

# پیش‌بینی قیمت‌های نقدی گاز طبیعی به کمک مدل‌های غیرخطی ناپارامتریک

نرگس صالح‌نیا<sup>۱</sup> محمد علی فلاحی<sup>۲</sup> احمد سیفی<sup>۳</sup> محمد حسین مهدوی عادل<sup>۴</sup>

تاریخ پذیرش: ۱۳۹۲/۰۳/۱۲

تاریخ دریافت: ۱۳۹۱/۰۸/۱۱

## چکیده

پیش‌بینی دقیق قیمت‌های نقدی گاز طبیعی از اهمیت زیادی برخوردار است زیرا می‌تواند در تصمیم‌گیری‌های نظارتی هر دو جانب عرضه و تقاضای گاز طبیعی مفید واقع شود. لذا در این مطالعه، آزمون گاما جهت قیمت‌های گاز، به‌عنوان یک ابزار غیرخطی و ناپارامتریک استفاده شد تا بتوان بهترین ترکیب ورودی‌ها را قبل از کالبراسیون و آزمون مدل انتخاب نمود. آزمون گاما دارای مدل‌های غیرخطی متعددی مانند رگرسیون خطی موضعی (LLR)، رگرسیون خطی موضعی پویا (DLLR) و شبکه‌های عصبی مصنوعی (ANN) می‌باشند. بدین منظور از قیمت‌های نقدی روزانه، هفتگی و ماهانه‌ی گاز هنری‌هاب از ۱۱۹۹۷/۷ تا ۲۰۱۲/۳/۲۰ استفاده شد. مقایسه‌ی نتایج نشان داد که مدل DLLR از ضریب همبستگی بالاتر و میانگین مربعات خطای پایین‌تر از LLR برخوردار بوده و پیش‌بینی‌های بهتری را بدست می‌دهد. مدل ANN نشان می‌دهد که هرچه دوره‌ی پیش‌بینی کوتاه‌تر باشد نتایج دقیق‌تری را داراست. بنابراین، مدل پیش‌بینی قیمت‌های نقدی روزانه با روش ANN می‌تواند به‌عنوان یک مدل مناسب در نظر گرفته شود. بعلاوه، مدل‌های ANN در مقایسه با مدل‌های LLR و DLLR دارای عملکرد بالاتری است و دقت

۱. دانشجوی دکتری، دانشگاه فردوسی مشهد، نویسنده مسئول، Email: salehnian@gmail.com

۲. استاد گروه اقتصاد، دانشگاه فردوسی مشهد، Email: falahi@um.ac.ir

۳. استادیار گروه اقتصاد، دانشگاه فردوسی مشهد، Email: spring0@um.ac.ir

۴. استاد گروه اقتصاد، دانشگاه فردوسی مشهد، Email: mh-mahdavi@um.ac.ir

بالاتری را جهت پیش‌بینی روند قیمت‌های گاز در مقیاس‌های زمانی متفاوت بدست می‌دهد اما این دسته از مدل‌ها از توانایی لازم جهت پیش‌بینی شوک‌های قیمتی بازار برخوردار نمی‌باشند.

**واژگان کلیدی:** گاز طبیعی، قیمت نقدی، آزمون گاما، مدل غیرخطی ناپارامتریک.

**JEL:** Q42, Q47, C14, C45.

#### ۱. مقدمه

گاز طبیعی یکی از منابع انرژی است که از دهه‌ی ۱۹۷۰ به سرعت در حال رشد بوده است و مطالعات نشان می‌دهند که در آینده نیز تقاضا جهت استفاده از آن در سراسر مناطق دنیا در حال افزایش خواهد بود. مصرف گاز طبیعی به دلایل فراوانی از جمله پراکندگی توزیع آن در جهان، پایین بودن هزینه‌های استخراج، قابل رقابت بودن قیمت آن با سایر انواع انرژی به لحاظ ارزش حرارتی آن، ایجاد آلاینده‌گی زیست‌محیطی کمتر در مقایسه با سایر سوخت‌های فسیلی و مزایای اقتصادی ناشی از استفاده از آن به طور قابل ملاحظه‌ای در حال افزایش است.

به عقیده کارشناسان مسائل انرژی، گاز طبیعی انرژی برتر در قرن ۲۱ بوده و در صورت توسعه تکنولوژی و ایجاد زمینه‌ی مناسب جهت استفاده گسترده‌تر از آن در واحدهای تولید فرآورده‌های با ارزش و باکیفیت نفتی، اهمیت این سوخت پاک در قرن حاضر دو چندان خواهد شد. علاوه بر مصارف مرتبط با انرژی، گاز طبیعی به عنوان یک ماده خام، به‌ویژه در صنایع پتروشیمی، از مصارف غیرانرژی نیز برخوردار است. از اینرو، با کاهش ذخایر نفتی جهان و افزایش تقاضای انرژی، جانشینی گاز طبیعی به جای نفت بسیار محتمل به نظر می‌رسد. انتظار می‌رود که استفاده از گاز طبیعی با نرخ متوسطی معادل ۲/۸٪ در سال رشد داشته و تا سال ۲۰۲۵ این میزان بالغ بر ۱۲۷۶ تریلیون فوت مکعب شود. تحقیقات نشان می‌دهند که سهم گاز طبیعی از مصرف انرژی نیز از ۲۳٪ به ۲۸٪ خواهد رسید.

گام‌های اولیه در جهت آزادسازی بازار گاز طبیعی در سال ۱۹۷۸ برداشته شد و قیمت‌های سرچاهی گاز طبیعی در سال ۱۹۸۹ مورد مقررات‌زدایی واقع شد که این خود گامی مهم برای این بازار تلقی می‌گردد. سرانجام، بازار

گاز طبیعی به یکی از مهمترین بازارهای جهانی از نقطه‌نظر شفافیت قیمت‌ها تبدیل شد (آلبریچ<sup>۱</sup>، ۲۰۱۱). یکی از پیامدهای حرکت به سوی بازارهای کمتر نظارت شده آن است که قیمت‌های گاز طبیعی متلاطم‌تر گشته چراکه قیمت‌ها نوسان کرده و فعالان بازار در پاسخ به این عوامل واکنش نشان می‌دهند. تلاطم قیمت همچنین منجر به افزایش نااطمینانی و ارزش بالقوه‌ی پیش‌بینی‌های حاصل از قیمت‌های گاز طبیعی در تاریخی معین، می‌گردد. به دلیل این عوامل، بررسی و پیش‌بینی جهت تغییر قیمت یکی از انگیزه‌های مهم در انجام پژوهش حاضر می‌باشد.

بنابراین، ارائه‌ی مدل‌هایی برای پیش‌بینی دقیق قیمت‌های گاز طبیعی و سمت و سوی تغییرات قیمت امری مهم بوده زیرا این پیش‌بینی‌ها را می‌توان در تعیین طیف گسترده‌ای از تصمیم‌گیری‌های نظارتی هر دو جانب عرضه و تقاضای گاز و یا برای کلیه‌ی فعالان بازار مورد استفاده قرار داد. این مقادیر همچنین به‌عنوان متغیرهای مهمی در برنامه‌ریزی جهت تعیین ظرفیت تولید برق و نیز تحلیل هزینه-فایده برنامه‌های جانب تقاضا و کارآیی انرژی مدنظر قرار می‌گیرند. باید توجه داشت که روش‌های مختلفی بر پایه مسایل تنوریک و اصول تجارت برای قیمت‌گذاری گاز طبیعی در جهان تعریف شده‌است که برخی از آنها عبارتند از: الف) قیمت‌گذاری بر اساس هزینه؛ که قیمت‌گذاری بر اساس هزینه تمام‌شده گاز در سرچاه انجام می‌شود؛ ب) قیمت‌گذاری بر اساس ارزش؛ در این روش قیمت گاز در نقطه احتراق، در مقایسه با هزینه بهترین سوخت جایگزین تعیین می‌گردد؛ ج) قیمت‌گذاری بر اساس برابری؛ در این دیدگاه ارزش حرارتی همه‌ی حامل‌های انرژی یکسان هستند؛ د) قیمت‌گذاری بر اساس معادل قیمت سرمرز؛ بر اساس این مفهوم، کلیه گاز وارداتی - صرف نظر از منبع عرضه‌کننده - در نقطه‌ای که وارد شبکه خطوط لوله می‌شود، دارای قیمت یکسانی خواهد بود. ه) قیمت‌گذاری بر اساس هزینه جایگزینی؛ بر اساس این مفهوم، صادرکنندگان آن دسته از حامل‌های انرژی را که در طی دوره قرارداد گاز، در دسترس بوده و یا به عنوان سوخت جانشین برای گاز می‌توانند مطرح باشند، مدنظر قرار می‌دهند، مانند: گاز ناشی از زغال سنگ و ... (رحیمی، ۱۳۸۶؛ هنینگ و همکاران<sup>۲</sup>، ۲۰۰۳). تجارت گاز نسبت به تجارت نفت از دو ویژگی خاص شامل: ۱) وابستگی شدید صادرکنندگان و واردکنندگان به یکدیگر به دلیل نبود فرصت ذخیره‌سازی

1. Albrecht  
2 - Henning et al.

حداقل برای دوره قرارداد و ۲) ریسک بالای سرمایه‌گذاری، برخوردار است. با توجه به این دو ویژگی خاص، باید از قراردادهای مدت‌دار (یا ابزار مشتقه) جهت پوشش ریسک ناشی از تغییر قیمت استفاده شود و قیمت‌ها در قراردادها به گونه‌ای تنظیم شوند که ضمن حفظ امنیت عرضه گاز، قیمت این فرآورده با تغییرات احتمالی در قیمت سایر سوخت‌ها به ویژه سوخت‌های رقیب، به گونه‌ای تغییر یابد که منافع عرضه‌کنندگان حفظ گردد. بازارهای گاز امروزه تقریباً حالت رقابتی دارند و بازارهای مهم دنیا عبارت از بازارهای گاز اروپا بخصوص انگلستان، ژاپن در جنوب شرق آسیا، بازار آمریکای شمالی و کانادا می‌باشند. قیمت گاز طبیعی در اروپا بجز بازار انگلیس که قیمت گاز در آن توسط مکانیسم بازار تعیین می‌شود، عمدتاً توسط قراردادهای خرید و فروش بلندمدت و در نتیجه مذاکرات، تعیین می‌شود.

علاوه بر این، قیمت‌های گاز در قراردادهای بلندمدت خط لوله گاز این منطقه معمولاً با فرآورده‌هایی چون گازوئیل و نفت کوره در ارتباطند، زیرا این فرآورده‌ها اصلی‌ترین رقبای گاز طبیعی به ترتیب در بازارهای خانگی و تجاری، صنعتی و تولید برق می‌باشند. در برخی موارد، از سایر حامل‌های انرژی نظیر زغال‌سنگ و الکتریسیته و یک شاخص تورم عمومی جهت تعدیل فرمول خط لوله گاز استفاده می‌شود. در ژاپن میان قیمت‌های گاز و قیمت‌های فرآورده‌های نفتی، حداقل در کوتاه مدت ارتباطی وجود ندارد و عرضه - تقاضای گاز تنها بر اساس قیمت تعیین می‌شود. با وجود چنین حالتی، قیمت‌های نفت و گاز هنوز با هم ارتباط دارند. در این میان، امکان جانشینی نفت خام، قیمت سوخت‌های نفتی و امکانات فنی مصرف‌کننده‌ها برای تغییر دادن سوخت‌ها، در زمره عوامل هستند که در بلندمدت بر قیمت گاز تأثیر می‌گذارند. قیمت - گذاری گاز در آمریکای شمالی بر اساس قیمت برخی مراکز منسجم عرضه (Hubs) انجام می‌شود که یکی از مشهورترین آنها نری‌هاب در لوئیزیانای شمالی مستقر است. عوامل مؤثر بسیار زیادی بر قیمت - گذاری گاز در Henry Hub تأثیر می‌گذارد و این امر تأثیر قابل ملاحظه‌ای بر قیمت عرضه گاز به مصرف‌کنندگان نهایی خواهد داشت؛ البته خاطر نشان می‌شود که شاخص قراردادهای منعقد شده قیمت‌های بازار اسپات یا نقدی (Spot) است.

امروزه بازارهای نقدی بخش مهمی از تجارت بین‌المللی حامل‌های انرژی را دربردارند. این بازارها با تغییر دادن روند تجارت، زمینه‌ای را فراهم می‌آورند که هزینه فرصت دارایی‌های بلااستفاده، به خوبی مشاهده شود. به دلیل این که این بازارها، قیمت‌های جاری را مشخص می‌کنند، بنگاه‌ها و فعالان اقتصادی به راحتی تغییرات کوچک در ارزش دارایی‌های خود را مشاهده کرده و برای عرضه به این بازارها،

تصمیم‌های لازم را می‌گیرند. خریداران نیز در این بازارها به راحتی قیمت پیشنهادی عرضه‌کننده را با قیمت بازار اسپات مقایسه می‌کنند (منصور کیایی، ۱۳۸۷؛ اداره انرژی آمریکا، ۲۰۰۲؛ برون و ماین<sup>۱</sup>، ۲۰۰۸).

بطور کلی، اتخاذ تصمیم براساس پیش‌بینی‌های انجام شده می‌تواند در کلیه سطوح و بخشهای اقتصاد، امری قابل توجه باشد. در بخش انرژی، که عمدتاً به دلیل وجود عدم قطعیت در تولید، دارای ریسک در تصمیم‌گیری می‌باشد، تصمیم‌گیرندگان نیازمند اطلاعاتی راجع به پیامدهای آتی محتمل می‌باشند. لذا، دسترسی به پیش‌بینی‌های دقیق از قیمت‌های نقدی از یک، دو، سه و ... ماه قبل و یا بیشتر در اتخاذ تصمیم‌های تولیدی سودآور، خرید، و برنامه‌ریزی بسیار ضروری به نظر می‌رسد. تاکنون مطالعات زیاد و روشهای متعددی در رابطه با پیش‌بینی قیمت صورت پذیرفته است. یکی از روشهای مناسب در این زمینه مبتنی بر نوع داده‌های موجود، وضعیت مطلوب و سطح جزئیاتی است که در تخمین‌ها مدنظر می‌باشند. برخی از مطالعات انجام شده در این حوزه شامل تخمین تقاضای گاز طبیعی (آزاده و همکاران<sup>۲</sup>، ۲۰۱۰)، تعیین قیمت گاز طبیعی در بخش خانگی (دادک و همکاران<sup>۳</sup>، ۲۰۰۶)، و روند قیمت بلندمدت گاز طبیعی (مک‌اووی و مشکین<sup>۴</sup>، ۲۰۰۰) می‌باشند. وو و همکاران<sup>۵</sup> (۲۰۰۵) با استفاده از مدل رگرسیونی تعدیل جزئی به ارزیابی روند قیمت گاز طبیعی در کالیفرنیا پرداختند. سرلتیس و رنجل‌روئیز<sup>۶</sup> (۲۰۰۴) نیز با استفاده از قیمت‌های نقدی بازار هنری‌هاب و آلبرتا به بررسی ارتباط متقابل موجود بین این دو بازار پرداخته و دریافتند که از زمان مقررات‌زدایی به بعد قیمت‌های گاز طبیعی آمریکای شمالی بطور گسترده‌ای با کمک روند قیمت‌های موجود در هنری‌هاب قابل تعریف است. قابل توجه اینکه برخی مطالعات تورش به سمت پایین استفاده از قیمت‌های آتی گاز طبیعی نایمکس در پیش‌بینی قیمت‌های نقدی می‌باشند و این مسئله بیانگر وجود ریسک می‌باشد (والس<sup>۷</sup>، ۱۹۹۵؛ مجتهدی و موثق<sup>۸</sup>، ۲۰۰۵ و موثق و مجتهدی، ۲۰۰۵).

