فصلنامه پژوهش‌های اقتصادی ایران/ سال دهم/ شماره ۳۴/ بهار ۱۳۸۷/ صفحات ۱-۲۹

ارزیابی آثار مسایل انتخاب و درونزایی بر بازده آموزش: یک تحلیل تطبیقی

دکتر ابوالقاسم نادری*

تاریخ ارسال:۸۶/۱۰/۴ تاریخ پذیرش: ۸۷/۲/۱۷

چکیده

هدف اصلی این پژوهش، تحلیل و ارزیابی آثار "مسایل انتخاب و درونزایی" بر بازده آموزش است که از تصمیم افراد برای مشارکت در بازار کار و کسب تحصیلات رسمی ناشی می‌شود. برای این منظور، از تحلیل‌های آماری متناسب یعنی رگرسیون شرطی (برای مسأله انتخاب) و فن متغیر ابزاری (برای مسأله درونزایی)، و از داده‌های استخراج‌شده از آمارگیری ویژگی‌های اقتصادی و اجتماعی خانوار استان تهران (سال ۱۳۸۰) و سرشماری جمعیت سال ۱۹۹۱ آمریکا استفاده کرده‌ایم. یافته‌ها نشان می‌دهند که: ۱) آموزش یک سرمایه‌گذاری پُر بازده است، ۲) لحاظ‌کردن مسایل انتخاب و درونزایی در تحلیل‌های آماری، نتایج تجربی بازده آموزش را متأثر کرده و آن را به طور قابل ملاحظه‌ای افزایش می‌دهد و ۳) نتایج نمونهٔ استان تهران به طور کامل با یافته‌های حاصل از نمونهٔ جامعه آمریکا منطبق است. بدین ترتیب، الگوواره یافته‌ها، مشابه و مطابق با تحلیل‌های نظری است؛ یعنی برآوردهای تجربی بازده آموزش- که مسایل انتخاب و درونزایی را نادیده گرفته باشند- تورش‌دار و ناسازگارند.

طبقه‌بندی JEL: C3, I2, J2, J3

واژگان کلیدی: سرمایه‌گذاری آموزشی، بازده آموزش، مسأله انتخاب، مسأله درونزایی، رگرسیون شرطی، متغیر ابزاری.

---

* استادیار دانشگاه تهران

e-mail: anadery@ut.ac.ir

مقدمه

مطالعهٔ رفتار تخصیص منابع افراد، موضوع اصلی علم اقتصاد است. در چند دههٔ گذشته، آموزش جایگاه ویژه‌ای در تصمیم‌های تخصیص منابع افراد پیدا کرده است؛ اقتصاددانان همواره به دنبال یافتن پاسخ مناسبی بوده‌اند که تبیین‌کننده رفتار مزبور باشد. در این رابطه، دو نوع بحث یعنی مبانی نظری و مبانی تحلیلی (روش‌شناسی)مرتبط هستند که به کمک آنها، امکان ارائه پاسخ مناسب فراهم می‌شود. در حوزه مبانی نظری، وجه سرمایه‌گذاری بودن آموزش به عنوان پاسخ مناسب مقبولیت‌یافته و در حوزه مبانی روش‌شناسی، تحلیل هزینه- فایده به عنوان روش مناسب به کارگرفته شده است؛ اما فرایند عملی انجام تحلیل هزینه-فایده (و ارزیابی بازده آموزش)، موضوع انتقادهای گسترده‌ای بوده است. یکی از مهم‌ترین و اساسی‌ترین انتقادها که متوجه تحلیل‌های مرسوم ارزیابی نرخ بازده بوده، از ناحیه مسایل انتخاب و درونزایی است که نتایج تحلیل‌ها یعنی ارزیابی بازده آموزش را به شدت متأثر می‌کند.

مسایل "انتخاب و درونزایی"،[1] فرایند تصمیم افراد در خصوص سرمایه‌گذاری آموزشی و تأثیر آن بر تحلیل‌های آماری و اقتصادسنجی ارزیابی بازده را مورد توجه قرار می‌دهد. بحث اصلی در مسالهٔ انتخاب این است که فرایند آموزش و ادامه تحصیل به‌گونه‌ای است که تنها عده خاصی از افراد قادر به ادامه تحصیل هستند؛ مهم‌تر اینکه افراد دارای تحصیلات رسمی، نمونه‌ای تصادفی از کل جامعه نیز نبوده، بلکه در یک فرایند خود- انتخابی تصمیم‌گرفته‌اند تا در فرایند آموزش رسمی، ادامه تحصیل دهند. برای تبیین مسألهٔ خود- انتخابی در چارچوب مبانی علم اقتصاد، تحلیل منطقی این است که تصمیمات یا رفتار تخصیص منابع افراد مبتنی بر محاسبات عقلایی است؛ آنهایی که مزیت نسبی‌شان را از طریق تحصیلات رسمی تشخیص داده‌اند، مسیر تحصیل را و در مقابل، افرادی که مزیت نسبی خود را در ادامه ندادن تحصیلات تشخیص داده‌اند، ورود به بازار کار را انتخاب کرده‌اند. این فرایند خود- انتخابی، ماهیت تصادفی بودن ملاحظات و مفروضات تحلیل‌های مرسوم آماری و اقتصادسنجی را-که مسألهٔ انتخاب را نادیده می‌گیرند- به شدّت تحت‌تأثیر قرار می‌دهد که در نتیجه آن، نتایج تحلیل‌های آماری متکی به روش حداقل مربعات از لحاظ سازگاری و تورش متأثر می‌سازد.

برای مقابله با آثار مسایل انتخاب و درونزایی، رویکرد مبانی آمار و اقتصادسنجی (با بهره‌گیری از رگرسیون شرطی و متغیر ابزاری) به عنوان یکی از رویکردهای مناسب[2] به کارگرفته شده است. این

1. Selectivity & Endogeneity

۲.شایان توجه است که استفاده از فنون آمار و اقتصادسنجی تنها رویکرد مقابله با مسألهٔ انتخاب نیست؛ طراحی و اجرای پژوهش‌های آزمایشگاهی (Experimental Research) تصادفی نیز به عنوان یک رویکرد اساسی و مهم قابل طرح است. اما در حوزه مطالعات علوم انسانی و اجتماعی، امکان طراحی و اجرای پژوهش‌های آزمایشگاهی به‌راحتی مقدور نبوده و بخش اصلی مطالعات، غیرآزمایشگاهی است. مهم‌تر اینکه داده‌ها و اطلاعات آماری مورد استفاده در پژوهش‌های ارزیابی بازده آموزش، به طور کلی برگرفته از شواهد واقعی و محقق‌شده (Ex Post) است که حاصل یک فرایند غیرآزمایشگاهی و مستخرج از شرایط مبتلا به مسأله "خود- انتخابی" است. از همین رو،

رویکرد به خصوص از اوایل دهه ۱۹۷۰ میلادی موضوع پژوهش‌های متعددی در خارج از کشور بوده[۱] اما تا جایی‌که نگارنده اطلاع دارد، هیچ پژوهشی برای اقتصاد ایران در این خصوص انجام نشده است.[۲] مهم‌تر اینکه، در سطح مطالعات بین‌المللی نیز آثار این مسأله بر نتایج تحلیل‌های تجربی به صورت مقایسه‌ای و تطبیقی مورد ارزیابی قرار نگرفته است. از همین رو، در این پژوهش، تأثیر مسایل انتخاب و درونزایی بر بازده آموزش در ایران ارزیابی و الگوواره یافته‌های ایران با یافته‌های حاصل از یک نمونه مستقل از جامعه آمریکا به طور تطبیقی تحلیل می‌شوند.

در ادامه مباحث، مبانی نظری و روش‌شناسی مسایل انتخاب و درونزایی برای ارزیابی بازده آموزش (از منظر افراد) تبیین و ارائه می‌شود. سپس، داده‌های آماری استخراج‌شده از آمارگیری‌های مرکز آمار ایران (آمارگیری از ویژگی‌های اقتصادی- اجتماعی خانوار، سال ۱۳۸۰، استان تهران) و نیز نمونه آماری از جامعه آمریکا (به‌دست آمده از سرشماری جمعیت سال ۱۹۹۱ آمریکا) تشریح و توصیف می‌شود. در پایان، تحلیل‌های آماری برای دو جامعه ایران و آمریکا به طور تطبیقی ارائه و یافته‌های حاصله مورد بحث و تفسیر قرار می‌گیرند.

استفاده از مبانی آمار و اقتصادسنجی برای رفع این مشکل به عنوان رویکرد مناسب پیش‌روی تحلیل‌گران قرار دارد.

۱. مسأله خود- انتخابی ابتدا توسط توبین (Tobin, 1958) در زمینه انتخاب شغل و حرفه افراد مورد توجه قرار گرفت، اما از دهه ۱۹۷۰ به عنوان یک جریان اصلی در حوزه‌های مختلف کاربرد الگوهای اقتصادسنجی از جمله ارزیابی بازده آموزش، به کارگرفته شده است. برخی از پژوهش‌های برجسته در خصوص ارزیابی بازده آموزش که آثار خود- انتخابی و درونزایی را مورد مطالعه و ارزشیابی قرار داده‌اند، عبارتند از: گرونا (Gronau, 1974)، هازمن و وایز (Hausman &Wise, 1977)، هکمن (Heckman, 1979)، ویلیس و روزن (Willis & Rosen, 1979)، کنی و دیگران (Kenny, *et. al.*,1979)، فرازیس (Frazis, 1993)، کارد (Card, 2001)، بلک و اسمیت (Black & Smith, 2004)، هارمون و دیگران (Harmon, *et. al.*, 2000)، مورتی (Moretti, 2004)، هامرمش و دونالد (Hamermesh & Donald, 2004)، هکمن و دیگران (Heckman, *et. al.*, 2005) و بلزیل (Belzil, 2007).

۲. به طور کلی، مطالعات انجام شده برای ارزیابی بازده آموزش در ایران از روش‌های ساده تحلیل هزینه- فایده استفاده کرده‌اند (نظیر مطالعات رضوی، ۱۳۵۷؛ پورحسین، ۱۹۷۹، به نقل از: ساخاروپولوس، ۱۹۸۱؛ افشاری، ۱۳۷۷؛ و صالحی، ۱۳۸۴) که ملاحظات روش‌شناسی کمتر مورد توجه قرار گرفته است. البته، مطالعات نادری (۱۳۸۰، ۱۳۸۳) و نادری و میس (Naderi & Mace, 2003) استثناهایی هستند که محور اصلی آنها، ملاحظات روش‌شناسی است، اما این ملاحظات به ارزشیابی آثار "گروه‌بندی‌ها و ساختار سلسله‌مراتبی" بر برآوردهای بازده آموزش محدود می‌شود.

۱. مبانی و الگوی نظری

با پیدایش اقتصاد آموزش، نظریه سرمایه انسانی به عنوان نظریه غالب مطرح شد؛ بر مبنای این نظریه، آموزش یک سرمایه‌گذاری مولّد است که برای افراد بازده قابل‌توجهی به همراه دارد. به بیان دیگر، افراد به منظور تأمین منافع آتی از طریق آموزش بر روی خود سرمایه‌گذاری می‌کنند. از همین رو، انتظار می‌رود که افراد دارای آموزش و تحصیلات بالاتر، از درآمد و دریافتی بیشتری نیز برخوردار باشند و از طریق آن بازده قابل توجهی نصیب‌شان شود.[۱] برای ارزیابی توان پیش‌بینی و توضیح‌دهندگی نظریه سرمایه انسانی باید مبانی تحلیلی و روش‌شناسی مناسبی به کارگرفته شود. از همین رو، انتخاب نوع روش ارزیابی بازده آموزش- که به نتایج دقیق و معتبر بینجامد- اهمیت اساسی خواهد داشت. یکی از روش‌های مرسوم، الگوی دریافتی مینسر است[۲] که به عنوان روش میان‌بُر مشهورشده و در سه دهه گذشته به طور گسترده‌ای در تحقیقات تجربی به کارگرفته شده است.

