

۱. مقدمه

۱.۱. هدف از تحلیل هدف اصلی این تحلیل، بررسی و درک پویایی‌ها و ارتباطات متقابل بین چندین متغیر کلیدی اقتصادی مرتبط با بازار مسکن تهران است. علاوه بر این، تلاش شده است تا با استفاده از یک مدل سری زمانی چند متغیره، پیش‌بینی‌هایی برای این متغیرها در آینده نزدیک ارائه شود. درک اینکه چگونه این متغیرها بر یکدیگر تأثیر می‌گذارند، می‌تواند به تصمیم‌گیری‌های آگاهانه‌تر در زمینه سیاست‌گذاری و سرمایه‌گذاری کمک کند.

۱.۲. داده‌های مورد استفاده داده‌های مورد استفاده در این تحلیل شامل مشاهدات ماهانه برای متغیرهای زیر از [تاریخ شروع داده‌ها، مثلاً فروردین ۱۳۹۶] تا [تاریخ پایان داده‌ها، مثلاً مرداد ۱۴۰۳] برای بازار مسکن تهران می‌باشد:

- قیمت (تومان): میانگین قیمت هر متر مربع مسکن.
- CPI: شاخص قیمت مصرف‌کننده، معیاری برای تورم عمومی.
- تعداد معاملات: تعداد معاملات ثبت شده مسکن.
- شاخص اجاره: شاخصی برای نشان دادن روند تغییرات اجاره‌بها.
- نرخ ارز (ریال): نرخ برابری ریال ایران در مقابل یک ارز خارجی مرجع (مثلاً دلار آمریکا).

این داده‌ها از فایل `microeconomics_project new.csv` استخراج شده‌اند.

۱.۳. اهمیت تحلیل تاثیرات متقابل متغیرهای اقتصادی انتخاب شده به ندرت به صورت مجزا عمل می‌کنند و معمولاً تاثیرات پیچیده و متقابلی بر یکدیگر دارند. به عنوان مثال، تغییرات نرخ ارز می‌تواند بر تورم (CPI) و در نتیجه بر قیمت مسکن و شاخص اجاره تأثیر بگذارد. همچنین، تعداد معاملات می‌تواند تحت تأثیر قیمت مسکن و انتظارات تورمی باشد. درک این روابط متقابل برای تحلیل دقیق بازار و پیش‌بینی روندهای آینده ضروری است.

۲. مدل - VAR (Vector Autoregression) مبانی و کاربرد

۲.۱. معرفی مدل VAR: مدل VAR یا "رگرسیون خودگردان برداری" یک مدل آماری قدرتمند برای تحلیل سری‌های زمانی چند متغیره است. در این مدل، هر متغیر در سیستم به عنوان یک ترکیب خطی از مقادیر گذشته (وقفه‌ها یا Lags) خود و مقادیر گذشته تمام متغیرهای دیگر موجود در سیستم مدل‌سازی می‌شود. به عبارت دیگر، VAR به ما اجازه می‌دهد تا ببینیم چگونه تاریخچه تمام متغیرهای منتخب بر مقدار فعلی و آینده هر یک از آن‌ها تأثیر می‌گذارد.

۲.۲. دلایل انتخاب مدل VAR برای این تحلیل: مدل VAR به دلایل زیر برای این تحلیل انتخاب شد:

- **مدل‌سازی تأثیرات متقابل:** این مدل به طور ذاتی برای بررسی و کمی‌سازی روابط دوطرفه و بازخوردی بین متغیرهای مختلف طراحی شده است.
- **پیش‌بینی همزمان VAR:** قادر است پیش‌بینی‌هایی را برای تمام متغیرهای موجود در سیستم به صورت همزمان تولید کند، که با توجه به وابستگی آن‌ها به یکدیگر مفید است.
- **ابزارهای تحلیلی قدرتمند VAR:** ابزارهایی مانند آزمون علیت گرنجر و توابع واکنش به ضربه را فراهم می‌کند که به درک عمیق‌تر دینامیک سیستم کمک می‌کنند.

۲.۳. فرضیات و محدودیت‌های مدل VAR

- **خطی بودن VAR:** فرض می‌کند که روابط بین متغیرها خطی است.
- **ایستایی (Stationarity):** متغیرهای ورودی به مدل VAR باید ایستا باشند، یعنی میانگین، واریانس و ساختار خودهمبستگی آن‌ها در طول زمان ثابت باشد. در غیر این صورت، نتایج مدل قابل اتکا نخواهد بود.
- **حساسیت به تعداد وقفه‌ها:** انتخاب تعداد صحیح وقفه‌ها برای مدل بسیار مهم است و می‌تواند بر نتایج تأثیر بگذارد.
- **محدودیت با داده‌های کم:** با افزایش تعداد متغیرها و وقفه‌ها، تعداد پارامترهای مدل به سرعت افزایش می‌یابد. با حجم داده محدود (مانند داده‌های این پروژه با حدود ۹۰ مشاهده ماهانه)، این موضوع می‌تواند منجر به عدم قطعیت در تخمین پارامترها و پیش‌بینی‌ها شود.

