فصلنامهٔ علمی ـ پژوهشی رفاه اجتماعی، سال هفدهم، بهار ۹۶، شماره ٦٤
Social Welfare Quarterly, Vol 17, spring 2017, No 64

شناسایی متغیرهای مؤثر بر سهم درآمد خانوارها از یارانه در شهر تهران

Identification of Effective Variables on the share of Household Income From Direct Subsidies in Tehran

لیدا کلهری*، زهره فلاح محسن‌خانی**

مقدمه: *قانون هدفمندکردن یارانه‌ها در سال ۱۳۸۷ به صورت لایحه از سوی دولت نهم جمهوری اسلامی ایران ارائه شد و سرانجام با اعمال تغییراتی به تصویب مجلس شورای اسلامی رسید. قانونی که قرار بود از محل درآمدهای آن رونق تولید و رشد اقتصادی صورت بگیرد با اجرای نادرست عملاً به معضل جدی دولت تبدیل شد. در اجرای فاز دوم این قانون، حذف یارانه‌بگیران پردرآمد مباحث زیادی را به خود اختصاص داد که سرانجام مجلس در حین بررسی لایحه بودجه سال ۹۵ کل کشور تصویب کرد که ۲٤ میلیون یارانه‌بگیر از فهرست دولت حذف شوند. به‌منظور اجرای صحیح این مصوبه می‌بایست خانوارهای پردرآمد به درستی شناسایی شوند. بر این اساس ٤ معیار «شرایط اقلیمی»، «بعد خانوار»، «محل سکونت» و «میزان درآمد» برای شناسایی این خانوارها پیشنهاد شد. در صورت معناداری این متغیرها در مدلهای ریاضی و آماری می‌توان میزان تأثیر این متغیرها را بر خانوارهای پردرآمد مشخص کرد.*

روش: *در مطالعات کاربردی معمولاً برای تحلیل داده‌هایی که مرتبط با مجموعه‌ای از متغیرهای تبیینی هستند، مدلهای رگرسیونی*

Lida Kalhori[1], Zohreh Fallah Mohsenkhani[2]

Introduction: The law on targeted subsidies was presented as a bill by the ninth government of the Islamic Republic of Iran in 2008 and eventually with amendments was passed by the Islamic Consultative Assembly.

The law that was supposed to be the revenues to increase productivity and economic growth, became a serious problem for the government due to false implementation. In the second phase of implementing this law, the removal of the wealthy people who receive direct subsidy led to a lot of discussions that, at the end of the parliament's review of the budget bill for the year 2016, the removal of 24 million people ,who receive subsidies, from the government's list was approved. For a proper implementation of this directive, high-income households should be identified correctly. Therefore, the four criteria of "climatic conditions", "household size", "area of living" and "level of income" were proposed to identify these households. If these variables are significant in mathematical and sta-

* کارشناسی ارشد آمار، پژوهشکده آمار
** کارشناسی ارشد آمار، پژوهشکده آمار، (نویسنده مسئول)، <zohrehf@yahoo.com>

1. M.A in statistics
2. M.A in statistics
<zohrehf@yahoo.com>

tistical models, the impact of these variables on high-income households can be determined.

Method: In applied studies, regression models are usually used to analyze data that are related to a set of explanatory variables. The fitting of these models is done by assuming the normality of the response variable or its transformation, along with the constant of the variance of the response variable and the incompatibility of the error components. In some cases, the response variable may be limited in the interval (0,1) and, by fitting the regression model, predictions are obtained outside the defined interval.

In these cases the use of normal regression models are not appropriate and beta regression model is suggested. This model is based on the assumption that response variable has Beta distribution, and the mean of the response variable is linked by a linear predictor with unknown coefficients and a link function to a set of explanatory variables. If the response variable gets the values of zero and one, an augmented Beta model that is mixture of a Beta distribution with two degenerated distributions at 0 and 1 has been suggested. Since in this study the variable of the response, that is the share of household income from the subsidy,

استفاده می‌شود. برازش این مدلها با فرض نرمال بودن متغیر پاسخ یا تبدیلی از آن همراه با ثابت بودن واریانس متغیر پاسخ و ناهمبسته بودن مؤلفه‌های خطا انجام می‌شود. در مواردی ممکن است متغیر پاسخ به بازه (۱و۰) محدود شود و با برازش مدل رگرسیونی، پیش‌گوییهایی خارج از بازه تعریف شده به‌دست آید. در این حالت به‌کار گیری مدلهای رگرسیونی معمول مناسب نیست و استفاده از مدل رگرسیون بتا پیشنهاد شده است. این مدل مبتنی بر این فرض است که متغیر پاسخ دارای توزیع بتا است و میانگین متغیر پاسخ از طریق یک پیشگوی خطی با ضرایب نامعلوم و یک تابع پیوند به مجموعه‌ای از متغیرهای تبیینی مرتبط می‌شود. در صورتی‌که متغیر پاسخ مقادیر صفر و یک را نیز اختیار کند، استفاده از مدل بتای افزوده که از آمیختن توزیع بتا روی بازه (۱و۰) و دو توزیع تباهیده در صفر و یک ایجاد می‌شود، پیشنهاد شده است. از آنجایی که در این مطالعه متغیر پاسخ مورد بررسی یعنی سهم درآمد خانوار از یارانه می‌تواند مقادیری در بازهٔ بسته [۰,۱] اختیار کند، برای بررسی متغیرهای مؤثر بر آن مدل رگرسیون بتای افزوده استفاده شده است.

داده‌های مورد استفاده، اطلاعات اخذ شده از آمارگیری هزینه و درآمد خانوار ایران است که توسط مرکز آمار ایران اجرا شده است. هدف اصلی آمارگیری از هزینه و درآمد خانوار اندازه‌گیری متوسط هزینه خوراک، غیرخوراک

و کل هزینه‌های خانوارهای شهری و روستایی در ایران است. اطلاعات به‌دست آمده از این آمارگیری برای محاسبه خط فقر و مطالعه ناهمگونی در تسهیلات و درآمد خانوار به‌کار می‌رود. جامعه آماری در این آمارگیری، تمام خانوارهایی است که ساکن نقاط شهری و روستایی هستند. مدل رگرسیون ارائه شده به‌منظور شناسایی متغیرهای مؤثر بر نرخ پاسخ با استفاده از داده‌های سال ۱۳۹۰ در شهر تهران به‌کار گرفته شده است. معیار انحراف اطلاع برای ارزیابی مدل استفاده شده است.