الگوی دریافتی مینسر یک تابع نیمه‌لگاریتمی است و رویکرد کلی آن مبتنی بر مبانی تحلیل هزینه-فایده است؛ یعنی جریان‌های منافع و هزینه‌های آموزش مورد مقایسه و سودمندی یا بازدهی آموزش مورد ارزشیابی قرار می‌گیرد. این ارزشیابی در مناسبات علمی و شرایط ایده‌آل (دست‌کم به لحاظ داده‌های آماری مورد نیاز) بایستی به طور معقولی، مبانی تصمیمات سرمایه‌گذاری آموزشی افراد

---

۱. یک نکته مهم و اساسی، توجه به این پرسش است که چرا آموزش به افزایش دریافتی (درآمد) افراد منجر می‌شود. به بیان دیگر، چرا کارآفرینان اقتصادی و کارفرمایان مایل هستند به افراد دارای تخصص و تحصیلات بالاتر، حقوق و دستمزد بیشتری پرداخت نمایند؟ در چارچوب نظریه سرمایه انسانی، آموزش بهره‌وری افراد را افزایش می‌دهد؛ از همین رو، افراد دارای آموزش بیشتر، بهره‌وری بیشتری داشته و لذا بنگاه‌ها و سازمان‌ها مایلند به افراد با تحصیلات بالاتر حقوق و دستمزد بیشتری بپردازند. البته، وجه بهره‌ورزایی آموزش (از ناحیه طرفداران نظریه‌های سرند، علامت‌دهنده و فیلتر) مورد مناقشات فکری گسترده بوده است که این وجه در بین عاملان و تصمیم‌گیران، تنها منافع کل جامعه و همچنین بنگاه‌ها و سازمان‌ها را متأثر می‌کند؛ اما از منظر افراد، آموزش چه از طریق ارتقای بهره‌وری و چه با ایفای نقش سرند (فیلتر و علامت‌دهنده)، دارای منافع مورد انتظار است. برای توضیح بیشتر در مورد مبانی نظریه سرمایه انسانی و چالش‌های نظری آن، برای نمونه، رجوع شود به: نفیسی (ویراستار، ۱۳۸۰) و نادری (۱۳۸۳؛ فصول ۳-۵).

۲. مینسر (اقتصاددان لهستانی‌تبار مقیم آمریکا) مطالعات اولیه را در نیمه دوم دهه ۱۹۵۰ میلادی در چارچوب رساله دکتری خود انجام داد و نتایج به‌دست آمده را در سال ۱۹۵۸ منتشر کرد (Mincer, 1958). اما الگوی نهایی وی که به صورت یک تابع نیمه‌لگاریتمی (در بردارنده دو جزء اساسی سرمایه انسانی یعنی سرمایه‌گذاری از طریق آموزش رسمی و از طریق آموزش ضمن-کار) در سال ۱۹۷۴ ارائه شده (Mincer, 1974) و با عنوان **الگوی دریافتی مینسر** (Mincerian Earnings Function) معروف شد. برای توضیح بیشتر، ر.ک. به: هکمن و دیگران (Heckman, *et. al.,* 2003).

را مشخص و تبیین نماید. تابع دریافتی مینسر، ابزار فنی لازم را برای ارزیابی بازده آموزش فراهم‌کرده و دارای دو جزء اساسی سرمایه‌گذاری انسانی یعنی آموزش‌های رسمی و آموزش‌های غیررسمی، است.

همان‌گونه که اشاره شد، مبانی تحلیلی نظریه سرمایه انسانی بیان می‌کند کـه افـراد رفتـار عقلایـی دارند و برای اتخاذ یک تصمیم، فواید و هزینه‌های آن را مقایسه می‌کنند. به طور منطقی، تصمیماتی که دارای منفعت خالص بیشتری باشند، اولویت بیشتری در تخصیص منـابع خواهنـد داشـت. در خـصوص تصمیم فرد برای آموزش رسمی نیز همین قاعده صادق است؛ یعنی یک فرد منفعت خـالص تحـصیلات رسمی را در مقایسه با گزینه‌های دیگر (مانند ورود به بازار کار یا سپرده‌گذاری) ارزیابی کرده و تخصیص منابع در هر یک از گزینه‌ها را تاجایی ادامه خواهد داد که جریان تنزیل‌شده منـافع تـصمیم‌هـا یکـسان باشد. ارزش فعلی خالص تصمیم یا گزینه "بازار کار و تحصیل نکردن"، با این فرض کـه فـرد مـورد نظـر بتواند n سال در بازار کار اشتغال داشته و به طور مستمر و پیوسته کسب درآمد نماید، بـه صـورت زیـر خواهد بود:

$$NPV_{0} = Y_{0}\int_{0}^{n} e^{-rt}\,dt = \frac{Y_{0}}{r}(1 - e^{-rn}) \qquad (۱)$$

که در آن، r و Y_0 به ترتیب، نرخ تنزیل و دریافتی متوسط سالانه است. به همین صورت، گزینه ادامه تحصیل (آموزش رسمی) با فرض اینکه تنها از ناحیه درآمدهای از دست‌رفته هزینه بر فرد تحمیل می‌شود و منافع آموزش فقط از ناحیه دریافتی باشد، دارای ارزش فعلی خالص زیر خواهد بود:

$$NPV_{S} = Y_{S}\int_{S}^{S+n} e^{-rt}\,dt = \frac{Y_{S}}{r}(e^{-rS} - e^{-r(S+n)}) \qquad (۲)$$

که در آن، S و Y_S به ترتیب، میزان تحصیلات و دریافتی متوسط سالانه یک فرد با S سال تحصیل است. مبنای تصمیم‌گیری فرد آن است که $NPV_0 = NPV_S$. بدین ترتیب خواهیم داشت:

$$\begin{aligned} Y_{S} &= Y_{0}e^{rS} \\ \ln(Y_{S}) &= \ln(Y_{0}) + rS \end{aligned} \qquad (۳)$$

رابطه ۳، ارتباط بین سرمایه‌گذاری آموزشی و دریافتی فرد را نشان می‌دهد. این رابطه در واقع همان تابع معروف مینسر برای آموزش رسمی[1] است که تفاوت دریافتی افراد را با اتکا به نظریه سرمایه انسانی تبیین و تحلیل می‌نماید و ضریب متغیر تحصیلات، نرخ بازده سرمایه‌گذاری آموزشی را به دست می‌دهد.

همان‌گونه که مینسر (۱۹۷۴) نیز تأکیدکرده، باید توجه کرد که اولاً بسیاری از عناصر تصمیم‌گیری به طور کامل و دقیق قابل سنجش به ارقام پولی نیستند و ثانیاً منافع منتسب به آموزش (به‌ویژه درآمد) تنها ناشی از سرمایه‌گذاری آموزشی نیست، بلکه عوامل مختلفی این منافع را تحت تأثیر قرار می‌دهند

1. Schooling Earnings Function

که البته برخی از این عوامل (مانند آموزش غیررسمی، یادگیری ضمن‌کار، آموزش و یادگیری در محیط خانواده و جز اینها) جزء عناصر سرمایه انسانی محسوب می‌شوند. در این رابطه، مینسر الگوی خود را به‌گونه‌ای تدوین می‌کند که بتواند نیم‌رخ دریافتی[1] به صورت یک الگوی ریاضی-اقتصادسنجی مشخص نماید و همین نکته باعث نام‌گذاری این تابع به "تابع دریافتی" شده است. بدین ترتیب، مینسر الگوی ۳ را با لحاظ کردن سرمایه‌گذاری‌های انسانی از طریق آموزش و یادگیری ضمن-کار توسعه داده و به صورت زیر ارائه کرده است :[2]

$$\ln(Y_S) = \ln(Y_0) + rS + \beta_1 X + \beta_2 X^2 \qquad (۴)$$

و یا:

$$y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + u_i \qquad (۵)$$

که در آن، y_i لگاریتم دریافتی، S_i تحصیلات رسمی ، X_i سایر عوامل مؤثر بر دریافتی (تجربه کاری)، و u_i جمله خطا برای فرد i است.

این الگو مبنای بسیاری از تحلیل‌های تجربی ارزیابی بازده آموزش بوده که میزان بازده با استفاده از ضریب متغیر S_i قابل استخراج است.[3] پرسش اساسی و قابل طرح این است که با فرض درست "مشخص شدن" قسمت‌های دیگر الگو، آیا این ضریب برآورد مناسب و درستی از نرخ بازده آموزش می‌باشد یا خیر؟ پاسخ درست منوط به ماهیت رفتار افراد تحت مطالعه است؛ اگر وضعیت به‌گونه‌ای درنظرگرفته شود که افراد دارای تحصیلات رسمی، در صورت شرکت نکردن در برنامه‌های آموزش هم دریافتی بیشتری می‌داشتند، پاسخ این پرسش، منفی است؛ یعنی در اینجا مسأله توانمندی‌های ذاتی و تأثیر حذف آن بر جمله اخلال الگو (u_i) و سایر شرایط مفروض، مسأله‌ساز می‌شود. به بیان دیگر، چنانچه رفتار و انتخاب آگاهانه افراد بر مبنای مزیت نسبی‌شان شکل‌گرفته باشد، در چنین شرایطی، با مسایل "انتخاب و درونزایی" مواجه هستیم. بدین روی، مقدار ضریب β_1 متفاوت از میزان واقعی نرخ بازده آموزش شده و لذا برآورد حداقل مربعات با تورش مواجه خواهد بود.

---

1. Earnings Profile

۲. برای پرهیز از طولانی شدن مباحث و به لحاظ اینکه هدف اصلی این نوشته، پرداختن به مسایل روش‌شناسی و اقتصادسنجی ناشی از مسالهٔ خود- انتخابی است، مبانی ریاضی برای بسط الگوی دریافتی مینسر به تفصیل در اینجا ارائه نشده است. برای توضیحات بیشتر در این رابطه، برای نمونه، ر.ک. به: مینسر (Mincer, 1974, 1979) و نادری (۱۳۸۳: ۸۵-۸۷).

۳. شایان توجه است که تابع دریافتی مینسر با مفروضات خاصی بازده آموزش را به دست می‌دهد. با نقض این مفروضات، استفاده از ضریب سال‌های تحصیل به عنوان بازده آموزش مورد مناقشه است. برای توضیح بیشتر در خصوص برخی از نکات اساسی و مرتبط، برای نمونه ر.ک. به: ساخاروپولوس و لیارد (Psacharopoulos & Layard, 1979).

۲. مسایل انتخاب و خود- انتخابی در آموزش

۲-۱. ماهیت و ابعاد مسأله

انتخاب و خود- انتخابی، ابعاد و زوایای متعددی داشته که در چهار دسته به صورت زیر قابل تقسیم هستند:

- انتخاب نمونه آماری[۱] ،
- خود- انتخابی[۲] در زمینه انتخاب آموزش[۳] (نوع، مقطع و مؤسسه آموزشی و به ویژه در تحصیلات عالی)، انتخاب (عدم) ورود به بازار کار، انتخاب نوع شغل و انتخاب محل‌کار یا مؤسسه/ سازمان، انتخاب مهاجرت و محل آن[۴]،
- مسأله درونزایی[۵]،
- تأثیر برنامهٔ عمل[۶].

یکی از علل مهم مسأله انتخاب، به "نمونه‌گیری" و نتیجهٔ کار آمارگیران مربوط می‌شود؛ یعنی زمانی که آمارگیران ناگزیر به انتخاب نمونه هستند و این نمونه، یک نمونه تصادفی و معرف نمایندهٔ

---

1. Sample Selection
2. Self-selection

۳. وجود برخی قوانین و مقررات نظیر قانون آموزش اجباری، انتخاب افراد را تحت تأثیر قرار می‌دهد و در چنین شرایطی، مسأله خودانتخابی کمرنگ می‌شود.

۴. مسأله خود- انتخابی در ابعاد مختلف زندگی انسان‌ها (نظیر انتخاب آموزش، فعالیت و شغل، محل زندگی، همسر، دوست، و همسایه) بروز و ظهور دارد؛ در این مطالعه، تنها به مسایل بازار کار به ویژه آموزش پرداخته شده است.

5. Endogeneity

۶. در زندگی شخصی و اجتماعی دنیای امروز، برنامه‌ها، سیاست‌ها و روش‌های مختلفی به‌کار گرفته می‌شود و در غالب موارد، درجه و میزان اثربخشی آن مورد سنجش و ارزیابی قرار می‌گیرد. برای مثال، سیاستگذاران درصدد آن هستند که تأثیر یک برنامه آموزشی (مانند تحصیلات دانشگاهی، آموزش‌های ضمن کار وجز اینها) بر بهره‌وری یا دریافتی افراد را مورد سنجش و ارزیابی قرار دهند. همچنین، برنامه‌ریزان حوزه سلامت اصولاً تأثیر اجرای برنامه بهداشتی خاصی را بر سلامت و بهداشت جامعه تحلیل و ارزشیابی می‌نمایند. مسئولین آموزش اصولاً روش‌ها و فرایندهای خاص و جدیدی را به‌کارمی‌گیرند و آثار آنها را مطالعه می‌کنند. برای ارزیابی آثار این قبیل برنامه‌ها، سیاست‌ها و روش‌ها، ابزار و فنون مختلفی در اختیار تحلیل‌گران قرار دارد که یکی از ابزارهای مهم و مؤثر در این خصوص، الگوهای آماری است که عمدتاً از فن متغیر مجازی بهره‌گرفته می‌شود. فن متغیر مجازی (Dummy Variable Technique) که یک روش ساده برای ارزیابی‌های مورد بحث می‌باشد، با مشکلات آماری مواجه است که نتیجه تحلیل‌ها را تحت تأثیر قرار می‌دهد. مسایل روش‌شناسی ارزیابی آثار برنامه عمل (Treatment Effect) برای تحصیلات دانشگاهی، به تفصیل در نادری (۱۳۸۷) آمده است.