۳. آماده‌سازی داده‌ها

۳.۱. بارگذاری، پاکسازی اولیه و انتخاب متغیرها داده‌ها از فایل CSV خوانده شده و ستون‌های تاریخ بر

اساس ترتیب زمانی صحیح مرتب شدند. پنج متغیر اصلی ذکر شده در بخش ۱.۲ برای ورود به مدل VAR انتخاب شدند. مقادیر غیر عددی (مانند کما در اعداد) حذف و داده‌ها به نوع عددی تبدیل شدند. ردیف‌هایی از داده‌ها که در آن‌ها حداقل یک مقدار ناموجود (NaN) وجود داشت، حذف شدند تا از بروز خطا در مدل‌سازی جلوگیری شود، زیرا مدل VAR نمی‌تواند با مقادیر ناموجود کار کند.

۳.۲. بررسی و اعمال ایستایی (Stationarity) همانطور که اشاره شد، ایستایی یک پیش‌نیاز اساسی برای مدل VAR است.

• ۳.۲.۱. آزمون **Augmented Dickey-Fuller (ADF)** برای بررسی ایستایی هر سری زمانی، از آزمون ADF استفاده شد. در این آزمون، فرضیه صفر (H_0) این است که سری دارای ریشه واحد است (یعنی ایستا نیست). اگر مقدار **p-value** آزمون کمتر از یک سطح معناداری مشخص (معمولاً ۰.۰۵) باشد، فرضیه صفر رد شده و نتیجه می‌گیریم که سری ایستا است.

• ۳.۲.۲. تبدیلات اعمال شده (لگاریتم و تفاضل‌گیری) بر اساس نتایج آزمون ADF روی داده‌های سطح (اصلی):

○ برای متغیر 'قیمت (تومان)' که معمولاً دارای رشد نمایی و واریانس ناهمسان است، ابتدا تبدیل لگاریتم طبیعی (**ln**) اعمال شد تا واریانس پایدارتر شده و رشد به حالت خطی نزدیک‌تر شود.

○ سپس، سری لگاریتمی قیمت و سایر متغیرها که در سطح ایستا نبودند، یک یا چند بار تفاضل‌گیری (**Differencing**) شدند (یعنی اختلاف مقدار فعلی از مقدار قبلی محاسبه شد). هدف از این کار حذف روند و ایستا کردن میانگین سری بود.

• ۳.۲.۳. نتایج ایستایی برای هر متغیر (بر اساس خروجی شما)

○ 'قیمت (تومان)': پس از تبدیل لگاریتمی و یک بار تفاضل‌گیری (**log-diff1**)، ایستا شد ($p\text{-value} = 0.009$).

- 'CPI' پس از یک بار تفاضل گیری (diff1)، ایستا شد (p-value بسیار کوچک).
 - 'تعداد معاملات': پس از یک بار تفاضل گیری (diff1)، ایستا شد. (p-value = 0.017)
 - 'شاخص اجاره': پس از یک بار تفاضل گیری (diff1)، ایستا شد (p-value بسیار کوچک).
 - 'نرخ ارز (ریال)': پس از یک بار تفاضل گیری (diff1)، ایستا شد (p-value بسیار کوچک).
- بنابراین، تمام متغیرهای ورودی به مدل VAR پس از این تبدیلات ایستا شدند. دیتافریم نهایی ایستا شده (df_stationary) شامل این سری‌های تبدیل شده است که برای برازش مدل استفاده می‌شود.

۴. ساخت و برازش مدل VAR

- ۴.۱. انتخاب تعداد وقفه بهینه (Optimal Lag Order) تعداد وقفه‌ها (p) در مدل VAR نشان می‌دهد که مقادیر گذشته تا چند دوره قبل، بر مقادیر فعلی متغیرها تأثیر می‌گذارند. انتخاب وقفه صحیح بسیار مهم است.
- با توجه به حجم داده محدود (حدود ۸۷ مشاهده پس از ایستا سازی)، تعداد وقفه‌های زیاد می‌تواند منجر به بیش‌برازش و کاهش درجه آزادی شود.
 - از تابع `select_order()` در `statsmodels` برای بررسی معیارهای اطلاعاتی مختلف مانند AIC (Akaike Information Criterion) و BIC (Bayesian Information Criterion) برای تعداد وقفه‌های مختلف (مثلاً از ۰ تا ۶) استفاده شد.
 - بر اساس خروجی شما، معیار AIC وقفه ۱ را به عنوان وقفه بهینه پیشنهاد داد (`best_lag = 1`). این انتخاب با توجه به محدودیت داده، منطقی به نظر می‌رسد.