- 
- 1 - Brown and Mine
  - 2 - Azadeh et al.
  - 3 - Dudek et al.
  - 4 - MacAvoy and Moshkin
  - 5 - Woo et al.
  - 6 - Serletis and Rangel-Ruiz
  - 7 - Walls
  - 8 - Movassagh and Modjahedi

در بازارهای انرژی، مدل‌های از پایین به بالای<sup>۱</sup> زیادی وجود دارند که دربرگیرنده‌ی مباحث بنیادین جانب عرضه و تقاضا می‌باشند (در این زمینه می‌توان به پژوهش‌های انجام گرفته توسط فلتن و لمینگ<sup>۲</sup>، ۲۰۰۳؛ کامباروگلو و مدلنر<sup>۳</sup>، ۲۰۰۳؛ مارتینسن و همکاران<sup>۴</sup>، ۲۰۰۳ اشاره کرد). مادامیکه، بیشتر تحلیل‌گران و مدلسازان از این دسته از مدل‌ها بهره می‌گیرند، مدل‌های سری زمانی که تنها نیازمند دستیابی به قیمت‌های بازار می‌باشند، نسبت به داده‌های مورد نیاز جهت مدل‌های از پایین به بالا با محدودیت کمتری به لحاظ دستیابی به داده مواجهند. پیش‌بینی قیمت‌های کالاهای انرژی مرتبط با هم با استفاده از هر دو دسته تکنیک‌های خطی و غیرخطی را می‌توان در پژوهش انجام شده توسط مالیاریسا و مالیاریسب (۲۰۰۸) مشاهده کرد. نتایج آنها نشان می‌دهد که پیش‌بینی‌های غیرخطی برای نفت خام، نفت حرارتی، بنزین و گاز طبیعی بهترین تخمین‌ها را بدست داده در حالی که استفاده از مدل خطی برای پروپان، دارای کمترین خطا می‌باشد. از اینرو مشاهده می‌شود که پژوهش‌های فراوانی در خصوص تحلیل و پیش‌بینی قیمت کالاهای انرژی وجود دارند.

بطور کلی روش‌های مورد استفاده در این مطالعات را می‌توان به دو دسته تقسیم کرد: مدل‌های ساختاری و روش‌های داده‌ای. برخی از مدل‌های ساختاری در ترسیم و طراحی بازار جهانی نفت و تحلیل تلاطم قیمت نفت برحسب تعادل عرضه و تقاضا بکار رفته‌اند (بکن<sup>۵</sup>، ۱۹۹۱؛ الفارس<sup>۶</sup>، ۱۹۹۱؛ هانتینگتون<sup>۷</sup>، ۱۹۹۴؛ زمانی<sup>۸</sup>، ۲۰۰۴؛ یانگ و همکاران<sup>۹</sup>، ۲۰۰۲). استفاده از مدل‌ها و روش‌های داده‌ای شامل مدل‌های خطی مانند خودهمبسته میانگین متحرک (ARMA)، خودهمبسته واریانس ناهمسانی شرطی (ARCH) (سادورسکای<sup>۱۰</sup>، ۲۰۰۲؛ مورانا<sup>۱۱</sup>، ۲۰۰۱؛ بوچانانان و همکاران<sup>۱۲</sup>، ۲۰۰۱؛ ابریشمی و همکاران، ۱۳۸۶)،

- 
- 1 - Bottom-up models
  - 2 - Fleten & Lemming
  - 3 - Kumbarouglu & Madlener
  - 4 - Martinsen et al.
  - 5 - Bacon
  - 6 - Al Faris
  - 7 - Huntington
  - 8 - Zamani
  - 9 - Yang et al.
  - 10 - Sadorsky
  - 11 Morana
  - 12 . Buchanan et al.

مدل‌های غیرخطی مانند شبکه عصبی مصنوعی (میرمیرانی و لی<sup>۱</sup>، ۲۰۰۴؛ مشیری<sup>۲</sup>، ۲۰۰۴؛ یو و همکاران<sup>۳</sup>، ۲۰۰۶) و مدل‌های رگرسیون برداری پشتیبان (ژای و همکاران<sup>۴</sup>، ۲۰۰۶)، می‌باشد.

مطالعات نشان می‌دهند که عوامل متعددی از جانب عرضه و تقاضای گاز طبیعی می‌توانند سبب تاثیرگذاری بر قیمت آن گردند؛ در جانب عرضه، تغییر ذخیره‌سازی گاز طبیعی، تولید، واردات یا قیود و محدودیت‌های تحویل و عرضه گاز از آن جمله می‌باشند. در بین این عوامل، سطوح ذخیره‌سازی به دلیل ایجاد شرایط مصونیت از ریسک فیزیکی در طی دوره‌های با تقاضای بالا، از اهمیت بسیار زیادی برخوردار است. اختلالات ایجاد شده از جانب شرایط سخت آب و هوایی، وقایع ناگوار در بهره‌برداری، یا تعمیرات برنامه‌ریزی شده نیز می‌توانند باعث ایجاد شرایط سخت کوتاه‌مدت در عرضه‌ی گاز گردند.

همچنین یکی دیگر از عوامل تاثیرگذار از جانب عرضه، توسعه‌ی ذخایر گاز شل آمریکا به دلیل وجود فناوری‌های جدید حفاری، مانند شکست هیدرولیکی و حفاری افقی می‌باشد که باعث افزایش میزان ذخایر گاز برآوردی آمریکا در سال ۲۰۰۹ تا ۳۵٪ گردید که این خود بر قیمت گاز تاثیر داشته و می‌تواند موجب کاهش آن گردد (استودیوکر<sup>۵</sup>، ۲۰۱۱). در جانب تقاضا نیز، تغییرات درجه حرارت یکی از قویترین عوامل تاثیرگذار در کوتاه‌مدت بوده و اساساً نوسانات قیمت گاز طبیعی در آمریکا، عموماً سیکلی/ فصلی می‌باشند. در طی ماههای سرد، این مسئله می‌تواند منجر به ایجاد فشار روبه بالا بر قیمت‌ها گردد. اگر شرایط آب و هوایی غیرقابل انتظار یا سختی بوقوع پیوندد، تاثیر بر قیمت‌ها تشدید می‌گردد چراکه عرضه اغلب قادر نمی‌باشد که در برابر واکنش‌های کوتاه‌مدت تقاضا، بسرعت عکس‌العمل نشان دهد. همچنین درجه حرارت بالاتر از سطح نرمال در طول تابستان می‌تواند منجر به استفاده از ذخایر گاز بیشتر جهت تامین سوخت نیروگاه‌های مصرف‌کننده آن که به تولید نیروی الکتریسیته اشتغال دارند، گردد. همچنین، افزایش قیمت یا وقوع اختلال عرضه‌ی هرگونه از بازارهای سوخت رقیب، می‌تواند به افزایش تقاضای گاز منجر گردد. مثلاً زمانی که قیمت محصولات نفتی افزایش می‌یابد، بخش صنعتی تقاضای گاز طبیعی را افزایش

- 
1. Mirmirani and Li
  2. Moshiri
  3. Yu et al.
  4. Xie et al.
  5. Studebaker

می‌دهد. بعلاوه، میزان فعالیت اقتصادی به‌عنوان عامل مهمی است که بازارهای گاز طبیعی را تحت‌تأثیر قرار می‌دهد. زمانیکه اقتصاد رشد می‌کند، تقاضای افزایش یافته جهت کالاها و خدمات بخش‌های تجاری و صنعتی باعث افزایش تقاضای گاز طبیعی می‌گردند. البته، قیمت نقدی گاز طبیعی نیز بطور پیچیده‌ای با میزان اعتبار مجوز انتشار گازهای گلخانه‌ای مرتبط می‌باشد. بطور خاص، می‌توان انتظار داشت که قیمت گاز طبیعی همگام با افزایش قیمت مجوز انتشار گازهای گلخانه‌ای افزایش یابد. عوامل تعیین‌کننده‌ی تقاضا در بلندمدت نیز نقش بسیار مهمی را جهت جایگاه آتی گاز طبیعی بازی می‌کنند. این عوامل شامل دورنمای قوانین و سیاست‌گذاری‌های مرتبط با تغییر اقلیم در آمریکا نیز می‌باشد (واین، ۲۰۱۰؛ اداره انرژی آمریکا، ۲۰۱۱؛ دهکردی و پاشنگ، ۱۳۸۸).

علیرغم وجود مطالعات فراوانی که با استفاده از مدل‌های کلاسیک سری زمانی، مدل‌های مالی و تکنیک‌های غیرخطی مانند شبکه‌های عصبی مصنوعی به مدلسازی و پیش‌بینی قیمت انواع کالاهای انرژی پرداخته‌اند، هنوز هم سؤالات بسیار زیادی وجود دارند که باید بدنبال یافتن پاسخی برای آنها بود؛ مثلاً، تعداد داده‌های مورد نیاز جهت انجام بهینه‌ترین و دقیق‌ترین پیش‌بینی چقدر است؟ کدامیک از داده‌های ورودی به مدل، مرتبط با پیش‌بینی و کدامیک غیرمرتبط با آن می‌باشند؟ بعلاوه، ادبیات موجود در زمینه‌ی پیش‌بینی قیمت‌های گاز طبیعی بر دو دسته‌ی اصلی از مدل‌های خطی، تک‌معادله‌ای و خلاصه‌شده‌ی اقتصادسنجی متمرکز شده‌اند. اولین دسته از این مدل‌ها، مدل‌های مالی می‌باشند که بطور مستقیم برآمده از تئوری اقتصاد مالی و بر مبنای فرضیه‌ی بازار کارآمد (MEH) بوده، درحالی‌که مدل‌های متعلق به دسته‌ی دوم، مدل‌های ساختاری بوده و به بررسی تأثیر عوامل و متغیرهای واقعی بازار گاز طبیعی بر قیمت‌های گاز طبیعی می‌پردازند. هر دو دسته از مدل‌های مالی و ساختاری اغلب از تصریح‌سازی‌های سری زمانی استفاده می‌کنند.

به‌هرحال، در مقاله‌ی حاضر، به دلیل توسعه و پیشرفت فناوری‌های مدرن محاسباتی و با کمک از یک الگوریتم جدید سعی در ارائه‌ی دسته‌سومی از این مدل‌ها را داریم که روشی ناپارامتریک بوده و آزمون



گاما<sup>۱</sup> (GT) نامیده می‌شود (استفانسون و همکاران<sup>۲</sup>، ۱۹۹۷؛ کنکار<sup>۳</sup>، ۱۹۹۷؛ تسو<sup>۴</sup>، ۱۹۹۹ و گودس‌دالیویرا<sup>۵</sup>، ۱۹۹۹).

روش‌های ناپارامتریک جهت پیش‌بینی سری‌های زمانی را می‌توان به‌عنوان حالت خاصی از برآورد رگرسیون ناپارامتریک تحت وابستگی در نظر گرفت. در این زمینه نیز مطالعات زیادی انجام شده که می‌توان به پژوهش‌های گایورفی و همکاران<sup>۶</sup> (۱۹۸۹)، هاردل و ویو<sup>۷</sup> (۱۹۹۲)، هارت<sup>۸</sup> (۱۹۹۱ و ۱۹۹۶)، مسری و جوستیم<sup>۹</sup> (۱۹۹۵)، هاردل و همکاران (۱۹۹۷ و ۱۹۹۸) و بسک<sup>۱۰</sup> (۱۹۹۸) اشاره کرد. ایوانز<sup>۱۱</sup> (۲۰۰۲)، رمسان و همکاران<sup>۱۲</sup> (۲۰۰۸) و کمپ<sup>۱۳</sup> (۲۰۰۶) به اثبات و ارائه‌ی آزمون گاما پرداخته‌اند. این تکنیک جدید به مدلسازان کمک می‌کند تا با اعمال بهترین ترکیب از داده‌های ورودی، خروجی موردنظر را تعیین نموده و با کمک داده‌ها بطور مستقیم و کارآ به برآورد خروجی، آماره‌ی گاما، از میانگین کمترین مربعات خطا بر روی خروجی و از طریق یک مدل هموار<sup>۱۴</sup> بپردازند. با بررسی آماره‌ی گاما جهت انتخاب‌های متفاوت از متغیرهای ورودی به مدل، می‌توان انتخابی را برگزید که میانگین مربعات خطای انتظاری را کمینه سازد.

در پژوهش حاضر نیز سعی بر آن است تا با استفاده از این روش، داده‌های سری زمانی مورد بررسی قرار گیرند. بعلاوه، تکنیک گاما به وسیله‌ی برآورد نزدیکترین مدل هموار که قادر به برازش بر داده‌های مشاهده نشده می‌باشد، به حل مشکل آموزش بیش از حد که تقریباً کلیه‌ی تکنیک‌های مدل‌سازی غیرخطی مانند ANN دچار آن هستند، می‌پردازد (رمسان و همکاران، ۲۰۰۸).

مدل‌سازی داده‌های سری زمانی، نیازمند ساخت مدلی بر مبنای انتخاب مقادیر گذشته تا  $m$  عدد (که

- 
1. Gamma Test
  2. Stefánsson et al.
  3. Koncar
  4. Tsui
  5. Guedes de Oliveira
  6. Gyorfí et al.
  7. Hardle & Vieu
  8. Hart
  9. Masry & Tjostheim
  10. Bosq
  11. Evans
  12. Remesan et al.
  13. Kemp
  14. smooth model

اغلب آن را بعد تعبیه‌سازی<sup>۱</sup> می‌نامند) بوده تا ورودی‌های مدل مشخص شود. خروجی مدل نیز، مقدار جاری سری زمانی می‌باشد. بنابراین، تعبیه‌سازی یک سری زمانی عبارت از انتخاب مقادیر گذشته‌ای است که جهت پیش‌بینی مقدار جاری از طریق مدل ساخته شده از روی داده‌ها، مورد استفاده قرار می‌گیرد. یک فرآیند تعبیه‌سازی معمولی<sup>۲</sup> از کلیه‌ی مقادیر گذشته تا  $m$  استفاده می‌کند. مبنای اصلی این شیوه، اولین بار جهت مدل‌سازی سیستم‌های پویا و با استفاده از تعبیه‌سازی معمولی توسط تیکنر<sup>۳</sup> (۱۹۸۱) مورد مطالعه قرار گرفت. روش تعبیه‌سازی غیرمعمولی<sup>۴</sup> به انتخاب زیرمجموعه‌هایی از  $m$  مقدار گذشته پرداخته و مادامیکه  $m$  انتخاب می‌شود تعداد  $2^m - 1$  تعبیه‌سازی غیرمعمولی وجود دارد. جاد<sup>۵</sup> (۱۹۹۸) پیشنهاد می‌کند که تعبیه‌سازی غیرمعمولی، اغلب مدل بهتری را بدست می‌دهد. لذا، ملاحظه می‌شود که انتخاب یک تعبیه‌سازی غیرمعمولی مناسب جهت پیش‌بینی مدل‌های سری زمانی، حائز اهمیت فراوانی است.

همان‌گونه که قبلاً نیز بیان شد، یکی از مشکلات کلیدی در ارائه‌ی یک مدل هموار از داده‌های سری زمانی، تعیین متغیرهای وقفه‌دار جهت پیش‌بینی یک خروجی موردنظر می‌باشد. تسو و همکاران (۲۰۰۲) نیز به بیان آزمون گاما به عنوان یک روش ناپارامتریک پرداخته‌اند که می‌تواند جهت انتخاب ترکیبی از متغیرهای وقفه‌دار که بهترین مدل هموار از یک خروجی را بدست می‌دهند، بکار رود. این محققان به این نتیجه رسیدند که آزمون گاما، ابزاری مؤثر در تعیین تعبیه‌سازی‌های غیرمعمولی سری‌های زمانی می‌باشد. بعلاوه، شنگ و هایندمن<sup>۶</sup> (۲۰۰۹)، از یک مدل ناپارامتریک در پیش‌بینی سری زمانی فصلی التینو از سال ۱۹۵۰ تا ۲۰۰۸ بهره گرفته و چهار روش پویا جهت بهبود میزان دقت پیش‌بینی نقطه‌ای مدل پیشنهاد می‌کنند. روش‌های پویای پیشنهادی و مدل پیش‌بینی آنها مبتنی بر داده و به لحاظ محاسباتی بسیار سریع می‌باشند و از اینرو می‌توان آنها را در عمل مورد استفاده قرار داد. هدف اصلی این پژوهش، استفاده از تکنیک گاما در انتخاب تعبیه‌سازی‌های غیرمعمولی جهت داده‌های سری زمانی و ارائه‌ی یک مدل پیش‌بینی مناسب جهت سری زمانی قیمت‌های نقدی گاز طبیعی می‌باشد. بدین منظور، از سه روش

- 
1. embedding dimension
  2. regular embedding
  3. Takens
  4. irregular embedding
  5. Judd
  6. Shang and Hyndman

مدل‌سازی رگرسیون خطی<sup>۱</sup> (LLR) موضعی، رگرسیون خطی موضعی پویا<sup>۲</sup> (DLLR) و شبکه‌های عصبی مصنوعی (ANN) که به دلیل قابلیت انعطاف‌پذیری بالای نرم‌افزار متلب (Matlab) در تغییر نرون‌ها و لایه‌ها در این نرم‌افزار مورد آموزش قرار داده شده، استفاده شده است.