جامع‌الاطرافی از کل جامعه تحت مطالعه نیست. انتخاب جامعه ثروتمندان یا فقرا، گزینش یک منطقه جغرافیایی خاص، بررسی ویژگی‌های یک جامعه در زمانی خاص و موارد دیگر از مصادیق واقعی چنین انتخاب‌هایی محسوب می‌شوند. با توجه به اینکه یکی از اهداف و کاربردهای اصلی علم آمار، استخراج ویژگی‌های جامعه از ویژگی‌های نمونه است، زمانی این مهم قابل دستیابی است که نمونهٔ مورد نظر از ویژگی‌های یک نمونه تصادفی برخوردار بوده و دارای شرایط توزیعی مشخصی نظیر توزیع "نرمال" باشد؛ در غیراین‌صورت، تحلیل‌های آماری و نتایج آن دقت و اعتبار لازم نخواهد داشت.

مورد دیگر، ناشی از تصمیم و انتخاب موضوع یا فرد مورد مطالعه است. اصولاً زندگی انسان با "انتخاب" همراه است که در ابعاد و حوزه‌های مختلفی مصداق پیدا می‌کند. یک فرد تصمیم می‌گیرد که چه آموزشی را انتخاب کند و تا چه سطح یا مقطعی ادامه تحصیل دهد. انتخاب فرد، حتی نوع مؤسسه آموزشی و چه بسا معلم یا استاد را نیز شامل می‌شود؛ اصولاً افراد به دنبال تحصیل در مؤسسات آموزشی با کیفیت هستند و برای چنین انتخابی حتی حاضرند منابع (مالی) قابل توجهی را اختصاص دهند. به همین طریق، افراد تحت‌تعلیم اصولاً معلمان یا استادان برتر را ترجیح داده و برمی‌گزینند. افزون بر این، انتخاب‌ها پس از پایان آموزش نیز "نتایج و پیامدهای" سرمایه‌گذاری آموزشی را به شدّت متأثر می‌نماید؛ ورود به بازار کار اولین مرحله انتخاب یک فرد تحصیلکرده است. به بیان دیگر، همه افراد تحصیلکرده لزوماً وارد بازار کار نمی‌شوند. از بین افرادی که وارد بازار کار می‌شوند، عده‌ای ممکن است، به طور مثال، یک شغل دولتی را انتخاب نمایند؛ دسته‌ای نیز احتمالاً به جوامع دیگری مهاجرت می‌کنند. در مراحل بعدی، انتخاب سمت شغل، محل کار و حتی مؤسسه یا سازمانی که فرد علاقه‌مند به اشتغال در آن است نیز جزء بسته تصمیم افراد است. همه این موارد، به انتخاب فرد نه به مسایل نمونه‌گیری (که توسط افراد دیگر انجام می‌گیرد) مربوط است؛ از همین رو، عبارت خود- انتخابی به آن گفته می‌شود. نتیجه اساسی چنین انتخاب‌هایی این است که از یک سوی افراد را به سمت بهره‌گیری از مزیت‌های نسبی‌شان هدایت می‌کند و از سوی دیگر، توزیع مشاهدات یا نمونه‌های انتخاب‌شده را به شدّت تحت تأثیر قرار داده و لذا تحلیل‌های آماری و نتایج آن را متأثر می‌سازد.

موضوع دیگری که ارتباط بسیار نزدیکی با مقوله انتخاب دارد، مسأله درونزایی است که در نتیجهٔ آن، ارتباط بین متغیر توضیحی و جمله خطای الگوی آماری غیرصفر است. اصولاً مسأله درونزایی ناشی از حذف یا نادیده‌گرفتن یک متغیر تأثیرگذار (مانند بهره‌هوشی) است که یک یا چند متغیر مستقل (مانند تحصیلات) و همچنین متغیر وابسته (مانند دریافتی) را تحت‌تأثیر قرار می‌دهد. برای نمونه در یک الگوی دریافتی، متغیر آموزش که به عنوان یک متغیر اصلی تلقی می‌شود، در نتیجهٔ فرایند انتخاب یادشده، با جمله خطای الگو ارتباط پیدا می‌کند و لذا یکی از فروض اساسی تحلیل‌های آماری کلاسیک (یعنی $Cov(x,e)=0$) نقض می‌شود؛ تحلیل آثار مسایل انتخاب و درونزایی از منظر اقتصادسنجی در قسمت بعد به تفصیل ارائه می‌شود.

۲-۲. مسألهٔ انتخاب و اقتصادسنجی: جایگاه، آثار و نحوه مقابله با مشکلات آن

مسالهٔ "انتخاب"، نتایج تحلیل‌های آماری متکی به روش حداقل مربعات را از لحاظ سازگاری و تورش متأثر می‌سازد. برای مقابله با آن، به لحاظ ماهیت غیرآزمایشگاهی پژوهش‌های مرتبط و همچنین این نکته که داده‌ها و اطلاعات آماری مورد استفاده در پژوهش‌های اقتصادی به‌طور کلی برگرفته از شواهد واقعی و محقق‌شده (Ex Post) است، از رویکرد مبانی علم آمار و اقتصادسنجی استفاده می‌شود.

از لحاظ مبانی اقتصادسنجی، همان‌گونه که قبل از این اشاره شد، مسألهٔ انتخاب محورها و حوزه‌های متعددی را در بردارد؛ در این قسمت، مسایل روش‌شناسی سه حوزهٔ مشخص یعنی داده‌های بریده، داده‌های سانسورشده، و داده‌های مبتلا به خود- انتخابی مورد بررسی و تحلیل قرار می‌گیرند.

۲-۲-۱. داده‌های بریده[1]

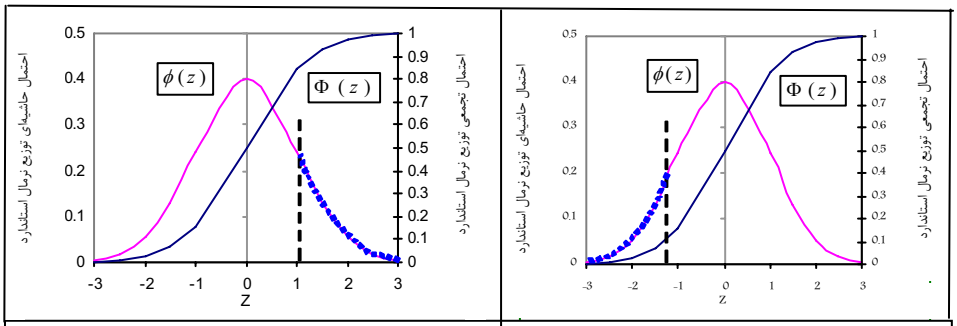
بریدگی و نقصان داده‌ها زمانی رخ می‌دهد که نمونهٔ انتخابی، بخش یا زیرمجموعه‌ای از یک جامعه را شامل شود. به‌طور مثال، چنانچه ویژگی‌های اقتصادی قشر خاصی از خانواده‌ها (مانند وضع درآمد یا مصرف فقرا یا ثروتمندان) در یک جامعه جمع‌آوری شود، چنین داده‌هایی به لحاظ تفاوت توزیع آن با توزیع جامعه مورد هدف، برای مطالعه ویژگی‌های اقتصادی کل جامعه یادشده چندان سودمند نخواهد بود.

بریدگی داده‌ها در ارتباط با توزیع‌های مرسوم به ویژه توزیع نرمال مورد بررسی و تحلیل قرار می‌گیرد که این بریدگی به طور کلی از ناحیه بالا یا پایین "نمودار توزیع نرمال" است. چنانچه هدف مطالعه، بررسی وضعیت درآمد (یا مصرف) فقرا باشد، به‌طور طبیعی تعیین خط فقر یا سقف درآمد مطرح است؛ در این صورت، بریدگی داده‌ها از بالا شکل می‌گیرد. در مقابل، وقتی وضعیت ثروتمندان مطالعه و تحقیق می‌شود، مسأله بریدگی از پایین مطرح است. قسمت "الف" نمودار ۱، وضعیت بریدگی داده‌ها از پایین و قسمت "ب" آن، وضعیت بریدگی داده‌ها از بالا را نشان می‌دهند. بریدگی از بالا، در واقع بیانگر وضعیتی است که مشاهدات و نمونه انتخابی، نشان‌دهنده بخشی از جامعه و قسمت سمت چپ توزیع نرمال را پوشش می‌دهد. در مقابل، بریدگی از پایین، بیانگر وضعی است که توزیع نمونه در قسمت سمت راست توزیع جامعه (نرمال) قرار می‌گیرد. به دلیل وجود این قبیل بریدگی‌ها، نتایج تحلیل‌های مبتنی بر نمونه‌های بریده قابل‌تعمیم به کل جامعه نیست. به بیان دیگر، توزیع داده‌های نمونه، "مشروط" است.

---

1. Truncated Data

نمودار-۱. توزیع احتمال نرمال استاندارد تصادفی



(ب) بریدگی از بالا | (الف) بریدگی از پایین

چنانچه متغیر تصادفی y، توزیع نرمال با میانگین $\beta'\mathbf{x}$ ($\mathbf{x}$ شامل متغیرهای تأثیرگذار می‌باشد) و واریانس σ^2 داشته باشد، یعنی:

$$y_i|\mathbf{x}_i \sim N(\beta'\mathbf{x}_i, \sigma^2) \qquad (۶)$$

در شرایط نبود بریدگی، میانگین این متغیر برابر رابطه زیر خواهد بود:

$$E[y_i|\mathbf{x}_i] = \beta'\mathbf{x}_i \qquad (۷)$$

اما با داده‌های بریده، توزیع به صورت شرطی و میانگین آن تابعی غیرخطی از $\mathbf{x}_i$ و β و به صورت زیر است:[1]

$$E[y_i|\mathbf{x}_i, y_i > a] = \beta'\mathbf{x}_i + \sigma\{[\phi((a-\beta'\mathbf{x}_i)/\sigma)]/[1-\Phi((a-\beta'\mathbf{x}_i)/\sigma)]\}$$
$$E[y_i|\mathbf{x}_i, y_i > a] = \beta'\mathbf{x}_i + \sigma\lambda_i(\alpha_i), \qquad (۸)$$
$$\text{جایی‌که}: \alpha_i = (a-\beta'\mathbf{x}_i)/\sigma$$
$$\lambda_i = [\phi((a-\beta'\mathbf{x}_i)/\sigma)]/[1-\Phi((a-\beta'\mathbf{x}_i)/\sigma)]$$

بدین ترتیب، ملاحظه می‌شود که میانگین توزیع داده‌های بریده، متفاوت از توزیع داده‌های نابریده است؛ تأثیر حاشیه‌ای تغییر $\mathbf{x}_i$ به صورت:

$$\frac{\partial E[y_i|\mathbf{x}_i]}{\partial \mathbf{x}_i} = \beta + \sigma\left(\frac{d\lambda_i}{d\alpha_i}\right)\frac{\partial \alpha_i}{\partial \mathbf{x}_i}$$
$$= \beta + \sigma(\lambda_i^2 - \alpha_i\lambda_i)\left(\frac{-\beta}{\sigma}\right) = \beta(1-\lambda_i^2+\alpha_i\lambda_i) \qquad (۹)$$

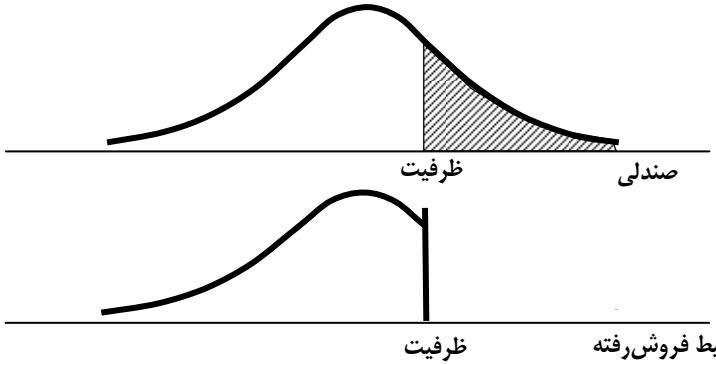
1. Maddala (1983: 365) and Greene (1993: 687).

خواهد بود که متفاوت از میزان حاصل از الگوی متعارف یعنی β است.