***یافته‌ها**: این مقاله به منظور شناسایی خانوارهای پردرآمد، مدل‌بندی سهم درآمد خانوار از یارانه در شهر تهران را با استفاده از نتایج آمارگیری از هزینه و درآمد خانوار مرکز آمار ایران در سال ۱۳۹۰ ارائه می‌کند. پس از بررسی متغیرهای کمکی مختلف، متغیرهای تبیینی دهک وزنی درآمد و بعد خانوار به ترتیب با ضرایب منفی و مثبت به مدل رگرسیونی بتای افزوده وارد شدند. نقشه فضایی سهم درآمد خانوار از یارانه ناهمگنی بین مناطق شهرداری تهران را نشان می‌داد، که این اثر با ضریب مثبت در مدل آماری رگرسیون بتای افزوده با لحاظ اثرات فضایی معنادار شد. بنابراین در این مدل‌بندی نیز دهک وزنی درآمد، بعد خانوار و محل سکونت خانوارها به عنوان متغیرهای معنادار بر سهم درآمد خانوار از یارانه شناسایی شدند.*

***بحث**: یافته‌های این پژوهش معیارهای پیشنهاد شده توسط کارشناسان اقتصادی*

can take values in the closed interval [1.0], to examine the effective variables on that additional beta regression model has been used. The data used in this study were collected from Iran's household income and household expenditure survey (HIES) which has been implemented by the Statistical Center of Iran.. The main objective of the household income and expenditure survey is to measure the average of food-expenditure, non-food expenditure, and total expenditure of the urban and rural households in Iran. Information provided by HIES has been applied to calculate poverty line, study the impurity in household income and facilities. Statistical population in this survey is all of the households which are settled in urban and rural areas. The presented regression model to understand the effective covariates on response rate is applied by the data set of Tehran city. Deviance Information Criterion (DIC) is used for model evaluation.

Finding: To distinguish high-income households, this paper presented modeling of the share of household income from subsidies in Tehran city using the results of Household Income and Expenditure survey conducted by the Statistical Center of Iran in 1390. After checking different covariates, deciles of house-

hold income and household size were entered into the augmented beta regression model with negative and positive coefficients respectively. Spatial map of Tehran city shows heterogeneity among different districts of Tehran municipality, which is significant in the spatial augmented beta regression model with a positive coefficient. Hence, decile of income, household size, and area of living were identified as significant variables on the proportion of household income from subsidies.

Discussion:. The findings of this study confirm the criteria proposed by economic experts to identify high-income households. Since information about area of living is more accessible in comparison to other covariates, level of income and household size, it is suggested to imply an area of living geographically for grouping households in order to make a decision about getting the direct subsidies.

To access the information about the area of household living, it is suggested to use Instruments and Landed Property Registration Organization data base. Since decile of income is a significant variable on share of household income from subsidies, information about status of ownership and area of housing unit which

برای شناسایی خانوارهای پردرآمد را تأیید می‌کند. اطلاعات محل سکونت خانوار در مقایسه با دو متغیر دیگر یعنی میزان درآمد و بعد خانوار قابل حصول‌تر است، پیشنهاد می‌شود محل سکونت خانوار به لحاظ جغرافیایی، در گروه‌بندی خانوارها به منظور تصمیم‌گیری در خصوص دریافت یارانه مورد توجه قرار بگیرد.

برای دست‌یابی به اطلاعات محل سکونت خانوار، پیشنهاد می‌شود از پایگاه اطلاعات ثبتی سازمان اسناد و املاک کشور کمک گرفته شود. از آن‌جا که دهک درآمدی به عنوان متغیر مؤثر بر سهم درآمد خانوار از یارانه معنادار شد، می‌توان اطلاعات وضعیت تملک و متراژ واحد مسکونی را نیز از پایگاه اطلاعات ثبتی سازمان اسناد و املاک کشور به‌دست آورد وبرای شناسایی خانوارهای پردرآمد از آن استفاده کرد. همچنین متغیرهای دیگری مانند تعداد و قیمت خودروی خانوار یا دارابودن جواز کسب نیز برای این منظور قابل استفاده است.

یکی دیگر از متغیرهای مؤثر بر سهم درآمد خانوار از یارانه، بعد خانوار (تعداد اعضای خانوار) است که از طریق سرشماری به‌هنگام می‌شود. به‌منظور برنامه‌ریزی درخصوص تخصیص یارانه نقدی به خانوارها، ضرورت دارد که این متغیرها همزمان با دهک درآمدی و محل سکونت خانوار مورد توجه قرار بگیرد تا سرانه

درآمد خانوار از یارانه ملاک تصمیم‌گیری قرار بگیرد. با اتصال پایگاههای مذکور و به‌کارگیری مدل پیشنهادی در این مطالعه، شناسایی خانوارهای پردرآمد امکان‌پذیر است.

کلیدواژه‌ها: *رگرسیون بتای افزوده، یارانه، آمارگیری هزینه و درآمد خانوارهای شهری*

تاریخ دریافت: ۹۵/۶/۲۰

تاریخ پذیرش:۹۶/۴/۲۸

are available in Instruments and Landed Property Registration Organization data base can be used as an indicator to identify wealthy people. Moreover,, Numbers and prices of automobiles, which belong to a household and having a business license can also be used for this purpose.

Another effective variable on share of household income from subsidies is household size, which is updated through population and housing census. In order to plan for the allocation of subsidies to households it is necessary to consider household size, decile on income and area of household living simultaneously, for a decision to be taken in regard to per capita household income from subsidies criterion. Identification of wealthy households is being possible by linking these data bases.

Keywords: augmented Beta regression, direct subsidies, urban household income and expenditure survey

مقدمه

یارانه به کمکی گفته می‌شود که دولتها برای کمک به مردم کشورشان به آنها پرداخت می‌کنند و به دو صورت نقدی و غیر نقدی می‌باشد (نسیبی‌پور، ۱۳۶۹) که البته مطالعات بسیاری در این حوزه انجام شده است (کلانتری، ۱۳۷۵). قانون هدفمندکردن یارانه‌ها در سال ۱۳۸۷ به صورت لایحه از سوی دولت نهم جمهوری اسلامی ایران ارائه شد که در آن فرایند دادن یارانه‌ها تغییر کرد به‌طوری که بخشی از یارانه‌های غیر نقدی حذف شده به صورت نقدی به مردم پرداخت شود و سایر درآمد این کار صرف کارهای عمرانی و فرهنگی شود. هدف از پرداخت یارانه به‌طور عمده کاهش فقر و کمک به توزیع متناسب درآمدها در جامعه است. فاز اول این طرح با موفقیت اجرا