## ۲. مواد و روش‌ها

### ۱.۲. منطقه مطالعاتی و داده‌های مورد استفاده

کشور آمریکا دارای بازارهای نقدی متعدد به‌همراه فروشندگان و خریدارانی است که گاز طبیعی و سایر خدمات را مورد مبادله قرار می‌دهند. یک بازار نقدی معمولاً دارای خط‌لوله‌هایی است که مناطق مختلف را بهم متصل ساخته و از هزینه‌های مبادلاتی پایینی نیز برخوردار می‌باشد. بسیاری از مراکز خدمات عام‌المنفعه، مشاورین، و سایر نهادهایی مانند اداره انرژی آمریکا (EIA) سعی در پیش‌بینی قیمت‌های انرژی‌هاب دارند. از اینرو، هدف این مقاله نیز پیش‌بینی قیمت‌های نقدی گاز طبیعی انرژی‌هاب و براساس داده‌های تاریخی این بازار از ۷ ژانویه ۱۹۹۷ تا ۲۰ مارس ۲۰۱۲ می‌باشد. خصوصیات آماری قیمت‌های نقدی روزانه‌ی گاز طبیعی در جدول ۱ ارائه شده است.

در مدل‌سازی کاربردی و تجربی پژوهش حاضر، از سری زمانی قیمت‌های نقدی روزانه، هفتگی و ماهانه در طی دوره‌ی زمانی مذکور استفاده شده است. انرژی‌هاب، بزرگترین و سیال‌ترین نقطه‌ی قیمت‌گذاری در دنیا و به‌عنوان مبنایی جهت قراردادهای آتی گاز طبیعی و سایر ابزار مشتقه می‌باشد. داده‌های مورد نیاز این مطالعه از سایت گروه CME<sup>۳</sup> و نیز سایت EIA<sup>۴</sup> جمع‌آوری شده است. داده‌های قیمت نقدی انرژی‌هاب برحسب دلار بر میلیون بی‌تی‌یو گزارش شده‌اند.

1- Local Linear Regression

2- Dynamic Local Linear Regression

3 - Chicago Mercantile Exchange(CME) Group Website:( <http://www.cmegroup.com>).

4 - Energy Information Administration(EIA) Web Page:(<http://tonto.eia.gov/dnav/ng/hist/mgwhhda.htm>).

## جدول ۱. خصوصیات آماری قیمت نقدی روزانه گاز طبیعی هنری هاب

از ۲۰۱۲/۳/۲۰ تا ۱۹۹۷/۰۱/۰۷

بازه زمانی	تعداد مشاهدات	میانگین	میانه	حد اکثر	حد ادکل	انحراف استاندارد	جولگی	کشیدگی	جاک-بوا
روزانه	۳۸۰۳	۴/۸۵۴	۴/۳۹۰	۱۸/۴۸	۱/۰۵۰	۲/۴۲۶	۱/۱۴۰	۴/۶۷۹	۱۲۷۱/۱۱۲
هفتگی	۷۹۲	۴/۸۶۵	۴/۳۸۸	۱۴/۴۹	۱/۳۴۰	۲/۴۳۳	۱/۱۱۶	۴/۴۶۰	۲۳۴/۷۲۱
ماهانه	۱۸۲	۴/۸۸۱	۴/۴۲۶	۱۳/۴۲	۱/۷۲۲	۲/۴۱۳	۱/۰۵۵	۴/۱۹۸	۴۴/۶۲۶

منبع: نتایج محاسبات تحقیق

## ۲.۲. مدل سازی داده‌ای و روش ناپارامتریک

در طی ۷ سال گذشته، انقلاب شگرفی در خصوص مدل‌سازی و پیش‌بینی به کمک روش‌های داده‌ای، ناپارامتریک و غیرخطی رخ داده است. تکنیک‌های مدل‌سازی با استفاده از روش‌های داده‌ای مستقیماً بدنبال ایجاد مدل‌هایی از یک سیستم با استفاده از مجموعه اندازه‌گیری‌های انجام شده از رفتار سیستم و بدون در نظر گرفتن هرگونه دانشی از قوانین منطقی یا معادلاتی می‌باشند که به تعیین این رفتار می‌انجامند. این دسته از مدل‌ها طیف گسترده‌ای را تشکیل می‌دهند که بازه‌ی آنها از توابع منطقی تا سیستم‌های مبتنی بر قوانین تعریف شده و مدل‌های احتمالی مرتبط با توابع پارامتری، می‌باشد. اگر فرض شود که مجموعه‌ای از داده‌های ورودی-خروجی بصورت زیر در دست باشد:

$$\{x_1(i), \dots, x_m(i), y_i\} = \{(x_i, y_i) | 1 \leq i \leq M\} \quad (1)$$

که بردار  $X = (x_1, \dots, x_m) \in R^m$  به‌عنوان ورودی مدل بوده که محدود به مجموعه‌ای بسته و کراندار بصورت  $y \in R$ ،  $C \subset R^m$  خروجی مدل و  $M$  نیز نشان‌دهنده‌ی تعداد سری مجموعه مشاهده شده از پدیده مورد بررسی می‌باشد. سؤالی که عموماً مطرح می‌شود آن است که خروجی تا چه میزانی توسط ورودی تعیین می‌شود؟ روشی که بطور سنتی جهت مدل‌سازی داده‌ای بکار می‌رود، در نظر گرفتن فروضی خاص در رابطه با شکل رابطه‌ی موجود بین ورودی  $x \in R^m$  و خروجی متناظر آن به‌صورت  $y \in R$  و سپس تلاش در جهت یافتن "بهترین برازش" جهت پارامترهای موجود در رابطه‌ی فرضی در مقایسه با داده‌های مشاهده شده می‌باشد.

این روش منجر به مطالعه‌ی آمار ناپارامتریک گشت (در این زمینه می‌توان به پژوهش بیتس و واتس<sup>۱</sup> (۱۹۸۸) اشاره کرد). پی و پیترسون<sup>۲</sup> (۱۹۹۴)، راه‌حل جدید ناپارامتریکی را برای حل این مسئله‌ی عمومی ارائه کرده‌اند؛ در صورتیکه  $f$  بیانگر مدل بهینه‌ی سیستم تحت بررسی باشد، در این صورت بجای لحاظ پیش‌فرض‌هایی خاص درباره‌ی شکل پارامتریک  $f$ ، فرض می‌شود که  $f$  متعلق به دسته‌ی عمومی از توابع بوده که بر روی فضای ورودی‌ها بطور یکنواخت و پیوسته باشد. در راستای هدف اصلی از مدل‌سازی، مدل به‌صورت الگوریتمی فرض می‌شود که با استفاده از ساختار داده‌ای که از مجموعه رابطه‌ی (۱) حاصل آمده و می‌تواند جهت پیش‌بینی خروجی  $Y$  متناظر با بردار  $X$  که قبلاً عنوان شد، بکار رود. بسط رابطه‌ی مذکور در قالب تحلیل غیرخطی داده‌ها، یک الگوریتم ناپارامتریک در برآورد خطا را تشکیل می‌دهد که به آزمون گاما معروف است (استفانسون و همکاران، ۱۹۹۷). اثبات ریاضیاتی این الگوریتم توسط ایوانز و همکاران، ۲۰۰۲ و ایوانز و جونز<sup>۳</sup>، ۲۰۰۲ ارائه شده است.

### ۳.۲. توصیف آزمون گاما

آزمون گاما یک ابزار مدل‌سازی غیرخطی است که به کمک آن می‌توان ترکیب مناسب از پارامترهای ورودی برای مدل‌سازی داده‌های خروجی و ایجاد یک مدل هموار را بررسی نمود. همچنین یک ابزار توسعه‌یافته برای تخمین میانگین مربعات خطای حاصل از مدل‌سازی پدیده‌های مختلف با استفاده از مجموعه داده‌های مشاهده‌ای از آن پدیده است. به کمک بررسی این آزمون جهت انواع مختلف انتخاب‌های صورت گرفته از ورودیها، می‌توان به نمایش چگونگی انتخاب بهترین گزینه از ورودیها برای مدل‌سازی تابع هدف خروجی موردنظر پرداخت. نرم‌افزار WinGamma با هدف انجام فرآیند فوق طراحی شده که در انجام این تحقیق نیز از آن استفاده شده است (جونز، ۲۰۰۱)؛ ذیلاً به برخی از مفاهیم و اصطلاحات جهت فهم هرچه بیشتر این آزمون اشاره می‌گردد:

**مدل:** در این فرآیند، مدل داده‌ای همواری مفروض است که تابعی دیفرانسیل پذیر از ورودی‌های  $X = (x_1, \dots, x_m)$  به ازای هر خروجی است. فرض می‌شود که داده‌ها را بتوان به واسطه‌ی یک مدل

1. Bates and Watts
2. Pi & Peterson
3. Evans & Jones

ناشناخته از  $f$  نشان داد، بطوری که:

$$y = f(x_1 \dots x_m) + r \quad (۲)$$

که،  $f$  نشان‌دهنده‌ی تابع همواری است که برای مدل‌سازی داده‌ها استفاده می‌شود و  $r$  نشان‌دهنده‌ی متغیر تصادفی است که برای نمایش خطا بکار می‌رود با فرض آنکه میانگین توزیعی که  $r$  از آن پیروی می‌کند برابر با صفر و واریانس خطا کراندار باشد. آزمون گاما نشان‌دهنده‌ی آن بخش از واریانس داده‌های خروجی است که نمی‌تواند توسط مدل هموار محاسبه گردد.

**آماره گاما:** آزمون گاما به ارائه‌ی تخمینی از  $\text{Var}(r)$  می‌پردازد که اغلب به آن "مقدار گاما" نیز می‌گویند که بصورت عرض از مبدا عمودی خط رگرسیونی برازش شده بوده و بیانگر بهترین تخمین از واریانس  $r$  می‌باشد. بعلاوه، این تخمین را می‌توان برحسب  $O(M \log M)$  بار نیز آن را بدست آورد که برطبق آنچه که ذکر شد ثابت ضمنی بوده و به بعد  $m$  از فضای مجموعه داده‌های ورودی وابسته است. آماره‌ی گاما با  $\Gamma$  نشان داده می‌شود.

**آزمون گاما:** الگوریتمی جهت محاسبه‌ی واریانس خطای  $\text{Var}(r)$  مرتبط با یک خروجی خاص بوده و به نمایش چگونگی تغییر تخمین آماره‌ی گاما در حین استفاده از داده‌های بیشتر جهت تخمین آن می‌پردازد. سرانجام، اگر از تعداد داده‌ی کافی استفاده شود، آماره‌ی گاما به سمت واریانس واقعی خطا جهت خروجی‌ای که این آماره برای آن محاسبه شده، مجانب می‌گردد. آزمون گاما همچنین به محاسبه‌ی فواصل میانگین مربعات  $p$  امین و نزدیکترین همسایگی ( $1 \leq k \leq \gamma$ ) و  $\gamma(p)$  مربوطه می‌پردازد. اگرچه آزمون گاما تابعی ناشناخته از  $f$  بوده، اما می‌تواند  $\text{Var}(r)$  را بطور مستقیم از داده‌ها برآورد نماید. لذا، با فرض وجود نمونه‌ای از داده‌ها به صورت  $(x(i), y(i))$  که  $x(i) = (x_1(i), \dots, x_m(i))$  و  $1 \leq i \leq M$  در صورتی که  $N[i, p]$  نیز شاخصی از  $p$  امین نزدیکترین همسایگی به  $x_i$  و  $p$  نیز شماره‌ی همسایگی نزدیک باشد، لذا:

$$\delta_M(p) = \frac{1}{M} \sum_{i=1}^M |x_{N[i,p]} - x_i|^2 \quad (۳)$$

که،  $||$  بیانگر فاصله‌ی اقلیدسی و  $\gamma(p)$  نیز برابر است با:

$$\gamma_M(p) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i,p]} - y_i)^2 \quad (1 \leq p \leq p_{\max}) \quad (۴)$$

سپس با ایجاد رابطه‌ی رگرسیون خطی بین  $p$  مجموعه نقاط  $(\delta_M(p), \gamma_M(p)) (1 \leq p \leq p_{\max})$  معادله خط را به صورت زیر بدست می‌آید:

$$\gamma = A\delta + \Gamma \quad (5)$$

عرض از مبدا عمودی خط رگرسیونی  $(\delta(p), \gamma(p))$  به "آماره‌ی گاما" یا  $\bar{\Gamma}$  معروف است.  $\bar{\Gamma}$  عبارت از حد  $\gamma$  است وقتی  $\delta \rightarrow 0$ ، که به لحاظ تئوری معادل  $\text{Var}(r)$  می‌باشد. همچنین گرادینت (A) شاخصی از پیچیدگی مدل بوده به این صورت که مقدار بیشتر گرادینت بیانگر پیچیدگی بیشتر مدل می‌باشد (جونز، ۲۰۰۱). باید توجه داشت که آزمون گاما، ابتدا با استفاده از بردار مقادیر ورودی  $x_i (1 \leq i \leq M)$  و ایجاد یک kd-tree به تخمین  $\text{Var}(r)$  برحسب  $O(M \log M)$  بار پرداخته و سپس با استفاده از kd-tree حاصله به ایجاد مجموعه‌ای از  $k$  امین  $(1 \leq k \leq p)$  نزدیکترین همسایگی  $X_{N[i,k]} (1 \leq i \leq M)$  از  $X_i$  می‌پردازد. نمایش گرافیکی خط رگرسیونی موجود در رابطه ۵، اطلاعات مفیدی را بدست می‌دهد. اول اینکه، عرض از مبدا عمودی  $\Gamma$  از محور  $\gamma$  یا همان گاما، به ارائه‌ی تخمینی از بهینه‌ترین مقدار میانگین مجذور خطا با استفاده از یک تکنیک مدل‌سازی جهت توابع هموار ناشناخته از متغیرهای پیوسته می‌پردازد (ایوانز و جونز، ۲۰۰۲). دوم، گرادینت، شاخصی از پیچیدگی مدل است هرچه مقدار گرادینت بیشتر باشد، مدل نیز دارای پیچیدگی بیشتری است.

**همسایگی نزدیک<sup>۱</sup>:** بیانگر شاخصی از  $k$  امین نزدیکترین همسایگی است که متناسب با آزمون گاما می‌باشد. هنگامی که به تخمین آماره‌ی گاما اقدام می‌شود،  $p_{\max}$  باید متناسب با حجم مجموعه داده‌ها انتخاب شود. با توجه به تحقیقات انجام گرفته به کمک آزمون گاما، تعداد همسایگی‌های نزدیک بهتر است کوچکتر از ۳۰ لحاظ گردد. معمولاً انتخاب از ۱۰ تا ۲۰، انتخاب مناسبی است که خیلی از محققین نیز آن را پیشنهاد می‌کنند.

**تعبیه‌سازی<sup>۲</sup>:** انتخاب مقادیر گذشته‌ی یک سری زمانی مورد استفاده جهت پیش‌بینی مقادیر جاری است.

**V-ratio:** یکی دیگر از معیارهای مهمی که می‌توان با استفاده از این آزمون بدست آورد V-ratio است. این معیار بدون بعد بوده و دارای مقادیری بین بازه‌ی ۰ تا ۱ است. هرچه این معیار به صفر نزدیکتر باشد نمایانگر دقت

1. Near Neighbour

2. Embedding

بالای مدل برای یافتن خروجی‌های مطلوب از ورودی‌هاست. در واقع اگر مقدار V-ratio از یک کم شود، مقدار ضریب تعیین را نشان می‌دهد. بنابراین:

$$V_{ratio} = \frac{\Gamma}{\sigma^2(y)} \quad (6)$$

که،  $\sigma^2(y)$  واریانس خروجی  $y$  است.