۲-۲-۲. داده‌های سانسورشده[1]

داده‌های سانسورشده، به یک مسأله عمومی در بسیاری از پژوهش‌های اقتصاد خُرد تبدیل شده است (Greene, 1993: 691). ویژگی اصلی داده‌های سانسورشده، این است که "مشاهدات" شامل یک نمونهٔ تصادفی از کل جامعه بوده، اما برخی از مقادیر متغیر وابسته، ارقامی متفاوت از واقعیت اختیار کرده و به یک مقدار مشخص تبدیل یا سانسور شده است.[2] مخارج مصرفی خانوار در مورد کالاهای مختلف (به ویژه کالاهای بادوام)، عرضه نیروی کار زنان، امور خانوادگی، بازداشت‌های بعد از آزادی از زندان، فروش بلیط سینما و جز اینها از جمله مهم‌ترین حوزه‌ها در مطالعات اقتصادی هستند که مسأله داده‌های سانسورشده در آنها مورد تشخیص و اذعان قرار گرفته است. برای نمونه، در فروش بلیط سینما، کل ظرفیت و بلیط‌های فروش‌رفته در مقایسه با تقاضا برای ورود به سینما مورد بررسی قرار می‌گیرد. اطلاعات موجود تنها در مورد بلیط‌های فروش‌رفته یا تقاضای بالفعل در اختیار است و میزان تقاضای بالقوه، همان‌گونه که نمودار ۲ نشان می‌دهد، مشخص نیست. از همین رو، کمیت متغیر برای ارقام واقعی، به صورت پیوسته و برای مشاهدات دیگر، به صورت ناپیوسته است؛ توزیع مناسب برای چنین داده‌هایی، "توزیع نرمال سانسورشده" است که ترکیبی از توزیع‌های پیوسته و ناپیوسته است.

نمودار -۲. توزیع سانسورشده جزئی برای تقاضای صندلی سینما



مأخذ: گرین (Greene, 1993: 692).

---

1. Censored Data

۲. در داده‌های بریده، نمونهٔ مطالعه‌شده تنها بخشی از جامعه را شامل می‌شود و هیچ اطلاعاتی در مورد متغیر وابسته (و متغیرهای مستقل) از بخش غیرنمونه گردآوری نمی‌شود. برای توضیح بیشتر، ر.ک. به: مادالا (Maddala, 1983: 166).

مبنای سانسورشدگی متغیر وابسته، یک مقدار ثابت (مانند a که می‌تواند هر کمیت حقیقی اختیار کند) در نظر گرفته می‌شود. یعنی، مقادیر متغیر وابسته برای موارد مشاهده‌شده، برابر مقادیر واقعی اما برای مقادیر مشاهده‌نشده، برابرa در نظر گرفته می‌شود. الگوی آماری متناسب با داده‌های سانسورشده به الگوی رگرسیونی سانسورشده یا به اصطلاح توبیت[۱] معروف است و شکل کلی آن به صورت تابع شاخص[۲] بیان می‌شود:

$$y_i^* = \beta' \mathbf{x}_i + u_i \quad ; \quad \begin{cases} y_i = a & if \quad y_i^* \le a \\ y_i = y_i^* & if \quad y_i^* > a \end{cases} \qquad (۱۰)$$

که y_i^* در واقع یک متغیر مکنون یا ناپیدا[۳] است.

با داده‌های سانسورشده، توزیع آماری مقادیر (شامل موارد قابل‌مشاهده و غیرقابل‌مشاهده) به صورت شرطی و میانگین آن تابعی غیرخطی از $\mathbf{x}_i$ و β ، و به قرار زیر است:[۴]

$$E(y|\mathbf{x}_i) = \Pr ob(y^* \le a) \times E(y^* | y^* \le a) + \Pr ob(y^* > a) \times E(y^* | y^* > a) \qquad (۱۱)$$

$$= \Pr ob(y^* \le a) \times a + \Pr ob(y^* > a) \times E(y^* | y^* > a)$$

$$E[y_i|\mathbf{x}_i] = \Phi(\alpha_i) \times a + (1 - \Phi(\alpha_i))\{\beta' \mathbf{x}_i + \sigma[\phi(\alpha_i)/(1 - \Phi(\alpha_i))]\}$$

$$E[y_i|\mathbf{x}_i] = \Phi(\alpha_i) \times a + (1 - \Phi(\alpha_i))(\beta' \mathbf{x}_i + \sigma\lambda_i)\,,$$

$$\text{جایی‌که}: \ \alpha_i = (a - \beta' \mathbf{x}_i)/\sigma \ , \ \ \lambda_i = \phi(\alpha_i)/(1 - \Phi(\alpha_i))$$

اما میانگین متغیر مکنون (غیرقابل‌مشاهده) عبارت است از:

$$E[y_i^* | \mathbf{x}_i] = \beta' \mathbf{x}_i \qquad (۱۲)$$

تأثیر حاشیه‌ای تغییر $\mathbf{x}_i$ برای هر یک از این وضعیت‌ها به صورت زیر خواهد بود:

۱.همان‌طور که در متون اقتصادسنجی آمده، رگرسیون توبیت (Tobit) ابتدا توسط توبین (Tobin, 1958) معرفی شده که وی در تحلیل‌های خود از الگوی پروبیت (Probit) بهره گرفته است. از همین رو، نام‌گذاری آن مرکب از بخشی از نام‌های توبین و پروبیت است؛ ر.ک. به: مادالا (Maddala, 1983: 151).

2. Index

3. Latent

۴. برای متغیر تصادفی $y^* \sim N(\mu, \sigma^2)$ میانگین توزیع نرمال شرطی سانسورشده برای $a = 0$، (یعنی $y = y^* \ \ if \ \ y^* > 0, \quad y = 0 \ \ if \ \ y^* \le 0$) از قضیه زیر پیروی می‌نماید:

$$E(y) = \Pr ob(y^* \le 0) \times E(y^* | y^* \le 0) + \Pr ob(y^* > 0) \times E(y^* | y^* > 0)$$

$$= \Pr ob(y^* \le 0) \times 0 + \Pr ob(y^* > 0) \times E(y^* | y^* > 0) = \Pr ob(y^* > 0) \times E(y^* | y^* > 0)$$

برای توضیح بیشتر، برای نمونه، ر.ک. به: گرین (Greene, 1993: 692-3).

$$\frac{\partial E[y_i^*|\mathbf{x}_i]}{\partial \mathbf{x}_i} = \beta \qquad (۱۳)$$

$$\frac{\partial E[y_i|\mathbf{x}_i]}{\partial \mathbf{x}_i} = \Pr ob[y_i^* > a]\frac{\partial E[y_i^*|\mathbf{x}_i, y_i^* > a]}{\partial \mathbf{x}_i} + E[y_i^*|\mathbf{x}_i, y_i^* > a]\frac{\partial \Pr ob[y_i^* > a]}{\partial \mathbf{x}_i} + a\frac{\partial \Pr ob[y_i^* \le a]}{\partial \mathbf{x}_i}$$

۲-۲-۳. داده‌های خود- انتخابی[1]

همان‌طور که پیش از این بیان شد، بشر در زندگی اجتماعی- اقتصادی خود همواره با انتخاب مواجه است که در نتیجهٔ همین انتخاب‌ها، گروه‌بندی‌ها و قشربندی‌های مختلفی در جوامع شکل می‌گیرد. دارندگان تحصیلات عالی، شاغلین، معلمان، دانشجویان، کارگران عضو اتحادیه کارگری و جز اینها نمونه‌های بارزی از گروه‌ها و زیرجامعه‌های یک کشور هستند که در نتیجهٔ "انتخاب" شکل‌گرفته‌اند. دارندگان تحصیلات عالی، تصمیم‌گرفته‌اند که در دانشگاه، ادامه تحصیل دهند؛ افراد شاغل، انتخاب کرده‌اند که وارد بازارکار شده و به حرفه مورد علاقه‌شان اشتغال یابند؛ عده‌ای از کارگران عضویت در اتحادیه کارگری را انتخاب کرده‌اند و جز اینها. بدین ترتیب، عناصر یک اجتماع، شاخه-شاخه شده و هر شاخه زیرجامعه‌ای را تشکیل می‌دهد که پژوهشگران در حالت کلی با این زیرجامعه‌ها سروکار دارند. نکته مهم و اساسی این است که وقتی داده‌های آماری برای انجام یک پژوهش جمع‌آوری می‌شوند، این داده‌ها معمولاً دربردارندهٔ یک نمونه تصادفی از زیرجامعه (نه از کل جامعه) هستند؛ یعنی نمونه‌ها تنها یک یا چند زیرجامعه را شامل می‌شود. از همین رو، قابلیت تعمیم نتایج به کل جامعه در معرض اشکال و انحراف جدّی است. با این حال، اشکال مورد بحث از دو حالت یعنی از نوع داده‌های سانسورشده و یا از نوع داده‌های بُریده، خارج نیست. تفاوت اساسی با مباحث قبلی، در این است که منشأ اصلی سانسورشدگی یا بُریدگی، انتخاب اعضا یا عناصر (زیر)جامعه است. از همین رو، این منشأ بایستی در الگوی تحت برازش لحاظ شود که در اصطلاح به صورت یک متغیر درونزا (نه برونزا) مورد توجه قرار گیرد.

چنانچه تابع اصلی برازش‌شده به صورت زیر در نظر گرفته شود:

$$y_i = \beta'\mathbf{x}_i + u_i \qquad (۱۴)$$

---

1. Self-selection Data

که y_i (لگاریتم) دریافتی و $\mathbf{x}_i$ شامل ویژگی‌های شاغلین (از جمله تحصیلات آنها) در یک جامعه فرضی است. مطابق تحلیل‌های کلاسیک ارزیابی بازده آموزش، می‌توان شکل تعمیم‌یافته تابع دریافتی مینسر را مشخص و برآورد کرد. در این صورت، تابع دریافتی به صورت زیر خواهد بود:

$$y_i = \beta_0 + \beta_1 S_i + \sum_{k=2}^{K} \beta_k x_{ki} + u_i \qquad (۱۵)$$

که متغیر S_1 (جایگزین x_1) نشان‌دهنده سال‌های تحصیل و ضریب آن (β_1) بیانگر بازده آموزش خواهد بود.

در چارچوب مباحث مسألهٔ انتخاب، دو بحث جدّی را می‌توان مطرح و مورد تحلیل قرار داد: ۱) تصمیم برای ورود به بازار کار و ۲) تصمیم در مورد نوع و میزان تحصیلات. تصمیم برای ورود به بازار کار، یک انتخاب مهم است که تحت تأثیر عوامل متعددی (نظیر سطح حقوق و دستمزد، نوع شغل، و محیط کار) است. برای الگوسازی مشارکت در بازار کار اصولاً از متغیر دو گزینه‌ای (صفر و یک که «صفر» برای واردنشدن به بازار کار و «یک» برای مشارکت در بازار کار) استفاده می‌شود. بر این اساس، تابع مشارکت در بازار کار را می‌توان به صورت زیر تعریف کرد:

$$z_i^* = \gamma' \mathbf{w}_i + e_i \qquad (۱۶)$$

که در آن، $\mathbf{w}_i$ بردار عوامل مؤثر بر مشارکت در بازار کار (یعنی z) و $e \sim N(0,\ \sigma_e^2)$ است. می‌توان z_i^* را به صورت یک متغیر ناپیدا (مکنون) در نظر گرفت که برای مقادیر مثبت آن، $z = 1$ و برای سایر مقادیر $z = 0$ اختیار نماید. با این مبانی و ابزار تحلیلی، می‌توان احتمال مشارکت در بازار کار و سهم عوامل آن را مورد تحلیل و ارزیابی قرار داد.

