předpoklad nepřítomnosti autokorelace reziduí v modelu. Whiteův test heteroskedasticity, tedy *p-hodnota* testu heteroskedasticity, je 0,211673, tedy větší než 0,05 a nulovou hypotézu „ H_0 : Není zde heteroskedasticita“ nelze zamítnout na hladině významnosti 0,05. Náhodná složka modelu je homoskedastická, předpoklad nepřítomnosti heteroskedasticity v modelu tedy není porušen. Test normality reziduí, chí kvadrát test, ukazuje normální rozdělení chyb, *p-hodnota* na hladině významnosti 0,05 je 0,4475. Graficky je normalita reziduí v tomto modelu zobrazena na Grafu, Obr. 10.

Obr. 10 Normalita reziduí, chí kvadrát test



Zdroj: SW GRETL

Po odstranění další nejméně statisticky významné proměnné je evidentní, že se hodnota R^2_{ADJ} zlepšuje. Po odstranění další vysvětlující proměnné již docházelo ke snižování hodnot adjustovaného koeficientu determinace R^2_{ADJ} a k výraznému snižování hodnot koeficientu determinace R^2 . Hodnota adjustovaného koeficientu determinace R^2_{ADJ} byla při použití nezpožděných proměnných 0,966842, tedy 96,69%, hodnoty postupně s odstraňováním nejvíce statisticky nevýznamných proměnných stoupaly: 97,21%, 97,32%, 97,40%. Je tedy evidentní, že zjištěné zpoždění vysvětlujících proměnných model vylepšil na hodnotu adjustovaného koeficientu determinace R^2_{ADJ} 97,40%. Zpětnou hřebenovou regresí bylo tedy zjištěno, že jako statisticky významné vysvětlující proměnné je vhodné použít proměnné x_1 , x_9 , x_{15} , x_{16} , x_{21} a x_{22} . Tab. 3 vybrané vysvětlující proměnné obsahuje, včetně zjištěného zpoždění, Tab. 4 pak obsahuje data vybraných proměnných.

Do nejužšího výběru se dostaly i proměnné x_{15} , x_{16} , tedy indexy nabídkových cen bytů a indexy realizovaných cen bytů. Výsledná vysvětlovaná proměnná je House price index, která samozřejmě z cen nemovitostí vychází, nicméně se jedná o index obsáhlejší, jak je vysvětleno v kap. 2.1. tohoto článku.

Tab. 3 Optimální vysvětlující proměnné a jejich zpoždění

Označení	Popis vysvětlující proměnné	Zkratka	Optimální zpoždění (kvartální)
y	<i>House price index</i>	HPI	x
x_1	<i>Hrubý domácí produkt</i>	HDP	1
x_9	<i>Průměrná úroková sazba</i>	US	5
x_{15}	<i>Indexy nabídkových cen bytů</i>	NC	1
x_{16}	<i>Indexy real.cen nových bytů</i>	RC	1
x_{21}	<i>Domácí realizovaná poptávka</i>	RP	1
x_{22}	<i>Index dostupnosti bydlení</i>	DB	1