نشد و به گفته برخی کارشناسان موجب شد خانوارهای بیشتری زیر خط فقر قرار بگیرند. مرحله دوم طرح هدفمندی یارانه‌ها از سال ۱۳۹۳ و با ثبت نام از متقاضیان این مرحله آغاز شد. در این مرحله علی رغم درخواست دولت از خانوارهای پردرآمد برای انصراف از دریافت یارانه نقدی، بیشتر خانوارها برای دریافت یارانه ثبت‌نام کردند. پس از بحثها و بررسیهای بسیار، مجلس تصویب کرد که باید ۲٤ میلیون یارانه‌بگیر از فهرست دولت حذف شوند. برای اجرای صحیح این مصوبه ابتدا لازم بود خانوارهای پردرآمد به درستی شناسایی شده و از جامعه یارانه‌گیران حذف شوند. علی‌رغم این که متغیرهای بسیاری به عنوان ملاکی برای شناسایی خانوارهای پردرآمد درنظر گرفته شدند که از میان آنها می‌توان به شغل سرپرستان خانوار اشاره کرد، در عمل مشخص کردن خانوارهای پردرآمد با استفاده از آنها میسر نشد. کارشناسان اقتصادی ٤ معیار «شرایط اقلیمی»، «بعد خانوار»، «محل سکونت» و «میزان درآمد» را برای شناسایی این خانوارها مناسب و عملی معرفی کردند. مسأله‌ای که پاسخ به آن اهمیت دارد آن است که آیا واقعا متغیرهای مذکور به‌صورت معنادار بر نقش دریافت یارانه در رفاه خانوار اثر دارد یا خیر، هدف از این مطالعه بررسی این سوال و پاسخ به آن است. اگر این متغیرها با استفاده از مدلهای ریاضی و آماری مناسب نیز معنادار شوند می‌توان به واسطه این متغیرها خانوارهای پردرآمد را شناسایی کرد، منظور از این خانوارها، خانوارهایی است که قطع رایانه آنها مشکلات اقتصادی برای خانوار ایجاد نمی‌کند. بنابراین با به‌کارگیری مدل آماری مناسب بر داده‌های حاصل از طرح آمارگیری از هزینه و درآمد خانوارها در سال ۱۳۹۰، معنادار بودن متغیرهای پیشنهاد شده مؤثر برای شناسایی خانوارهای پر درآمد مورد بررسی قرار خواهد گرفت. با افزایش درآمد خانوار، یارانه نقش کمتری

در کل درآمد خانوار ایفا می‌کند بنابراین می‌توان برای شناسایی متغیرهای مؤثر در تعیین خانوارهای پردرآمد، سهم درآمد خانوار از یارانه را به عنوان متغیر پاسخ در نظر گرفته و آن را مدل‌بندی کرد. لازم به ذکر است متغیرهای دیگری مانند تعداد افراد شاغل خانوار، تعداد افراد تحصیل‌کرده خانوار، سطح زیربنای محل سکونت و تملک اتومبیل شخصی نیز در مطالعه مورد بررسی قرار خواهند گرفت. بررسی رابطه متغیرهای پاسخ و متغیرهای تبیینی معمولاً با فرض آن که متغیر پاسخ دارای توزیع نرمال است، با استفاده از مدلهای رگرسیونی انجام می‌شود. به‌کارگیری یک مدل آماری مستلزم آن است که فرضیات اساسی برای برازش آن مدل برقرار باشد و در غیر اینصورت مدل برازش شده از اعتبار لازم برخوردار نیست.

آثار هدفمندسازی یارانه‌های انرژی در بخش کشاورزی توسط بخش پژوهشهای مجلس شورای اسلامی مورد بررسی قرار گرفت. به‌منظور شناخت سهم هزینه‌های انرژی در میان خانواده‌های شهری و روستایی، بر حسب میزان درآمدهای کسب شده به دهکهای اول تا دهم تقسیم‌بندی شده است. نتایج نشان داد که در خانوارهای شهری و روستایی هر چه سطح درآمد پایین‌تر باشد، سهم هزینه انرژی در مجموع هزینه‌های خانوار افزایش می‌یابد (مرکز پژوهشهای مجلس شورای اسلامی، ۱۳۸۷). بررسی آثار پرداخت رایانه نقدی و حذف یارانه سه دهک خانوار بر تولید و توزیع درآمد توسط دفتر مطالعات اقتصادی مجلس شورای اسلامی انجام شده است. نتایج نشان داد با افزایش تورم ناشی از قطع یارانه غیرنقدی، بخشی از رفاه حاصل از پرداخت یارانه نقدی کاهش می‌یابد که میزان این کاهش در خانوارهای کم‌درآمد بیشتر است (دفتر مطالعات اقتصادی مجلس شورای اسلامی، ۱۳۹۳). کرمی و همکاران (۲۰۱۲)

ارزیابی اثرات تغییر نحوه اختصاص یارانه مواد غذایی در ایران را در دو حالت حذف تدریجی یارانه طی دوره سه تا پنج ساله و یا حذف یکباره یارانه بررسی کردند. در هر دو حالت رفاه خانوارها کاهش می‌یابد به‌ویژه در خانوارهای شهری، البته خانوارهای واقع در چارکهای درآمدی بالا کمتر تحت تأثیر قرار می‌گیرند. صبوحی (۲۰۰۱) ارزیابی تأثیر کاهش یارانه‌های انرژی بر هزینه‌های خانوار را بررسی کرده است. آذربایجانی و همکاران (۲۰۱۲) اثرات کاهش یارانه انرژی و پرداخت یارانه نقدی بر رفاه خانوارها در ایران را تحلیل کردند. پیرایی و سیف (۱۳۸۹) تأثیر هدفمندسازی یارانه‌ها بر رفاه اجتماعی در ایران را بررسی نمودند. احمدی و همکاران (۱۳۹۳) بررسی ترکیب مخارج مصرفی خانوارهای شهری در ایران را تحت سناریوهای مختلف قیمت حاملهای انرژی است بررسی کردند. عباسیان و اسدی بگی (۱۳۹۲) اثر هدفمندسازی یارانه‌های انرژی با رفاه اجتماعی را با مدل‌بندی ارزیابی نمودند. خسروی نژاد و همکاران (۱۳۹۲) از نظریه تقاضا در برآورد معادلات تقاضا استفاده کردند و معیار تغییر جبرانی جهت ارزیابی تغییر رفاه مصرف‌کنندگان را ارائه نمودند.

عموما مدل‌بندیهایی که برای مطالعه یارانه انجام شده است و بخشی از آنها در مطالعات مروری ارائه شده است، به‌صورت کلان داده بوده است در صورتی که در این مقاله بر اساس ریز داده‌ها (در سطح خانوار) مدل‌بندی انجام شده است.

هدف این مطالعه، مدل‌بندی سهم درآمد خانور از یارانه است. با توجه به این که متغیر پاسخ از جنس نسبت است و مقادیر آن در بازه (۱و۰) تغییر می‌کند، استفاده از مدلهای رگرسیونی استاندار غیرواقع‌گرایانه است چون وقتی حوزه مقادیر متغیر پاسخ بازه (۱و۰) باشد استفاده از این مدلها ممکن است منجر

به پیشگویی مقادیری خارج از این بازه شود. یک رهیافت برای حل این مسأله استفاده از تبدیلهایی مناسب بر متغیر پاسخ است، به گونه‌ای که مقادیر آن در مجموعه اعداد حقیقی واقع شوند و سپس مدلها بر داده‌های تبدیل یافته برازش شوند، که در این‌صورت تفسیر پارامترهای مدل دشوار است(فراری و کریباری، ۲۰۰٤).