نکته مهم و اساسی در مورد تحلیل دریافتی این است که دریافتی فرد زمانی قابل مشاهده است که وی مشارکت در بازار کار را انتخاب کرده باشد. بدین ترتیب، متغیر وابسته وقتی مجسم شده و شکل می‌گیرد که مشارکت افراد در بازار کار نیز رخ داده باشد. همین مسأله، باعث بریدگی مشاهدات می‌شود. چرا که وضعیت دریافتی افرادی مورد تحلیل قرار می‌گیرد که وارد بازار کار شده باشند؛ سایر افراد (آنهایی که در بازار کار مشارکت ندارند) مورد آمارگیری و تحلیل قرار نمی‌گیرند؛ به تعبیر دقیق‌تر، داده‌های آماری در مورد ویژگی‌های کار آنها در دسترس نمی‌باشد. برای لحاظ‌کردن نقش یا تأثیر بریدگی در داده‌های آماری (مطابق با رویهٔ گرین، Greene, 1993: 709 در مورد مسالهٔ بریدگی داده‌ها)، تابع دریافتی به صورت شرطی و امید ریاضی آن به قرار زیر خواهد بود:

$$E[y_i \,|\, y_i \ is\ observed\ ] = E[y_i \,|\, z_i^* > 0] \qquad (۱۷)$$
$$= \beta' \mathbf{x}_i + E[u_i \,|\, z_i^* > 0]$$
$$= \beta' \mathbf{x}_i + E[u_i \,|\, e_i > -\gamma' \mathbf{w}_i]$$
$$= \beta' \mathbf{x}_i + \rho \sigma_u \lambda_i(\alpha_e)$$

جایی که : $\alpha_e = -\gamma' \mathbf{w}/\sigma_e ; \quad \lambda_i = \phi(\alpha_e)/\Phi(\alpha_e)$

و تابع دریافتی شرطی به صورت زیر خواهد بود:

$$y_i \,|\, y_i \ is\ observed\ = y_i \,|\, z_i^* > 0 = \beta' \mathbf{x}_i + \rho \sigma_u \lambda_i(\alpha_e) + v_i \qquad (۱۸)$$
$$= \beta' \mathbf{x}_i + \beta_\lambda \lambda_i(\alpha_e) + v_i$$

بدین ترتیب، تابع دریافتی شرطی بالا را می‌توان به کمک روش حداقل مربعات معمولی برآورد کرده و به برآوردهای سازگار دست‌یافت. نادیده‌گرفتن مسألهٔ خودانتخابی به لحاظ مبانی آمار و اقتصادسنجی آثاری همانند آثار متغیر مرتبط حذف‌شده دارد که در واقع، λ_i در این رابطه ایفای نقش می‌نماید. تأثیر حاشیه‌ای متغیر(های) مستقل شامل دو جزء یکی تأثیر مستقیم و دیگری تأثیر از ناحیه احتمال مشارکت در بازار کار (λ_i) خواهد بود. بدین ترتیب، تأثیر کل عبارتست از:

$$\frac{\partial E[y_i \,|\, z_i^* > 0]}{\partial x_{ik}} = \beta_k - \lambda_k \left(\rho \sigma_u / \sigma_e\right) \delta_i(\alpha_e) \qquad (۱۹)$$

جایی که: $\delta_i(\alpha_e) = \lambda_i^2 + \alpha_i \lambda_i$

نوع دیگری از خود- انتخابی، تصمیم‌گیری در مورد نوع و سطح تحصیلات است؛ یعنی متغیر آموزش یک متغیر تصمیم است و افراد به طور آگاهانه نوع و سطح تحصیلات خود را انتخاب می‌نمایند. عوامل مختلفی انتخاب نوع و سطح تحصیلات را تحت تأثیر قرار می‌دهند؛ سواد و تحصیلات والدین، توانمندی‌های ذاتی، شرایط محیطی- اجتماعی، بازده مورد انتظار و جز اینها از جمله مهم‌ترین عوامل تاثیرگذار هستند. با توجه به این قبیل عوامل، افراد میزان و نوع آموزش مورد علاقه خود را (یعنی S^*) مشخص و سپس آن را محقق می‌سازند؛ آنگاه است که فرد می‌تواند با سطح تحصیلات بالاتر وارد بازار کار شده و درآمد کسب نماید. بدین ترتیب، زمانی متغیر وابسته (دریافتی متناسب با تحصیلات جدید) قابل مشاهده خواهد بود که انتخاب برای تحصیل نیز محقق شده باشد (یعنی $S^* > 0$).[۱] برای تبیین ریاضی این نوع خود- انتخابی، فرض می‌شود که متغیر آموزش (S) بر اساس رابطه زیر تعیین می‌شود:

۱.تابع تصمیم‌گیری آموزشی در متون علمی به شکل‌های مختلفی از جمله احتمال کسب یک درجه تحصیلی مشخص (به شکل متغیر وابسته صفر و یک) و میزان تحصیلات مورد علاقه (به شکل متغیر پیوسته، قابل سنجش با سال تحصیل) به کارگرفته شده است. در این پژوهش، میزان تحصیلات مورد علاقه (S^*) به عنوان ملاک

$$s_i^* = \eta' \mathbf{z}_i + e_i \qquad (۲۰)$$

$$y_i \ is \ observed \ \ if \ \ s_i^* > 0$$

با ادغام یا جایگذاری رابطه (تحصیلات) در رابطه (تابع) دریافتی، این فرض اساسی را می‌توان در نظر گرفت که جملات اخلال توابع دریافتی و تصمیم آموزشی توزیع نرمال مشترک دارند به‌طوری‌که:

$$(u, e) \sim bN(0,\ 0,\ \sigma_u^2,\ \sigma_e^2,\ \rho)$$

و در نتیجه، تابع دریافتی به صورت تابع دریافتی شرطی تبدیل شده و میانگین آن به صورت زیر خواهد بود:

$$E[y_i \,|s_i^* > 0] = \beta' \mathbf{x}_i + E[u_i | \eta' \mathbf{z}_i + e_i > 0] = \beta' \mathbf{x}_i + E[u_i | e_i > -\eta' \mathbf{z}_i] \qquad (۲۱)$$

$$= \beta' \mathbf{x}_i + \rho \sigma_u \lambda_i(\alpha_e)$$

جایی که: $\alpha_e = -\eta' \mathbf{z}_i / \sigma_e \quad , \quad \lambda_i = \phi(\eta' \mathbf{z}_i / \sigma_e) / \Phi(\eta' \mathbf{z}_i / \sigma_e)$

و رابطهٔ برآوردشده به صورت زیر خواهد بود:

$$y_i | s_i^* > 0 = \beta' \mathbf{x}_i + \rho \sigma_u \lambda_i(\alpha_e) + v_i \qquad (۲۲)$$

فرایند تحلیلی مزبور وجه خود- انتخابی مسأله را مرتفع می‌سازد؛ اما مسأله همزمانی (بین تأثیرپذیری آموزش و دریافتی از توانمندی‌های ذاتی) یا درونزایی (برخی از متغیرهای توضیحی نظیر آموزش به عنوان یک متغیر تصمیم) وجود دارد که نیازمند تحلیل و بررسی بیشتری است. مسألهٔ درونزایی باعث نقض برخی از فروض کلاسیک تحلیل‌های آماری نظیر نبود رابطهٔ بین متغیرهای توضیحی و جمله خطای الگو (یعنی $Cov(u,\ x) = ۰$) می‌شود؛ این مسأله در مورد آموزش شدیدتر است. چرا که برای ارزیابی بازده آموزش، در عمل وضعیت افراد تحصیلکرده با افراد تحصیل‌نکرده (یا با تحصیلات کمتر) مقایسه می‌شود. این مقایسه تنها در صورتی بی‌اشکال است که ویژگی‌های این‌دو گروه (به‌ویژه ویژگی‌های ناپیدای آنها) یکسان باشند. در واقعیت، اصولاً چنین شرایطی وجود ندارد؛ مهم‌تر اینکه برخی از ویژگی‌های ناپیدا (مانند توانمندی‌های ذاتی افراد) متغیر وابسته و متغیر تصمیم (یعنی آموزش) را تحت تأثیر قرار می‌دهند. در چنین وضعیتی می‌توان ادعا کرد که افراد دارای تحصیلات بیشتر، حتی در صورت برخورداری از تحصیلات کمتر، می‌توانستند دارای دریافتی بالاتری باشند؛ چرا که این افراد به طور ذاتی تواناتر هستند. برعکس، افراد دارای تحصیلات کمتر، حتی با داشتن تحصیلات بالاتر، به دلیل توانایی ذاتی کمتر، دارای دریافتی کمتری (در مقایسه با سایر دارندگان تحصیلات بالا)

مناسب‌تری تلقی شده است؛ این مسأله زمانی به عنوان یک متغیر تصمیم تبلور می‌یابد که میزان تحصیل ورای تحصیلات اجباری (قانونی) در نظر گرفته شود. چرا که عموماً تحصیلات اجباری در بستر انتخاب و تصمیم (مبتنی بر هزینه-فایده) افراد و خانواده‌ها قرار ندارد.

می‌بودند. در هر حال، از قلم انداختن متغیرهایی مانند توانمندی‌های ذاتی (قابل سنجش، برای مثال، با بهره هوشی-IQ) سبب نقض فرض "استقلال جمله اخلال و متغیر آموز" می‌شود.

در متون علمی مرتبط[۱]، برای در نظرگرفتن مسالهٔ همزمانی یا رفع مشکل درونزایی متغیر توضیحی، استفاده از فن متغیر ابزاری[۲] به عنوان یک راه اساسی پیشنهاد و در پژوهش‌های گسترده‌ای به کارگرفته شده است. برای ارزیابی بازده آموزش با فن متغیر ابزاری بایستی دو تابع، یعنی تابع دریافتی و تابع تصمیم آموزش، در نظر گرفته شود که متغیرهای تابع دریافتی شامل آموزش، پیشینه خانوادگی، تجربه کاری و محل و نوع کار، و متغیرهای تابع تصمیم آموزش شامل متغیرهایی چون پیشینه خانوادگی، انگیزه‌های اقتصادی و توانمندی‌های ذاتی در نظر گرفته می‌شود. اما برای سادگی، فرض می‌شود تابع دریافتی اصلی دارای یک متغیر (یعنی آموزش) و تابع آموزش نیز تنها دارای یک متغیر (مانند تحصیلات والدین یا همسر) در سمت راست است:

$$y_i = \beta_s S_i + u_i$$
$$S_i = \gamma_z Z_i + e_i \qquad (۲۳)$$

به طوری که : $Cov(S,\ u) \neq 0;\ \ Cov(Z,\ u) = 0;\ \ Cov(Z,\ S) \neq 0$

گفتنی است که هدف اصلی، برآورد β_s بوده که با توجه به مسألهٔ درونزایی ناشی از $Cov(S,\ u) \neq ۰$، به طور مستقیم با روش حداقل مربعات معمولی از برازش y بر روی S قابل حصول نیست و در عمل صورت خلاصه شده تابع دریافتی بایستی برآورد شود. یعنی:

$$y_i = \beta_s S_i + u_i = \beta_s(\gamma_z Z_i + e_i) + u_i = \theta_Z Z_i + v_i \qquad (۲۴)$$

و نتایج به‌دست آمده برای ارزیابی β_s مورد استفاده قرار گیرد که به تخمین‌زن متغیر ابزاری معروف است:

$$\beta_s^{IV} = \frac{\theta_Z}{\gamma_Z} = \frac{Cov(y,Z)}{Cov(Z,Z)} \Big/ \frac{Cov(S,Z)}{Cov(Z,Z)} = \frac{Cov(y,Z)}{Cov(S,Z)} = \beta_s + \frac{Cov(u,Z)}{Cov(S,Z)} \qquad (۲۵)$$

دقت برآوردهای فن متغیر ابزاری، به شدّت و قوّت رابطه بین Z و u بستگی دارد. مطابق تحلیل نظری،[۳] چنانچه $Cov(Z,\ u) = ۰$ باشد، مقدار برآورد متغیر ابزاری دقیقاً برابر ضریب واقعی خواهد بود. اما در واقعیت، این‌گونه نیست، یعنی $Cov(Z,\ u) \neq ۰$ و لذا $\beta_s^{IV} \neq \beta_s$ و تخمین‌زن متغیر ابزاری

---

۱. ر.ک. برای نمونه، به: دوهرتی (Dougherty, 2002:252) و وولدریج (Wooldridge, 2002).

2. Instrumental Variable

۳. متغیر ابزاری مناسب، همان‌گونه که دوهرتی (Dougherty, 2002:254) تأکید می‌کند، بایستی دارای ویژگی‌های مشخصی باشد، مانند: ۱) بین جمله خطای الگوی اصلی و متغیر مجازی رابطه‌ای شد، یعنی: $Cov(Z,\ u) = 0$؛ ۲) بین متغیر مجازی و متغیر درونزا رابطه قوی وجود داشته باشد، یعنی: $Cov(Z,\ X) \neq 0$ و ۳) متغیر مجازی به طور مستقیم بر متغیر درونزا و به طور غیرمستقیم بر متغیر اصلی (y) تاثیر داشته باشد.

مساوی با مقدار واقعی به علاوه یک جمله خطا خواهد بود. با این حال، در نمونه‌های بزرگ، جمله خطا از بین می‌رود و بدین ترتیب، $\beta_s^{IV} = \beta_s$.[1]

بنابراین، برای در نظر گرفتن هر دو مسالهٔ خود- انتخابی و درونزایی، استفاده همزمان از فنون آماری این دو حوزه ضروری است که در رابطه زیر تصریح شده است:

$$y_i \left| y_i \ is \ observed \right. = y_i \left| z_i^* > 0 \right. = \beta' \mathbf{x}_i + \beta_s S_i^* + \rho\sigma_u \lambda_i(\alpha_e) + v_i \qquad (۲۶)$$

$$= \beta' \mathbf{x}_i + \beta_s S_i^* + \beta_\lambda \lambda_i(\alpha_e) + v_i$$

جائیکه : $S_i^* = \gamma_z Z_i + e_i$

۲-۳. روش‌های برآورد

برای برآورد الگوهای مبتلا به مسألهٔ انتخاب، روش‌های خاصی مناسب است. روش حداقل مربعات معمولی همان‌گونه که تحلیل شد، نتایج سازگار و بدون‌تورش فراهم نمی‌کند. دو روش مشخص برای برآورد این توابع که با مسأله بریدگی و سانسور مواجه‌اند، در متون علمی ارائه و به‌کار گرفته شده‌اند: روش دو مرحله‌ای هکمن[2] و روش حداکثر درستنمایی.

روش دومرحله‌ای:

چنانچه تابع زیر که با مسأله بریدگی یا سانسور (از جهت انتخاب نمونه[3]) مواجه باشد را در نظر بگیریم:

$$y_i \left| y_i \ is \ observed \right. = y_i \left| z_i^* > 0 \right. = \beta' \mathbf{x}_i + \beta_\lambda \lambda_i(\alpha_i) + v_i \qquad (۲۷)$$

که در مقایسه با توابع مرسوم دارای عنصر جدید λ_i و بیانگر مبنای انتخاب (بریدگی یا سانسور داده‌ها) است. برای برآورد تابع مورد نظر در روش دومرحله‌ای، ابتدا بایستی مقدار λ_i برای همه مشاهدات برآورد شود که این مهم با فنون آماری مانند پروبیت (Probit)، و با توجه به ماهیت غیرخطی بودن آن (یعنی $\lambda_i = \phi(\alpha_i) / \Phi(\alpha_i)$)، با روش حداکثر درستنمایی قابل انجام است (مرحله اول). در مرحله دوم، با استفاده از داده‌های آماری مربوط به متغیر $\hat{\lambda}_i$ همراه با متغیرهای دیگر گرفته شده در این تابع به صورت:

---

1. Dougherty (2002: 253)

۲. روش دومرحله‌ای معروف به روش هکمن (Heckman)، توسط جیمز هکمن با ارائه مقاله سال ۱۹۷۹ وارد جریان اصلی علمی مباحث شد. البته، همان‌گونه که هکمن اذعان کرده، این روش در ابتدا توسط گرونا (Gronau, 1974) به‌کارگرفته شده است ولی با توجه به نقش هکمن در تسهیل محاسبات و برآوردها، به روش دومرحله‌ای هکمن مشهور شده است. برای توضیح بیشتر، ر.ک. برای نمونه به، مادالا (Maddala, 1983) و دوهرتی (Dougherty, 2002: 298).

3. Sample Selection

$$y_i \left| y_i \ \textit{is observed} \ = y_i \right| z_i^* > 0 = \beta' \mathbf{x}_i + \beta_\lambda \hat{\lambda}_i(\alpha_i) + v_i \qquad (۲۸)$$

با روش حداقل مربعات برآورد می‌شود؛ نتایج به‌دست‌آمده، سازگار و بدون‌تورش خواهند بود.[۱]

درصورتی که همه مشاهدات شامل مشاهدات غیرصفر ($y_i > 0$) و مشاهدات صفر ($y_i \le 0$) در تحلیل لحاظ شود، داریم:

$$E(y_i|\mathbf{x}_i) = \Pr(y_i = 0) \times E(y_i|y_i = 0) + \Pr(y_i > 0) \times E(y_i|y_i > 0) \qquad (۲۹)$$
$$= \Phi_i \beta' \mathbf{x}_i + \sigma \phi_i$$

و بنابراین:
$$y_i|\mathbf{x}_i = \Phi_i \beta' \mathbf{x}_i + \sigma \phi_i + v_i$$

در این شرایط نیز، ابتدا مقادیر Φ_i و ϕ_i برآورد شده و سپس رابطه $y_i|\mathbf{x}_i$ با استفاده از حدقل‌مربعات معمولی برآورد می‌شود.

روش حداکثر درستنمایی:

در روش حداکثر درستنمایی، ابتدا بایستی تابع چگالی احتمال مشترک متغیرهای مطالعه‌شده را مشخص و پس از آن تابع درستنمایی را تشکیل داد. تابع چگالی نرمال استاندارد برای مشاهدات غیرصفر (به این فرض که نقطه بریدگی a باشد) به صورت زیر قابل ارائه است:

$$f(y_i \left| y_i^* > a \right.) = \frac{1}{\sigma} \phi\left(\frac{y_i - \beta' \mathbf{x}_i}{\sigma} \right) \Bigg/ \left[1 - \Phi\left(\frac{a - \beta' \mathbf{x}_i}{\sigma} \right) \right] \qquad (۳۰)$$

و تابع درستنمایی و لگاریتم آن به صورت:

$$L = \prod_{i=1}^{n} \frac{1}{\sigma} \phi\left(\frac{y_i - \beta' \mathbf{x}_i}{\sigma} \right) \Bigg/ \left[1 - \Phi\left(\frac{a - \beta' \mathbf{x}_i}{\sigma} \right) \right]$$

$$\ln L = \frac{-n}{2} \left[\ln(2\pi) + \ln \sigma^2 \right] - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \beta' \mathbf{x}_i \right)^2 \qquad (۳۱)$$
$$- \sum_{i=1}^{n} \ln\left[1 - \Phi\left(\frac{a - \beta' \mathbf{x}_i}{\sigma} \right) \right]$$

برای مشخص‌کردن مقادیر پارامترهای مورد نظر (β و σ^2)، از رابطه لگاریتم درستنمایی نسبت به پارامترهای یادشده، مشتق گرفته و طبق روال مرسوم، برابر صفر قرار می‌دهیم.[۲]

---

۱. به رغم سازگار و بدون‌تورش بودن، برآوردهای حاصله لزوماً کارآیی لازم را نداشته و با مشکل واریانس ناهمسانی مواجه هستند. برای توضیح بیشتر در این رابطه، ر.ک. به: گرین (Greene, 2003, ch. 22).

۲. با توجه به روال شناخته‌شده و برای پرهیز از طولانی‌شدن مباحث، فرمول‌ها و محاسبات مربوط ارائه نمی‌شود؛ البته بایستی توجه داشت که به دلیل وجود بخش غیرخطی تابع، محاسبات ویژه‌ای برای مشتق‌گیری لازم است. برای توضیح بیشتر در این رابطه، ر.ک. برای نمونه، به: گرین (Greene, 1993: ch. 12).

چنانچه مشاهدات نامشهود در تحلیل‌ها اضافه شود، یعنی یک تحلیل مبتنی بر داده‌های با توزیع سانسورشده اجرا شود، با این فرض که مبنای نامشهود بودن داده‌ها نقطه صفر است، تابع درستنمایی (بر اساس Greene, 1993: 696) به صورت زیر خواهد بود:

$$L = \prod_{y>0} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(y_i - \beta'\mathbf{x}_i)^2}{2\sigma^2} \right) \prod_{y=0} \left[1 - \Phi\left(\frac{\beta'\mathbf{x}_i}{\sigma} \right) \right]$$

$$\ln L = \sum_{y>0} \frac{-1}{2} \left[\ln(2\pi) + \ln\sigma^2 + \frac{(y_i - \beta'\mathbf{x}_i)^2}{2\sigma^2} \right] + \sum_{y=0} \ln\left[1 - \Phi\left(\frac{\beta'\mathbf{x}_i}{\sigma} \right) \right] \qquad (۳۲)$$

برای مشخص‌کردن مقادیر پارامترهای مورد نظر (β و σ^2)، از رابطهٔ لگاریتم درستنمایی نسبت به پارامترهای یادشده مشتق گرفته و برابر صفر قرار می‌دهیم. بدین ترتیب، مقادیر β و σ^2 از روابطه شرایط اولیه حداکثرسازی به صورت زیر به‌دست می‌آید:

$$\beta = (\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'Y_1 - \sigma(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_0'\lambda_0 \qquad (۳۳)$$
$$= \beta_{LS} - \sigma(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_0'\lambda_0$$

تخمین‌زن حداقل مربعات= β_{LS} ؛ جایی که: $\lambda_i = \phi_i/(1-\Phi_i)$

$$\sigma^2 = \frac{Y_1'(Y_1 - \beta'\mathbf{x}_1)}{N_1} \qquad (۳۴)$$

که در آن، Y_1 و $\mathbf{x}_1$ به ترتیب مقادیر y_i و $\mathbf{x}_i$ برای مشاهدات $y_i > 0$ ، و $\mathbf{x}_0$ و λ_0 به ترتیب مقادیر $\mathbf{x}_i$ و λ_i برای مشاهدات $y_i = 0$ می‌باشند.

برای برآورد ضرایب با توجه به عناصر غیرخطی الگو، اصولاً از روش نیوتن در یک فرایند تکرار تا حصول به همگرایی مقادیر برآورد استفاده می‌شود.[۱]

---

۱. در منابع علمی مرتبط، روش‌های چندمرحله‌ای با اتکا به روش نیوتن-رافسون (Newton-Raphson) برای حصول به همگرایی (تقارب) بایسته پیشنهاد شده است. برای نمونه، فیر (Fair, 1977) یک روش ۵ مرحله‌ای برای این منظور پیشنهاد کرده است (ر.ک. به: Maddala, 1983: 154). به لحاظ اینکه در بسیاری از نرم‌افزارهای آماری و اقتصادسنجی روش و مراحل برآورد به طور حرفه‌ای و آسانی تعبیه شده است، از ارائه توضیحات تفصیلی در این خصوص پرهیز می‌شود؛ برای توضیح بیشتر، ر.ک. به: گرین (Greene, 1993: ch. 12).

۳. داده‌ها و اطلاعات آماری

برای مطالعه مباحث و مسایل یادشده به طور تجربی، دو نمونه آماری یکی از جامعه استان تهران (استخراج‌شده از آمارگیری ویژگی‌های اقتصادی- اجتماعی مرکز آمار ایران) و دیگری از جامعه آمریکا (استخراج‌شده از سرشماری جمعیت سال ۱۹۹۱)، استفاده کرده‌ایم.[۱] در ابتدا، ویژگی‌های مهم این دو نمونه توصیف و سپس تحلیل‌های آماری ارائه می‌شود.

برخی ویژگی‌های دو نمونه، در جدول ۱ ارائه شده است. همان‌طور که ملاحظه می‌شود، کل مشاهدات نمونهٔ جامعه ایران شامل ویژگی‌های ۲۳۵۰ نفر است که از این تعداد، ۱۰۷۳ نفر شاغل هستند. میانگین دریافتی ساعتی در سال ۱۳۸۰ برای کل نمونه ۳۵۶۷ ریال (با انحراف معیار ۷۹۸۵) و برای شاغلین ۷۸۱۳ ریال (با انحراف معیار ۱۰۳۲۱) بوده است.

مشاهدات استخراج‌شده از سرشماری جمعیت آمریکا شامل ۵۶۳۴ فرد متأهل است که از این تعداد، ۴۷۸۵ نفر شاغل مرد هستند. میانگین دریافتی ساعتی در نمونه مردان متأهل آمریکا در سال ۱۹۹۱، برابر ۱۲/۵ دلار (با انحراف معیار ۱۷) بوده است.

میانگین سن در هر دو جامعه، مشابه و حدود ۴۲ سال است. از ویژگی‌های جالب توجه، وضعیت تحصیلات است؛ همان‌طور که ارقام نشان می‌دهند، میانگین سال‌های تحصیل در نمونهٔ استان تهران ۷/۷۵ سال برای کل نمونه و ۸/۹۵ سال برای شاغلین بوده که ارقام مشابه برای نمونه مردان متأهل آمریکا در حد قابل توجهی بالاتر و به ترتیب ۱۳/۱ و ۱۳/۴ سال بوده است. در مقابل، میزان تجربه کاری بالقوه در نمونه استان تهران تا حدی (یعنی دامنه ۲۶ تا ۲۸ سال در مقایسه با ۲۱ تا ۲۳ سال) بالاتر است.

---

۱. از آنجا که هدف اصلی این پژوهش، کاوش مسایل روش‌شناسی در مورد دو مسأله انتخاب و درونزایی برای دو جامعه متفاوت (توسعه‌یافته و درحال توسعه) است، تلاش کرده‌ایم دو نمونهٔ آماری قابل مقایسه (از لحاظ ویژگی‌ها و متغیرهای دردسترس) استفاده شود؛ این دو نمونهٔ تنها گزینه‌هایی بوده‌اند که دسترسی بدان‌ها میسر بوده است.

جدول -۱. برخی از ویژگی‌های جامعهٔ آماری افراد متأهل در ایران و آمریکا

جامعه آماری:	ایران (۱۳۸۰)				آمریکا (۱۹۹۱)			
	کل نمونه (۲۳۵۰)		شاغلین (۱۰۷۳)		کل نمونه (۵۶۳۴)		شاغلین (۴۷۸۵)	
متغیر	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار	میانگین	انحراف معیار
دریافتی ساعتی (ریال/دلار)	۳۵۶۷	۷۹۸۵	۷۸۱۳	۱۰۳۲۱	۱۲/۴۵ [۳]	۱۷/۰۲	۱۲/۴۵	۱۷/۰۲
تعداد افراد خانواده[۱]/ فرزندان	۴/۲۲	۱/۵۱	۴/۲۱	۱/۴۱	۰/۵۹	۰/۴۹	۰/۶۳	۰/۴۸
سن (سال)	۴۲/۴۷	۱۳/۴۱	۴۱/۶۴	۱۰/۸۶	۴۲/۴۵	۱۱/۲۳	۴۱/۱۱	۱۰/۲۴
تحصیلات (سال)	۷/۷۵	۵/۰۱	۸/۹۵	۴/۸۵	۱۳/۱۵	۲/۹۸	۱۳/۳۸	۲/۸۴
تجربه کار (سال)[۲]	۲۸/۷۳	۱۵/۶۰	۲۶/۶۸	۱۲/۶۴	۲۳/۳۱	۱۱/۷۶	۲۱/۷۳	۱۰/۵۹
تحصیلات همسر (سال)	۷/۷۴	۵/۰۱	۸/۴۸	۴/۹۳	۱۲/۹۸	۲/۶۲	۱۳/۱۳	۲/۵۶

توضیح: داده‌های آماری ایران برخی از ویژگی‌های یک نمونه از افراد متأهل (همسردار) جامعهٔ استان تهران در سال ۱۳۸۰ را نشان می‌دهد که از نتایج آمارگیری از ویژگی‌های اقتصادی- اجتماعی خانوار مرکز آمار ایران استخراج شده است. داده‌های آماری جامعه آمریکا مربوط به نمونه‌ای تصادفی از خانوارها است که از نتایج "سرشماری جمعیت سال ۱۹۹۱" آن کشور استخراج و از سایت [http://www.msu.edu/~ec/faculty/wooldridge] گرفته شده است.

[۱] برای نمونهٔ استان تهران، تعداد افراد خانواده (که شامل فرزندان، زن و شوهر، و برخی از بستگانی که با آنها زندگی می‌کنند) است. برای نمونه افراد متأهل آمریکا، شاخص تعداد افراد فقط شامل تعداد فرزندان است.

[۲] شامل تجربه کاری بالقوه بوده که از رابطه «سن - ۶- سال‌های تحصیل» محاسبه شده است.

[۳] تعداد مشاهدات برای متغیر دریافتی ساعتی فقط شامل شاغلین (یعنی ۴۷۸۵ نفر) است.

۴. تحلیل‌های تجربی

برای ارزیابی آثار مسایل انتخاب و درونزایی، الگوهای رگرسیونی مختلفی برآورد شده و نتایج در جداول ۲ و ۳، ارائه شده است. هدف اصلی از تحلیل‌های این قسمت، حصول این نکته اساسی است که آیا این مسایل نتایج تحلیل‌های آماری (یعنی ارزیابی *بازده آموزش*) را تحت تأثیر قرار می‌دهد یا خیر. مبنا و پایه تحلیل‌ها برای انجام مقایسه، برآوردهای مبتنی روش حداقل مربعات معمولی (OLS) یعنی الگوهای ۱ و ۲ این جدول است. همان‌طور که ارقام ضرایب متغیر آموزش (سال تحصیل) نشان می‌دهد، میزان بازده آموزش در سال ۱۳۸۰ برای شاغلین استان تهران حدود ۹ درصد ارزیابی می‌شود؛ رقم مشابه برای جامعه مردان متأهل آمریکا در سال ۱۹۷۵ حدود ۸/۲ درصد است.[۱]

۱. با توجه به هدف اصلی تحلیل‌های آماری این قسمت (یعنی کاوش مسایل روش‌شناسی پیرامون ارزیابی آثار مسایل خود- انتخابی و نمونه- انتخابی بر بازده آموزش)، سعی کرده‌ایم که الگوهای دریافتی برازش‌شده به

در الگوی ۳، مسأله نمونه- انتخابی[1] مورد تحلیل و ارزیابی قرار می‌گیرد؛ به بیان دیگر، داده‌ها و مشاهدات برگرفته از دنیای واقعی، وضعیت ویژگی‌های شاغلین را نشان می‌دهد که با تحلیل میزان دریافتی و ویژگی‌های سرمایه انسانی، بازده آموزش برآورد می‌شود. اما در واقعیت، تنها بخشی از کل افراد جامعه، ورود به (یا مشارکت در) بازار کار را انتخاب کرده‌اند و لذا نمونه استخراج‌شده از شاغلین، نمونه تصادفی از کل افراد جامعه نیست. برای مقابله با این مسألهٔ آماری، از روش هکمن که در قسمت‌های قبل به تفصیل بیان شد، استفاده شده و نتایج به‌دست آمده در الگوی ۳ ارائه شده است. همان‌طور که نتایج نشان می‌دهند، مسالهٔ نمونه- انتخابی، بازده آموزش را تحت تأثیر قرار می‌دهد. در نمونه استان تهران تأثیر آن ناچیز بوده ولی میزان بازده را کاهش می‌دهد؛ در مقابل، برای نمونه مردان متأهل آمریکا میزان بازده آموزش را قدری افزایش می‌دهد.

مسأله دوم، خود انتخابی و درونزایی ناشی از آن است که الگوی ۴ به آن می‌پردازد. برای این منظور، از فن متغیر ابزاری استفاده شده است. یعنی فرض‌شده که تحصیل افراد، یک متغیر تصمیم است؛ در بین کل افراد جامعه، تنها یک عده خاص آموزش را انتخاب کرده و موفق به ادامه تحصیل می‌شوند. در اینجا نیز نکته مهم و اساسی، اشاره به این بحث است که افراد منتخب، که داده‌های آماری (به ویژه در خصوص دریافتی و تحصیلات) فقط در مورد آنها جمع‌آوری و تحلیل می‌شود، نمونه تصادفی از کل افراد جامعه نبوده بلکه ناشی از یک فرایند انتخاب یا رفتار منطقی است که به شدت تصادفی بودن نمونه مورد انتظار در تحلیل‌های آماری را خدشه‌دار می‌کند. افزون بر این، آموزش و تحصیل افراد متأثر از عوامل دیگری نیز می‌باشد که عموماً غیرقابل مشاهده‌اند، و میزان دریافتی (متغیر وابسته) را به‌طور مستقیم تحت تأثیر قرار می‌دهند. حاصل این شرایط، حذف شدن نشانگرهای مربوط به ویژگی‌های مؤثر غیرقابل مشاهده و لذا ایجاد ارتباط و همبستگی بین متغیرهای توضیحی "پیدا و آشکار" نظیر آموزش است. یکی از مناسب‌ترین راه‌های مقابله با این مسأله، استفاده از متغیر ابزاری است؛ یعنی اولاً متغیر هدف (که در اینجا آموزش است) به جای برونزا دیدن، به صورت درونزا در نظرگرفته می‌شود؛ ثانیاً این متغیر وابسته به متغیر یا متغیرهایی که بیشترین ارتباط را با آموزش و کمترین ارتباط مستقیم را با متغیر وابسته (مانند دریافتی) داشته باشد، تعریف می‌شود. برای این مقصود، تحصیلات همسر فرد مورد مطالعه، به عنوان متغیر ابزاری مناسب انتخاب شده و نتایج حاصله در قالب الگوی ۴ ارائه شده

ساده‌ترین شکل آن (یعنی الگوی پایه مینسر، در بردارنده متغیرهای آموزش و تجربه کاری) باشد. با این حال، لحاظ کردن متغیرهای تأثیرگذار دیگر (مانند جنسیت، محل زندگی، نوع شغل و جز اینها) یافته‌های تحقیق را تغییر نمی‌دهد. افزون بر این، «واریانس ناهمسانی» یکی از مشکلات مبتلا به داده‌های مقطعی است و براساس توضیحات پانویس شماره ۳۴، در تحلیل مسایل انتخاب و درونزایی نیز مطرح است که می‌تواند نتایج آزمون فرضیه را متأثر سازد. در این رابطه، بررسی‌ها و آزمون‌های آماری بایسته انجام و دریافتیم که واریانس ناهمسانی و رفع آن هیچ تأثیری بر نتایج تجربی این تحقیق ندارد.

1. Sample-selection

است. همان‌طور که ملاحظه می‌شود، میزان نرخ بازده آموزش برای نمونه استان تهران حدود ۸ درصد (از ۹/۳ به ۱۰) و برای نمونه آمریکا حدود ۱۸ درصد (از ۸/۲ به ۹/۷) افزایش می‌یابد![1]

جدول -۲. ارزیابی آثار مسایل نمونه- انتخابی و خود- انتخابی بر بازده آموزش برای شاغلین استان تهران

الگو:	۱ (OLS)		۲ (OLS)		۳ (S.S.)(۱)		۴ (End.)(۲)		۵ (E&S.S.)(۳)	
متغیر	ضریب	آماره t	ضریب	آماره t	ضریب	آماره t	ضریب	آماره t	ضریب	آماره t
آموزش (سال)	۰/۰۹۳	۱۸/۱	۰/۰۸۸	۱۶/۳	۰/۰۸۶	۱۵/۸	۰/۱۰۰۱	۱۲/۲	۰/۱۰۰۴	۱۲/۳
تجربه (سال)	۰/۰۷۰	۱۴/۸	۰/۰۵۶	۸/۱	۰/۰۵۴	۷/۸	۰/۰۵۷۰	۸/۳	۰/۰۵۵۶	۸/۰
مربع تجربه	-۰/۰۰۱	-۱۰/۰	-۰/۰۰۱	-۶/۳	-۰/۰۰۱	-۶/۰	-۰/۰۰۰۶	-۶/۱	۰/۰۰۰۶	-۵/۸
عرض از مبدأ	۶/۶۱۸	۷۸/۷	۶/۹۱۹	۵۶/۶	۷/۰۱۷	۵۴/۳	۶/۷۵۵۴	۴۶/۰	۶/۸۱۹۲	۴۳/۹
تعداد مشاهدات	۱۴۹۹		۱۰۷۳		۲۳۵۰		۱۰۷۳		۲۳۵۰	
مشاهدات سانسورنشده					۱۰۷۳				۱۰۷۳	
ضریب تعیین- $\overline{R}^2$	۰/۲۵۹		۰/۲۳۲۳		-		۰/۲۲۸۴		۰/۲۳۰۵	
آماره F	۱۷۵/۸		۱۰۹/۱۰		-		۷۰/۴۰		۵۸/۹۸	
آماره خی‌دو Wald					۳۰۴					

توضیح: متغیر وابسته لگاریتم دریافتی ساعتی است. داده‌های آماری به جامعهٔ استان تهران در مقطع سال ۱۳۸۰ مربوط می‌شود که از نتایج آمارگیری از ویژگی‌های اقتصادی- اجتماعی خانوار مرکز آمار ایران استفاده شده است. الگوی ۱ برای کل مشاهدات (تمام افراد شاغل) و الگوهای ۲ تا ۵ برای شاغلین سرپرست خانوار دارای همسر برآورد شده است.

(۱) الگوی ۳ با در نظر گرفتن مسأله نمونه- انتخابی برای «مشارکت در بازار کار» برآورد شده است. مشارکت در بازار کار تابعی از تحصیلات، تعداد افراد خانواده و جنسیت در نظر گرفته شده است.

(۲) الگوی ۴ با در نظر گرفتن مسأله درونزا بودن (یا خود- انتخابی) برای متغیر تحصیلات برآورد شده است؛ برای رفع این مسأله، از تحصیلات همسر به عنوان «متغیر ابزاری» استفاده شده است.

(۳) الگوی ۵ مشتمل بر مسایل انتخاب (اشتغال در بازار کار) و درونزایی (تصمیم آموزش) یعنی ادغام الگوهای ۳و ۴ است.

---

۱. در سال‌های گذشته و در نتیجهٔ تحولات سریع فناوری (به‌ویژه در حوزه IT) انتظار می‌رود نرخ بازده آموزش در جوامعی مانند آمریکا، افزایش قابل توجهی تجربه کرده باشد. با این حال، نرخ بازده سال ۱۹۹۱حاصل از الگوی ۵ جدول ۳ بسیار نزدیک و قابل مقایسه با برآوردهای مطالعات اخیر آن کشور (۱۰٪ ، از: ساخاروپولوس و پاترینوس، ۲۰۰۴:۱۳۴) است.

الگوهای ۳ و ۴، مسایل مورد بحث را به طور جداگانه مورد ارزیابی قرار می‌دهند. در الگوی ۵، لحاظ کردن دو مسالهٔ "انتخاب و درونزایی" به طور همزمان مورد تحلیل و ارزیابی قرار می‌گیرد. تأثیر همزمان مسایل خود- انتخابی و نمونه- انتخابی به صورت افزایش بازده آموزش برای نمونه شاغلین استان تهران و همچنین برای نمونه شاغلین جامعه آمریکا است. بدین ترتیب، نادیده گرفتن این مسایل (یعنی اتکا به برآوردهای حداقل مربعات معمولی) به کم‌برآوردی نرخ بازده در نمونه‌های مطالعه‌شده (استان تهران و جامعه آمریکا) منجر می‌شود؛ از همین رو، تحلیل‌های تجربی ادعای تحلیل‌های نظری مسالهٔ "انتخاب" را (به‌ویژه از حیث تورش‌دار بودن برآوردهای حداقل مربعات) به روشنی تأیید می‌نماید.

جدول -۳. ارزیابی آثار مسایل نمونه- انتخابی و خود- انتخابی بر بازده آموزش برای جامعه آمریکا

الگو	۲ (OLS)		۳ (S.S.)[۱]		۴ (End.)[۲]		۵ (E&S.S.)[۳]	
متغیر	ضریب	آماره t	ضریب	آماره t	ضریب	آماره t	ضریب	آماره t
آموزش (سال)	۰/۰۸۲۴	۲۸/۵	۰/۰۸۳۴	۲۰/۸	۰/۰۹۷۴	۱۹/۱	۰/۱۰۷۳	۱۳/۹
تجربه (سال)	۰/۰۳۵۳	۱۲/۱	۰/۰۳۵۵	۱۱/۸	۰/۰۳۴۰	۱۱/۵	۰/۰۳۶۹	۱۲/۲
مربع تجربه	-۰/۰۰۰۵	-۸/۹	-۰/۰۰۰۵	-۸/۳	-۰/۰۰۰۵	-۷/۹	-۰/۰۰۰۶	-۸/۹
عرض از مبدأ	۰/۹۸۴۱	۱۹/۷	۰/۹۵۹۲	۱۱/۴	۰/۷۸۲۹	۱۰/۴	۰/۵۳۶۹	۳/۸
تعداد مشاهدات	۴۰۱۱		۴۸۶۰		۴۰۱۱			
مشاهدات سانسورنشده			۴۰۱۱				۴۰۱۱	
ضریب تعیین- $\overline{R}^2$	۰/۲۰۲۳		-		۰/۱۹۷		۰/۱۹۴	
آماره **F**			-		۱۹۲		۱۷۸	
آماره خی‌دو Wald			۷۲۶					

توضیح: متغیر وابسته، لگاریتم دریافتی ساعتی برای شاغلین مرد متاهل آمریکا (از نتایج سرشماری جمعیت سال ۱۹۹۱ آن کشور) است.

[۱] الگوی ۳ با در نظر گرفتن مسأله نمونه- انتخابی از جهت «انتخاب مشارکت در بازار کار» برآورد شده است؛ مشارکت در بازار کار تابعی از تحصیلات و تعداد فرزندان در نظر گرفته شده است.

[۲] الگوی ۴ با در نظر گرفتن مسأله درونزا بودن (یا خود- انتخابی) برای متغیر «تحصیلات» برآورد شده است؛ برای رفع این مسأله، از تحصیلات همسر به عنوان متغیر ابزاری استفاده شده است.

[۳] الگوی ۵ مشتمل بر مسایل انتخاب (اشتغال در بازار کار) و درونزایی (تصمیم آموزش) یعنی ادغام الگوهای ۳و ۴ است.

۵. جمع‌بندی و نتیجه‌گیری

ارزیابی بازده آموزش، سابقه طولانی در تحقیقات اقتصاد آموزش دارد؛ اما، صحت و دقت ارزیابی‌ها از ناحیه مسایل روش‌شناسی متأثر از ماهیت تصمیم‌گیری‌های آموزشی، مورد مناقشه جدّی بوده است. یکی از مهم‌ترین مناقشات که از نیمه دوم دهۀ ۱۹۷۰ میلادی به عنوان یک جریان اساسی مطرح شده، مسایل "انتخاب و درونزایی" و تأثیر آن بر برآوردهای بازده آموزش (حاصل از روش حداقل مربعات) است. به‌رغم اقبال و توجه گسترده، به طور مقایسه‌ای آثار تجربی این مسایل مطالعه‌نشده و مهم‌تر اینکه، در مطالعات تجربی ایران، این مسایل به طور کلی مورد غفلت قرار گرفته است. برای رفع این نقصان‌ها، این پژوهش با به‌کارگیری مبانی نظری-تحلیلی مسایل انتخاب و درونزایی (یعنی رگرسیون شرطی و فن متغیر ابزاری) و داده‌های استخراج شده از آمارگیری مرکز آمار در مورد ویژگی‌های اقتصادی و اجتماعی خانوار (سال ۱۳۸۰ استان تهران) و همچنین یک نمونه از سرشماری جمعیت ۱۹۹۱ جامعه آمریکا، تأثیر مسایل انتخاب و درونزایی بر برآوردهای بازده آموزش برای شاغلین استان تهران را مورد ارزیابی قرار داده و نتایج به‌دست آمده را با یافته‌های نمونۀ آماری جامعه آمریکا مورد مقایسه قرار داده است. یافته‌ها نشان می‌دهند که: ۱) آموزش یک سرمایه‌گذاری پُر بازده است، ۲) لحاظ کردن مسایل انتخاب و درونزایی بازده آموزش را مطابق تحلیل‌ها و انتظارات نظری، متأثرکرده و آن‌را به طور قابل ملاحظه‌ای (حدود ۸ درصد در ایران و ۱۸ درصد در آمریکا) افزایش می‌دهد، و ۳) نتایج نمونۀ استان تهران کاملاً منطبق با یافته‌های نمونه استفاده شده از جامعه آمریکا است؛ بدین ترتیب، الگوواره یافته‌ها مشابه و مطابق با تحلیل‌های نظری است. یعنی برآوردهای بازده آموزش که مسایل انتخاب و درونزایی را نادیده‌گرفته تورش‌دار و ناسازگار هستند و لذا سیاستگذاری بر پایه آنها می‌تواند گمراه‌کننده باشد و آثار نامطلوب جدّی بر تخصیص منابع در یک جامعه به جای گذارد. به بیان دیگر، نادیده‌گرفتن آثار این مسایل و لذا اتکای تصمیم‌ها و سیاستگذاری‌ها به نرخ بازدهی که با کم‌برآوردی مواجه است، باعث کم‌توجهی به انباشت مناسب سرمایه انسانی در جامعه خواهد شد و این مسأله می‌تواند فرایند توسعه در تمام سطوح (کلان، منطقه و خُرد) را کُند و بلکه متوقف نماید. از همین رو، هدف اصلی در حوزه سیاستگذاری باید بر کشف و درک "واقعیت‌ها" متمرکز باشد که مسایل روش‌شناسی از قبیل مسایل انتخاب ودرونزایی، همان‌طور که نشان داده شد، نقشی تعیین‌کننده در این رابطه دارند.

منابع

افشاری، زهرا (پاییز۱۳۷۷). بررسی اثر تحصیلات عالی و تجربه بر دریافتی‌های شاغلان (بخش خصوصی-دولتی) در ایران با استفاده از مدل مینسر. فصلنامه پژوهش و برنامه‌ریزی در آموزش عالی. شماره ۱۷.

رضوی، حسین (۱۳۵۷). بازده اقتصادی سرمایه‌گذاری آموزشی در سطوح مختلف. تهران. سازمان برنامه و بودجه.

صالحی، محمدجواد (بهار و تابستان ۱۳۸۴). محاسبه بازده سرمایه انسانی در ایران. فصلنامه پژوهش و برنامه‌ریزی آموزش عالی. شماره ۳۵ و ۳۶ صفحات ۱۳۹-۱۶۶.

مرکز آمار (۱۳۸۱). آمارگیری از ویژگی‌های اجتماعی اقتصادی خانوار. تهران. مرکز آمار ایران.

نادری، ابوالقاسم (فروردین و اردیبهشت ۱۳۸۰). نقد مبانی روش‌شناسی مطالعات تجربی نظریه سرمایه انسانی و ارائه راه حل مناسب. مجله برنامه و بودجه، شماره ۶۰ و ۶۱ . صفحات۴۱-۸۴.

نادری، ابوالقاسم (۱۳۸۳). تحلیل تطبیقی بازده سرمایه انسانی در بازار کار آموزش و پرورش ایران. مجموعه مقالات همایش ملی مهندسی اصلاحات در آموزش و پرورش (۲۰ و ۲۱ خرداد ۱۳۸۱). تهران. پژوهشکده تعلیم و تربیت.

نادری، ابوالقاسم (۱۳۸۳). اقتصاد آموزش. تهران. نشر یسطرون.

نادری، ابوالقاسم (۱۳۸۷). بازده آموزش عالی: ارزیابی آثار مسالهٔ خود- انتخابی. فصلنامه پژوهش و برنامه‌ریزی در آموزش عالی (*زیر چاپ*).

نفیسی، عبدالحسین (ویراستار)(۱۳۸۰). دانشنامه اقتصاد آموزش و پرورش، ج. ۱. تهران: پژوهشکده تعلیم و تربیت.

Belzil, Christian (2007). The Return to Schooling in Structural Dynamic Models: A Survey. European Economic Review 51:PP 1059–1105.

Black, Dan A., & Jeffrey A. Smith (2004). How Robust is the Evidence on the Effects of College Quality? Evidence from Matching. J. of Econometrics 121 (2004):PP 99 – 124.

Card, D. (2001) .Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. Econometrica Vol. 69 (5):PP 1127-1160.

Dougherty Christopher (2002). Introduction to Econometrics. Oxford, Oxford Un. Press.

Fair, R. (1977). A Note on Computation of the Tobit Estimator. Econometrica, 45:PP 1723–27.

Greene, W. H. (1993, 2003). Econometric Analysis. New Jersey, Prentice-Hall, Inc.

Gronau, R. (1974). Wage Comparisons: A Selectivity Bias. J. of Pol. Economy, 82(6):PP 1119-43.

Hamermesh, Daniel S. & Stephen G. Donald (2004). The Effect of College Curriculum on Earnings: Accounting for Non-Ignorable Non-Response Bias. NBER Working Paper No. 10809, September, Cambridge, MA.

Frazis, Harley (1993). Selection Bias and the Degree Effect. Journal of Human Resources, Vol. 28, No. 3. (Summer, 1993), pp 538-554.

Harmon, C., H. Oosterbeek, & I. Walker (2000). The Returns to Education: A Review of Evidence. Issues and Deficiencies in the Literature. London, Centre for Economics of Education, LSE.

Hausman, J. A., & D. A. Wise (1977). Social Experimentation, Truncated Distributions, and Efficient Estimation. Econometrica, 45(4):PP 919-938.

Heckman, James J., Lance J. Lochner, & Petra E. Todd (2005). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, Mimeo.

Heckman, James J., *et. al.* (2003). Fifty Years of Mincer Earnings Regressions. NBER, Working Paper (9732, May).

Heckman, J. (1979). Sample Selection Bias as a Specification Error. Econometrica 47(1):PP 153-61.

Kenny, L. W., L.F. Lee, G. S. Maddala, & R. P. Trost (1979). Returns to College Education: An Investigation of Self-selection Bias Based on the Project Talent Data. Int'l Economic Review, 20 (3):PP 775-789.

Maddala, G. S. (1983). Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, Cambridge University Press.

Mincer, Jacob (1979). Human Capital and Earnings. Economic Dimensions of Education. D. M. Windham , National Academy of Education:PP 1-31.

Mincer, Jacob (1974). Schooling, Experience, and Earnings. New York, Columbia Un. Press.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. J. Pol. Economy, 66(4, August):PP 281-302.

Moretti, Enrico (2004). Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-sectional Data. J. of Econometrics 121:PP 175 – 212.

Naderi, A. and J. Mace (2003). Education and Earnings: A Multilevel Analysis. Economics of Education Rev. 22(2):PP 143-56.

Psacharopoulos, G. (1981). Returns to Education: An Updated International Comparison. In: The Economic Value of Education: Studies in the Economics of Education. M. Blaug. Hants, Edward Elgar: PP 321-41.

Psacharopoulos, G. and H. A. Patrinos (2004).Returns to Investment in Education: A Further Update. Education Economics 12(2):PP 111-134.

Psacharopoulos, G. & Richard Layard (1979). Human Capital and Earnings: British Evidence and a Critique. Reprinted in: (1992) The Economic value of Education:

Studies in the Economics of Education. M. Blaug. Hants, Edward Elgar Publication, Ltd.:PP 458-503.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. Econometrica 26:PP 24-36.

Willis, R. J. and S. Rosen (1979). Education and Self-Selection. J. of Political Economy 87 (5, part 2): S7 -S36.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press.