**M-تست:** روشی جهت تعیین این امر است که آیا آماره‌ی گاما مقدار  $\text{Var}(r)$  را بطور معتبری برآورد می‌کند یا خیر. این معیار به کمک محاسبه‌ی آماره‌ی گاما جهت زیرمجموعه‌ای از داده‌های موجود، حاصل می‌آید. بدین وسیله در محاسبات پیاپی از آماره‌ی گاما، مقدار  $M$  را بطور تدریجی افزایش داده تا اینکه یا کلیه‌ی داده‌ها مورد استفاده قرار گیرند و یا آماره‌ی گاما به مقدار ثابتی همگرا شود. این تابع خاص، شرط باقی‌ماندن نمونه‌ی داده‌های موردنظر در فضای ورودی‌ها در یک مجموعه بسته کراندار مانند  $C$  را تضمین می‌نماید (با این شرط، نمونه‌ی موردنظر در فاصله‌ی  $[0, 1]$  مقید می‌شود).

**میانگین مربعات خطا (MSEError):** اگر  $y(i) (1, 2, \dots, M)$  مجموعه‌ای از مقادیر خروجی و  $y^*(i)$  مجموعه‌ای از پیش‌بینی‌های انجام گرفته برای  $y(i)$  باشد در این صورت میانگین مربعات خطای پیش‌بینی‌ها به صورت زیر خواهد بود:

$$\text{MSEError} = \frac{1}{M} \sum_{i=1}^M (y^*(i) - y(i))^2 \quad (7)$$

**خطای استاندارد (SE):** عبارت از خطای استاندارد خط رگرسیونی بوده و به صورت زیر محاسبه می‌شود:

$$\text{SE}(\Gamma) = \sqrt{\frac{1}{n-2} \sum_{i=1}^{p_{\max}} (\Gamma(i) - \bar{\Gamma})^2} \quad (8)$$

که،  $i$  عبارت از  $i$  امین مقدار نقطه‌ای از رگرسیون گاما و  $\bar{\Gamma}$  نیز میانگین آنهاست. اگر مقدار خطای استاندارد نزدیک به صفر باشد، با اطمینان بیشتری می‌توان پذیرفت که مقدار آماره‌ی گاما نشانگر تخمینی از واریانس خطا بر روی یک خروجی مفروض باشد.



### ۱.۳.۲. شناسایی مدل

در طی مرحله‌ی شناسایی مدل، مجموعه‌ی ورودی‌هایی که دارای حداقل مقدار آماره‌ی گاما می‌باشند برای یک خروجی خاص انتخاب می‌شوند. به این منظور باید در فرآیند مدل‌سازی به کمک نرم‌افزار، از عمل ماسک یا mask استفاده کرد به این نحو که اگر متغیر مورد نظر باید در مدل وارد شود با عدد "۱" و در غیر این صورت با "۰" نشان داده می‌شود؛ مثلاً برای سه متغیر ورودی موجود، ماسک ۱۰۱ بیان می‌کند که متغیرهای اول و سوم باید وارد مدل شده و متغیر دوم را نباید در مدل لحاظ کرد. به منظور دستیابی به بهترین شکل انتخاب ورودی‌ها، باید به کمک تعبیه‌سازی غیرمعمولی، ماسکی را برگزید که دارای حداقل مقدار  $|\bar{F}|$  باشد. متاسفانه تعداد ماسک‌هایی که از این روش بدست می‌آیند برابر با  $2^m - 1$  است (ماسک‌هایی که کلیه‌ی مقادیر آنها صفر است، لحاظ نمی‌شوند) و از اینرو انجام آزمون گاما جهت بررسی هر یک از آنها می‌تواند زمان‌بر باشد اما نرم‌افزار Win Gamma از سرعت بسیار بالایی برخوردار بوده و می‌تواند این کار را در زمان اندکی انجام دهد. کاربردی‌ترین تکنیک‌های انجام شناسایی مدل عبارت از تعبیه‌سازی کامل، ژنتیک الگوریتم، هیل کلایمینگ (Hill Climbing)، تعبیه‌سازی متوالی و تعبیه‌سازی فزاینده می‌باشند.

### ۲.۳.۲. ساخت مدل

در مطالعه‌ی حاضر، به کمک آزمون گاما به بررسی انواع مختلف ترکیب‌های حاصل از متغیرهای وقفه‌دار به عنوان ورودی‌های مدل اقدام شد تا اهمیت آنها بر پیش‌بینی قیمت گاز طبیعی مورد بررسی قرار گیرد. انواع معنی‌داری از این ترکیب‌ها وجود دارند که باید با مشاهده‌ی مقدار گامای متناظر با هر یک از آنها نسبت به انتخاب بهترین مورد اقدام کرد. از آنجائیکه آزمون گاما بر اساس ایجاد یک درخت K بعدی از داده‌های ورودی (در حال آموزش) عمل می‌کند، از این ساختار داده‌ای می‌توان جهت جستجوی سریع نزدیکترین همسایگی‌های  $p_{max}$  استفاده کرد. سپس، خط رگرسیونی ساده‌ای را باید بر اساس همسایگی‌های حاصله ایجاد کرد و از این مدل جهت پیش‌بینی نقاط مورد بررسی استفاده نمود. انتخاب مقدار  $p_{max}$  در روش رگرسیون خطی موضعی اساساً با انتخاب آماره‌ی گاما مرتبط نمی‌باشد. مقدار بهینه‌ی  $p_{max}$  در رگرسیون خطی موضعی وابسته به بعد تعبیه‌سازی و مقدار خطاست. اگر داده‌ها دارای خطا (نویزی) باشند به مقدار  $p_{max}$  بزرگتری نیاز است. در این تحقیق همچنین از تکنیک‌های رگرسیون خطی موضعی پویا و شبکه‌ها عصبی مصنوعی هم استفاده شده است.

قابل ذکر است که قبل از انجام مدل‌سازی، داده‌های موجود به دو بخش تقسیم شده‌اند؛ داده‌های آموزش (۷۰٪ داده‌ها) و داده‌های مورد آزمون (۳۰٪ داده‌ها)؛ در قسمت بعد، هر یک از این تکنیک‌ها تشریح می‌شوند. باید توجه داشت که اگرچه تکنیک گاما قادر به مدل‌سازی ANN می‌باشد اما در این پژوهش از قابلیت‌های نرم‌افزار متلب استفاده شده زیرا این نرم‌افزار از انعطاف‌پذیری بالایی به لحاظ تغییر در لایه‌ها و نرون‌ها در طی فرآیند مدل‌سازی برخوردار است.

**رگرسیون خطی موضعی (LLR):** این تکنیک یکی از روش‌های رگرسیونی ناپارامتریک بوده که بطور گسترده‌ای مورد مطالعه قرار گرفته است. این روش، عملکردهای بسیار مناسبی را در مسائل مرتبط با توابع هموار و پیش‌بینی‌های با ابعاد پایین، بدست می‌دهد. مزیت استفاده از LLR آن است که این روش نیازمند سری‌های زمانی بلندمدت جهت توسعه‌ی یک مدل پیش‌بینی نمی‌باشد؛ بموجب این روش می‌توان به ارائه‌ی موضعی یک مدل منطقی و قابل اعتماد آماری جهت مقادیر اندکی از داده‌های نمونه‌ای دست یافت. همچنین خاطر نشان می‌شود که LLR قادر به ایجاد پیش‌بینی‌های بسیار دقیق جهت داده‌های با چگالی بالا در فضای ورودی‌ها نیز می‌باشد که این خود به جذابیت استفاده از این روش می‌افزاید. این دسته از مدل‌ها را همچنین می‌توان با در نظر گرفتن داده‌های جدید مورد آموزش، پیوسته به‌روز نمود که شبکه‌های عصبی در مقایسه با آنها از این قابلیت برخوردار نمی‌باشند.

**رگرسیون خطی موضعی پویا (DLLR):** این روش اساساً مشابه با LLR بوده اما نسبت به آن از خصوصیت دیگری نیز برخوردار است؛ مادامیکه داده‌ی جدیدی برای اولین بار مشاهده می‌شود می‌توان آن را در مدل وارد نمود. تاثیر این خاصیت را می‌توان با آغاز انجام فرآیند مدل‌سازی به‌مراه داده‌های در حال آموزش اندک مشاهده نمود و سپس آزمون را بر روی تعداد زیادی از داده‌ها اجرا کرد. زمانی که داده‌ی جدیدی جهت آزمون وارد می‌شود، DLLR پیش‌بینی‌های بهتری را بدست خواهد داد. در واقع، نرم‌افزار Win Gamma همچنین دارای گزینه‌ای برای انجام این تکنیک می‌باشد. این روش نیز برای پیش‌بینی سری‌های زمانی مفید بوده که در تحقیق حاضر نیز از آن استفاده شده است (جونز، ۲۰۰۴).

**شبکه‌های عصبی مصنوعی (ANN):** مدل ANN عبارت از گروهی از نرون‌های مصنوعی بوده دارای ارتباط متقابل با هم می‌باشند و از مدل‌های ریاضی جهت پردازش اطلاعات استفاده می‌کند. به بیان تجربی،

شبکه‌های عصبی عبارت از ابزاری آماری جهت مدل‌سازی غیرخطی داده‌ها می‌باشد. بیشتر مدل‌های ANN دارای سه لایه یا بیشتر می‌باشند: یک لایه ورودی که جهت ارائه داده‌ها به شبکه بکار می‌رود؛ یک لایه خروجی که جهت تولید واکنش‌های مناسب به ورودی‌های مفروض بکار رفته و یک یا چند لایه پنهان که از توابع فعالیت مختلفی به فراخور اسلوب مسئله در نرون‌ها استفاده می‌کند. توانایی شبکه‌ی عصبی در پردازش اطلاعات برآمده از فرآیند یادگیری بوده که عبارت از تطبیق وزن‌های رابط بگونه‌ای است که بتواند خروجی مناسبی را بدست دهد (سادهییر و همکاران<sup>۱</sup>، ۲۰۰۲؛ لاسیو و همکاران<sup>۲</sup>، ۲۰۰۷؛ سایینکو و همکاران<sup>۳</sup>، ۱۹۸۹).

تحقیقات نشان می‌دهد که این دسته از مدل‌ها نسبت به اکثر مدل‌های پارامتریک غیرخطی کمتر دچار مشکل عدم تصریح می‌شوند (بارون<sup>۴</sup>، ۱۹۹۱). منابع بسیار زیادی جهت مطالعه‌ی این دسته از مدل‌ها وجود دارد (هیگین<sup>۵</sup>، ۱۹۹۹؛ آزاده و همکاران، ۲۰۱۲؛ پرز<sup>۶</sup>، ۲۰۰۶ و ویکی‌پدیا<sup>۷</sup>، ۲۰۰۸، "شبکه‌های عصبی مصنوعی"، شکل ۱).



شکل ۱. شبکه عصبی چندلایه

منبع: ویکی‌پدیا، شبکه‌ی عصبی مصنوعی، ۲۰۰۸

1. Sudheer et al.
2. Lucio et al.
3. Cybenko et al.
4. Barron
5. Haykin
6. Perez
7. Wikipedia

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

به منظور ارزیابی مدل ساخته شده جهت پیش‌بینی قیمت گاز طبیعی از معیار مجذور میانگین مربعات خطا (RMSE) جهت مراحل آموزش و آزمون و نیز خروجی‌های نرم‌افزار WinGamma استفاده شده است. اگر مقادیر شناخته شده‌ی  $y$  و  $y^*$  به ترتیب عبارت از مقادیر مشاهدات و پیش‌بینی‌ها باشند معادله‌ی RMSE برای خطای استاندارد به صورت زیر می‌باشد:

$$RMSE = \sqrt{\frac{1}{n-2} \left[ \sum (y^* - \bar{y}^*)^2 - \frac{[\sum (y - \bar{y})(y^* - \bar{y}^*)]^2}{\sum (y - \bar{y})^2} \right]} \quad (9)$$

### ۳. نتایج و بحث

#### ۱.۳. انتخاب ورودی‌های مدل با استفاده از آزمون گاما

همانطور که قبلاً نیز عنوان شد، دستیابی به متغیرهای قیمت گاز طبیعی مؤثر در دوره‌های قبل و ترکیبات مختلف آنها جهت پیش‌بینی قیمت، کاری دشوار و زمان‌بر می‌باشد. بنابراین، آزمون گاما بطور قابل ملاحظه‌ای منجر به کاهش زمان لازم برای مدل‌سازی می‌گردد. بدین منظور در قسمت‌های بعد فرآیندهای لازم جهت تعیین ورودی‌های مؤثر به مدل تشریح می‌شوند.

##### ۱.۱.۳. تعیین نزدیکترین همسایگی

پارامتر همسایگی نزدیک در آزمون گاما عبارت از عاملی قابل تعدیل بر مبنای اندازه‌گیری خطای داده‌ای و میزان حجم نمونه است. بر طبق نتایج حاصل از مطالعات محققین، پیشنهاد شده است که در آزمون گاما، تعداد همسایگی‌های نزدیک  $p_{max}$  باید برابر با ۱۰ لحاظ شود (جونز و همکاران، ۲۰۰۲؛ استفانسون و همکاران، ۱۹۹۷). این مسئله در صورتی درست است که با در نظر گرفتن انحنا تابع مورد استفاده، تعداد داده‌ی کافی موجود باشد.

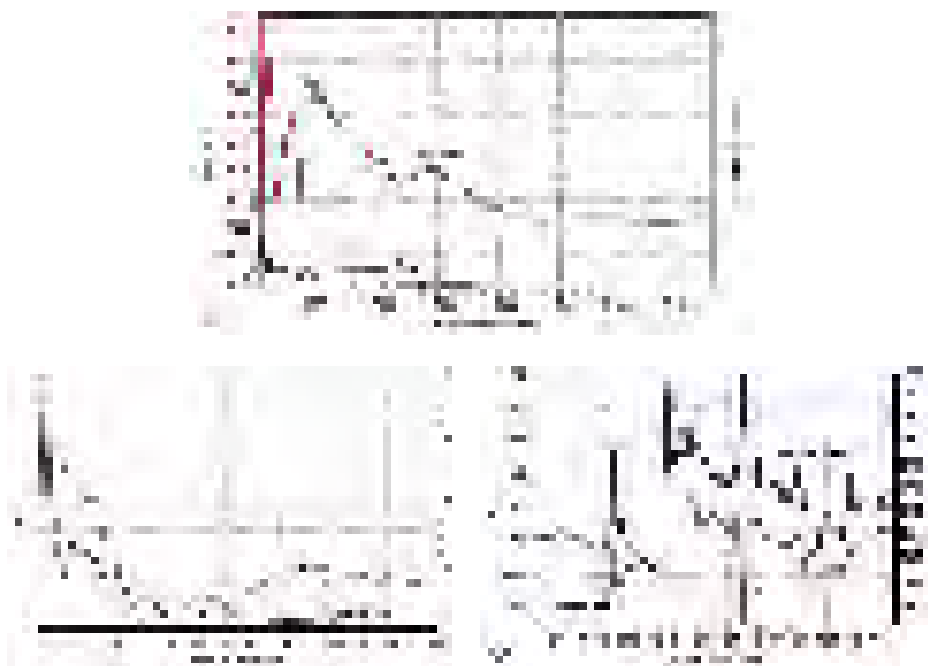
اما مسئله این است که از مناسب بودن تعداد داده‌های موجود جهت اجرای یک آزمون گاما در  $p_{max}$  برابر با ۱۰ آگاهی کافی وجود نداشته و اگر داده‌ها دارای نویز باشند، برای دستیابی به مقدار گامای قابل اطمینان به تعداد بیشتری از همسایگی‌های نزدیک نیاز می‌باشد. همچنین اگر حجم نمونه هم بالا باشد نیاز به تعداد بسیار زیادی از همسایگی‌های نزدیک خواهد بود. از آنجایی که دستیابی به متغیرهای وقفه دار مؤثر جهت پیش‌بینی قیمت گاز طبیعی، دشوار و زمان‌بر است، آزمون گاما قبل از ارائه‌ی مدل، داده‌های ورودی را مشخص می‌کند و تا حد

زیادی باعث کاهش زمان لازم جهت ساخت مدل می‌گردد. از اینرو، در وهله‌ی اول، کلیه‌ی داده‌ها وارد نرم‌افزار Win Gamma شده و سری زمانی قیمت‌های گاز طبیعی مورد بررسی قرار گرفت و تلاش شد تا بهترین تعبیه‌سازی ممکن بدست آید (به عبارتی، تعبیه‌سازی‌ای که دارای نزدیکترین مقدار  $\Gamma$  به صفر باشد). اما قبل از انتخاب بهترین تعبیه‌سازی، میزان همسایگی نزدیک تعیین شد و سپس تعداد ورودی‌ها مشخص گردید. در این پژوهش، همسایگی‌های نزدیک مختلفی مورد آزمون قرار گرفت و مقدار مناسبی از  $p_{max}$  برای مجموعه داده‌های متفاوت انتخاب شد. نتایج نشان می‌دهند که  $p_{max}$  برابر با ۱۰، ۱۵ و ۱۰ به ترتیب برای داده‌های روزانه، هفتگی و ماهانه مناسب می‌باشند؛ در این مقادیر، نمودار خطای استاندارد هموار می‌شود.

