

الگوسازی غیرخطی و پیش‌بینی درآمدهای مالیات بر مشاغل در اقتصاد ایران (کاربرد شبکه‌های عصبی مصنوعی و مقایسه آن با الگوهای رگرسیون خطی و سری زمانی)

محمد علی فلاحي

استادیار دانشکده علوم اداری و اقتصادی، گروه اقتصاد، دانشگاه فردوسی مشهد.

Email: falahi@ferdowsi.um.ac.ir

حمید خالوزاده

استادیار دانشکده مهندسی برق، گروه کنترل، دانشگاه صنعتی خواجه نصیرالدین طوسی

Email: h_khaloozadeh@kntu.ac.ir

سعیده حمیدی علمداری

کارشناس ارشد اقتصاد Email: Hamidi_Saeedeh@yahoo.com

تاریخ دریافت: ۸۴/۸/۲۳ تاریخ تصویب: ۸۴/۱۱/۴

چکیده

در این مقاله، با هدف دستیابی به پیش‌بینی‌های دقیق‌تر، سه نوع الگوی رگرسیون خطی، سری‌زمانی و شبکه عصبی مصنوعی طراحی و برآورد شده است. در ابتدا ماهیت ساختاری سری‌زمانی مورد نظر از جهت خطی، غیرخطی و تصادفی بودن با استفاده از آزمون نمای لیپانوف بررسی شده است. نتایج این آزمون، وجود آشوب ضعیفی را در سیستم نشان می‌دهد و بیانگر امکان استفاده از الگوسازی غیرخطی برای پیش‌بینی دقیق‌تر کوتاه مدت است. در مرحله بعد با استفاده از رگرسیون خطی، الگوی سری زمانی و شبکه‌های عصبی مصنوعی، ضمن انجام پیش‌بینی درآمدهای مالیات بر مشاغل طی دوره زمانی ۱۳۷۹-۱۳۸۳، عملکرد این سه روش با یکدیگر مقایسه شده است. نتایج برآورد الگوها نشان دهنده عملکرد بهتر الگوی شبکه عصبی مصنوعی می‌باشد.

طبقه‌بندی JEL: E62; C53; C20; C45

: مالیات بر مشاغل، پیش‌بینی، آشوب، نمای لیپانوف، رگرسیون خطی، الگوی

سری‌زمانی، شبکه عصبی مصنوعی، ایران.

۱- مقدمه

اهمیت مالیات به‌طور عام و مالیات بر مشاغل به‌طور خاص در تأمین بودجه و نقش ابزاری آن در جهت‌دهی به اهداف کلان اقتصادی موجب می‌گردد تا انجام پیش‌بینی‌های دقیق‌تر بتواند دولت را در تدوین برنامه‌ها و مآل‌رسیدن به اهداف خود بهتر یاری رساند. بخش مشاغل، یکی از بخش‌های تأمین‌کننده مالیات می‌باشد و می‌تواند سهم زیادی در تأمین درآمدهای مالیاتی دولت داشته باشد. پیش‌بینی میزان وصولی درآمدهای مالیاتی از این بخش با توجه به نظام مالیاتی موجود و رویه‌های اجرایی و بدون در نظر گرفتن ظرفیت بالقوه بخش مشاغل، نقش بسیار مهمی در برنامه‌ریزی‌های اقتصادی دولت دارد. پیش‌بینی‌های دقیق از درآمدهای مالیات بر مشاغل به دولت کمک می‌کند تا برنامه‌ریزی‌های مطمئن تری انجام داده و میزان مشارکت این بخش در تأمین هزینه‌های عمومی دولت به‌طور دقیق تری مشخص شود. در مقاله حاضر برای انجام پیش‌بینی درآمدهای مالیات بر مشاغل از الگوی رگرسیون خطی^۱، الگوی سری زمانی باکس-جنکینز^۲ (ARIMA) و شبکه عصبی مصنوعی^۳ (الگوی غیرخطی) استفاده شده است. دوره برآورد در هر سه الگو سال‌های ۱۳۴۷-۱۳۷۸ و دوره پیش‌بینی سال‌های ۱۳۷۹-۱۳۸۳ می‌باشد. پس از مقدمه پیشینه تحقیق ارائه و به‌دنبال آن متغیرهای تحقیق معرفی شده است. برای پرهیز از پدیده رگرسیون کاذب، مانایی این متغیرها با استفاده از آزمون‌های دیکی-فولر و پرون بررسی گردیده است. همچنین برای بررسی امکان پیش‌بینی درآمدهای مالیات بر مشاغل، آزمون‌های آشوب با استفاده از آزمون نمای لیپانوف به‌عمل آمده است. در پایان ضمن بهره‌گیری از الگوهای سه‌گانه فوق جهت پیش‌بینی، کارایی آن‌ها با استفاده از معیارهای مربوطه مقایسه شده است.

۲- مروری بر مطالعات انجام شده در ایران

در مطالعات انجام شده توسط شهشهانی و داوولینگ (۱۳۵۶)، شهشهانی (۱۳۵۷)، محمدی (۱۳۶۳)، عیسی‌زاده روشن (۱۳۷۶)، فهیم یحیایی (۱۳۷۰) و ملایی پور و

1- Linear Regression.

2- Box- Jenkins.

3- Artificial Neural Network.

فتحانی (۱۳۸۱) از روش‌های خطی برای پیش‌بینی درآمدهای مالیاتی استفاده شده است. محدودیت اطلاعات در بخش مشاغل موجب شده است تا مطالعات انجام شده در این بخش بسیار اندک باشد. روش الگوسازی و پیش‌بینی درآمدهای مالیاتی در این مطالعات مبتنی بر الگوهای متداول رگرسیون خطی است.

۳- معرفی متغیرها و بررسی مانایی آن‌ها

متغیرهای مورد استفاده در مقاله حاضر عبارتند از:

LTJ: لگاریتم مالیات بر مشاغل

LR: لگاریتم ارزش افزوده گروه بازرگانی، رستوران و هتلداری

با توجه به این که متغیرهای تحقیق به صورت اسمی با مشکل درجه جمعی^۱ بیشتر از ۱ روبرو می‌باشند، لذا برای پرهیز از مشکلات و محدودیت‌های آن و نیز تفسیر روشن تر نتایج از لگاریتم طبیعی متغیرها استفاده شده است.^۲

نتایج آزمون مانایی^۳ (آزمون ریشه واحد دیکی- فولر^۴) سری‌های زمانی لگاریتم درآمدهای مالیات بر مشاغل و لگاریتم ارزش افزوده گروه بازرگانی، رستوران و هتلداری طی دوره زمانی ۱۳۴۷-۱۳۷۸ (دوره براورد) در جدول ۱ نشان داده شده است. نتایج

جدول ۱- آزمون ریشه واحد متغیرهای تحقیق طی دوره ۱۳۴۷-۱۳۷۸

| با روند زمانی | بدون روند زمانی | با روند زمانی | بدون روند زمانی | |
|---------------|-----------------|---------------|-----------------|--------------|
| -۴/۸۵۶۴ | -۳/۸۷۹۰ | -۱/۳۵۳۷ | ۱/۴۳۱۵ | LTJ |
| -۴/۱۶۷۹ | -۴/۰۷۱۸ | -۲/۰۳۶۲ | ۰/۷۲۵۱۵ | LR |
| -۳/۶۱۱۹ | -۲/۹۹۰۷ | -۳/۶۰۲۷ | -۲/۹۸۵۰ | مقدار بحرانی |

منبع: محاسبات تحقیق

1- Integration.