مرور منابع موجود در حوزه مدلهای آماری نشان داد که در مدل‌بندی متغیرهای پاسخ که روی بازه (۱و۰) تعریف شده‌اند، برای اجتناب از اعمال تبدیل به داده‌های اولیه و با در نظر گرفتن خواص انعطاف‌پذیری توزیع بتا[1]، استفاده از مدل رگرسیون بتا[2] مناسب است. در این مدل متغیر پاسخ دارای توزیع بتای بازپارامتریده[3] است که شامل دو پارامتر میانگین و دقت[4] است که می‌توانند ثابت یا متغیر باشند (فراری و کریباری[5]، ۲۰۰٤).

علاوه بر آن باید دقت داشت که برخی از خانوارها یارانه نقدی دریافت نمی‌کنند یا سهم بسیار بالایی از درآمد آنها از محل یارانه نقدی است، بنابراین مقدار متغیر پاسخ برای آنها به‌ترتیب برابر با صفر و یک است. در این شرایط تکیه‌گاه توزیع بتا نمی‌تواند داده‌ها را به‌طور کامل پوشش دهد، بنابراین لازم است توزیعی در نظر گرفته شود که علاوه بر (۱و۰) برای مقادیر {۱و۰} نیز قابل تعریف باشد. در مواجهه با قرارگرفتن نسبتها در بازه [۱و۰] مدل بتای افزوده[6] صفر و یک پیشنهاد شده است (گالویس[7] و همکاران، ۲۰۱٤). این پیشنهاد شامل یک توزیع سه قسمتی است به‌گونه‌ای که در نقاط صفر و یک تباهیده[8]

1. beta distribution
2. beta regression
3. reparametrized beta distribution
4. precision parameter
5. Ferrari and Cribari
6. augmented beta model
7. Galvis
8. degenerate

و در بازه (۱و۰) دارای چگالی بتا است. این مدل را مدل رگرسیون بتای افزوده گویند که برای مدل‌بندی مشاهداتی به کار می‌رود که از واحدهای مستقل به دست آمده‌اند.

با تغییر ساختار داده‌ها، ممکن است وابستگی بخشی از ماهیت داده‌ها باشد و لازم است تمهیدات لازم برای مدل‌بندی آنها اتخاذ شود(فیگورا[1] و همکاران، ۲۰۱۳). یکی از اهداف این مطالعه بررسی اثر محل سکونت بر سهم درآمد خانوار از یارانه است، این وابستگی در صورت وجود، ناشی از موقعیت جغرافیایی قرار گرفتن واحدهای نمونه است. بنابراین اثر فضایی[2] نیز باید در مدل وارد شود که برای این منظور از مدل (کپدا[3] و همکاران، ۲۰۱۲) استفاده شده است.

هدف از این مطالعه بررسی معنادار بودن متغیرهای «شرایط اقلیمی»، «بعد خانوار»، «محل سکونت» و «میزان درآمد» بر سهم درآمد خانوارها با استفاده از یک مدل آماری مناسب است. سهم درآمد خانوارها از یارانه با استفاده از اطلاعات طرح آمارگیری از هزینه و درآمد خانوار محاسبه شد. لازم به ذکر است که در این مطالعه خانوارهای شهر تهران به‌عنوان جامعه هدف در نظر گرفته شدند. با توجه به نوع داده‌ها مدل رگرسیون بتای افزوده با اثرات فضایی برای تحلیل سهم درآمد خانوار از یارانه پیشنهاد شده است که یک رویکرد جدید برای بررسی این موضوع است. همچنین برازش مدل با رهیافت بیزی انجام شده و یافته‌های حاصل از آن ارائه شده است. نتایج حاصل از برازش مدل، بیان‌گر آن است که متغیرهای دهک درآمدی، بعد خانوار و محل سکونت خانوار بر سهم درآمد خانوار از یارانه نقدی به صورت معنادار اثر دارد. در بخش پایانی

1. Figueroa
2. spatial effect
3. Cepeda

بحث و نتیجه‌گیری ارائه شده است.

روش

در مسائل کاربردی معمولاً برای شناخت ارتباط میان متغیرهای مستقل و متغیر پاسخ و شکل این روابط از مدلهای رگرسیونی استفاده می‌شود. برازش این مدلها با فرض نرمال بودن متغیر پاسخ یا تبدیلی از آن، ثابت بودن واریانس متغیر پاسخ و ناهمبسته بودن مؤلفه‌های خطا انجام می‌شود. حال اگر متغیر پاسخ به بازه (۱و۰) محدود شود، استفاده از مدل رگرسیونی، منجر به پیش‌گوییهایی خارج از بازه تعریف شده خواهد شد. در این استفاده از نوعی مدل گرسیونی که در آن متغیر پاسخ دارای توزیع بتا باشد پیشنهاد شده است (پائولینو[1]، ۲۰۰۱). برای مدل‌بندی متغیرهای پاسخ از جنس نسبت، فرض می‌شود که متغیر پاسخ از توزیع بتا با تابع چگالی

$$\pi(y;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}y^{\alpha-1}(1-y)^{\beta-1} \quad 0<y<1\ ,\ \alpha,\beta>0$$

پیروی می‌کند، که میانگین و واریانس آن به ترتیب عبارتند از؛

$$E(Y)=\frac{\alpha}{\alpha+\beta} \quad \text{و} \quad Var(Y)=\frac{\alpha\beta}{(\alpha+\beta)^2\times(\alpha+\beta+1)}$$

با توجه به این که در مدلهای رگرسیونی، الگوی رفتار میانگین متغیر پاسخ

1. Paolino

مشروط بر متغیرهای تبیینی مورد بررسی قرار می‌گیرد، توزیع بتای بازپارامتریده برای مطالعه نسبتها پیشنهاد شد(فراری و کریباری، ۲۰۰٤). آنها پارامترهای توزیع بتا را به‌گونه‌ای بازنویسی کردند که مدل رگرسیونی براساس میانگین متغیر پاسخ معین شود. برای این منظور با قرار دادن $\phi = \alpha + \beta$ و $\mu = \frac{\alpha}{\alpha+\beta}$ توزیع بتا به صورت:

$$\pi(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \qquad 0 < y < 1$$

بازپارامتریده می‌شود، که در آنان μ میانگین و ϕ پارامتر دقت است. آنها با در نظر گرفتن الگویی خطی برای متغیرهای تبیینی به برازش مدل پرداختند. در این مدلها پارامتر میانگین از طریق تابع پیوند[1] مناسب به صورت

$$g(\mu_i) = \sum_{i=1}^{k} x_i \beta_i = \eta_i$$

به متغیرهای تبیینی مربوط می‌شوند.