Zdroj: Vlastní tvorba

Tab. 4 Data vybraných vysvětlujících proměnných

	y	x_1	x_9	x_{15}	x_{16}	x_{21}	x_{22}
Čtvrtletí / rok	House Price Index	Hrubý domácí produkt	Průměrná úroková sazba	Indexy nabídkových cen bytů	Indexy real.cen nových bytů	Dom. realizovaná poptávka	Index dostupnosti bydlení
1q/2008	97,20	78,35	259,48	92,73	95,55	47,91	159,62
2q/2008	101,20	83,62	258,55	99,27	99,66	47,91	156,82
3q/2008	102,70	83,55	269,32	104,61	99,85	47,91	156,24
4q/2008	102,10	84,07	265,11	104,70	101,12	47,91	155,23
1q/2009	99,90	78,26	267,45	101,80	103,28	-55,17	154,58
2q/2009	96,70	82,03	258,55	96,37	102,98	-55,17	149,98
3q/2009	95,90	81,20	264,64	96,65	102,40	-55,17	146,04
4q/2009	95,20	82,67	262,30	95,53	99,46	-55,17	135,50
1q/2010	95,20	80,86	253,40	94,32	97,90	19,65	128,19
2q/2010	95,40	81,72	239,34	93,94	98,78	19,65	115,89
3q/2010	95,10	82,68	226,70	93,47	97,51	19,65	111,35
4q/2010	95,20	82,35	207,03	92,54	97,11	19,65	109,51
1q/2011	95,40	82,60	197,66	90,85	97,02	-6,20	108,71
2q/2011	95,90	83,33	196,25	88,98	98,19	-6,20	110,74
3q/2011	95,30	85,58	189,70	88,42	96,14	-6,20	110,77
4q/2011	94,40	85,87	170,96	87,67	94,28	-6,20	107,29
1q/2012	93,90	85,50	169,09	88,80	94,47	-36,19	107,92
2q/2012	94,00	85,03	170,49	90,57	94,77	-36,19	104,76
3q/2012	93,80	85,45	164,40	89,82	93,59	-36,19	102,82
4q/2012	93,70	85,54	152,22	90,39	95,26	-36,19	100,81
1q/2013	93,60	85,23	149,88	90,20	93,50	2,14	100,33
2q/2013	94,10	86,15	139,58	90,39	94,08	2,14	99,91
3q/2013	94,00	85,98	140,52	91,13	93,99	2,14	100,26
4q/2013	93,80	88,72	143,79	92,16	94,47	2,14	101,91
1q/2014	95,00	88,20	140,98	92,91	95,75	43,55	101,97
2q/2014	95,80	94,81	131,62	94,13	95,16	43,55	100,46
3q/2014	96,60	95,81	124,12	95,35	95,45	43,55	99,87
4q/2014	97,30	96,29	114,29	94,69	96,04	43,55	99,44
1q/2015	98,30	98,66	106,32	96,28	97,70	100,00	100,26
2q/2015	99,40	99,70	97,42	98,81	100,44	100,00	98,62
3q/2015	100,60	100,53	98,36	101,15	100,83	100,00	99,20
4q/2015	101,70	103,07	97,89	103,77	101,03	100,00	101,92
1q/2016	103,00	104,27	96,02	106,11	101,32	31,28	102,68
2q/2016	105,10	105,10	88,99	109,01	104,25	31,28	102,54
3q/2016	107,70	105,70	86,65	111,16	104,65	31,28	101,64
4q/2016	112,80	105,95	84,31	114,06	110,42	31,28	106,86
1q/2017	116,10	108,28	88,06	116,21	115,31	41,55	112,85
2q/2017	117,20	109,20	95,08	119,39	119,12	41,55	121,32

Zdroj: www.cnb.cz, www.czso.cz, přepočít na bazické indexy, tabulka: vlastní tvorba

2.5. Vytvoření zpožděného vícenásobného regresního modelu

Zpětnou hřebenovou regresí bylo zjištěno, že statisticky nejvýznamnější je pro tvorbu modelu 6 vysvětlujících proměnných. Testy ukazují, že je možno vytvořit regresní model. Dle vzorce

$$y = a + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n, \quad (5)$$

kde a, b – určují svoji hodnotou vlastnosti dané přímky,
 y – je hodnota závisle proměnné,
 $x_1 \dots x_n$ – je hodnota nezávisle proměnné,

Závisle proměnná, tedy vysvětlovaná proměnná, je v tomto případě House price index (HPI), jehož hodnoty je cílem predikovat, neboť tento index nejlépe ukazuje růst cenové hladiny rezidenčních nemovitostí. Jako nezávisle proměnné, tedy vysvětlující proměnné, jsou zvoleny x_1 Hrubý domácí produkt (HDP), x_9 Průměrná úroková sazba (US), x_{15} Indexy nabídkových cen bytů, celá ČR NC), x_{16} Indexy realizovaných cen nových bytů, celá ČR RC), x_{21} Domácí realizovaná poptávka (RP) a x_{22} Index dostupnosti bydlení (DB).

Hodnoty do modelu byly určeny v posledním výpočtu, viz Obr. 9, data jsou uvedena v Tab. 4. Pro přehled je v Tab. 5 uveden přehled relevantních vysvětlujících proměnných, zjištěné optimální zpoždění a zjištěný koeficient.