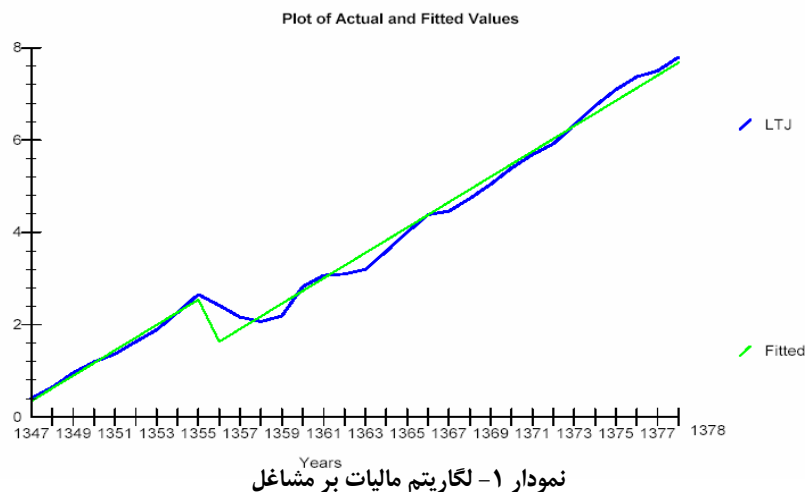
۲- جهت آشنایی بیشتر با تجزیه و تحلیل اقتصادسنجی متغیرهای (I(۲) به Neils Haldrup (1998) مراجعه شود.

3- Stationarity.

4- Dickey-Fuller Unit Root Test.

نشان‌دهنده نامانایی این دو سری زمانی می‌باشد که با یک بار تفاضل‌گیری مانا می‌گردند. لازم به ذکر می‌باشد که مقدار وقفه بهینه با استفاده از معیار شوارتز-بیزین^۱، انتخاب شده است.^۲

پرون^۳ (۱۹۸۹) خاطر نشان می‌سازد در شرایطی که انتظار شکست ساختاری وجود داشته باشد، لازم است مراقبت ویژه‌ای در زمان استفاده از آزمون ریشه واحد اعمال کرد چرا که آزمون دیکی-فولر در جهت رد نکردن ریشه واحد اریب دارد. ملاحظه روند زمانی سری‌های زمانی تحقیق که در نمودارهای ۱ و ۲ نشان داده شده است حکایت از وجود یک شکست ساختاری از نوع تغییر عرض از مبدأ (الگوی A پرون) در سال ۱۳۵۵ دارد. خط رسم شده بر روی این دو نمودار نشان‌دهنده روند زمانی است.



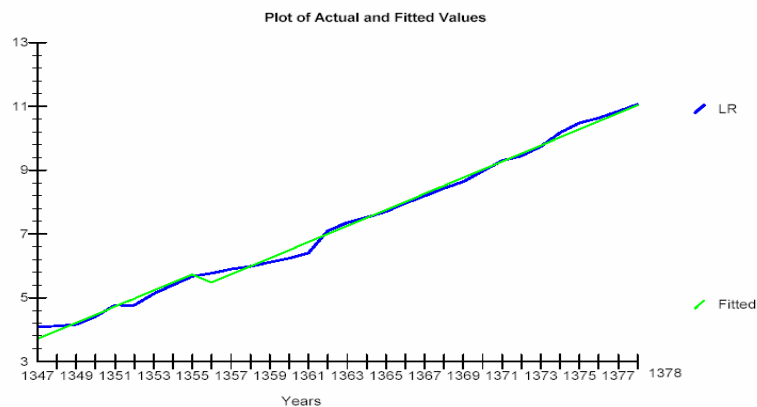
1- Schwarz Bayesian Criterion.

۲- محاسبات با استفاده از نرم افزار Microfit 4.0 انجام شده است.

3- Pierre Perron.

برای آزمون شکست ساختاری از نوع تغییر در عرض از مبدأ باید از الگوی اول پرون (الگوی A) به صورت زیر استفاده نمود:

$$y_t = \mu + \theta DU_t + \beta t + dD(TB)_t + \alpha y_{t-1} + \sum_{i=1}^k c_i \Delta y_{t-i} + \varepsilon_t \quad (1)$$



نمودار ۲- لگاریتم ارزش افزوده گروه بازرگانی، رستوران و هتلداری

TB: زمان شکست ساختاری

$D(TB)$: متغیر مجازی شکست ساختاری است که در زمان $t = TB + 1$ مساوی

یک و در غیر این صورت صفر می باشد.

DU : متغیر مجازی و در زمان‌های $t > TB$ مساوی یک و در غیر این صورت صفر

است.

در روش پرون، فرض صفر یعنی وجود ریشه واحد در برابر فرض مقابل یعنی روند

مانا با قیود اعمال شده بر پارامترهای الگو به صورت زیر آزمون می گردد:

$$H_0: \alpha = 1, \theta = 0, \beta = 0, d \neq 0$$

(۲)

$$H_1: \alpha < 1, \theta \neq 0, \beta \neq 0, d = 0$$

جدول ۲ نتایج آزمون پرون را برای دو متغیر تحقیق نشان می دهد:

جدول ۲- بررسی شکست ساختاری در متغیرهای تحقیق با استفاده از آزمون پرون

| LR | | LTJ | | متغیر ضریب |
|----------|--------------|----------|--------------|---------------|
| آماره t | ضریب برآوردی | آماره t | ضریب برآوردی | |
| ۴/۷۹۷ | ۱/۹۶۹ | ۱/۲۹۱ | ۰/۰۶۸ | μ |
| -۳/۴۸۰ | -۰/۳۳۱ | -۴/۸۶۵ | -۰/۷۱۵ | θ |
| ۴/۵۴۷ | ۰/۱۴۶ | ۵/۱۱۴ | ۰/۱۴۱ | β |
| ۱/۳۱۴ | ۰/۱۶۶ | ۱/۰۵۹ | ۰/۱۹۸ | d |
| ۳/۴۶۰ | ۰/۴۳۹ | ۵/۱۸۷ | ۰/۵۱۰ | α |
| - | - | ۱/۸۵۶ | ۰/۲۳۵ | c_1 |
| ۰/۹۹۸ | | ۰/۹۹۸ | | R^2 |
| ۰/۹۹۸ | | ۰/۹۹۷ | | \bar{R}^2 |
| ۲/۰۱۶ | | ۱/۸۳۷ | | D-W |
| ۳۲۰۷/۹۴۰ | | ۲۱۳۰/۴۳۲ | | F |

منبع: محاسبات تحقیق

مقدار آماره آزمون بر اساس صحت فرضیه H_0 برابر است با $\tau = \frac{\hat{\alpha} - 1}{s.e.(\hat{\alpha})}$ و مقدار

بحرانی بر اساس نسبت تعداد مشاهدات قبل از وقوع شکست به کل مشاهدات $\lambda = \frac{TB}{T}$ تعیین می‌گردد. مقدار آماره آزمون یعنی τ برای ضریب برآورد شده α در الگوهای مربوط به LR و LTJ به ترتیب -۴/۹۷۶ و -۴/۴۰۵ می‌باشد. مقدار بحرانی برای $\lambda = ۰/۳$ (نزدیکترین کمیت به $\lambda = ۰/۲۸$ در مورد حاضر) در سطح ۱ درصد، ۲/۵ درصد، ۵ درصد و ۱۰ درصد به ترتیب برابر -۴/۳۹، -۴/۰۳، -۳/۷۶ و -۳/۴۶ است. با توجه به این که قدر مطلق مقدار محاسبه شده کمیت آماره آزمون در هر دو متغیر تحقیق از قدر مطلق تمامی مقادیر بحرانی در سطوح مختلف معنی‌دار بودن بزرگتر است لذا H_0 رد می‌شود. علاوه بر این برقراری دیگر شرایط فرض یک یعنی $\theta \neq 0, \beta \neq 0, d = 0$ نیز مؤید رد فرض صفر می‌باشد. بنابراین هر دو سری زمانی تحقیق یعنی LR و LTJ دارای ریشه واحد نیستند و در نتیجه روند-مانا هستند.

۴- بررسی آشوب در سری زمانی درآمدهای مالیات بر مشاغل در ایران

برای بررسی امکان پیش‌بینی درآمدهای مالیات بر مشاغل، در ابتدا لازم است آزمون‌های آشوب به عمل آید که در این تحقیق از آزمون نمای لیاپانوف^۱ استفاده می‌گردد. در صورت تایید وجود آشوب، سیستم با استفاده از الگوسازی غیرخطی قابلیت پیش‌بینی کوتاه‌مدت را داراست. در واقع چنانچه نمای لیاپانوف کوچک باشد قابلیت پیش‌بینی برای دوره‌ای که با عکس نمای لیاپانوف متناسب است ممکن خواهد بود. پیش‌بینی بلندمدت به دلیل وجود ویژگی حساسیت سیستم‌های آشوب‌گونه نسبت به شرایط اولیه بسیار مشکل و یا می‌توان گفت که غیرممکن است و به این ترتیب تنها در کوتاه مدت پیش‌بینی امکان پذیر می‌باشد.

۴-۱- تخمین نمای لیاپانوف

نظریه آشوب بینش جدیدی را از دنیای واقعی به ما ارائه می‌دهد. رفتار آشوبگونه تنها در سیستم‌های غیرخطی پویا اتفاق می‌افتد. در چنین سیستم‌هایی، وقوع حتی یک تغییر بسیار جزئی در شرایط اولیه منجر به تغییر اساسی در رفتار سیستم می‌گردد (گارلیوسکاس^۲). برای تشخیص وجود آشوب در سری‌های زمانی آزمون‌های مختلفی وجود دارد که مهمترین آن‌ها تخمین بزرگترین نمای لیاپانوف است. نمای لیاپانوف میانگین نرخ رشد هم‌گرایی یا واگرایی مسیرهای نزدیک بهم در فضای ایجاد شده توسط بردارهای m حافظه است (خاکی صدیق، کارولوکس و خالوزاده). هر سیستمی با داشتن حداقل یک نمای لیاپانوف مثبت، سیستمی آشوبگونه است. نمای لیاپانوف با دو روش تابع معین^۳ و بعد جانسانی^۴ قابل محاسبه و ارزیابی می‌باشد.^۵

1- Lyapunov Exponents.

2- A. Garliauskas, 1999.

3- Deterministic Function.

4- Embedding Dimension.

۵- برای اطلاع بیشتر از روش تابع معین به Ellner and Turchin (1995)، Serletis and Shintani (2003)،

و Kendall (2001) و بعد جانسانی به خاکی صدیق و دیگران (۱۳۷۷)، احراری (۱۳۸۱)، قدیمی و مشیری (۱۳۸۱)،

مشیری و فروتن (۱۳۸۲) و حمیدی (۱۳۸۴) مراجعه کنید.

در اینجا بررسی نمای لیاپانوف به روش بعد جانشانی صورت می‌گیرد. برای محاسبه نمای لیاپانوف، ابتدا ماتریس‌هایی که دارای m سطر و $N-m+1$ ستون هستند با استفاده از N داده اسکالرسری زمانی تشکیل می‌گردد. از میان این ماتریس‌ها تمامی جفت بردارهایی که در رابطه زیر صدق می‌کنند، مشخص می‌شود.

$$r_0(m; i, j) = \|x_i - x_j\| \leq \varepsilon \quad (3)$$

ε یک مقدار کوچک مثبت است. محاسبه فوق طی n مرحله زمانی انجام می‌شود.

$$r_n(m; i, j) = \|x_{i+n} - x_{j+n}\| \quad (4)$$

سپس میزان واگرایی نقاط نزدیک به هم محاسبه می‌گردد. اگر نقاط نزدیک به هم به ازاء n های بزرگتر از صفر در فضای m بعدی از یکدیگر واگرا شوند، $d(m; i, j)$ بزرگتر از یک خواهد بود.

$$d(m; i, j) = \frac{r_n}{r_0} = \frac{\|x_{i+n} - x_{j+n}\|}{\|x_i - x_j\|} \quad (5)$$

در نهایت نمای لیاپانوف بر اساس رابطه زیر محاسبه می‌شود:

$$\lambda(m, n) = \frac{1}{N(N-m-1)} \sum \log d_n(m; i, j) \quad (6)$$

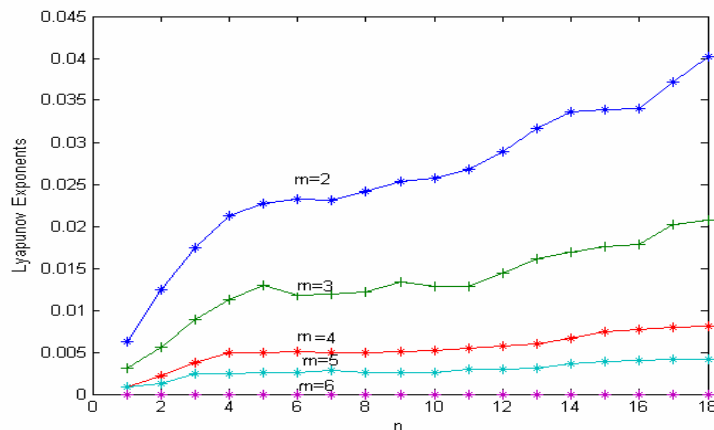
نتایج حاصل از تخمین نمای لیاپانوف برای سری زمانی مالیات بر مشاغل طی دوره زمانی ۱۳۴۷-۱۳۸۳ نشان دهنده هم‌گرایی پایدار مقادیر تخمینی نمای لیاپانوف برای n دوره می‌باشد. نمای لیاپانوف محاسبه شده برای ابعاد جانشانی ۲ تا ۵ بزرگتر از صفر می‌باشد (جدول ۳ و نمودار ۳). هر چقدر که بتوان در ابعاد جانشانی بالاتر به λ مثبت کم‌تری دست یافت می‌توان نتیجه گرفت که سیستم از آشوب ضعیف‌تر و حساسیت نسبت به شرایط اولیه کمتری برخوردار است. در نتیجه می‌توان از اطلاعات گذشته به‌میزان بیشتری برای پیش‌بینی استفاده کرد. لازم به ذکر است که در صورت وجود آشوب در سیستم، پیش‌بینی برای بلندمدت امکانپذیر نمی‌باشد و فقط می‌توان برای کوتاه‌مدت پیش‌بینی‌های لازم را انجام داد.

جدول ۳- نتایج آزمون نمای لیاپانوف برای سری زمانی درآمدهای مالیات بر مشاغل طی ۱۳۴۷-۱۳۸۳

| λ | بعد | ۲ | ۳ | ۴ | ۵ |
|---------------------|------|--------|--------|-------|---|
| سری زمانی اصلی | ۰/۰۳ | ۰/۰۱۶۵ | ۰/۰۰۸۸ | ۰/۰۰۴ | |
| سری زمانی بهم ریخته | ۰/۰۵ | ۰/۰۳۶۳ | ۰/۰۱۵ | ۰/۰۰۹ | |

منبع: محاسبات تحقیق

برای اطمینان در محاسبه نمای لیاپانوف از آزمون بهم ریختگی^۱ استفاده شده است. در این روش ابتدا سری اصلی به صورت تصادفی بهم ریخته می‌شود تا یک سری جدید ایجاد گردد. سپس با محاسبه نمای لیاپانوف برای سری جدید، اگر مقدار نمای لیاپانوف (برای سری جدید) بزرگتر از نمای لیاپانوف سری اصلی باشد، سری زمانی اصلی غیرتصادفی، معین^۲ و دارای نظم خواهد بود. البته این نظم ممکن است تابع پیچیده و عموماً غیرخطی باشد که توسط روش‌های خطی آمار کلاسیک قابل کشف نباشد. همان‌طور که در جدول ۳ مشاهده می‌شود مقادیر نمای لیاپانوف برای سری زمانی بهم ریخته بیشتر از سری زمانی اصلی می‌باشد، که نشان دهنده غیرتصادفی و معین بودن سری زمانی مربوطه می‌باشد.



نمودار ۳- مقادیر نمای لیاپانوف در ابعاد ۲ تا ۶ برای سری زمانی درآمدهای مالیات بر مشاغل

1- Shuffle Test.

2- Deterministic.

۵- الگوسازی و پیش‌بینی درآمدهای مالیات بر مشاغل

نتایج آزمون نمای لیپانوف نشان دهنده وجود آشوب ضعیف در سیستم می‌باشد، بنابراین از طریق الگوسازی غیرخطی می‌توان برای کوتاه‌مدت پیش‌بینی دقیق‌تری انجام داد. در این قسمت برای تأیید نتایج آزمون نمای لیپانوف، پیش‌بینی بر اساس برآورد الگوی مربوطه به دو شیوه خطی و غیرخطی صورت می‌گیرد و در نهایت عملکرد این دو روش با هم مقایسه می‌شود. در بسیاری از مطالعات انجام شده همچون مطالعات فلاریوف و آورهن کوف^۱ (۱۹۹۹)، پالیت و پوپویک^۲ (۲۰۰۰)، ویریلی و فرشلبن^۳ (۲۰۰۰)، قدیمی و مشیری (۱۳۸۱)، راعی و چاوشی (۱۳۸۲) و خالوزاده و خاکی صدیق (۱۳۸۲) برتری روش شبکه عصبی نسبت به روش‌های خطی اشاره شده است. در مجموعه مطالعات انجام شده توسط وایت^۴ و همکاران او مانند لی، وایت و گرنجر^۵ (۱۹۹۳) و سوانسن و وایت^۶ (۱۹۹۷) برتری روش‌های غیر خطی و از جمله شبکه عصبی مصنوعی نسبت به روش‌های خطی در بسیاری از سری‌های زمانی اقتصادی نشان داده شده است.

۵-۱- برآورد الگوی رگرسیون خطی و پیش‌بینی درآمدهای مالیات بر مشاغل

نتایج برآورد الگوی رگرسیون درآمدهای مالیات بر مشاغل در جدول ۴ آمده است: همان‌طور که مشاهده می‌گردد ضریب LR_t علامت مورد انتظار را دارد و با توجه به مقدار آماره t ، در سطح بسیار بالایی معنی‌دار است. ضریب تعیین R^2 برابر ۰/۹۹ و نشان دهنده قدرت توضیح‌دهندگی بسیار بالای الگو می‌باشد. آماره دوربین - واتسون (D-W) عدم وجود مشکل خودهمبستگی پیاپی و آماره F معنی‌دار بودن کل معادله در سطح اطمینان بسیار بالایی را نشان می‌دهد. برای رفع خودهمبستگی سریالی

1 - G.F. Flareiov and E.O. Averehenkov.

2- A. Palit and D. Popovic.

3- F. Virili and B. Freisleben.

4 -Halbert White.

5 -T-H Lee, H. White and C.W.J. Granger.

6- N. R. Swanson and H. White.

اجزای اخلاص، از الگوی خودبازگشت مرتبه سوم استفاده شده است که ضرایب مربوطه همگی معنی‌دار هستند.

جدول ۴- برآورد الگوی رگرسیون خطی درآمدهای مالیات بر مشاغل طی دوره ۱۳۴۷-۱۳۷۸

| متغیر | ضریب برآوردی | آماره t |
|---------------|--------------|---------------------|
| c | -۳/۵۲۳ | -۵/۸۲۵ |
| LR | ۱/۰۰۱ | ۱۴/۱۱۰ |
| AR(1) | ۱/۰۶۴ | ۵/۶۱۹ |
| AR(2) | -۰/۶۶۷ | -۲/۲۶۶ |
| AR(3) | ۰/۳۵۲ | ۱/۸۱۵ |
| $R^2 = ۰/۹۹۲$ | | $\bar{R}^2 = ۰/۹۹۱$ |
| D-W = ۱/۹۸۸ | | F = ۷۷۶/۷۸۱ |

منبع: محاسبات تحقیق

در جدول ۵ مقادیر واقعی و پیش‌بینی شده درآمدهای مالیات بر مشاغل با استفاده از الگوی رگرسیونی فوق برای دوره زمانی ۱۳۷۹ - ۱۳۸۳ آورده شده است.

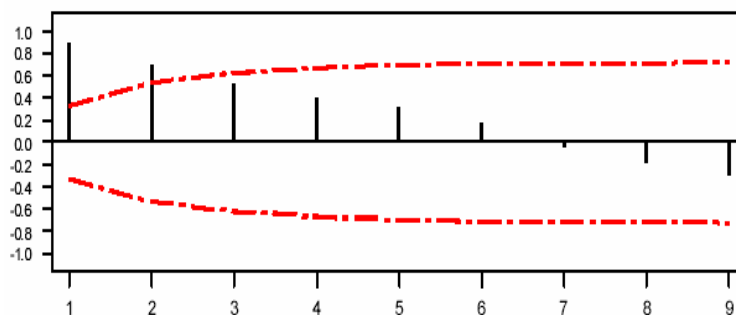
جدول ۵- مقادیر واقعی و پیش‌بینی شده درآمدهای مالیات بر مشاغل با استفاده از الگوی رگرسیون خطی (۱۳۷۹-۱۳۸۳)

| سال | ۱۳۷۹ | ۱۳۸۰ | ۱۳۸۱ | ۱۳۸۲ | ۱۳۸۳ |
|-------------------|-------|-------|-------|-------|-------|
| مقادیر واقعی | ۷/۹۸۱ | ۸/۱۳۲ | ۸/۲۷۸ | ۸/۳۳۷ | ۸/۵۷۳ |
| مقادیر برآوردی | ۷/۷۴۵ | ۷/۹۰۱ | ۸/۱۰۳ | ۸/۲۸۲ | ۸/۴۸۳ |
| قدر مطلق درصد خطا | ۲/۹۶ | ۲/۸۴ | ۲/۱۱ | ۰/۶۶ | ۱/۰۵ |

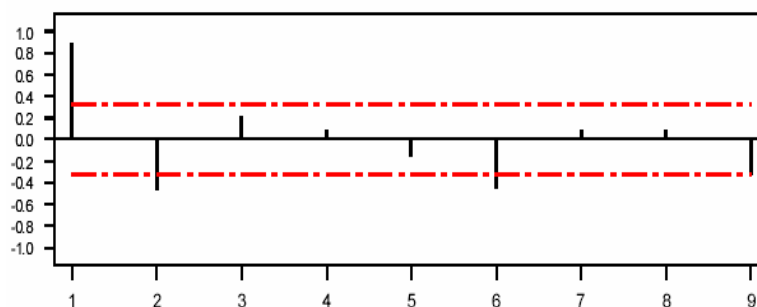
منبع: محاسبات تحقیق

۵-۲ برآورد الگوی $ARIMA$ و پیش بینی درآمدهای مالیات بر مشاغل

بر اساس روش باکس-جنکینز، برای تعیین p و q در الگوی $ARIMA$ می توان از تابع خودهمبستگی (ACF) ^۱ و تابع خودهمبستگی جزئی (PACF) ^۲ نمونه سری زمانی استفاده نمود. از آنجا که سری LTJ زمانی دارای روند می باشد (نمودار ۱) ابتدا باید روند زدایی شود. نمودارهای ۴ و ۵، ACF و PACF پسماندهای الگوی روند زدایی شده LTJ را نشان می دهد.



نمودار ۴- تابع خودهمبستگی (ACF) پسماندهای الگوی روند زدایی شده LTJ



نمودار ۵- تابع خودهمبستگی جزئی (PACF) پسماندهای الگوی روند زدایی شده LTJ

1- Autocorrelation Function.

2-Partial Autocorrelation Function.

با توجه به نمودارهای فوق ملاحظه می‌گردد که ACF پسماندهای الگوی روند زدایی شده LTJ به‌طور نمایی تنزل می‌نماید و مقدار PACF بعد از دو تأخیر قطع می‌شود که این رفتار در ابتدای امر نشان‌دهنده الگوی خودبازگشت از مرتبه دوم می‌باشد یعنی $AR(2)$ یا به عبارت دیگر $ARIMA(3,0,0)$ می‌باشد.^۱

نتایج حاصل از برآورد الگوی $AR(2)$ نشان از معنی‌داری این الگو دارد اما نیکویی برازش رد می‌شود چرا که مقدار آماره چی دو باکس - پی یرس^۲ اصلاح شده $۲۹/۴$ می‌باشد که در مقایسه با سطح بحرانی (چی دو با ۱۰ درجه آزادی) $۱۸/۳$ ، فرض صفر مبنی بر نیکویی برازش رد می‌شود (پیوست ۱). لذا با توجه به این که مقدار PACF در سه وقفه معنی دار است و همچنین با توجه به معیارهای AIC و SBC که در جدول ۶ ذکر شده است، الگوی $AR(3)$ در نظر گرفته شد. نتایج به دست آمده نیکویی برازش را رد نمی‌کند (پیوست ۲). بررسی ACF و PACF پسماندهای این الگو اغتشاش محض بوده و الگوی خاصی را نشان نمی‌دهد (پیوست ۳). لذا الگوی فوق مناسب به نظر می‌رسد.^۳

جدول ۶- معیارهای انتخاب AIC و SBC برای تعیین مرتبه الگوی AR

| معیار آزمون | P | ۱ | ۲ | ۳ | ۴ |
|-------------|---|--------|--------|--------|--------|
| AIC | | -۰/۴۷۶ | -۰/۶۹۰ | -۰/۲۹۰ | -۰/۵۹۰ |
| SBC | | -۰/۳۴۴ | -۰/۵۱۲ | -۰/۰۵۵ | -۰/۳۱۸ |

منبع: محاسبات تحقیق

۱- جهت آشنایی با نحوه تشخیص صحیح نوع الگوی سری زمانی و مرتبه آن به ویلیام دبلیو.اس.وی (۱۳۷۶) یا Walter Enders (2004) مراجعه شود.

۲- Box - Pierce.

۳- محاسبات و نمودارهای ACF و PACF با کمک نرم افزار Minitab انجام شده است.

نتایج حاصل از برآورد الگوی $AR(3)$ برای سری زمانی LTJ در جدول ۷ آمده است:^۱

جدول ۷- برآورد الگوی سری زمانی درآمدهای مالیات بر مشاغل طی دوره ۱۳۴۷-۱۳۷۸

| متغیر | ضریب برآوردی | آماره t |
|---------------|--------------|---------------------|
| c | -۱/۶۷۲ | -۰/۵۷۱ |
| t | ۰/۲۹۲ | ۳/۳۴۸ |
| AR(1) | ۱/۴۶۱ | ۷/۶۳۸ |
| AR(2) | -۰/۹۱۴ | -۲/۹۹۴ |
| AR(3) | ۰/۳۶۶ | ۱/۸۰۸ |
| $R^2 = ۰/۹۹۴$ | | $\bar{R}^2 = ۰/۹۹۳$ |
| D-W = ۲/۱۰۴ | | F = ۱۰۵۰/۵۷۹ |

منبع: محاسبات تحقیق

بر اساس نتایج مندرج در جدول بالا تمامی ضرایب از نظر آماری معنی دار هستند. ضریب تعیین R^2 برابر ۰/۹۹۵ و نشان دهنده قدرت توضیح‌دهندگی بسیار بالای الگو می‌باشد. آماره دوربین - واتسون (D-W) عدم وجود مشکل خودهمبستگی پیاپی و آماره F معنی‌دار بودن کل معادله در سطح اطمینان بسیار بالایی را نشان می‌دهد. در جدول ۸ مقادیر واقعی و پیش‌بینی شده درآمدهای مالیات بر مشاغل با استفاده از الگوی سری زمانی فوق برای دوره زمانی ۱۳۷۹ - ۱۳۸۳ منعکس شده است.

۱- محاسبات با استفاده از نرم افزار Eviews انجام شده است.