برای داده‌های مشاهده شده در بازه بسته [۱و۰] یک توزیع آمیخته سه قسمتی شامل دو توزیع تباهیده در صفر و یک، و یک توزیع بتا در بازهٔ (۱و۰) به صورت:

$$f(y|p_0,p_1,\mu,\phi) = \begin{cases} p_0 & y = 0 \\ p_1 & y = 1 \\ (1-p_0-p_1)h(y|\mu,\phi) & 0 < y < 1 \end{cases}$$

1. Link Function

پیشنهاد شده است، که در آن $h(y|\mu,\phi)$ تابع چگالی بتا و $p_0,p_1\geq 0$ و $0\leq p_0+p_1\leq 1$ هستند(گالویس و همکاران، ۲۰۱٤).

میانگین این توزیع $E(Y)=(1-p_0-p_1)\mu+p_1$ و واریانس آن:

$$Var(Y) = p_1(1-p_1) + (1-p_0-p_1)[\frac{\mu(1-\mu)}{1+\phi} + (p_0+p_1)\mu^2 - 2\mu p_1]$$

در بسیاری از زمینه‌های کاربردی با داده‌هایی مواجه می‌شویم که برحسب موقعیت قرار گرفتنشان به یکدیگر وابسته‌اند، این نوع وابستگی را وابستگی فضایی گویند. برای مطالعه اثرات فضایی در مدل رگرسیون بتا، مدل‌بندی همزمان پارامترهای میانگین و دقت با در نظر گرفتن یک الگوی همبستگی برای لوجیت میانگین و لگاریتم پارامتر دقت پیشنهاد شده است(کپدا[1] و همکاران، ۲۰۱۲). که در آن اثر فضایی با استفاده از حاصل ضرب ماتریس وزن در متغیر پاسخ به عنوان یک متغیر تبیینی در مدل اعمال شد. این مدل به صورت زیر ارایه شده است:

$$\text{logit}(\mu) = x\beta + \rho Wy$$
$$\log(\Phi) = z\delta + \lambda Wy$$

که در آن ρ و λ ضرایب رگرسیونی و W ماتریس وزن فضایی و Y بردار متغیر پاسخ است. برآورد پارامترهای مجهول با استفاده از رهیافت بیزی به دست می‌آیند. درصورتی که پارامتر دقت ثابت باشد، تنها پارامتر میانگین با استفاده از الگوی فضایی فوق مدل‌بندی می‌شود.

1. Cepeda

برای ارزیابی و انتخاب مدل مناسب از میان چند مدل رقیب، ملاک اطلاع انحراف[1] (DIC) استفاده می‌شود (اشپیگل هالتر[2] و همکاران، ۲۰۰۲). DIC اندازه‌ای از پیچیدگی و برازش مدل است. اگر ملاک انحراف بیزی بر اساس لگاریتم درست‌نمایی به‌صورت:

$$D(X,\theta)=-2\sum_i log\{\pi(y_i \mid x_i,\theta)\}+c$$

در نظر گرفته شود در آن c مقداری ثابت است،، ملاک DIC به صورت زیر تعریف می‌شود:

$$DIC=\bar{D}+p_D$$

که در آن میانگین پسین انحرافها و p_D تفاضل میانگین انحرافها و انحراف میانگینها است. مقدار کمتر آن برازش بهتر مدل را نشان می‌دهد.

$$\bar{D}=E_{\theta \mid y}(D)$$

۳- برازش مدل و یافته‌های تحقیق

حدود پنجاه سال است که آمارگیری از هزینه و درآمد خانوار توسط مرکز آمار ایران اجرا می‌شود. این آمارگیری علاوه بر هزینه، درآمد خانوارها را نیز شامل می‌شود و همه ساله انجام و نتایج آن استخراج و منتشر شده است. هدف اصلی آمارگیری از هزینه و درآمد خانوار اندازه‌گیری متوسط هزینه خوراک، غیر خوراک و کل هزینه‌های خانوارهای شهری و روستایی در ایران است. اطلاعات به‌دست آمده از این آمارگیری برای محاسبه خط فقر و مطالعه عدم همگونی

1. Deviance Information Criteria
2. Spiegelhalter

در تسهیلات و درآمد خانوار به‌کار می‌رود. جامعهٔ آماری در این آمارگیری، تمام خانوارهایی است که ساکن نقاط شهری و روستایی هستند.

اگر درآمد خانوار از یارانه کمتر از کل درآمد خانوار باشد، این نسبت عددی بین صفر و یک است. اگر کل درآمد خانوار از یارانه باشد، این نسبت یک و در صورتی که درآمد خانوار از یارانه صفر باشد یعنی خانوار یارانه دریافت نکند این سهم مقدار صفر را اختیار می‌کند. بنابراین متغیر پاسخ یعنی سهم درآمد خانوار از یارانه، متغیری پیوسته با تحققهایی در بازهٔ بسته [۱و۰] است. از آن جایی که در نمونه مورد مطالعه خانواری وجود نداشت که یارانه تنها منبع درآمدش باشد، داده‌ها شامل مقدار یک برای سهم درآمد خانوار از یارانه نبود. بنابراین برای مدل‌بندی داده‌ها رگرسیون بتای افزوده در بازه (۱و۰] در نظر گرفته شد.

با فرض ثابت بودن پارامترهای p_0 و ϕ و انتخاب تابع پیوند لوجیت برای میانگین سهم درآمد خانوار از یارانه از مدل رگرسیون بتای آمیخته افزوده برای مدل‌بندی داده‌ها استفاده می‌شود. اگر $x^{(1)}$ دهک وزنی درآمد، $x^{(2)}$ بعد خانوار، $x^{(3)}$ تعداد افراد شاغل خانوار ، $x^{(4)}$ تعداد افراد تحصیل‌کرده خانوار، $x^{(5)}$ سطح زیربنای محل سکونت و $x^{(6)}$ تملک اتومبیل شخصی به عنوان متغیرهای تبینی در این مدل قرار گیرند، لوجیت میانگین سهم درآمد خانوار از یارانه به صورت مدل ۱:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \beta_4 x_i^{(4)} + \beta_5 x_i^{(5)} + \beta_6 x_i^{(6)}$$

در نظر گرفته می‌شود، که در آن β_0 عرض از مبدا و $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ و β_6 ضرایب رگرسیونی مدل هستند. برآورد پارامترها با استفاده از روش بیزی انجام می‌شوند.