### ۲.۱.۳. تعیین تعداد ورودی مناسب جهت مدل‌سازی

برای تعیین تعداد ورودی‌های مناسب از  $M$ -تست استفاده شد و نمودار مقادیر گاما در مقابل مقادیر  $M$  ترسیم شد. با انجام  $M$ -تست به محاسبه‌ی دنباله‌ای از آماره‌های گاما  $\Gamma$  جهت تعداد فرایندهای نقاط  $M$  پرداخته می‌شود (کنکار، ۱۹۹۷).

بطور مشخص باید عنوان نمود که برای مقادیر کوچک  $M$ ، نمودار مذکور دارای نوسانات و تغییرپذیری‌هایی خواهد بود، اما زمانی که  $M$  افزایش می‌یابد، نمودار به مقدار ثابتی مجانب شده و در آن سطح تثبیت گشته که بیانگر مقدار واقعی واریانس خطا (نویز) می‌باشد. وقتی نمودار تثبیت می‌شود، دیگر نیازی به در نظر گرفتن مجموعه‌های بزرگتر  $M$  نمی‌باشد (شکل ۲). با توجه به شکل ۲ مشخص است که یک مدل تقریباً کامل را می‌توان با استفاده از مقادیر  $M$  برابر با ۲۲۰۰، ۷۳۰ و ۱۵۰ به ترتیب برای داده‌های روزانه، هفتگی و ماهانه بنا کرد زیرا مقدار واریانس گاما پس از این مقادیر در مقایسه با مقادیر واقعی، نسبتاً کوچک می‌باشد. در این مقادیر، نمودار مذکور تثبیت شده و از اینرو می‌توان اطمینان داشت که برآورد حاصله دقیق است. واضح است که اگر  $\text{Var}(r)$  بزرگ باشد در این صورت به منظور تثبیت نمودار  $M$ -تست (و آموزش مدل) به داده‌ی بیشتری نیاز خواهد بود؛ به عبارت دیگر می‌توان اظهار داشت که نرخ همگرایی متاثر از نحوه‌ی توزیع نویز یا خطاست. از اینروست که نتایج حاصل از  $M$ -تست می‌تواند در جهت مدل‌سازی مجموعه داده‌های پراکنده به کمک ANN مفید واقع شود.



شکل ۲. نتایج حاصل از M-تست جهت مجموعه داده‌های: الف) روزانه، ب) هفتگی و ج) ماهانه

منبع: نتایج نرم‌افزاری تحقیق

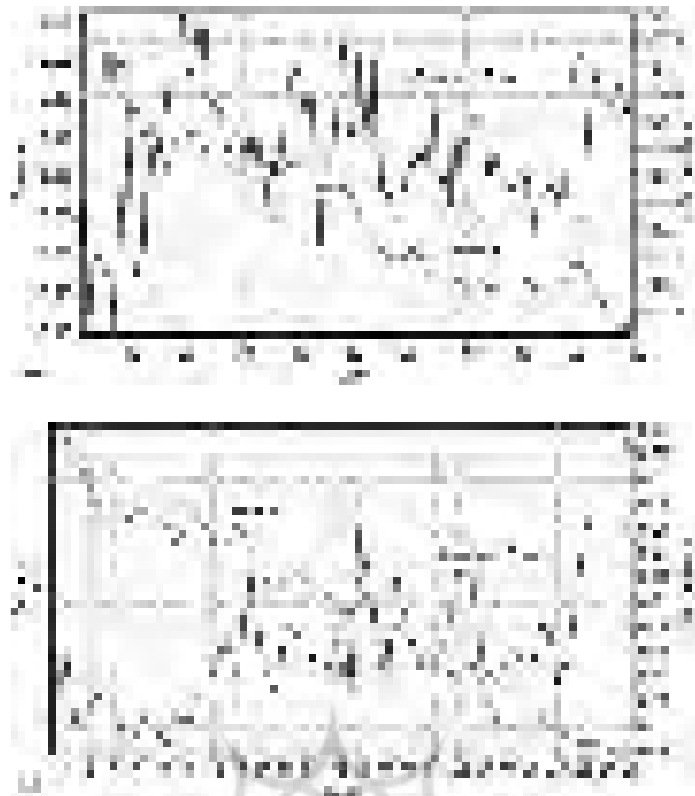
### ۳.۱.۳. انتخاب زیرمجموعه‌های مناسب از ورودی‌ها

ملاحظه شد که ترکیب خصوصیات خط رگرسیون گاما و M-تست می‌تواند برآوردی از  $\text{Var}(r)$  را به عنوان اطمینان از کیفیت خوب مدل بدست دهد. همچنین، این ترکیب، شاخصی از میزان داده‌ی مورد نیاز جهت ساخت یا آموزش مدل ایجاد شده در سطح مناسبی از میانگین مربعات خطا را نیز ایجاد می‌کند.

در بخش قبل M انتخاب شد اما آزمون گاما دارای دلالت‌های دیگری نیز می‌باشد: از آزمون گاما می‌توان جهت شناسایی مدل نیز بهره گرفت. هدف از شناسایی مدل جهت خروجی مورد انتظار، عبارت از انتخاب مجموعه‌ای از ورودی‌ها یا داده‌های وقفه‌دار مشابه بوده که بهترین مدل را برای خروجی تولید می‌کنند. حذف ورودی‌های غیرمرتبط در مدل به لحاظ عملی بسیار مهم بوده زیرا زمان لازم جهت آموزش

را برای مدل‌های شبکه‌های عصبی کاهش داده و منجر به بهبود نتایج مدل حاصله شده در حالی که ورود متغیرهای نامرتبط در فرآیند مدل‌سازی باعث افزایش خطای محاسباتی بر روی خروجی مطلوب می‌شود. از آنجائی که آزمون گاما، روشی نسبتاً سریع می‌باشد، امکان یافتن مجموعه داده‌هایی که مقدار آماره‌ی گاما را کمینه می‌سازند را فراهم می‌آورد و از اینرو "بهینه‌ترین انتخاب" از ورودی‌ها را حاصل می‌آورد. به دلیل آنکه نسبت به میزان داده‌های وقفه‌دار مورد نیاز جهت پیش‌بینی مدل قیمت‌های نقدی روزانه، هفتگی و ماهانه‌ی گاز طبیعی آگاهی کامل وجود ندارد، از اینرو در ابتدا از روش **تعییه‌سازی فراینده** استفاده شد. این روش می‌تواند در عین حال به تعیین بعد بهینه‌ی تعییه‌سازی نیز پردازد. این الگوریتم با لحاظ ماسک حاصل از اخیرترین ورودی یا سمت‌راست‌ترین ورودی در مدل‌های سری زمانی آغاز شده و مقدار گاما برای این ماسک را تولید می‌کند. این روش رفته‌رفته تعداد اجزاء موجود در ماسک را از سمت راست به چپ افزایش داده و برای هر ماسک جدیدی که به این طریق شکل می‌گیرد، آزمون گاما را انجام می‌دهد. بنابراین در این تحقیق، از الگوریتم تعییه‌سازی کامل جهت تعداد داده‌های موردنیاز روزانه، هفتگی و ماهانه که از  $M$ -تست بدست آمده، استفاده شده است. ترتیب نتایج حاصله با توجه به بهینه‌ترین مقادیر گاما (به عبارت دیگر با لحاظ نزدیکترین مقادیر  $\bar{\Gamma}$  به صفر) نشان می‌دهند که باید از ۱۸، ۵۰ و ۲۲ داده‌ی وقفه‌دار به ترتیب برای مقیاس‌های زمانی روزانه، هفتگی و ماهانه استفاده کرد (شکل ۳). این مقادیر نشان می‌دهند که برای ساخت یک مدل غیرخطی پیش‌بینی‌کننده جهت مقیاس‌های زمانی مورد نظر مناسب می‌باشند.





شکل ۳. تعداد داده‌های وقفه‌دار مناسب جهت مقیاس‌های زمانی: الف) روزانه، ب) هفتگی و ج) ماهانه

منبع: نتایج نرم‌افزاری تحقیق

از آنجائی که تعداد داده‌های وقفه‌دار مورد نیاز بسیار زیاد می‌باشند (مثلاً، ۵۰ ورودی وقفه‌دار برای مدل پیش‌بینی هفتگی)، تعداد ترکیبات ممکن نیز زیاد خواهد بود (تعداد  $1125899906842623 = 2^{50} - 1$  ترکیب)؛ بنابراین، اجرای آزمون گاما برای کلیه ترکیبات غیرعملی بوده و در بین این ترکیبات بهترین مورد را می‌توان با مشاهده مقدار گاما تعیین کرد که بیانگر معیاری از بهترین میزان MSE ممکن با استفاده از هرگونه روش مدل‌سازی جهت توابع هموار از متغیرهای پیوسته می‌باشد. لذا جهت حل این مسئله می‌توان از سه روش میان‌بر در شناسایی مدل استفاده کرد: ژنتیک الگوریتم، هیل کلایمینگ و تعبیه‌سازی فزاینده. این سه روش کمک می‌کنند تا بتوان بهترین ترکیب از داده‌های وقفه‌دار یا ورودی‌های مشابه از



مدل پیش‌بینی را انتخاب نمود. روش‌های ذکر شده به ارائه‌ی ترکیبات مختلفی از داده‌های وقفه‌دار به‌همراه کمترین مقادیر آماره‌ی گاما و کمترین سطح میانگین مربعات خطا می‌پردازند. در این مطالعه، ژنتیک الگوریتم چندین بار تکرار شده و جهت بهترین انتخاب مورد استفاده قرار گرفت. البته با تغییر ورودی‌های ملحوظ در مدل، ژنتیک الگوریتم پیوسته بطور مجدد اجرا شده است. تعبیه‌ی سازی‌های ممکنه جهت مدل مورد نظر بررسی شد. جدول ۲ به ارائه‌ی بهترین تعبیه‌ی سازی‌های بدست آمده می‌پردازد. اگر آزمون گاما بر روی کلیه‌ی ورودی‌ها در مقابل خروجی مورد نظر اجرا شود با توجه به ماسک‌های انتخابی می‌توان بهترین معیارهای عملکردی مانند مقادیر جدول ۲ را داشت.

تحلیل ژنتیک الگوریتم که در سناریوهای مختلف ایجاد شده را می‌توان در جدول ۲ ملاحظه کرد. ترکیبات مختلف تعبیه‌سازی در مقیاس‌های زمانی مختلف روزانه، هفتگی و ماهانه هیچگونه ترتیب خاصی را نشان نمی‌دهند. واضح است که در مقیاس زمانی ماهانه، ترکیبات ارائه شده بیانگر تاثیر اخیرترین ورودی‌ها بوده زیرا معنی‌دار می‌باشند. به هر حال، کلیه‌ی مدل‌ها به‌عنوان ساختارهای مناسب شناسایی شده و اولین پیش‌بینی‌ها دارای بهترین ساختار می‌باشند زیرا از سطح خطای (مقدار  $\Gamma$ ) پایینی برخوردارند. آماره‌های دیگری که در جدول ۲ ارائه شده، تصدیقی بر این مدعا می‌باشند؛ به‌عنوان مثال، مقدار پایین V-ratio (نشانگر وجود یک مدل هموار و دقیق)، برازش خط رگرسیونی با شیب ۰/۱۵۵۱، ۰/۳۲۲ و ۰/۴۷۲ به ترتیب برای مقیاس‌های زمانی روزانه، هفتگی و ماهانه (این مقادیر بقدر کافی پایین بوده و دال بر وجود یک مدل غیرخطی به‌همراه حداقل پیچیدگی می‌باشند) و خوبی برازش بدلیل مقدار پایین SE، تاییدکننده‌ی خوبی ساختارها می‌باشند.

## جدول ۲. بهترین ماسک‌های انتخابی و معیارهای ارزیابی عملکرد مربوطه جهت مقیاس‌های زمانی مختلف

پارامترها	روزانه				
	t	t+1	t+2	t+3	t+4
تعییه‌سازی	۱۱۰۱۰۱۱۱	۱۱۱۱۰۰۱۱۰۱۱	۱۱۰۱۰۱۰۱۱۱۰	۱۱۰۱۰۱۱۱۱۱۰۱۱۰۱	۰۰۱۰۱۰۱
آماره گاما	-۵۴۴/۷۰	۰/۰۰۰۱۲	۰/۰۰۰۸۱	۰/۰۰۰۳۱	۰/۰۰۰۲۵
گرادینت	۰/۱۵۵۱	۰/۱۱۲۹	۰/۱۳۵۴	۰/۱۳۲۶	۰/۲۵۰۸
خطای استاندارد	۰/۰۴۲۵	۰/۰۳۲۱	۰/۰۵۴۲	۰/۰۱۹۲	۰/۰۲۲۳
V Ratio	-۵۴۱/۵۶	-۵۴۳/۹۰	۰/۰۰۰۲۷	۰/۰۰۰۱۰	-۵۴۸/۱۷
پارامترها	هفتگی				
	t	t+1	t+2	t+3	t+4
تعییه‌سازی	۱۱۱۱۰۰۰۱۰۱۱۰۱۱۰۰	۱۱۱۰۰۰۰۰۰۰۱۱۰۱۱	۰۰۱۰۰۱۱۰۰۰۱۰	۰۰۰۰۱۰۱۱۰	۰۰۰۰۱۰۱۱۰۰۰
آماره گاما	-۵۴۱/۲۳	۰/۰۰۰۲۰	۰/۰۰۰۱۶	۰/۰۰۰۵۴	۰/۰۴۹۳۳
گرادینت	۰/۰۳۲۲	۰/۰۳۹۴	۰/۰۴۴۱	۰/۰۵۲۴	۰/۰۵۸۸
خطای استاندارد	۰/۰۴۲۷	۰/۱۱۲۱	۰/۰۶۵۷	۰/۰۹۶۸	۰/۰۸۶۳
V Ratio	-۶۴۲/۰۸۹	-۵۴۳/۴۶	-۵۴۲/۶۶	۰/۰۰۰۹۴	۰/۰۰۰۸۳۸
پارامترها	ماهانه				
	t	t+1	t+2	t+3	
تعییه‌سازی	۱۰۰۱۱۱۱۱۰۱۰۰۰۱۱۱۱	۱۱۱۱۰۰۱۰۰۱۰۱۱۱۱۰۱۰۱۱۱	۱۱۱۱۱۱	۱۱۱۱۱۱۱۰۱۰۱۱۱۱۱	
آماره گاما	-۵۴۱/۵۸	۰/۰۰۰۲۵	۰/۰۰۰۱۵	۰/۰۰۲۵۶	
گرادینت	۰/۰۴۷۲	۰/۰۶۷۵	۰/۰۸۶۱	۰/۰۶۵۱	
خطای استاندارد	۰/۱۰۲۸	۰/۱۴۴۳	۰/۱۰۰۰	۰/۱۸۳۵	
V Ratio	-۵۴۲/۸۴	-۵۴۴/۴۸	-۵۴۲/۴۵	۰/۰۰۰۴۷	

منبع: نتایج نرم‌افزاری تحقیق

### ۲.۳. شناسایی مدل

با فرض وجود داده‌ها، مدل‌های پیش‌بینی برای داده‌های مربوط به مقیاس‌های زمانی مختلف مورد نظر ساخته شدند. از آنجائیکه فرآیند شناسایی مدل بسیار گسترده و حجیم است در این بخش بطور خلاصه به نتایج آن پرداخته شده است. داده‌های وقفه‌دار در این تحقیق عبارت از متغیرهای ورودی کارآیی هستند که در بخش قبل

مورد بررسی قرار گرفتند و خروجی‌ها نیز عبارت از قیمت‌های نقدی روزانه، هفتگی و ماهانه‌ی گاز طبیعی هنری‌هاب می‌باشند که به ترتیب مدل‌های اول، دوم و سوم نامیده شده‌اند. قبلاً، انتخاب بهترین ترکیب ورودی‌ها جهت مدل‌سازی غیرخطی کاری دشوار بود. نتایج حاصل از آزمون گاما که قبلاً بطور مفصل به آن پرداخته شد، مدل‌سازان را قادر می‌سازد تا به کمک انتخابی ثابت و از پیش تعیین شده (به کمک مدل) از ورودی‌ها به مراحل کالیبراسیون و آزمون مدل پردازند. در مطالعه‌ی حاضر، سه نوع مدل ساخته شد: (۱) مدل رگرسیون خطی موضعی (LLR)، (۲) مدل رگرسیون خطی موضعی پویا (DLLR) و (۳) مدل شبکه عصبی مصنوعی (ANN)؛ نتایج صحت‌سنجی جهت کلیه‌ی مدل‌ها در جداول ۳ و ۴ ارائه گشته‌اند.

مدل ناپارامتریک حاصل از روش LLR مانند مدل شبکه‌های عصبی مصنوعی، نیازمند آموزش نمی‌باشد. اما، در این تحقیق مجموعه داده‌ها بطور تصادفی به دو دسته جهت آموزش و صحت‌سنجی مدل تقسیم شدند. عملکرد مدل‌های LLR در مقایسه با مدل‌های حاصل از تکنیک ANN قرار گرفتند. به منظور انجام این امر از آماره‌های عمومی مختلف استفاده شد تا بهترین مدل‌ها انتخاب شوند. در راستای ساخت مدل‌های LLR تعداد بهینه‌ی همسایگی‌های نزدیک به روش سعی و خطا تعیین شدند که اساساً این روش وابسته به سطح خطا یا نویز می‌باشد. روش ناپارامتریک مبتنی بر LLR نیازی به فرآیند آموزش ندارد. پس از تعیین تعداد بهینه‌ی همسایگی‌های نزدیک جهت LLR، با استفاده از آماره‌های عمومی که در جدول ۳ ارائه شده این روش در مقایسه با مدل DLLR قرار گرفت. مقایسه‌ی این دو روش حاکی از آن است که مدل DLLR ضریب همبستگی بالاتر و میانگین مربعات خطای پایین‌تری نسبت به LLR داشته و از اینرو پیش‌بینی‌های بهتری را بدست خواهد داد. مطلب قابل توجه دیگر آن است که برای دوره‌های بلندمدت‌تر مانند دوره‌های ماهانه (در مقایسه با روزانه و هفتگی)، روش DLLR دارای میانگین مربعات خطای پایین‌تر و ضریب همبستگی بالاتری از LLR نسبت به سایر مقیاس‌های زمانی می‌باشد؛ با توجه به جدول ۳ ملاحظه می‌گردد که مقادیر میانگین مربعات خطا برای دوره‌های پیش‌بینی ماهانه‌ی  $t$ ،  $t+1$ ،  $t+2$  و  $t+3$  بر طبق روش LLR به ترتیب برابر با  $۳/۸۶۵$ ،  $۱۲/۴۷۵$ ،  $۱۳/۰۵۰$  و  $۱۰/۵۷۳$  و بر طبق روش DLLR به ترتیب برابر با  $۲/۵۹۳$ ،  $۲/۴۳۷$ ،  $۱/۸۲۷$  و  $۱/۵۱۱$  می‌باشد. بعلاوه، مادامی که به سمت آینده‌ی دورتر پیش می‌رویم این مقادیر کمتر نیز می‌شوند. بنابراین، در دوره‌های زمانی بلندمدت‌تر به کمک روش DLLR نسبت به روش LLR می‌توان به پیش‌بینی‌های دقیق‌تری دست یافت.

جدول ۳. مقایسه‌ی بهترین ماسک‌های انتخابی و معیارهای عملکرد LLR و DLLR جهت مقیاس‌های زمانی مختلف

روزانه - داده‌های آموزش						
LLR						پارامترها
t+5	t+4	t+3	t+2	t+1	t	
۱۱۱۱۱۱۰۱۰۱۱۱۱	۰۰۱۰۱۰۱۱۰۱۱۱۰۱	۱۱۰۱۰۱۱۱۱۰۱۱۰	۱۱۰۱۰۱۰۱۱۱۰۱۱۱۰۱۰۱	۱۱۱۱۰۰۱۱۰۱۱۱۱۰۱۱۰۱	۱۱۰۱۰۱۱۱۱۱۱۱۰۰۰۰۱۰	تعییه‌سازی
۱۱۱۰۱	۱۱۰۰	۱۱۰۱				نزدیکترین همسایگی
۵	۵	۵	۵	۵	۵	میانگین مربعات خطا
۰/۵۱۴۱۶	۰/۷۴۱۷۲	۰/۴۳۲۰۲	۰/۴۵۰۴۲	۰/۳۶۸۸۷	۰/۲۹۱۱۳	ضریب همبستگی
۰/۹۶۰	۰/۹۴۱	۰/۹۶۵	۰/۹۶۶	۰/۹۷۵	۰/۹۷۶	
DLLR						پارامترها
t+5	t+4	t+3	t+2	t+1	t	
۱۱۱۱۱۱۰۱۰۱۱۱۱	۰۰۱۰۱۰۱۱۰۱۱۱۰۱	۱۱۰۱۰۱۱۱۱۰۱۱۰	۱۱۰۱۰۱۰۱۱۱۰۱۱۱۰۱۰۱	۱۱۱۱۰۰۱۱۰۱۱۱۱۰۱۱۰۱	۱۱۰۱۰۱۱۱۱۱۱۱۰۰۰۰۱۰	تعییه‌سازی
۱۱۱۰۱	۱۱۰۰	۱۱۰۱				نزدیکترین همسایگی
۵	۵	۵	۵	۵	۵	میانگین مربعات خطا
۰/۱۶۶۶۸	۰/۳۵۳۳۹	۰/۱۶۵۴۶	۰/۱۷۱۱۹	۰/۱۴۵۰۰	۰/۱۳۹۷۷	ضریب همبستگی
۰/۹۸۶	۰/۹۶۹	۰/۹۸۶	۰/۹۸۵	۰/۹۸۷	۰/۹۸۸	
هفتگی - داده‌های آموزش						
LLR						پارامترها
t+4	t+3	t+2	t+1	t		
۰۰۰۰۱۰۱۱۰۱۰۱۰۱۰۱۱	۰۰۰۰۱۰۱۱۰۱۰۱۰۱۰۱۱	۰۰۱۰۱۱۰۰۰۱۰۰۱۰۱۱۱۰۰۰۱۱۱۱۰۱	۱۱۱۰۰۰۰۰۰۱۱۰۱۱۰۱۱۱۱	۱۱۱۱۰۰۰۱۰۱۱۰۱۱۰۰۰۱	۱۱۱۱۰۰۰۱۰۱۱۰۱۱۰۰۰۱	تعییه‌سازی
۱۰۱۰۱۰۱۰۱۱۱۰۱۰۰۰۱	۱۰۱۰۱۰۱۰۱۱۱۰۱۰۰۰۱	۰۰۱۱۰۱۱۱۱۰۱۱۰۰۰۱۰۰۱	۰۰۱۱۱۱۰۱۱۰۱۰۱۱۱۰۱۰۰۰۱	۰۱۱۱۱۰۰۰۱۱۱۱۱۰۰۰۱	۰۱۱۱۱۰۰۰۱۱۱۱۱۰۰۰۱	نزدیکترین همسایگی
۰۱۰۰۱۱۰۰۰۰۱۱	۰۱۰۰۱۱۰۰۰۰۱۱		۰۱۱۱۰۱۱۰	۰۰۰۰۰۱۰۰۰۱۱۱۱		میانگین مربعات خطا
۷	۷	۷	۷	۷	۷	ضریب همبستگی
۶/۹۸۰۶	۷/۸۲۲۰	۳/۰۳۶۲	۳/۵۶۱۹	۳/۴۳۱۷		
۰/۷۳۰	۰/۶۷۴	۰/۷۶۸	۰/۷۴۷	۰/۶۲۷		

ادامه جدول ۳. مقایسه‌ی بهترین ماسک‌های انتخابی و معیارهای عملکرد LLR و DLLR جهت مقیاس‌های زمانی مختلف

هفتگی - داده‌های آموزش					پارامترها
DLLR					
t+4	t+3	t+2	t+1	t	
.....۱۰۱۱۰۱۰۱۰۱۰۱۱۱	.....۱۰۱۱۰۱۰۱۰۱۰۱۱	.....۱۰۱۱۰۱۰۱۱۱۰۱۱۱۱۱۱۱۱۱۱۱	۱۱۱۰.....۱۱۰۱۱۰۱۱۱۱	۱۱۱۱۰۰۰۱۰۱۱۰۱۱۰۰۰۱	تعییه‌سازی
۰۱۰۱۰۱۰۱۱۱۰۱۰۱۰۰۱	۱۰۱۰۱۰۱۰۱۱۱۱۰۱۰۰۱	.....۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱	.....۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱	.....۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱۱	تزدیکترین همسایگی
۰۱۱۰۰۰۰۱۱	۰۱۰۱۱۰۰۰۰۱۱	۷	۷	۷	میانگین مربعات خطا
۰/۴۵۵۶۶	۰/۵۴۲۸۶	۰/۵۲۱۵۲	۰/۴۸۹۳۹	۰/۲۵۵۶۶	ضریب همبستگی
۰/۹۶۸	۰/۹۶۰	۰/۹۵۸	۰/۹۶۱	۰/۹۷۸	
ماهانه - داده‌های آموزش					پارامترها
LLR					
t+3	t+2	t+1	t		
۱۱۱۰۱۱۱۱۱۰۱۰۱۱۱۱۱۱۱۱۱۱۱	۱۱۰۰۱۰۱۰۰۰۰۱۱۱۱۱۱۱۱۱۱۱	۱۱۱۱۰۰۱۰۰۱۰۱۱۱۱۱۱۰۱۰۱۱۱	۱۰۱۱۱۱۱۱۰۱۰۰۰۱۱۱۱۱۱۱۱۱۱۱		تعییه‌سازی
۷	۷	۷	۷		تزدیکترین همسایگی
۱۰/۵۷۳	۱۳/۰۵۰	۱۲/۴۷۵	۳/۸۶۴		میانگین مربعات خطا
۰/۲۱۳	۰/۱۵۷	۰/۲۸۲	۰/۶۷۲		ضریب همبستگی
DLLR					پارامترها
t+3	t+2	t+1	t		
۱۱۱۰۱۱۱۱۰۱۰۱۱۱۱۱۱۱۱۱۱۱	۱۱۰۰۱۰۱۰۰۰۰۱۱۱۱۱۱۱۱۱۱۱	۱۱۱۱۰۰۱۰۰۱۰۱۱۱۱۱۱۰۱۰۱۱۱	۱۰۱۱۱۱۱۱۰۱۰۰۰۱۱۱۱۱۱۱۱۱۱۱		تعییه‌سازی
۷	۷	۷	۷		تزدیکترین همسایگی
۱/۵۱۱۲	۱/۸۲۷۴	۲/۴۳۷۵	۲/۵۹۳۲		میانگین مربعات خطا
۰/۷۷۳	۰/۷۵۵	۰/۸۳۸	۰/۸۶۲		ضریب همبستگی

منبع: نتایج نرم‌افزاری تحقیق

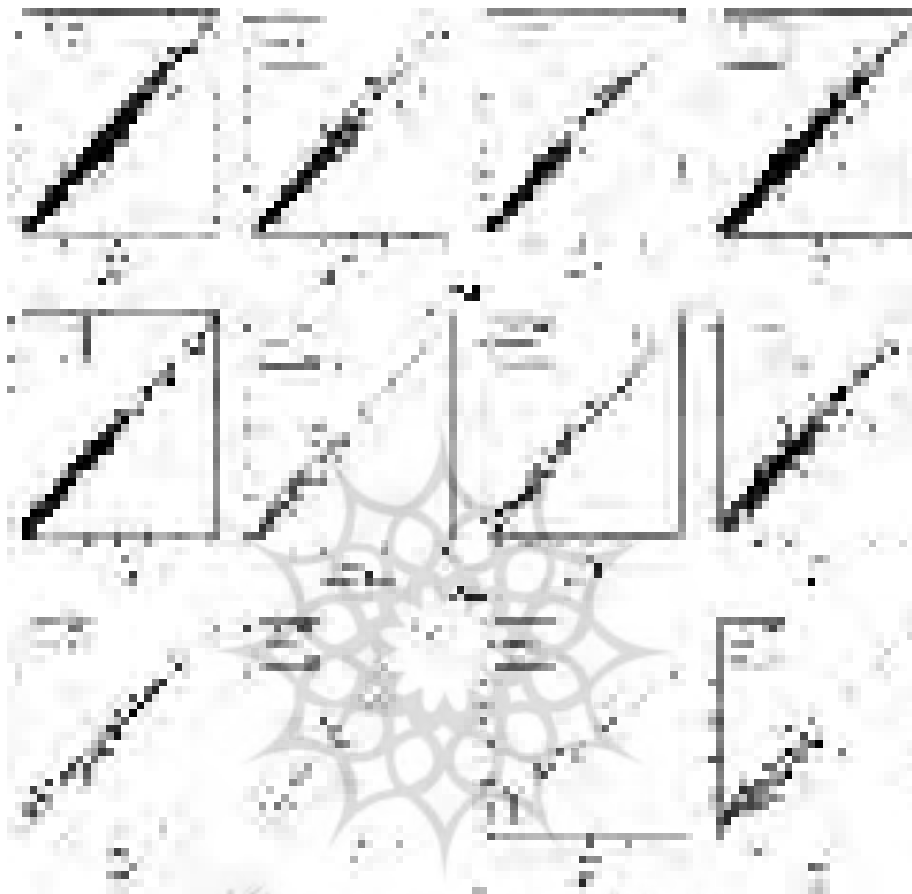
## جدول ۴. مقایسه‌ی عملکرد مدل‌های ANN جهت بهترین تعبیه‌سازی‌های انتخابی

روزانه						پارامترها
t+5	t+4	t+3	t+2	t+1	t	
۰/۲۰۳۰	۰/۱۴۳۸	۰/۲۰۲۸	۰/۱۹۲۲	۰/۰۸۷۴	۰/۱۱۵۱	آموزش میانگین
۰/۱۹۱۷	۰/۵۱۱۲	۰/۲۵۵۹	۰/۱۶۸۶	۰/۳۸۳۳	۰/۱۶۹۲	صحت‌سنجی مربعات خطا
۰/۳۳۳۲	۰/۲۲۳۰	۰/۲۹۶۴	۰/۳۳۱۵	۰/۱۴۶۵	۰/۳۳۶۶	آزمون
۰/۹۸۲۶	۰/۹۸۷۸	۰/۹۸۲۱	۰/۹۸۴۰	۰/۹۹۳۳	۰/۹۸۹۹	ضریب آموزش
۰/۹۸۳۱	۰/۹۵۳۸	۰/۹۸۰۰	۰/۹۸۴۶	۰/۹۶۲۶	۰/۹۸۵۸	صحت‌سنجی همبستگی
۰/۹۷۳۰	۰/۹۸۲۱	۰/۹۷۶۳	۰/۹۷۳۷	۰/۹۸۷۳	۰/۹۷۴۲	آزمون
هفتگی						پارامترها
t+4	t+3	t+2	t+1	t		
۰/۳۷۸۲	۰/۷۵۶۰	۰/۰۹۷۷	۰/۰۸۵۱	۰/۰۹۴۴	آموزش میانگین	
۰/۸۹۲۵	۰/۴۱۵۸	۰/۴۱۵۸	۰/۴۸۶۲	۰/۵۴۳۵	صحت‌سنجی مربعات خطا	
۱/۳۳۶۹	۰/۱۴۰۴	۰/۷۵۶۰	۱/۹۸۱۴	۰/۸۲۶۸	آزمون	
۰/۹۷۴۵	۰/۹۳۵۵	۰/۹۹۲۰	۰/۹۹۲۹	۰/۹۹۲۰	آموزش ضریب	
۰/۹۲۳۴	۰/۹۵۲۶	۰/۹۵۲۶	۰/۹۶۸۱	۰/۹۵۹۴	صحت‌سنجی همبستگی	
۰/۸۵۶۹	۰/۹۸۹۶	۰/۹۳۵۵	۰/۸۰۶۳	۰/۹۳۷۵	آزمون	
ماهانه						پارامترها
t+3	t+2	t+1	t			
۰/۵۸۴۳	۰/۳۵۲۶	۰/۶۵۶۸	۰/۴۲۳۵	آموزش میانگین		
۳/۰۲۸۲	۱/۸۱۹۰	۲/۰۳۳۲	۳/۰۵۹۴	صحت‌سنجی مربعات خطا		
۲/۶۵۸۲	۲/۸۶۰۴	۲/۸۹۳۰	۱/۲۷۷۳	آزمون		
۰/۹۵۵۱	۰/۹۶۶۸	۰/۹۳۳۹	۰/۹۸۳۱	آموزش ضریب		
۰/۸۰۱۰	۰/۸۴۶۴	۰/۸۹۰۰	۰/۸۷۹۳	صحت‌سنجی همبستگی		
۰/۷۹۷۳	۰/۸۲۹۶	۰/۸۵۸۰	۰/۸۱۸۲	آزمون		