جدول ۸- مقادیر واقعی و پیش‌بینی شده درآمدهای مالیات بر مشاغل با استفاده از الگوی سری زمانی (۱۳۷۹-۱۳۸۳)

| سال | ۱۳۷۹ | ۱۳۸۰ | ۱۳۸۱ | ۱۳۸۲ | ۱۳۸۳ |
|-------------------|-------|-------|-------|-------|-------|
| مقادیر واقعی | ۷/۹۸۱ | ۸/۱۳۲ | ۸/۲۷۸ | ۸/۳۳۷ | ۸/۵۷۳ |
| مقادیر برآوردی | ۸/۱۴۰ | ۸/۲۱۶ | ۸/۳۹۷ | ۸/۵۶۹ | ۸/۸۹۴ |
| قدر مطلق درصد خطا | ۱/۹۹ | ۱/۰۳ | ۱/۴۴ | ۲/۷۸ | ۳/۷۴ |

منبع: محاسبات تحقیق

۵-۳- برآورد الگوی شبکه عصبی و پیش‌بینی درآمدهای مالیات بر مشاغل برای یک دوره برون نمونه ای

با استفاده از شبکه عصبی مشتمل بر یک لایه پنهان^۱، تابع محرک^۲ غیرخطی و تعداد نرون‌های^۳ کافی در این لایه و یک تابع محرک خطی در لایه خروجی، می‌توان هر رابطه غیرخطی را تقریب زد. بنابراین برای طراحی الگوی درآمدهای مالیات بر مشاغل در ایران، یک شبکه عصبی با همان مجموعه متغیرهای ورودی در الگوی رگرسیون بالا، در نظر گرفته می‌شود.

شبکه مورد استفاده یک شبکه عصبی پیش‌خور^۴ با یک لایه پنهان و تعداد^۳ نرون در لایه پنهان می‌باشد. تابع محرک لایه پنهان، تابع غیرخطی تانژانت سیگموئید^۵ و تابع محرک لایه خروجی، تابع خطی پیورلین^۶ می‌باشد. از الگوریتم پس انتشار خطا^۷ نیز برای آموزش شبکه استفاده شده است به این صورت که خطای پیش‌بینی برای هر مشاهده، به وسیله محاسبه اختلاف خروجی شبکه با ارزش متغیرهای هدف اندازه‌گیری می‌شود، سپس بین لایه‌های موجود طبق روابط پس انتشار توزیع می‌گردد. به این

1- Hidden Layer .

2- Activation Function .

3- Neurons .

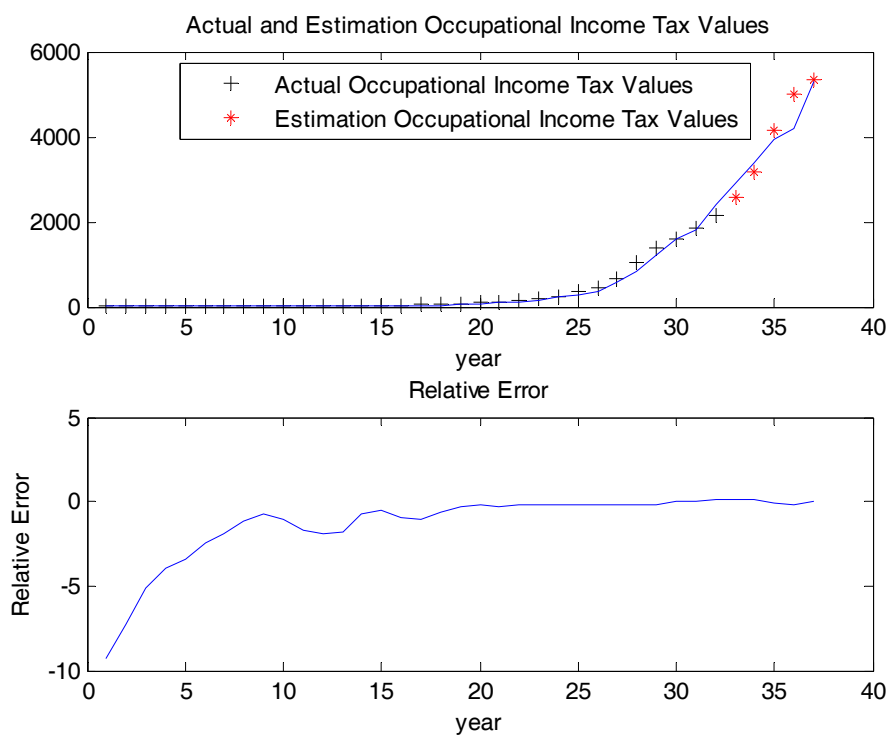
4- Feed Forward Neural Network .

5- Tansig .

6- Purelin .

7- Error Backpropagation .

ترتیب وزن‌های شبکه به‌گونه‌ای تعدیل می‌شود که مجموع مربعات خطا^۱ حداقل می‌شود. در اینجا شبکه از دو مجموعه تشکیل شده است. مجموعه آموزش^۲ تقریباً ۸۳ درصد از اطلاعات سری زمانی را به خود اختصاص داده و مربوط به سال‌های ۱۳۴۷-۱۳۷۸ می‌باشد. بعد از آموزش شبکه با این سری از اطلاعات، مجموعه آزمایش^۳ به آن اضافه می‌گردد. آنگاه میانگین خطای شبکه محاسبه و پارامترهای شبکه‌ای که دارای کم‌ترین میانگین خطاست، ذخیره می‌گردد. در این شرایط پیش‌بینی برای یک دوره بعد انجام شده است. نمودار ۶ پیش‌بینی یک دوره بعد درآمدهای مالیات بر مشاغل و منحنی خطای پیش‌بینی را نشان می‌دهد.



نمودار ۶- پیش‌بینی یک دوره بعد درآمدهای مالیات بر مشاغل و منحنی خطای پیش‌بینی

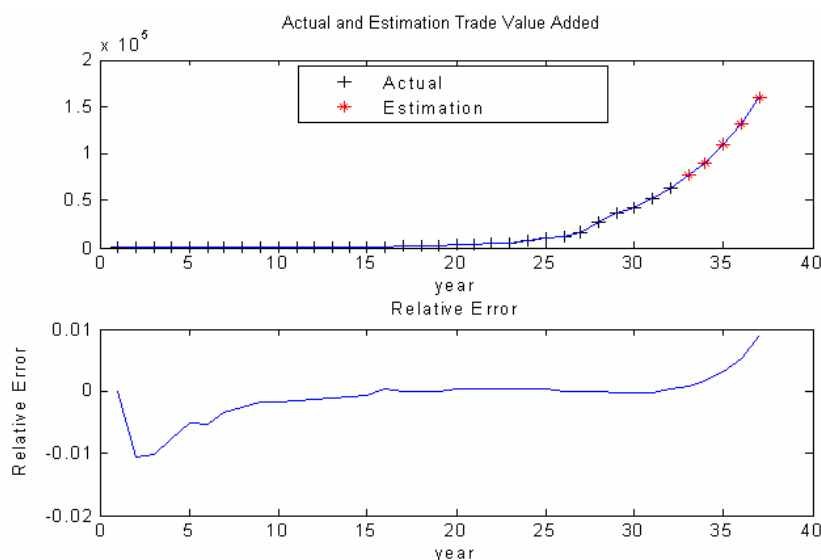
1- Sum of Squared Errors.

2- Training Set.

3- Testing Set.