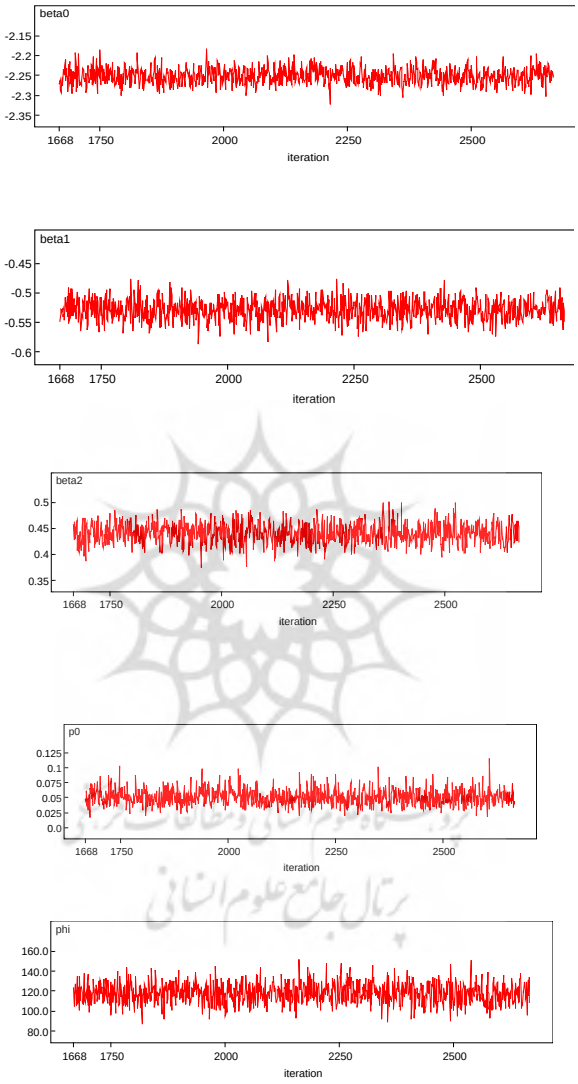
برای ضرایب رگرسیونی، توزیع پیشین نرمال با میانگین صفر و واریانس، (۱۰۰و ۰)N در نظر گرفته می‌شود (گالویس و همکاران، ۲۰۱٤). به دلیل این که اطلاعات قبلی در مورد توزیع پیشین در اختیار نیست، با بزرگ فرض کردن واریانس در واقع یک پیشین ناآگاهی بخش در نظر گرفته شده است. برای پارامتر ϕ توزیع پیشین گاما(۰/۰۱و۰/۰۱)G و برای پارامتر p_0 توزیع پیشین یکنواخت (۰,۱)U در نظر گرفته شده است. از آن جا که به دست آوردن توزیعهای پسینی حاشیه‌ای بسیار پیچیده است از الگوریتم زنجیر مارکوف مونت کارلو (MCMC) و نمونه‌گیر گیبز استفاده می‌شود. نمونه‌های پسینی از توزیعهای شرطی کامل پارامترها به دست می‌آید. محاسبات الگوریتم MCMC شامل نمونه‌گیری گیبز و الگوریتم متروپولیس هستینگ در نرم‌افزار WINBUGS انجام می‌شود، این نرم‌افزار آماری به صورت رایگان از اینترنت قابل نصب و اجرا است.

برآوردهای پارامترها با ۸۰۰۰۰ تکرار مونت کارلو در نظر گرفته شده است که بعد از صرف نظر کردن از ۵۰۰۰۰ نمونه اول، خلاصه‌های پسین به‌دست می‌آید. شکل۱ نمودار اثر۱ و شکل ۲ نمودار خودهمبستگی برآورد پارامترهای مدل در تکرارهای مختلف را نشان می‌دهد.

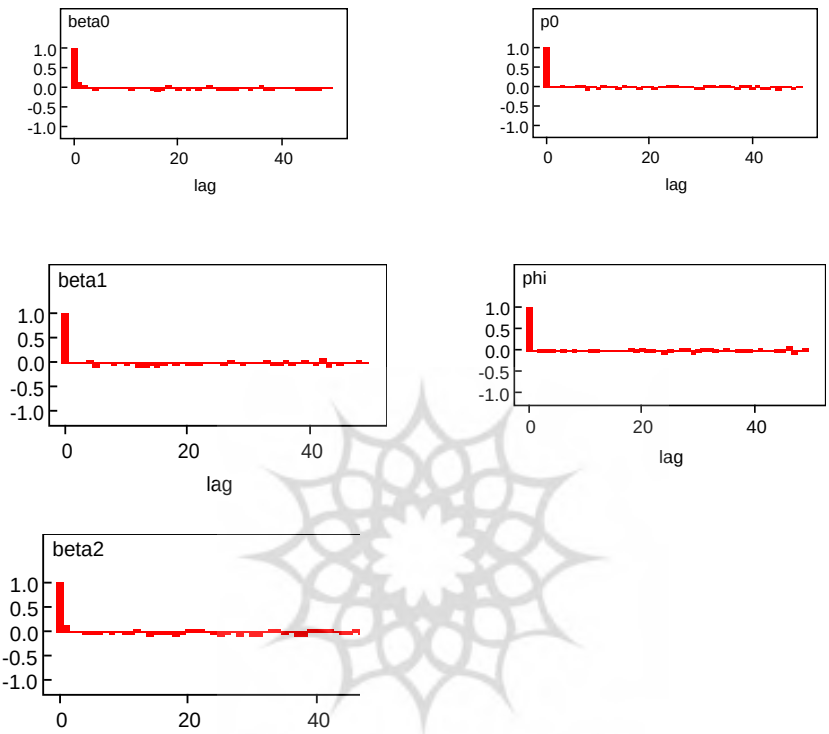
پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Trace

شکل ۱ـ نمودار اثر نمونه‌های تولید شده پارامترهای مدل



شکل ۲ـ نمودار خودهمبستگی نمونه‌های تولید شده پارامترهای مدل



نمودارهـای اثـر و خودهمبسـتگی نشـان می‌دهـد کـه زنجیـر هـر پارامتـر مانـا و ناهمبسـته اسـت کـه ایـن بررسـی بـرای رسـیدن بـه بـرآورد مناسـب پارامترهـای مـدل ضـروری اسـت. دهـک وزنـی درآمـد و بعـد خانـوار متغیرهایـی هسـتند کـه در مـدل وارد می‌شـوند. بنابرایـن مدلـی کـه ضرایـب رگرسـیونی آن بـا اطمینـان ۹۵٪ معنـادار هسـتند به‌صـورت زیـر اسـت. بـرآورد پارامترهـا بـه همـراه انحـراف اسـتاندارد و فاصلـه اطمینـان ۹۵٪ در جـدول ۱ آمـده اسـت.