Tab. 5 Vybrané vysvětlující proměnné, zpoždění a koeficient

Označení	Popis vysvětlující proměnné	Zkratka	Optimální zpoždění (kvartální)	Koeficient
y	House price index	HPI	x	-2,1721
x_1	Hrubý domácí produkt	HDP	1	0,183845
x_9	Průměrná úroková sazba	US	5	0,0142312
x_{15}	Indexy nabídkových cen bytů	NC	1	0,395358
x_{16}	Indexy real.cen nových bytů	RC	1	0,611055
x_{21}	Domácí realizovaná poptávka	RP	1	-0,0181625
x_{22}	Index dostupnosti bydlení	DB	1	-0,152422

Zdroj: Výpočty SW GRET, tabulka vlastní tvorba

Finální vícenásobný lineární regresní model je následující:

$$y = -2,17210 + 0,183845 * x_{11} + 0,0142312 * x_{95} + 0,395358 * x_{151} + 0,611055 * x_{161} - 0,0181625 * x_{211} - 0,152422 * x_{221} \quad (5)$$

kde y House price index
 x_{11} Hrubý domácí produkt se zjištěným zpožděním
 x_{95} Průměrná úroková sazba se zjištěným zpožděním
 x_{151} Indexy nabídkových cen bytů se zjištěným zpožděním
 x_{161} Indexy real.cen nových bytů se zjištěným zpožděním
 x_{211} Domácí realizovaná poptávka se zjištěným zpožděním
 x_{221} Index dostupnosti bydlení se zjištěným zpožděním

Při dosazení zkratk vysvětlujících proměnných je získán tento vícenásobný lineární regresní model, číslo za zkratkou jednotlivé vysvětlující proměnné značí zpoždění:

$$\text{HPI} = -2,17210 + 0,183845 \cdot \text{HDP1} + 0,0142312 \cdot \text{US5} + 0,395358 \cdot \text{NC1} + 0,611055 \cdot \text{RC1} - 0,0181625 \cdot \text{RP1} - 0,152422 \cdot \text{DB1} \quad (6)$$

kde *HPI* House price index
HDP1 Hrubý domácí produkt se zjištěným zpožděním
US5 Průměrná úroková sazba se zjištěným zpožděním
NC1 Indexy nabídkových cen bytů se zjištěným zpožděním
RC1 Indexy real.cen nových bytů se zjištěným zpožděním
RP1 Domácí realizovaná poptávka se zjištěným zpožděním
DB1 Index dostupnosti bydlení se zjištěným zpožděním

Po dosazení konkrétních dat do modelu je možno zjistit, a tedy predikovat vývoj cen rezidenčních nemovitostí. Z původně vyselektovaných 24 vysvětlujících proměnných bylo statisticky vybráno 6 nejvíce statisticky významných vysvětlujících proměnných za dodržení pravidel regresního modelu. Bylo také zjištěno optimální zpoždění těchto proměnných.

2.6. Ověření pravdivosti výsledků řešení problému

Pro ověření pravdivosti výsledků a platnosti vytvořeného modelu a pro verifikaci dat je možno provést výpočet pro zjištění výsledku, který je již znám, kdy vytvořeným modelem by měla být zjištěna stejná hodnota HPI.

Vybrán byl náhodně druhý kvartál roku 2016, tedy se modeluje predikce hodnoty HPI ve druhém kvartálu roku 2016. Hodnota HPI byla ve druhém kvartálu 2016 známa, ve výši 105,10, výsledek navrženého modelu by měl být obdobný s touto hodnotou. V Tab. 6 je graficky znázorněn postup výpočtu, hodnoty do výpočtu se musí uvažovat se zjištěným optimálním zpožděním.

Postup výpočtu:

Do modelu „ $\text{HPI} = -2,17210 + 0,183845 \cdot \text{HDP51} + 0,0142312 \cdot \text{US5} + 0,395358 \cdot \text{NC1} + 0,611055 \cdot \text{RC1} - 0,0181625 \cdot \text{RP1} - 0,152422 \cdot \text{DB1}$ “ byla dosazena konkrétní čísla, výpočet je graficky demonstrována v Tab. 6, kde jsou zvýrazněny hodnoty, které se do výpočtu dosazují, vyznačená je také hledaná hodnota HPI. Výpočet lze provést v programu MS Excel.