۴-۵- برآورد الگوی شبکه عصبی مصنوعی و پیش‌بینی بلندمدت درآمدهای مالیات بر مشاغل

پیش‌بینی صورت گرفته در بالا مربوط به یک دوره بعد است. برای پیش‌بینی بیشتر از یک دوره، از مقادیر پیش‌بینی شده به جای مقادیر واقعی استفاده می‌شود. به این منظور ابتدا ارزش افزوده بازرگانی، رستوران و هتلداری پیش‌بینی می‌گردد. در این حالت ساختار شبکه از نوع خود بازگشت مرتبه اول می‌باشد و تعداد ۳ نرون در لایه پنهان وجود دارد. تابع محرک لایه پنهان تانژانت سیگموئید، تابع محرک لایه خروجی خطی و یادگیری شبکه براساس قانون پس انتشار خطاست. نتایج پیش‌بینی برای یک دوره بعد با استفاده از این شبکه در نمودار ۷ نشان داده شده است.

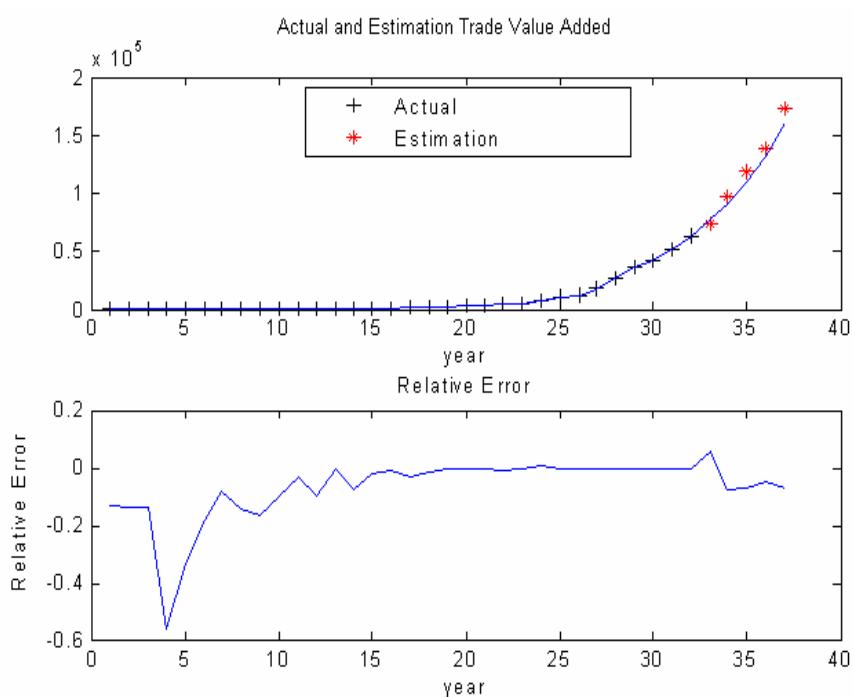


نمودار ۷- پیش‌بینی ارزش افزوده بازرگانی، رستوران و هتلداری و مقدار خطای پیش‌بینی برای یک دوره

سپس با استفاده از وزن‌ها و بایاس‌های^۱ به‌دست آمده در این مرحله، پیش‌بینی برای بلندمدت صورت می‌گیرد به این صورت که از خروجی هر دوره به‌عنوان

1- Bias.

ورودی برای دوره بعد استفاده می‌شود. به این ترتیب برای ۵ سال آخر یعنی ۱۳۷۹ - ۱۳۸۳ پیش‌بینی از ارزش افزوده بازرگانی، رستوران و هتلداری به‌عمل آمده است که نتایج آن در نمودار ۸ دیده می‌شود.



نمودار ۸- پیش‌بینی ارزش افزوده بازرگانی، رستوران و هتلداری و مقدار خطای پیش‌بینی برای ۵ دوره

حال با استفاده از مقادیر پیش‌بینی شده بازرگانی، رستوران و هتلداری، درآمد مالیات بر مشاغل طی افق زمانی ۱۳۷۹-۱۳۸۳ پیش‌بینی می‌گردد. نکته مهم و قابل توجه در این حالت آن است که این داده‌ها را به جای داده‌های واقعی به‌عنوان ورودی وارد الگوی شبکه عصبی نموده و درآمد مالیات بر مشاغل برای دوره t (سال ۱۳۷۹) پیش‌بینی می‌گردد. خروجی دوره t ورودی برای دوره بعد می‌باشد و مقدار درآمدهای

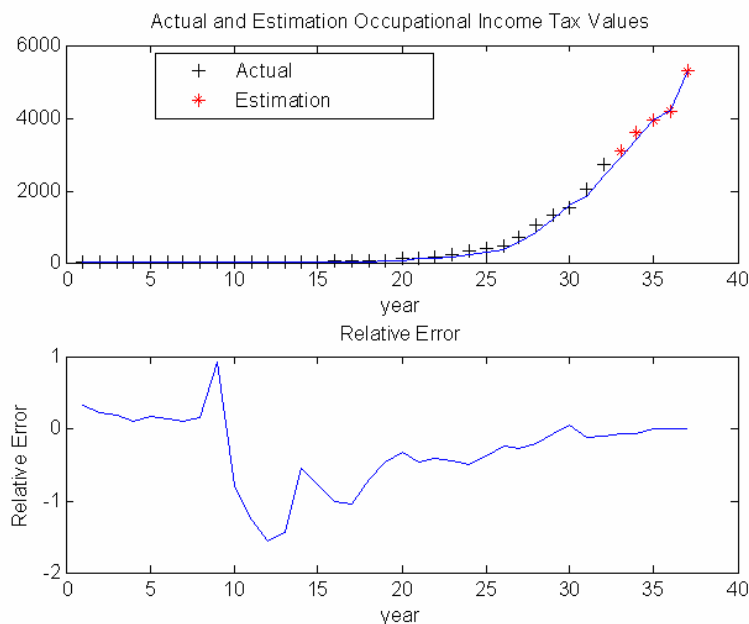
مالیات بر مشاغل برای دوره $t+1$ پیش‌بینی می‌شود. به این ترتیب برای هر ۵ دوره پیش‌بینی صورت می‌گیرد (جدول ۹ و نمودار ۹).

جدول ۹- مقادیر واقعی و پیش‌بینی شده درآمدهای مالیات بر مشاغل با استفاده از شبکه عصبی مصنوعی (۱۳۷۹ - ۱۳۸۳)

| | | | | | |
|-------|-------|-------|-------|-------|-------------------|
| ۸/۵۷۳ | ۸/۳۳۷ | ۸/۲۷۸ | ۸/۱۳۲ | ۷/۹۸۱ | مقادیر واقعی |
| ۸/۵۷۳ | ۸/۳۳۷ | ۸/۲۷۶ | ۸/۱۸۶ | ۸/۰۳۱ | مقادیر برآوردی |
| ۰/۰۰ | ۰/۰۰ | ۰/۰۲ | ۰/۶۶ | ۰/۶۳ | قدر مطلق درصد خطا |

منبع: محاسبات تحقیق

نمودار ۹ پیش‌بینی درآمدهای مالیات بر مشاغل و خطای ناشی از این پیش‌بینی را برای ۵ دوره زمانی نشان می‌دهد.



نمودار ۹- پیش‌بینی درآمدهای مالیات بر مشاغل و خطای پیش‌بینی برای ۵ دوره

۵-۵- مقایسه عملکرد الگوهای رگرسیون خطی، سری زمانی و شبکه عصبی برای پیش‌بینی درآمدهای مالیات بر مشاغل

برای بررسی عملکرد سه الگو، از معیارهای مجذور میانگین مربع خطا^۱ (RMSE) و میانگین مطلق خطا^۲ (MAE) استفاده می‌شود. مقایسه نتایج این دو معیار برای دوره پیش‌بینی الگوهای رگرسیون خطی، سری زمانی و شبکه عصبی نشان از عملکرد بهتر شبکه عصبی دارد (جدول ۱۰).

جدول ۱۰- مقادیر معیار RMSE و MAD برای سه الگوی رگرسیون خطی، سری زمانی و شبکه عصبی در دوره پیش‌بینی ۱۳۷۹-۱۳۸۳

| معیار | الگوی رگرسیون | الگوی سری زمانی | الگوی شبکه عصبی مصنوعی |
|------------------|---------------|-----------------|------------------------|
| RMSE (۱۳۸۳-۱۳۷۹) | ۰/۱۷۳ | ۰/۲۰۱ | ۰/۰۴۷ |
| MAE (۱۳۸۳-۱۳۷۹) | ۰/۱۵۷ | ۰/۱۸۲ | ۰/۰۲۲ |

منبع: محاسبات تحقیق

۶- نتیجه‌گیری

برای تعیین الگوی مناسب و با قابلیت پیش‌بینی دقیق‌تر، ماهیت ساختاری سری زمانی موردنظر از لحاظ مانایی و شکست ساختاری و نیز خطی، غیرخطی و تصادفی بودن مورد بررسی قرار گرفت. نتایج آزمون ریشه واحد دلالت بر نامانایی متغیرهای تحقیق دارد و درجه جمعی ۱ را نشان می‌دهد. اما با انجام آزمون شکست ساختاری پرون مشخص شد هر دو متغیر تحقیق روند-مانا هستند.

برای بررسی وجود آشوب در سیستم، از آزمون نمای لیاپانوف استفاده شد. نتایج آزمون بهم ریختگی نشان دهنده غیر تصادفی بودن سری زمانی مالیات بر مشاغل می‌باشد و آزمون نمای لیاپانوف آشوب ضعیفی را در سری زمانی مربوطه نشان

1- Root Mean Squared Error.

2- Mean Absolute Error.

می‌دهد. بنابراین با استفاده از الگوسازی غیرخطی می‌توان برای کوتاه مدت پیش‌بینی دقیقی را انجام داد.

برای پیش‌بینی درآمد مالیات بر مشاغل در ایران طی سال‌های ۱۳۷۹-۱۳۸۳، الگوی رگرسیون خطی، الگوی سری زمانی و سپس الگوی شبکه عصبی بر اساس داده‌های موجود در دوره ۱۳۴۷-۱۳۷۸ طراحی و برآورد گردید. جهت بررسی عملکرد این سه الگو، از معیارهای میانگین قدرمطلق انحراف و مجذور میانگین مربعات خطا استفاده شد. مقایسه نتایج این دو معیار، عملکرد بهتر الگوی شبکه عصبی مصنوعی در پیش‌بینی درآمدهای مالیات بر مشاغل نسبت به الگوی رگرسیون خطی و الگوی سری زمانی را نشان می‌دهد.

فهرست منابع

()

()

)

(

()

)

(

()

()

:

()

()

()

()

()

()

()

()

()

()

()

()

: () . .

- 16- Enders, Walter (2004), "Applied Econometric Time Series", John Wiley.
- 17- Ellner, S. and P. Turchin (1995), "Chaos in a Noisy World: New Methods and Evidence from Time Series Analysis", *American Naturalist*, Vol.145, pp.343-375.
- 18- Fllareiov, G.F and E.O. Averehenkov (1999), "Using Neural Nets for Time Series Forecasting", *IEEE*, pp. 249-253.
- 19- Garliauskas, A. (1999), "Neural Network Chaos and Computational Algorithms of Forecast in Finance", *IEEE*, pp. 638-643.
- 20- Haldrup, Neils (1998), "An Econometric Analysis of I(2) Variables", in Les Oxley and Michael McAleer (ed.), *Practical Issues in Cointegration Analysis*, Blackwell.
- 21- Kendall, E.B. (2001), "Nonlinear Dynamics and Chaos", *Encyclopedia of Life Sciences*, Vol. 13, pp.255-262.
- 22- Lee, Tae-Hwy, H. White and C.W.J. Granger (1993), "Testing for Neglected Nonlinearity in Time series Models: A Comparison of Neural Networks Methods and Alternative Tests" , *Journal of Econometrics*, Vol. 56, pp. 269-290.
- 23- Palit, A. and D. Popovic (2000), "Nonlinear Combination of Forecasts Using Artificial Neural Network, Fuzzy Logic and Neuro- Fuzzy Approaches", *IEEE*, pp. 566- 571.
- 24- Perron, Pierre (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, Vol. 57, pp. 1361-1401.
- 25- Serletis, A. and M. Shintani (2003), "No Evidence of Chaos but Some Evidence of Dependence in the US Stock Market", *Chaos, Solitons and Fractals*, Vol. 17, pp. 449- 454.
- 26- Swanson, N.R. and H. White (1997), "Forecasting Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models", *International Journal of Forecasting*, Vol. 13, pp. 439-461.
- 27- Virili, F. and B. Freisleben (2000), "Nonstaionarity and Data Preprocessing for Neural Network Predictions of an Economic Time Series", *IEEE*, pp. 129-134.

پیوست ۱: نتایج حاصل از برآورد الگوی AR(۲)

Final Estimates of Parameters

| Type | Coef | SE Coef | T | P |
|------|---------|---------|-------|-------|
| AR 1 | 1.4650 | 0.1741 | 8.41 | 0.000 |
| AR 2 | -0.4529 | 0.1839 | -2.46 | 0.020 |

Number of observations: 32

Residuals: SS = 1.17259 (backforecasts excluded)
MS = 0.03909 DF = 30

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

| Lag | 12 | 24 | 36 | 48 |
|------------|-------|-------|----|----|
| Chi-Square | 29.4 | 36.1 | * | * |
| DF | 10 | 22 | * | * |
| P-Value | 0.001 | 0.029 | * | * |

پیوست ۲: نتایج حاصل از برآورد الگوی AR(۳)

Final Estimates of Parameters

| Type | Coef | SE Coef | T | P |
|------|---------|---------|-------|-------|
| AR 1 | 1.5997 | 0.2146 | 7.46 | 0.000 |
| AR 2 | -0.9692 | 0.3807 | -2.55 | 0.016 |
| AR 3 | 0.3814 | 0.2350 | 1.62 | 0.115 |

Number of observations: 32

Residuals: SS = 1.38399 (backforecasts excluded)
MS = 0.04772 DF = 29

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

| Lag | 12 | 24 | 36 | 48 |
|------------|-------|-------|----|----|
| Chi-Square | 11.1 | 18.0 | * | * |
| DF | 9 | 21 | * | * |
| P-Value | 0.268 | 0.651 | * | * |

پیوست ۳: ACF و PACF پسماندهای الگوی سری زمانی (۳) AR