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1(\text{دهک وزنی درآمد}) + \beta_2(\text{بعد خانوار})$$

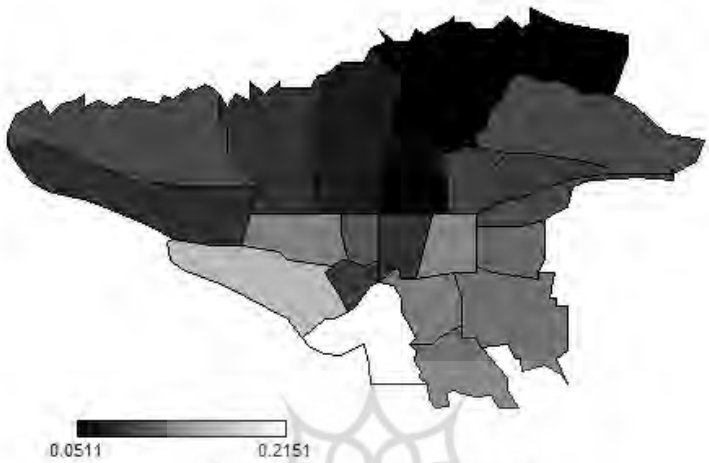
جدول ۱- برآورد پارامترها و انحراف استاندارد آنها به همراه فاصله اطمینان ۹۵٪

پارامتر	برآورد	انحراف استاندارد	فاصله اطمینان ۹۵٪	
β_0	-۲/۲۵	۰/۰۲	-۲/۲۱	-۲/۲٤
β_1	-۰/۵۳	۰/۰۲	-۰/٤۹	-۰/۵٦
β_2	۰/٤٤	۰/۰۲	۰/٤۸	۰/٤۰
p_0	۰/۰۵	۰/۰۱	۰/۰۸	/۰۳
ϕ	۱۱۸/٤	۱۰/۳٦	۱۳۷/۷	۹۷/۳٤

متغیر تبینی دهک وزنی درآمد با اثر منفی و متغیر تبینی بعد خانوار با اثر مثبت وارد مدل رگرسیونی میانگین می‌شوند و عرض از مبدا با اثر منفی وارد مدل می‌شود. نتایج نشان می‌دهد سهم درآمد خانوارها از یارانه برای خانوارهایی که در دهکهای بالایی درآمد قرار دارند، کاهش می‌یابد. همچنین با افزایش بعد خانوار، سهم بیشتری از درآمد خانوار از محل یارانه است. سهم خانوارهایی که رایانه دریافت نمی‌کنند ۵٪ برآورد شده است. همچنین پارامتر دقت مدل برابر با ۱۱۸/٤ برآورد شده است از آن‌جا که پارامتر دقت و واریانس رابطه عکس دارند، بالا بودن برآورد پارامتر دقت نشان‌گر کم بودن واریانس مدل است.

محل سکونت خانوار، متغیر دیگری است که بررسی اثر آن بر متغیر پاسخ در این مطالعه مورد نظر است. شکل ۳، نقشه فضایی سهم درآمد از یارانه به کل درآمد خانوار در مناطق تهران را نشان می‌دهد.

شکل ۳- نقشه فضایی سهم درآمد از یارانه به کل درآمد خانوار در مناطق تهران



شکل ۳ بیان‌گر آن است که سهم درآمد خانوار از یارانه در شهر تهران به‌صورت یکنواخت نیست و به منطقه زندگی خانوار بستگی دارد. برای بررسی معناداری این مشاهدات، از آماره موران[1] که معیاری برای آزمون خودهمبستگی فضایی است، استفاده شده است. P-مقدار به‌دست آمده از این آزمون مقداری نزدیک به صفر را اختیار می‌کند که نشان‌گر وجود همبستگی قوی فضایی متغیر پاسخ است. بنابراین می‌توان گفت معمولاً خانوارهایی که در جوار یکدیگر زندگی می‌کنند از لحاظ شرایط اقتصادی همگن‌تر هستند. در نتیجه مدل رگرسیون بتای افزوده با اثر فضایی برای تحلیل داده‌ها مورد استفاده قرار می‌گیرد. در این مدل علاوه بر متغیرهای تبینی اثر فضایی نیز در نظر گرفته می‌شود.

1. Moran

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1(\text{دهک وزنی درآمد}) + \beta_2(\text{بعد خانوار}) + \rho\left(\text{فضایی اثر}\right)$$

با لحاظ تمام فروض و توزیعهای پیشین، خلاصههای پسین که بیانگر ضرایب مدل رگرسیون بتای افزوده با لحاظ اثرات فضایی است، در جدول ۲ ارائه شده است:

جدول ۲- برآورد پارامترها و انحراف استاندارد آنها به همراه فاصله اطمینان ۹۵٪ در مدل با اثر فضایی

پارامتر	برآورد	انحراف استاندارد	فاصله اطمینان ۹۵٪	
β_0	-۲/۳۵	۰/۰۶	-۲/۲۳	-۲/٤٦
β_1	-۰/۵۳	۰/۰۲	-۰/٤۹	-۰/۵۷
β_2	۰/٤٤	۰/۰۲	۰/٤۸	۰/٤۰
ρ	۰/۰٤	۰/۰۲	۰/۰۹	۰/۰۱
p_0	۰/۰۵	۰/۰۱	۰/۰۸	۰/۰۳
ϕ	۱۱۹/۹	۱۰/۳۹	۱٤۱/۸	۱۰۰/٤

بر اساس نتایج جدول ۲ علاوه بر متغیرهای تبینی دهک وزنی درآمد و بعد خانوار، متغیر تبینی اثر فضایی نیز معنادار شده است. برآورد انحراف استاندار پارامترها در جداول ۱ و ۲ تفاوت محسوسی ندارند برای انتخاب مدل مناسب از ملاک اطلاع انحراف استفاده می‌شود. مقادیر DIC برای مدل۱ و مدل۲ به ترتیب ۵۳۵۲۷/٤ و ۵۳۵۲۵/۲ است که نشان می‌دهد مدل۲ از مدل۱ بهتر است، یعنی وارد کردن اثر فضایی مدل را بهبود می‌بخشد.