Tab. 6 Hodnoty použité do modelu, hodnoty zpožděného modelu, predikce 2q/2016

	y	x ₁	x ₉	x ₁₅	x ₁₆	x ₂₁	x ₂₂
Kvartál / rok	House Price Index	Hrubý domácí produkt	Průměrná úroková sazba	Indexy nabídkových cen bytů	Indexy real.cen nových bytů	Dom. realizovaná poptávka	Index dostupnosti bydlení
4q/2014	97,30	96,29	114,29	94,69	96,04	43,55	99,44
1q/2015	98,30	98,66	106,32	96,28	97,70	100,00	100,26
2q/2015	99,40	99,70	97,42	98,81	100,44	100,00	98,62
3q/2015	100,60	100,53	98,36	101,15	100,83	100,00	99,20
4q/2015	101,70	103,07	97,89	103,77	101,03	100,00	101,92
1q/2016	103,00	104,27	96,02	106,11	101,32	31,28	102,68
2q/2016	105,10	105,10	88,99	109,01	104,25	31,28	102,54
3q/2016	107,70	105,70	86,65	111,16	104,65	31,28	101,64
4q/2016	112,80	105,95	84,31	114,06	110,42	31,28	106,86
1q/2017	116,10	108,28	88,06	116,21	115,31	41,55	112,85
2q/2017	117,20	109,20	95,08	119,39	119,12	41,55	121,32
Koeficienty	-2,1721	0,183845	0,0142312	0,395358	0,611055	-0,0181625	-0,152422

Zdroj: www.cnb.cz, www.czso.cz, přepočten na bazické indexy, tabulka: vlastní tvorba

$$HPI_{2q/2016} = -2,17210 + 0,183845 \cdot 104,27 + 0,0142312 \cdot 106,32 + 0,395358 \cdot 106,11 + 0,611055 \cdot 101,32 - 0,0181625 \cdot 31,28 - 0,152422 \cdot 102,68 = 106,15$$

Výsledná hodnota HPI za 2q/2016 je na základě vybraného vícenásobného lineárního regresního modelu **HPI 2q/2016 = 106,15**. Hodnota reálná HPI ve 2q/2016 dle dat ČSÚ byla **HPI 2q/2016 = 105,10**.

Procentuální rozdíl, o kolik procent se liší predikce od opravdové hodnoty, je cca 1%. Výpočet je: $(106,15 - 105,1) / 105,1 = 0,0099$ (chyba je tedy přibližně 1 %).

Hodnoty vysvětlované proměnné jsou si velmi blízké a tím byla provedena verifikace modelu.

2.7. Výpočet predikce vývoje trhu rezidenčních nemovitostí

Po vytvoření modelu a jeho ověření je možno přistoupit k výpočtu predikce vývoje cen na rezidenčním trhu, tedy určení výše HPI v následujícím kvartálu. Postup výpočtu:

Do modelu „ $HPI = -2,17210 + 0,183845 \cdot HDP1 + 0,0142312 \cdot US5 + 0,395358 \cdot NC1 + 0,611055 \cdot RC1 - 0,0181625 \cdot RP1 - 0,152422 \cdot DB1$ “, byla dosazena konkrétní čísla, jak je demonstrováno v Tab. 7.

Tab. 7 Hodnoty použité do modelu, hodnoty zpožděného modelu, predikce 3q/2017

	y	x ₁	x ₉	x ₁₅	x ₁₆	x ₂₁	x ₂₂
Kvartál / rok	House Price Index (HPI)	Hrubý domácí produkt	Průměrná úroková sazba	Indexy nabídkových cen bytů, celá ČR	Indexy real.cen nových bytů, celá ČR	Dom. realizovaná poptávka	Index dostupnosti bydlení
4q/2014	97,30	96,29	114,29	94,69	96,04	43,55	99,44
1q/2015	98,30	98,66	106,32	96,28	97,70	100,00	100,26
2q/2015	99,40	99,70	97,42	98,81	100,44	100,00	98,62
3q/2015	100,60	100,53	98,36	101,15	100,83	100,00	99,20
4q/2015	101,70	103,07	97,89	103,77	101,03	100,00	101,92
1q/2016	103,00	104,27	96,02	106,11	101,32	31,28	102,68
2q/2016	105,10	105,10	88,99	109,01	104,25	31,28	102,54
3q/2016	107,70	105,70	86,65	111,16	104,65	31,28	101,64
4q/2016	112,80	105,95	84,31	114,06	110,42	31,28	106,86
1q/2017	116,10	108,28	88,06	116,21	115,31	41,55	112,85
2q/2017	117,20	109,20	95,08	119,39	119,12	41,55	121,32
3q/2017	?						
Koeficienty	-2,1721	0,183845	0,0142312	0,395358	0,611055	-0,0181625	-0,152422

Zdroj: www.cnb.cz, www.czso.cz, přepočít na bazické indexy, tabulka: vlastní tvorba

$$HPI_{3q/2017} = -2,17210 + 0,183845 \cdot 109,20 + 0,0142312 \cdot 88,99 + 0,395358 \cdot 119,39 + 0,611055 \cdot 119,12 - 0,0181625 \cdot 41,55 - 0,152422 \cdot 121,32 = 119,91$$

$$HPI_{3q/2017} = 119,91$$

Model tedy predikuje výrazný růst cen nemovitostí, hodnota bazického indexu byla zjištěna ve výši 119,91, růst oproti předchozímu kvartálu 2017 je predikován ve výši 2,32% (výpočet: $(119,91 - 117,20) / 117,20 = 2,32\%$).

V době dokončování tohoto článku byla čerstvě uveřejněna data k hodnotám HPI ke třetímu kvartálu roku 2017 a byla zpětně upravena data HPI za první a druhý kvartál roku 2017. Hodnoty druhého kvartálu byly upraveny na hodnotu 118,9, hodnota za 3. kvartál roku 2017 je ve výši 121,00. (Trading economics) Procentuální rozdíl, o kolik procent se liší predikce od opravdové hodnoty, je cca 1%. Výpočet je: $(119,91 - 121,0) / 121,0 = -0,009008$ (chyba je tedy přibližně 1 %). Verifikace modelu byla potvrzena i tímto výpočtem.

Predikce pro delší časové období, v horizontu roku 2018, 2019 bude provedena dle výpočtu demonstrovaném výše. Při hledání hodnoty HPI pro takto vzdálená období již do výpočtu vstupují predikovaná data a tedy výsledek je velmi citlivý na vstupní data. Pokud data vysvětlujících proměnných nebude možno dohledat, je možno alternativně přes jednoduchý lineární regresní model tato budoucí data zjistit.

Závěr

Článek se zabýval nalezením vhodného přístupu pro predikci vývoje cen na nemovitostním trhu. Jako vhodným se jeví vytvoření zpožděného vícenásobného lineárního regresního modelu. Tento model byl vytvořen a byla ověřena jeho platnost. Model je také vhodný pro užití v běžné praxi znalců a odhadců díky jednoduchosti výpočtu bez nutného vlastnictví jakéhokoliv výpočetního programu. Znalec či odhadce díky vytvořenému modelu je schopen predikovat vývoj nemovitostního trhu. Výsledek je vázaný na přesnost vstupních dat, což může být problematické, zvláště v případech snahy predikovat vývoj cen pro delší časové období. Zpětná hřebenová regrese při výběru podstatných veličin poukázala na relevantnost či nerelevantnost faktorů, které mohou vstupovat do predikce vývoje nemovitostního trhu. Byly provedeny statistické testy pro ověření splnění předpokladů regresních modelů, mezi kterými je nejvýznamnější pro tuto práci odstranění výskytu multikolinearity, Durbin-Watsonův test na zjištění přítomnosti autokorelace reziduí, Whiteův test pro zjištění přítomnosti heteroskedasticity, byl pozorován koeficient determinace R^2 pro zjištění výše procentuálního vysvětlení skutečnosti modelem a dále byl hodnocen adjustovaný koeficient determinace R^2_{ADJ} , který významněji reflektuje vhodnost odebrání vysvětlujících proměnných, což koeficient determinace R^2 neumí. Odstraňování statisticky méně významných proměnných skončilo u počtu 6 vysvětlujících proměnných, neboť při odstranění další statisticky nejméně významné proměnné již docházelo k výraznému snižování hodnoty koeficientu determinace i adjustovaného koeficientu. To znamená, že by model mohl začínat vysvětlovat menší procento skutečnosti, než bylo před odebráním této další proměnné. Výpočty byly prováděny ve statistickém software GRETl a podpůrně v MS Excel. Novost celkového přístupu spočívá ve využití matematicko-statistických metod, které v běžné praxi znalců a odhadců nejsou využívány.

Literatura:

- [1] MELOUN, M., MILITKÝ, J.: *Kompendium statistického zpracování dat*. Praha. ACADEMIA, 2002. ISBN: 9788024621968
- [2] RYBÁŘ, M.: *Regresní modely a jejich výuka*. Praha 2014. Rigorózní práce. Univerzita Karlova v Praze, Matematicko-fyzikální fakulta.
- [3] HEBÁK, P. *Regrese I.a II. část*. Praha 1998. Vysoká škola ekonomická v Praze, 1998. ISBN: 80-7079-909-9
- [4] ŘEZANKOVÁ, H., LÖSTER, T.: *Základy statistiky*. 1. vyd. Praha: Oeconomica, 2013. ISBN 978-80-245-1957-9
- [5] ČESKÝ STATISTICKÝ ÚŘAD. *Ceny bytů*. Dostupné online z: https://www.czso.cz/csu/czso/ceny_bytu121b-4cb6-a341-30579f827e0a?version=1.0
- [6] TRADING ECONOMICS. *Czech Republic House Price index*. [cit. 2018-20-01]. Dostupné na: <<https://tradingeconomics.com/czech-republic/housing-index>

Regresní model pro predikování vývoje cen na nemovitostním trhu

Darina Tauberová

ABSTRAKT

Článek se zabývá nalezením vhodného přístupu pro možnou predikci vývoje cen na nemovitostním trhu. Jako vhodným se jeví vytvoření zpožděného vícenásobného lineárního regresního modelu. Tento model byl vytvořen a byla ověřena jeho platnost. Model je také vhodný pro užití v běžné praxi znalců a odhadců díky jednoduchosti výpočtu bez nutného vlastnictví jakéhokoliv výpočetního programu. Znalec či odhadce díky vytvořenému modelu je schopen predikovat vývoj nemovitostního trhu. Výsledek je vázaný na přesnost vstupních dat. Zpětná hřebenová regrese při výběru podstatných veličin poukázala na relevantnost či nerelevantnost faktorů, které mohou vstupovat do predikce vývoje nemovitostního trhu. Byly provedeny statistické testy pro ověření splnění předpokladů regresních modelů, mezi kterými je nejvýznamnější odstranění výskytu multikolinearity, Durbin-Watsonův test na zjištění přítomnosti autokorelace reziduí, Whiteův test pro zjištění přítomnosti heteroskedasticity, byl pozorován koeficient determinace. Odstraňování statisticky méně významných proměnných skončilo u počtu 6 vysvětlujících proměnných. Výpočty byly prováděny ve statistickém software GRETl a podpůrně v MS Excel. Novost celkového přístupu spočívá ve využití matematicko-statistických metod, které v běžné praxi znalců a odhadců nejsou využívány.

Klíčová slova: Oceňování nemovitostí; Predikce vývoje cen; Regresní modely.

Regression model and prediction of real estate market development

ABSTRACT

The article deals with finding a suitable approach for prediction of real estate prices. A delayed multiple linear regression model appears to be appropriate. This model has been created and has been validated. The model is also suitable for use in routine practice by experts and appraisers, thanks to the simplicity of the calculation without ownership of any computing program. An expert or an appraiser, thanks to the created model, is able to predict the development of the real estate market. The result is bound to the accuracy of the input data. The backward regression in the selection of significant quantities has highlighted the relevance or inconsistency of factors that can enter the prediction of real estate market development. Statistical determinations were performed to verify compliance with the assumptions of regression models, among which the most significant removal of the occurrence of multi-collinearity, the Durbin-Watson test for detecting the presence of autocorrelation of residues, White's test to determine the presence of heteroskedasticity. Removing statistically less significant variables resulted in the number of 6 explanatory variables. The calculations were carried out in the GRETl statistical software and, in the alternative, in MS Excel. The novelty of the overall approach lies in the use of mathematical and statistical methods, which are not used in the usual practice of experts and appraisers.

Key words: Real estate valuation; Prediction of price development; Regression models.

JEL classification: C1