۴.۲. برازش مدل VAR و پارامتر روند (Trend) مدل VAR با استفاده از داده‌های ایستا شده (df_stationary) و تعداد وقفه بهینه انتخاب شده ($p=1$) برازش داده شد.

- پارامتر روند (trend): در مدل VAR، می‌توان یک عبارت ثابت ('c')، روند خطی ('ct')، یا هیچ کدام ('nc') را در معادلات لحاظ کرد. عبارت ثابت در مدل سازی روی داده‌های تفاضل گرفته شده، نقش یک

"روند پایه" یا "میانگین غیر صفر برای تغییرات" را ایفا می‌کند. در تحلیل فعلی، از `trend='c'` (پیش‌فرض) استفاده شد. انتخاب این پارامتر می‌تواند بر پیش‌بینی‌های بلندمدت تأثیرگذار باشد.

۴.۳. بررسی خلاصه نتایج رگرسیون مدل خروجی `var_model_fitted.summary()` نتایج رگرسیون را برای هر یک از ۵ متغیر (در مقیاس ایستا شده) به صورت جداگانه نشان می‌دهد. برای هر معادله:

- ضرایب: **(coefficient)** نشان‌دهنده میزان تأثیر وقفه اول ($L1$) هر متغیر بر متغیر وابسته آن معادله است.

- خطای استاندارد (**std. error**)، آماره **t (t-stat)** و مقدار احتمال (**prob**) یا **p-value** برای ارزیابی معناداری آماری هر ضریب استفاده می‌شوند. اگر **p-value** یک ضریب کوچک باشد (مثلاً کمتر از ۰.۰۵)، آن ضریب به لحاظ آماری معنادار است.

- مثال از خروجی شما :

- در معادله "قیمت(تومان)" (که در واقع تفاضل لگاریتم قیمت است)، ضریب $L1$ قیمت(تومان) برابر ۰.۲۹۲ و معنادار است. ($p=0.008$) این نشان می‌دهد که تغییرات گذشته قیمت، بر تغییرات فعلی قیمت تأثیر مثبت و معناداری دارد. سایر متغیرها در این معادله با وقفه ۱ تأثیر معناداری بر قیمت نداشته‌اند.

- در معادله "تعداد معاملات"، $L1$ قیمت(تومان) ضریب منفی و معناداری دارد (-۰.۴۶۴۰۶) که نشان می‌دهد افزایش در تغییرات لگاریتمی قیمت در دوره قبل، منجر به کاهش در تغییرات تعداد معاملات در دوره فعلی می‌شود.

- ماتریس همبستگی باقیمانده‌ها: **(Correlation matrix of residuals)** نشان‌دهنده همبستگی بین خطاهای مدل برای معادلات مختلف است. در حالت ایده‌آل، این همبستگی‌ها باید کم باشند.

۵. تحلیل دینامیک سیستم با مدل VAR

۵.۱. آزمون علیت گرنجر (**Granger Causality Test**) این آزمون بررسی می‌کند که آیا مقادیر گذشته یک سری زمانی (causing variable) به پیش‌بینی مقادیر آینده سری زمانی دیگر (caused variable) کمک می‌کند یا خیر، با در نظر گرفتن مقادیر گذشته خود متغیر وابسته.

• تفسیر خروجی شما :

- 'قیمت(تومان)' علت گرنجری 'تعداد معاملات' است: (p-value = 0.0000) تغییرات گذشته قیمت به پیش‌بینی تغییرات تعداد معاملات کمک می‌کند.
- 'قیمت(تومان)' علت گرنجری 'شاخص اجاره' است: (p-value = 0.0313) تغییرات گذشته قیمت به پیش‌بینی تغییرات شاخص اجاره کمک می‌کند.
- در سایر موارد بررسی شده، علیت گرنجر معناداری مشاهده نشد. (p-value > 0.05)
- خطای **object has no attribute 'statistic'** این خطا در آخرین اجرای شما نشان می‌دهد که نحوه استخراج آمار F از نتیجه آزمون نیاز به اصلاح دارد. در کد اصلاح شده قبلی، سعی کردم این را با دسترسی مستقیم به `test_result.summary().tables[0].data[1][2].strip()` برطرف کنم که ممکن است در نسخه‌های مختلف `statsmodels` متفاوت باشد. روش صحیح‌تر، استفاده از اتریبوت‌های مستقیم `test_result.statistic` (مانند `test_result.fvalue`) یا `test_result.pvalue` است.

۵.۲. توابع واکنش به ضربه (**Impulse Response Functions - IRFs**) ها نشان می‌دهند که چگونه یک شوک (یک انحراف معیار) ناگهانی و غیرمنتظره به یکی از متغیرهای مدل، در طول زمان بر خود آن متغیر و سایر متغیرهای سیستم تأثیر می‌گذارد.

• نحوه تفسیر نمودارهای IRF:

- محور افقی نشان‌دهنده تعداد دوره‌های زمانی (مثلاً ماه‌ها) پس از اعمال شوک است.
- محور عمودی میزان واکنش را نشان می‌دهد.

- خط ممتد نشان‌دهنده بهترین تخمین واکنش و ناحیه سایه‌دار (در صورت رسم) نشان‌دهنده فاصله اطمینان است. اگر فاصله اطمینان شامل صفر باشد، واکنش به لحاظ آماری معنادار نیست.

- می‌توان دید که آیا واکنش مثبت است یا منفی، آیا به سرعت میرا می‌شود یا پایدار است، و آیا نوسانی است یا خیر.

- **خطای قبلی در رسم:** `Axes' object is not iterable` در خروجی قبلی شما، با اصلاح حلقه پیمایش محورها در کد برطرف شد و اکنون باید ماتریس نمودارهای IRF به درستی نمایش داده شود.

۶. پیش‌بینی با مدل VAR

۶.۱. فرآیند پیش‌بینی با داده‌های ایستا شده مدل VAR برآزش داده شده، برای پیش‌بینی مقادیر آینده متغیرها در مقیاس ایستا شده (یعنی مقادیر تفاضل‌گرفته شده و/یا لگاریتمی شده) برای تعداد مراحل مشخص (مثلاً ۱۲ ماه) استفاده می‌شود.

۶.۲. چالش تبدیل معکوس پیش‌بینی‌ها به مقیاس اصلی این مرحله بسیار حیاتی و مستعد خطا است. پیش‌بینی‌های انجام شده در مقیاس ایستا باید به مقیاس اصلی و قابل تفسیر خود بازگردانده شوند.

- **مراحل تبدیل معکوس:**

1. بازگرداندن تفاضل‌گیری: با استفاده از جمع تجمعی (`cumsum`) و اضافه کردن آخرین مقادیر مشاهده شده صحیح قبل از هر مرحله تفاضل‌گیری.

2. بازگرداندن تبدیل لگاریتمی: با اعمال تابع نمایی (`np.exp()`) روی مقادیر بازگردانده شده از مرحله قبل (اگر متغیر لگاریتمی شده بود).

- **پیش‌بینی‌های غیرمنطقی در خروجی شما:** مقادیر بسیار بالا برای "قیمت (تومان)" و مقادیر منفی برای "شاخص اجاره" در پیش‌بینی‌های نهایی شما، قویاً نشان می‌دهد که تابع تبدیل معکوس `robust_inverse_transform` یا نسخه‌های قبلی آن به درستی عمل نکرده است. این

شایع‌ترین دلیل برای چنین پیش‌بینی‌های غلطی است. منطق این تابع در آخرین نسخه کد ارائه شده به شما، با دقت بیشتری بازنویسی شد تا این مشکل را برطرف کند.

۶.۳. نمایش و تفسیر نمودارهای پیش‌بینی در مقابل مقادیر واقعی این نمودارها به مقایسه بصری مقادیر واقعی مشاهده شده با مقادیر پیش‌بینی شده توسط مدل کمک می‌کنند.

- **محور X:** باید تاریخ‌های صحیح فارسی را برای کل دوره (واقعی و پیش‌بینی) نمایش دهد. (این مورد در آخرین کد اصلاح شد).

- **تفسیر :**

- آیا خط پیش‌بینی، روند کلی داده‌های واقعی را دنبال می‌کند؟
- آیا پیش‌بینی‌ها به سرعت از داده‌های واقعی واگرا می‌شوند یا نزدیکی قابل قبولی دارند؟
- فاصله بین مقادیر واقعی و پیش‌بینی شده (خطای پیش‌بینی) چقدر است؟
- با توجه به پیش‌بینی‌های فعلی شما، انتظار می‌رود که واگرایی شدیدی بین داده‌های واقعی و پیش‌بینی شده برای قیمت و شاخص اجاره در نمودارها مشاهده شود.