منبع: نتایج نرم‌افزاری تحقیق

بهترین ترکیب داده‌های ورودی برای مدل‌های اول، دوم و سوم که از آزمون گاما بدست آمده‌اند جهت ارائه‌ی مدل‌های ANN مورد استفاده قرار می‌گیرند. بنابراین، مدل‌های ANN ساخته شده با نرون‌های لایه‌های مخفی مختلف، آموزش داده شده و بهترین مقادیر برای تعدادی از نرون‌های لایه‌ی مخفی انتخاب شدند. مدل ANN با جعبه‌ابزار شبکه‌های عصبی نرم‌افزار متلب کالیبره شده است زیرا این نرم‌افزار از انعطاف‌پذیری بالایی به جهت تغییر لایه‌ها و نرون‌ها برخوردار است و برای اطلاعات بیشتر در این زمینه می‌توان به راهنمای این نرم‌افزار مراجعه کرد. برطبق پیش‌فرض جعبه‌ابزار متلب، ۷۰٪ داده‌ها به آموزش و بقیه‌ی آن بطور مساوی به آزمون و صحت‌سنجی اختصاص داده شده‌اند. نتایج مراحل آموزش، صحت‌سنجی و آزمون مدل شبکه عصبی مصنوعی را می‌توان در جدول ۴ برای مدل‌های مختلف روزانه، هفتگی و ماهانه ملاحظه نمود. هرچه دوره‌ی پیش‌بینی کوتاه‌تر باشد، مدل‌های ANN از نتایج دقیق‌تری

برخوردار خواهند بود. با توجه به جدول ۴ می‌توان مشاهده کرد که ضرایب همبستگی برای زمان‌های  $t$ ،  $t+1$  و  $t+2$  جهت دوره‌های زمانی روزانه، هفتگی و ماهانه در طی مرحله‌ی آزمون، به ترتیب عبارتند از:  $t$ :  $0/9742$ ،  $0/9375$  و  $0/8182$ ،  $t+1$ :  $0/9873$ ،  $0/8063$  و  $0/8580$  و  $t+2$ :  $0/9737$ ،  $0/9355$  و  $0/8296$  و مقادیر میانگین مربعات خطا نیز به ترتیب برابر با  $t$ :  $0/3366$ ،  $0/8268$  و  $1/2773$ ،  $t+1$ :  $0/1465$ ،  $0/9814$  و  $2/8930$  می‌باشند.



شکل ۴. مقادیر مشاهده شده‌ی قیمت نقدی در مقابل مقادیر پیش‌بینی شده در مدل‌های ANN

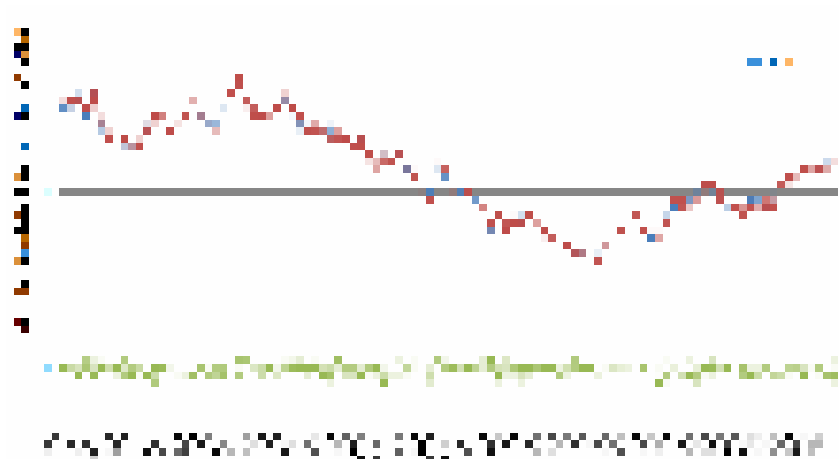
منبع: نتایج نرم‌افزاری تحقیق

پرتال جامع علوم انسانی

از اینرو، ملاحظه می‌شود که مدل پیش‌بینی قیمت‌های نقدی روزانه به کمک ANN می‌تواند تصویر روشنی را بدست دهد. همچنین، این نتایج را می‌توان با نتایج حاصل از مدل‌های LLR و DLLR نیز مقایسه کرد. واضح است که مدل ANN در مقایسه با مدل‌های LLR و DLLR از عملکرد بهتری برخوردار است. تحلیل مقایسه‌ای مدل‌های ANN و DLLR بر اساس آماره‌های محاسبه شده را نیز می‌توان انجام داد. اگرچه مقدار میانگین مربعات خطای مدل DLLR نسبت به مدل ANN کمتر می‌باشد اما انجام پیش‌بینی قیمت به کمک این مدل، غیرجذاب و دارای مقادیری از خطاست زیرا مدل‌های DLLR به محض اینکه داده‌ی مورد آموزش جدیدی وارد مدل شود به‌سادگی به‌روز و به‌نگام می‌شود درحالی‌که این مسئله برای مدل ANN اینگونه نمی‌باشد (باید توجه داشت که دوره‌های بلندمدت و طولانی جهت آموزش یا شروع مجدد مرحله‌ی آموزش ممکن است نیازمند اصلاح مدل بر مبنای داده‌های جدید باشد). بنابراین، مدل‌های DLLR تنها برای تحلیل سری زمانی مناسب است تا پیش‌بینی و مدل‌های ANN به‌عنوان بهترین مدل‌ها پیشنهاد شده‌اند و نتایج حاصل از آن هم این ادعا را تایید می‌کند.

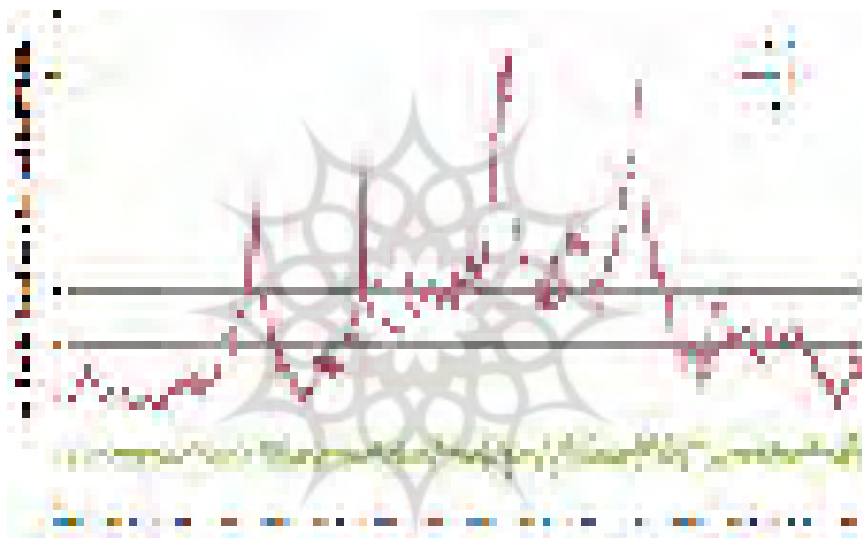
نتایج کامل‌تر مدل‌های ANN را می‌توان بطور گرافیکی در شکل ۴ مشاهده کرد. این شکل‌ها به نمایش نمودارهای پراکندگی محاسبه شده و مشاهده شده‌ی قیمت‌های نقدی روزانه، هفتگی و ماهانه‌ی گاز طبیعی در طول مراحل آموزش، صحت‌سنجی و آزمون و کل داده‌ها می‌پردازند. با تحلیل نتایج مشاهده شده مشخص می‌شود که مدل‌های ANN بدست آمده بر اساس داده‌های ورودی انتخاب شده از آزمون گاما برای پیش‌بینی کاملاً مناسب می‌باشند. اگرچه، مدل‌های ANN دورنمای واضحی را جهت پیش‌بینی روند قیمت‌های نقدی گاز طبیعی در مقیاس‌های زمانی مختلف بدست می‌دهند ولی با این حال توانمندی آنها جهت پیش‌بینی شوک‌های قیمتی بازار، قابل توجه نمی‌باشد. اما به‌هر جهت می‌توان در بخش زیر شاهد پیش‌بینی داده‌های نقدی قیمت روزانه، هفتگی و ماهانه با استفاده از مدل ANN بود که به‌عنوان بهترین مدل در مقایسه با دو دسته‌ی دیگر پیشنهاد شد. پیش‌بینی صورت گرفته برای مقیاس‌های زمانی روزانه در طول سال ۲۰۱۱ تا نوامبر ۲۰۱۲، هفتگی و ماهانه نیز برای کل دوره‌ی پیش‌بینی مشخص شده، ارائه گشته‌اند. با توجه به مقادیر خطای برآوردی واضح است که مدل ANN می‌تواند از وراثت انتخاب وقفه‌های بهینه به کمک آزمون گاما حاصل آمده، می‌تواند ابزار مناسبی برای پیش‌بینی قیمت‌ها در دوره‌های بعد تلقی گردد.





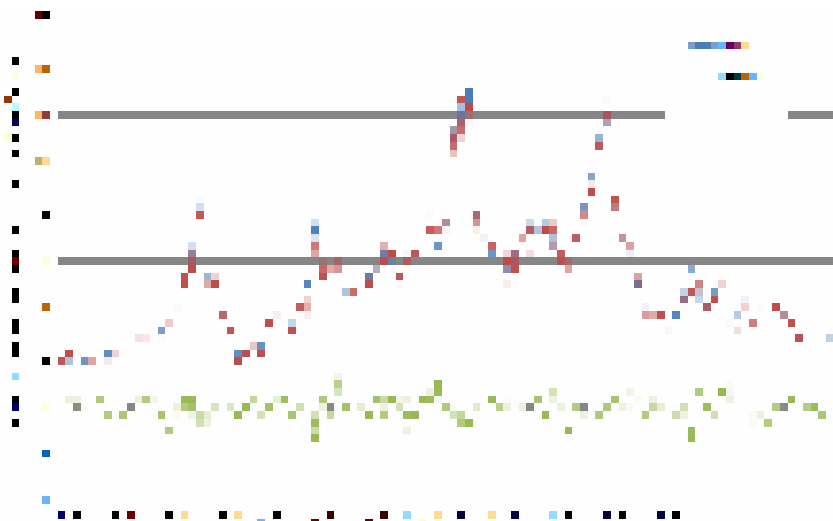
شکل ۵. مقادیر قیمت‌های نقدی روزانه پیش‌بینی شده، واقعی و میزان خطا

منبع: نتایج نرم‌افزاری تحقیق



شکل ۶. مقادیر قیمت‌های نقدی هفتگی پیش‌بینی شده، واقعی و میزان خطا

منبع: نتایج نرم‌افزاری تحقیق



شکل ۷. مقادیر قیمت‌های نقدی ماهانه پیش‌بینی شده، واقعی و میزان خطا

منبع: نتایج نرم‌افزاری تحقیق

#### ۴. نتیجه‌گیری

پیش‌بینی قیمت‌های نقدی گاز طبیعی، به‌عنوان یکی از منابع کلیدی انرژی برای اقتصاد دنیا، برای شرکت‌کنندگان بازار که به تصمیم‌گیری در خصوص تولید و بازاریابی می‌پردازند و نیز برای سیاست‌گذارانی که به تهیه و تنظیم برنامه‌های مرتبط با کالاهای انرژی و ارزیابی اثرات بازاری و قایع داخلی و بین‌المللی اشتغال دارند، بسیار مهم و ضروری به‌نظر می‌رسد. علیرغم وجود مطالعات فراوان در زمینه‌ی مدل‌سازی پیش‌بینی قیمت کالاهای انرژی که با استفاده از مدل‌های کلاسیک سری زمانی، مدل‌های مالی و تکنیک‌های غیرخطی مانند شبکه‌های عصبی مصنوعی انجام گرفته هنوز هم سؤالات زیادی وجود دارند که بی‌پاسخ مانده‌اند؛ به‌عنوان مثال، بمنظور انجام پیش‌بینی که دارای حداکثر دقت ممکنه باشد، به چه تعداد داده نیاز است؟ کدامیک از داده‌ها در فرآیند پیش‌بینی، مرتبط و کدامیک نامرتب می‌باشند؟ یکی از راه‌های ارائه‌ی پاسخ به این سؤالات، کمک گرفتن از آزمون گاما می‌باشد که در نرم‌افزار گاما ارائه شده است. بطور خلاصه، آزمون گاما، به لحاظ ریاضیاتی، ابزار مدل‌سازی غیرخطی و هموار به‌مراه انواع گسترده‌ای از کاربردهاست که به مدل‌سازان کمک می‌کند تا به انتخاب بهترین ترکیب از ورودی‌ها قبل از کالیبره‌سازی و آزمون مدل بپردازند؛ از این‌روست که این ابزار به کاهش

عدم قطعیت مرتبط با انتخاب داده‌های ورودی می‌انجامد. به‌هرحال، در این مطالعه، از یکی از توانمندی‌های قابل توجه آزمون گاما در انتخاب متغیرهای مرتبط در ساخت مدل‌های غیرخطی از قیمت‌های نقدی روزانه، هفتگی و ماهانه استفاده شده است. نتایج حاصله نشان می‌دهند که مقدار مناسب  $p_{max}$  برای مجموعه داده‌های مختلف روزانه، هفتگی و ماهانه به ترتیب برابر با ۱۰، ۱۵ و ۱۰ می‌باشد. تعداد داده‌های کافی و مورد نیاز جهت ساخت یک مدل قابل اطمینان با استفاده از  $M$ -تست برای داده‌های روزانه، هفتگی و ماهانه به ترتیب برابر با ۲۲۰۰، ۷۳۰ و ۱۵۰ می‌باشند. تعیین‌سازی فزاینده نشان داد که بمنظور پیش‌بینی قیمت‌های نقدی در مقیاس‌های زمانی روزانه، هفتگی و ماهانه به ۱۸، ۵۰ و ۲۲ داده‌ی وقفه‌دار نیاز است. ترکیبات مختلف تعیین‌سازی هیچگونه ترتیب خاصی را نشان نمی‌دهند. اما از نتایج می‌توان دریافت که برای مقیاس زمانی ماهانه می‌توان ترکیبات مشخص‌تری را بدست آورد در این مورد تاثیر ترکیب ورودی‌های اخیر معنی‌دار و واضح است. به‌هرحال، ساختار کلیه‌ی مدل‌ها مناسب و پیش‌بینی‌های اولیه‌ی دارای بهترین ساختار است زیرا دارای حداقل سطح خطا می‌باشند.

به کمک داده‌های موجود، مدل‌های پیش‌بینی  $LLR$ ،  $DLLR$  و  $ANN$  ساخته شدند. عملکرد حاصل از مدل  $LLR$  با استفاده از آماره‌های عمومی در مقایسه با عملکرد مدل  $DLLR$  قرار گرفت؛ مدل  $DLLR$  از ضریب همبستگی بالاتر و میانگین مربعات خطای پایین‌تری نسبت به  $LLR$  برخوردار است و از اینرو پیش‌بینی‌های بهتری را بدست می‌دهد. قابل توجه است که برای دوره‌ای بلندمدت‌تر مانند ماهانه (در مقایسه با روزانه و هفتگی)،  $DLLR$  دارای میانگین مربعات خطای پایین‌تر و ضریب همبستگی بالاتری در مقایسه با  $LLR$  نسبت به سایر بازه‌های زمانی می‌باشد. بنابراین، دستیابی به پیش‌بینی‌های دقیق‌تر در بازه‌های زمانی بلندمدت‌تر به کمک روش  $DLLR$  امکان‌پذیر خواهد بود. مدل  $ANN$  کالیبره شده به کمک ترکیبات مختلفی از نرون‌های موجود در لایه‌های پنهان گوناگون از طریق نرم‌افزار متلب مورد آموزش قرار گرفت و نشان داد که در طی مراحل آموزش، صحت‌سنجی و آزمون، هرچه دوره‌ی پیش‌بینی کوتاه‌تر باشد نتایج دقیق‌تری را می‌توان از مدل  $ANN$  بدست آورد. لذا، مدل پیش‌بینی قیمت‌های نقدی گاز طبیعی به کمک شبکه‌های عصبی مصنوعی، می‌تواند به‌عنوان مدلی مناسب تفسیر شود. مشاهده می‌شود که عملکرد مدل  $ANN$  در مقایسه با  $LLR$  و  $DLLR$  بهتر است. در طرف مقابل، مدل‌های  $LLR$  چنین نمی‌باشند. به‌هرحال، میزان میانگین مربعات خطا در مدل  $DLLR$  نسبت به مدل‌های  $ANN$  کمتر می‌باشد اما پیش‌بینی قیمت با این مدل ( $DLLR$ ) مناسب نیست زیرا دارای خطا می‌باشد و دلیل آن هم آن است که

مدل‌های DLLR بطور پویا خودتصحیح‌کننده می‌باشند و از اینرو پیش‌بینی دقیقی را نمی‌توان با تکیه بر آنها در مقایسه با مدل‌های ANN بدست آورد. لذا، مدل‌های DLLR تنها برای تحلیل سری‌های زمانی مناسبند تا انجام پیش‌بینی، و مدل‌های ANN بهترین گزینه‌ی پیشنهادی بوده و نتایج حاصله و نیز پیش‌بینی‌های انجام گرفته توسط آن برای دوره‌های بعدی، دلیلی بر این مدعاست. اگرچه مدل‌های ANN دورنمای واضحی را بدست داده و دارای دقت بالایی جهت پیش‌بینی روند قیمت‌های نقدی گاز طبیعی در مقیاس‌های مختلف زمانی می‌باشند اما با این حال، توانمندی آنها در پیش‌بینی شوک‌های قیمتی بازار چندان قابل ملاحظه نمی‌باشد. بنابراین، استفاده از ابزار دیگری مانند مدل‌های اقتصادسنجی مالی که مبتنی بر معادلات دیفرانسیل تصادفی می‌باشند به پیش‌بینی شوک‌های قیمتی و همچنین جهش‌های قیمتی ایجاد شده پردازند.

بطور کلی، روش شناختی توصیف شده در این مطالعه، دارای اشارات ضمنی معنی‌داری جهت سایر انواع مدل‌سازی‌های اقتصادی و مالی می‌باشد. اگر خطای ذاتی موجود در داده‌های ورودی بیش از توانمندی مدل‌سازی مدل باشد، صرف‌نظر از اینکه مدل چقدر مناسب است، ارائه و توسعه‌ی مدل کار دشواری خواهد بود. بدین منظور، آزمون گامای بکار گرفته شده در این تحقیق از توانایی زیادی برخوردار است و می‌تواند به مدل‌سازان و اقتصاددانان در حل مسائل مرتبط با عدم قطعیت در فرآیند مدل‌سازی مالی و اقتصادی کمک نماید.

## منابع و مأخذ

- ابریشمی، حمید، محسن مهرآرا و یاسمین آریانا (۱۳۸۶)، "ارزیابی عملکرد مدل‌های پیش‌بینی بی‌ثباتی قیمت نفت"، *مجله تحقیقات اقتصادی*، شماره ۷۸، بهار، صص ۲۱-۱.
- رحیمی، غلامعلی (۱۳۸۶)، "بررسی مکانیسم‌های قیمت‌گذاری گاز طبیعی در مناطق مختلف"، *فصلنامه مطالعات اقتصاد انرژی*، سال چهارم، شماره ۱۳، صص ۱۲۱-۹۶.
- مبینی‌دهکردی، علی و مریم پاشنگ (۱۳۸۸)، "شناسایی عوامل مؤثر بر عرضه و تقاضای جهانی نفت و گاز با استفاده از تکنیک AHP"، *فصلنامه مطالعات اقتصاد انرژی*، سال ششم، شماره ۲۰، صص ۱۱۸-۸۷.
- منصورکیایی، اسحاق (۱۳۸۷)، "تخمین رابطه میان قیمت نفت خام و گاز طبیعی مایع شده (LNG)"، *فصلنامه مطالعات اقتصاد انرژی*، سال پنجم، شماره ۱۸، صص ۱۲۱-۹۹.

- Albrecht, W. P. (2011). Price Transparency in the US Natural Gas Market. *Natural Gas Supply Association*.
- Al Faris, A. (1991). *The Determinants of Crude Oil Price Adjustment in the World Petroleum Market*, OPEC Rev, 15.
- Azadeh, A., S.M., Asadzadeh & A., Ghanbari.(2010). An Adaptive Network-Based Fuzzy Inference System for Short-Term Natural Gas Demand Estimation: Uncertain and Complex Environments. *Energ Policy*, 38, pp. 1529-1536.
- Azadeh, A., M., Sheikhalishahi & S., Shahmiri.(2012). A Hybrid Neuro-Fuzzy Approach for Improvement of Natural Gas Price Forecasting in Vague and Noisy Environments: Domestic and Industrial Sectors. *Intervational Conference on Trends in Industrial and Mechanical Engineering(ICTIME 2012)*, March 24-25, Dubai, pp. 123-127.
- Bacon, R. (1991). Modeling the Price of Oil. *Oxford Rev. Econ. Pol.*, 7(2), pp. 17-34.
- Barron, A. R.(1991). Universal Approximation Bonds for Superpositions of a Sigmoidal Function. Technical report No. 58, *Department of Statistics, University of Illinois, Urbana Champaign*.
- Bates, D. M. & D. G., Watts.(1988). Nonlinear regression analysis and its applications. *J. Wiley & Sons*, ISBN 0-8194-1845-5.
- Bosq, D.,(1998). *Nonparametric Statistics for Stochastic Processes*. Springer.
- Brown, S. P. A., and Mine K. Yücel.,(2008). What Drives Natural Gas Prices?, *The Energy Journal* , Vol. 29, No. 2, pp. 193-208.
- Buchanan, W. K., P., Hodges & J., Theis.(2001). Which Way the Natural Gas Price: An Attempt to Predict the Direction of Natural Gas Spot Price Movements Using Trader Positions. *J. Energ Econ*, vol. 23, no 3, pp. 279-293.
- Cybenko, G.,(1989). *Approximations by Superpositions of a Sigmoidal Function*. *Math Control Signals Syst.*, 2: pp. 303-314.
- Dudek, D.J., A.A. Golub & E.B., Strukova.(2006). Should Russia Increase Domestic Prices for Natural Gas? *Energ Policy*, Vol. 34 No. 13, pp 1659-1670.
- Energy Information Administration(EIA).(2002). Derivatives and Risk Management in the Petroleum, *Natural Gas, and Electricity Industries*.
- Energy Information Administration(EIA).(2011). *Shale Gas is a Global Phenomenon*.
- Evans, D.,(2002). Data Derived Estimates of Noise Using Near Neighbour Asymptotics. Ph.D. Thesis, Department of Computer Science, *Cardiff University*, Wales, UK.
- Evans, D. & A. J., Jones.(2002). A Proof of the Gamma Test. *Proc. Roy. Soc. Lond. Series A 458(2027)*, pp 2759-2799.

Evans, D., A. J., Jones & W. M., Schimidt.(2002). A Symptotic Moments of near neighbor distance distributions. *Proc. Roy. Soc. Lond. Series A 458(2028)*, pp. 2839-2849.

Fleten, S.E. & J., Lemming.(2003). Constructing Forward Price Curves in Electricity Markets. *Energ Econ*, 25 pp 409-424.

Guedes de Oliveira, A.,(1999). Synchronization of Chaos and Applications to Secure Communications. PhD thesis, Department of Computing, *Imperial College of Science, Technology and Medicine, University of London, UK*.

Gyorfi, L., W., Hardle, P., Sarda & P., Vieu.(1989). *Nonparametric Curve Estimation from Time Series. Springer*.

Hardle, W. & P., Vieu.(1992). Kernel Regression Smoothing of Time Series. *J Time Ser Anal.*, 13 pp 209-232.

Hardle, W., H., Lutkepohl & R., Chen.(1997). A Review of Nonparametric Time Series Analysis. *Int Stat Rev.*, 65, pp 49-72.

Hardle, W., A., Tsybakov & L., Yang.(1998). Nonparametric Vector Autoregression. *J Stat Plan Infer*, 68, pp 22-245.

Hart, J. D.,(1991). Kernel Regression Estimation with Time Series Errors. *J Roy Stat Soc, B*, 53, 1, pp 173-187.

Hart, J.D.,(1996). Some Automated Methods of Smoothing Time-dependent Data. *J Nonparametr Stat*, 6, pp. 115-142.

Haykin S.,(1999). *Neural networks – a comprehensive foundation*. 2nd ed. *Prentice Hall*.

Henning, B., Sloane, M., and DeLeon, M.(2003). *Natural Gas and Energy Price Volatility. American Gas Foundation*.

Huntington, H.G.,(1994). Oil Price Forecasting in the 1980s: What Went Wrong? *Energy J.*, 11(2) , pp 1-22.

Jones, A.J.,(2001). *The WinGamma User Guide*. Copyright: *University of Wales, Cardiff*, 77pp.

Jones, A.J.,(2004). New Tools in Nonlinear Modelling and Prediction. *Computational Management Science(CMS)*, pp. 109-149.

Jones, A.J., Tsui, A., de Oliveira, A.G.,(2002). Neural Models of Arbitrary Chaotic Systems: Construction and the Role of Time Delayed Feedback in Control and Synchronization. *Complexity int.*, 9.

Judd, K., Mees, A.,(1998). *Embedding As a Modeling Problem. Physica D*, 120: pp. 273-286.

Kemp, S. E.,(2006). Gamma test analysis tools for non-linear time series. Doctor of Philosophy thesis, Department of Computing & Mathematical Sciences, Faculty of Advanced Technology, *University of Glamorgan, Wales, UK*.

Koncar N.,(1997). Optimization Methodologies for Direct Inverse Neurocontrol. PhD thesis, Department of Computing, *Imperial College of Science Technology and Medicine, University of London, UK.*

Kumbarouglu, G., Madlener, R.,(2003). Energy and Climate Policy Analysis with the Hybrid Bottom-up Computable General Equilibrium Model SCREEN: the Case of the Swiss CO2 Act. *J. Energy Econ*, 25, pp. 409–424.

Lucio, P. S., Conde, F. C., Cavalcanti, I. F. A., Serrano, A. I., Ramos, A. M., Cardoso, A. O.,(2007). Spatiotemporal Monthly Rainfall Reconstruction via Artificial Neural Network(Case Study: South Brazil). *Journal of Advances in Geosciences(ADGEO)*, 10, pp. 67–76.

MacAvoy, P.W., Moshkin, N.V.,(2000). The New Trend in the Long Term Price of Natural Gas. *Resour Energy Econ*, Vol. 22, pp. 315-338.

Malliarisa, M. E., Malliarisb, S. G.,(2008). Forecasting Inter-related Energy Product Prices. *Europ. J. Finance*, Vol. 14 No. 6, September 2008, pp 453–468.

Martinsen, D., Linsen, J., Markewits, P., Voge, S.,(2003). CCS: A Future CO2 Mitigation Option for Germany?-A Bottom-up Approach. *Energy Econ.*, 25, pp 409–424.

Masry, E., Tjostheim, D.,(1995). Nonparametric Estimation and Identification of Nonlinear ARCH Time Series. *Economet Theor.*, 11, pp. 258-289.

Mirmirani, S., Li, H.C.,(2004). A Comparison of VAR and Neural Networks with Genetic Algorithm in Forecasting Price of Oil. *Applications of Artificial Intelligence in Finance and Economics: Adv E.*, 19 pp 203–223.

Modjtahedi, B., Movassagh, N.,(2005). Natural gas -Futures: Bias, Predictive Performances, and the Theory of Storage. *Energy Econ.*, 27: pp. 617-637.

Morana, C.,(2001). A Semi-Parametric Approach to Short-term Oil Price Forecasting. *Energy Econ.*, 23(3), pp. 325–338.

Moshiri, S.,(2004). Testing for Deterministic Chaos in Futures Crude Oil Price: Does Neural Network Lead to Better Forecast? *Economics, Working Paper, vol. 5.*

Movassagh, N., Modjtahedi, B.,(2005). Bias in Backwardation in Natural Gas Futures Prices. , *J. Futures Markets*, 25: pp. 281-308.

Perez, M.,(2006). Artificial Neural Networks and Bankruptcy Forecasting: *A State of the Art. Neural Comput Appl.*, 15: pp. 154–63.

Pi, H., Peterson, C.,(1994). Finding the embedding dimension and variable dependencies in time series. *Neural Comput.*, 6(3): pp 509–520.

Remesan, R., Shamim, M.A., Han, D.,(2008). Model Data Selection Using Gamma Test for Daily Solar Radiation Estimation. *Hydrol Process*, 22(21), pp. pp 4301–4309.

Sadorsky, P.,(2002). Time-varying Risk Premiums in Petroleum Futures Prices. *Energy Econ.*, 24(6), pp. 539–556.

Serletis, A., Rangel-Ruiz, R.,(2004). Testing for Common Features in North American Energy Markets. *Energy Econ.*, 26: pp 401-414.

Shang, H. L., Hyndman, J.,(2009). Nonparametric Time Series Forecasting with Dynamic Updating. 18th World IMAS/MODSIM Congress, Cairns, Australia 13-17 July, <http://mssanz.org.au/modsim09>.

Stefánsson, A., Koncar, N., Jones, A. J.,(1997). A Note on the Gamma Test. *Neural Comput Appl.*, 5, pp. 131–133.

Studebaker, P.(2011). Natural gas prices could be artificially low. *NewYork Times blog*.

Sudheer K. P., Gosain, A. K., Rangan, D. M., Saheb, S. M.,(2002). Modelling Evaporation Using an Artificial Neural Network Algorithm. *Hydrol Process*, pp. 189–202.

Takens, F.,(1981). Detecting Strange Attractors in Turbulence. In: Rand DA, Young LS, Eds., *Dynamical Systems and Turbulence: Lect Notes Math.*, 888, Springer-Verlag, pp 366–381.

Tsui, A. P. M.,(1999). Smooth Data Modeling and Stimulus Response via Stabilization of Neural Chaos. PhD thesis, *Department of Computing, Imperial College of Science Technology and Medicine, University of London, UK*.

Tsui, A. P. M., Jones, A. J., Guedes de Oliveira, A.,(2002). The Construction of Smooth Models using Irregular Embeddings Determined by a Gamma Test Analysis. *Neural Comput Appl.*, 10: pp 318–329.

Walls, W. D.,(1995). An Econometric Analysis of the Market for Natural Gas Futures *Energy J.*, 16: pp 71-83.

Wayne, L. J.(2010). Cool the Planet with Natural Gas. *The Wall Street Journal*.

Woo, C.K., Olson, A., Horowitz, I.,(2005). Market Efficiency, Cross Hedging and Price Forecasts: California's Natural-Gas Markets. *Energy J.*, Vol. 31, No. 8-9, pp. 1290-1304.

Xie, W., Yu, L., Xu, S.Y., Wang, S.Y.,(2006). A New Method for Crude Oil Price Forecasting Based on Support Vector Machines. *Lect Notes Comput Sc.* 3994, pp. 441–451.

Yang, C.W., Hwang, M. J., Huang, B.N.,(2002). An Analysis of Factors Affecting Price Volatility of the US Oil Market. *Energ Econ.*, 24, pp. 107–119.

Yu, L., Wang, S.Y., Lai, K.K.,(2006). Forecasting Foreign Exchange Rates and International Crude Oil Price Volatility. TEI@I Methodology. *Hunan University Press, Changsha*.

Zamani, M.,(2004). An Econometrics Forecasting Model of Short Term Oil Spot Price. *IIES Energy Economist*, 6<sup>th</sup> IAEE European Conference.