بحث

قانون هدفمند کردن یارانه‌ها و چگونگی اجرای آن به یکی از چالشهای اقتصاد ایران تبدیل شده است. دریافت یارانه در خانوارهای کم درآمد می‌تواند به بهبود وضعیت اقتصادی آنها کمک کند. از این رو شناسایی خانوارهای پردرآمد و حذف آنها از فهرست یارانه‌بگیران با هدف کاهش شکاف طبقاتی و برقراری توزیع عادلانه درآمد اهمیت دارد. تعیین متغیرهایی که از طریق آنها بتوان خانوارهای پردرآمد را مشخص کرد دشوار است، کارشناسان اقتصادی چهار معیار «شرایط اقلیمی»، «بعد خانوار»، «محل سکونت» و «میزان درآمد» را معیارهای عملی برای شناسایی این خانوارها معرفی کردند. در این مقاله با استفاده از اطلاعات طرح آمارگیری هزینه و درآمد سال ۱۳۹۰ در شهر تهران، سهم درآمد خانوار از یارانه با به کارگیری مدل رگرسیون بتای افزوده با اثرات فضایی بررسی شد. نتایج نشان‌دهنده آن است که دهک درآمدی خانوار، بعد خانوار و محل سکونت خانوار بر سهم درآمد خانوار از یارانه به‌صورت معنادار اثر دارد که با ملاکهای ارائه شده توسط کارشناسان اقتصادی مطابقت دارد. دسترسی به اطلاعات محل سکونت خانوار در مقایسه با دو معیار دیگر یعنی میزان درآمد و بعد خانوار ساده‌تر است. همچنین انتظار می‌رود میزان خطای آن در مقایسه با اطلاعات مربوط به درآمد کمتر باشد. از این رو پیشنهاد می‌شود محل سکونت خانوار به لحاظ جغرافیایی، در گروه‌بندی خانوارها به منظور تصمیم‌گیری در خصوص دریافت یارانه نیز مورد توجه قرار بگیرد. انتظار می‌رود این اطلاعات از پایگاه اطلاعات ثبتی سازمان اسناد و املاک کشور قابل حصول باشد، با استفاده از این اطلاعات علاوه بر منطقه محل سکونت، وضعیت تملک و متراژ

واحد مسکونی نیز مشخص می‌شود. متغیر دیگری که به صورت معنادار بر سهم درآمد خانوار از یارانه معنادار شد، دهک درآمدی است که می‌بایست برای شناسایی این متغیر از اطلاعات کمکی استفاده شود. پیشنهاد می‌شود از تعداد و قیمت خودوری خانوار و همچنین دارا بودن جواز کسب یا متراژ واحد مسکونی به عنوان ملاک شناسایی خانوارهای پر درآمد استفاده شود، زیرا این اطلاع در سازمان ثبت اسناد و املاک موجود است.

علاوه بر متغیرهای فوق‌الذکر بعد خانوار (تعداد اعضای خانوار) نیز بر سهم درآمد خانوار از یارانه اثر دارد زیرا با افزایش تعداد اعضای خانوار سهم درآمد خانوار از یارانه کاهش می‌یابد، این اطلاع با استفاده از سرشماری نفوس و مسکن به‌هنگام می‌شود. بنابراین برای برنامه‌ریزی درخصوص تخصیص یارانه نقدی به خانوارها، ضرورت دارد که این متغیر همراه با سایر متغیرهای شناسایی شده در مدل یعنی دهک درآمدی و محل سکونت خانوار مورد توجه قرار بگیرد تا سرانه درآمد خانوار از یارانه ملاک تصمیم‌گیری قرار بگیرد. با هماهنگی سازمانهای متولی این پایگاهها به منظور اتصال این اطلاعات و با به‌کارگیری مدل پیشنهادی در این مطالعه، شناسایی خانوارهای پردرآمد امکان‌پذیر است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

- ***آمارگیری از هزینه و درآمد خانوارهای شهری***، تهران، مرکز آمار ایران، (۱۳۹۰)
- احمدی، سید مهدی، پژویان، جمشید، غلامی، الهام (۱۳۹۳) هدفمند کردن قیمت حامل های انرژی و رفتار مصرفی خانوارهای شهری، دوره ۸، شماره ۲۷.
- ***بررسی آثار پرداخت یارانه نقدی و حذف یارانه سه دهک خانوارها بر تولید وتوزیع درآمد نهادی*** (۱۳۹۳) دفتر مطالعات اقتصادی مجلس شورای اسلامی.
- پیرایی، خسرو، سیف، سیدبهروز (۱۳۸۹) تأثیر هدفمندسازی یارانه ها بر رفاه اجتماعی در ایران، ***پژوهشنامه مالیات***، شماره ۹.
- خسروی نژاد، علی اکبر، خداداد کاشی، فرهاد، صحبتی، زهرا (۱۳۹۲) ارزیابی افزایش قیمت مواد غذایی بر رفاه خانوارهای شهری ایران، ***فصلنامه راهبرد اقتصادی***.
- کلانتری، عباس (۱۳۷۵) ***بررسی اقتصادی یارانه***، موسسه مطالعات و پژوهش های بازرگانی، چاپ سوم، تهران.
- فیروز نسیبی‌پور آذر (۱۳۶۹) ***بررسی انواع یارانه و جنبه‌های مختلف اقتصادی آنها***، مؤسسات مطالعه و پژوهشهای بازرگانی در گروه تحقیقات بازرگانی.
- عباسیان، عزت اله ، اسدبگی، زهرا (۱۳۹۲) ارتباط هدفمندسازی یارانه های انرژی با رفاه اجتماعی از مسیر رشد اقتصادی، ***فصلنامهٔ علمی پژوهشی رفاه اجتماعی***، سال دوازدهم، شماره ٤٤.
- Azarbaiejani, K., Hamedani, A. S., & Bahaloo, M. (2012). An analysis of the effect of energy subsidies decrease and transfer payments to households on social welfare in Iran. ***Interdiscipl J Contemp Res Bus,*** 4(1), 478.
- Cepeda, E. ,.D Urdinola, B. P. and Rodriguez,.D (2012), Double Generalized Spatial Econometric Models, ***Communications in Statistics-Simulation and Computation*** ,41, 671–685.
- Ferrari, S. and Cribari, F. (2004), Beta Regression for ModellingR ates and Proportions, ***Journal of Applied Statistics***, 31, 799-815.
- Figueroa-Zúñiga, J. I., Arellano-Valle, R. B. and Ferrari, S. L. (2013), Mixed Beta Regression: A Bayesian Perspective, ***Computational Statistics & Data Analysis***, 61, 137– 147.
- Galvis, M. D., Dipankar, B. and Victor, H. L. (2014), ***Augmented Mixed Beta Regression Models for Periodontal Proportion Data***, Preprinted, (In Press), Statistics in Medicine.

- Karami, A,. esmaili, A,. najafi, B. 2012. Assessing effects of alternative food subsidy reform in Iran, ***Journal of Policy Modeling***, Volume 34, Pages 788–799

- Paolino, P. (2001), Maximum Likelihood E stimation of Models with Beta-Distributed Dependent Variables, ***Political Analysis***, 9, 325-346.

- Saboohi.Y (2001), "An evaluation of the impact of reducing energy subsidies on living expenses of households", ***Journal of Energy policy***: (29) 245-252

- Spiegelhalter,D. J., Best,N.G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). J. R. ***Statist***. Soc. B, 64, 583–639.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی