7

The propensity score with continuous treatments

Keisuke Hirano and Guido W. Imbens[1]

7.1 Introduction

Much of the work on propensity score analysis has focused on the case in which the treatment is binary. In this chapter, we examine an extension to the propensity score method, in a setting with a continuous treatment. Following Rosenbaum and Rubin (1983a) and most of the other literature on propensity score analysis, we make an unconfoundedness or ignorability assumption, that adjusting for differences in a set of covariates removes all biases in comparisons by treatment status. Then, building on Imbens (2000) we define a generalization of the binary treatment propensity score, which we label the generalized propensity score (GPS). We demonstrate that the GPS has many of the attractive properties of the binary treatment propensity score. Just as in the binary treatment case, adjusting for this scalar function of the covariates removes all biases associated with differences in the covariates. The GPS also has certain balancing properties that can be used to assess the adequacy of particular specifications of the score. We discuss estimation and inference in a parametric version of this procedure, although more flexible approaches are also possible.

We apply this methodology to a data set collected by Imbens, Rubin, and Sacerdote (2001). The population consists of individuals winning the Megabucks

[1]Department of Economics, University of Arizona, Tuscon, Ariz., and Department of Economics and Department of Agricultural and Resource Economics, University of California, Berkeley. Financial support for this research was generously provided through NSF grants SES-0226164 and SES-0136789.

Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.
Edited by A. Gelman and X-L. Meng ISBN: 0-470-09043-X

lottery in Massachusetts in the mid-1980s. We are interested in the effect of the amount of the prize on subsequent labor earnings. Although the assignment of the prize is obviously random, substantial item and unit nonresponse led to a selected sample in which the amount of the prize is no longer independent of background characteristics. We estimate the average effect of the prize adjusting for differences in background characteristics using the propensity score methodology, and compare the results to conventional regression estimates. The results suggest that the propensity score methodology leads to credible estimates that can be more robust than simple regression estimates.

7.2 The basic framework

We have a random sample of units, indexed by $i = 1, \ldots, N$. For each unit i, we postulate the existence of a set of potential outcomes, $Y_i(t)$, for $t \in \mathcal{T}$, referred to as the unit-level dose–response function. In the binary treatment case, $\mathcal{T} = \{0, 1\}$. Here we allow $\mathcal{T}$ to be an interval $[t_0, t_1]$. We are interested in the average dose–response function, $\mu(t) = \mathrm{E}[Y_i(t)]$. For each unit i, there is also a vector of covariates X_i, and the level of the treatment received, $T_i \in [t_0, t_1]$. We observe the vector X_i, the treatment received T_i, and the potential outcome corresponding to the level of the treatment received, $Y_i = Y_i(T_i)$.

To simplify the notation, we will drop the i subscript in the sequel. We assume that $\{Y(t)\}_{t\in\mathcal{T}}$, T, X are defined on a common probability space, that T is continuously distributed with respect to Lebesgue measure on $\mathcal{T}$, and that $Y = Y(T)$ is a well-defined random variable (this requires that the random function $Y(\cdot)$ be suitably measurable).

Our key assumption generalizes the unconfoundedness assumption for binary treatments made by Rosenbaum and Rubin (1983), to the multivalued case:

Assumption 1 (Weak Unconfoundedness) $Y(t) \perp T|X$ *for all* $t \in \mathcal{T}$.

We refer to this as weak unconfoundedness, as we do not require joint independence of all potential outcomes, $\{Y(t)\}_{t\in[t_0,t_1]}$. Instead, we require conditional independence to hold for each value of the treatment.

Next, we define the generalized propensity score.

Definition 1 (Generalized Propensity Score) *Let* $r(t, x)$ *be the conditional density of the treatment given the covariates:*

$$r(t, x) = f_{T|X}(t|x).$$

Then the generalized propensity score is $R = r(T, X)$.

This definition follows Imbens (2000). For alternative approaches to the case with multivalued treatments, see Joffe and Rosenbaum (1999a, 1999b), Lechner (2001), and Imai and van Dyk (2004).

The function r is defined up to equivalence almost everywhere. By standard results on conditional probability distributions, we can choose r such that $R = r(T, X)$ and $r(t, X)$ are well-defined random variables for every t.

The GPS has a balancing property similar to that of the standard propensity score. Within strata with the same value of $r(t, X)$, the probability that $T = t$ does not depend on the value of X. Loosely speaking, the GPS has the property that

$$X \perp 1\{T = t\} | r(t, X).$$

This is a mechanical implication of the definition of the GPS, and does not require unconfoundedness. In combination with unconfoundedness, this implies that assignment to treatment is unconfounded given the generalized propensity score.

Theorem 1 (Weak Unconfoundedness Given The Generalized Propensity Score) *Suppose that assignment to the treatment is weakly unconfounded given pretreatment variables X. Then, for every t,*

$$f_T(t|r(t, X), Y(t)) = f_T(t|r(t, X)). \tag{7.1}$$

Proof. Throughout the proof, equality is taken as a.e. equality. Since $r(t, X)$ is a well-defined random variable, for each t we can define a joint law for $(Y(t), T, X, r(t, X))$. We use $F_X(x|\cdot)$ to denote various conditional probability distributions for X, and we use $f_T(t|\cdot)$ to denote conditional densities of T. Note that $r(t, X)$ is measurable with respect to the sigma-algebra generated by X. This implies, for example, that $f_T(t|X, r(t, X)) = f_T(t|X)$.

Using standard results on iterated integrals, we can write

$$\begin{aligned} f_T(t|r(t, X)) &= \int f_T(t|x, r(t, X))\, \mathrm{d}F_X(x|r(t, X)) \\ &= \int f_T(t|x)\, \mathrm{d}F_X(x|r(t, X)) \\ &= \int r(t, x)\, \mathrm{d}F_X(x|r(t, X)) = r(t, X). \end{aligned}$$

The left side of equation (7.1) can be written as:

$$f_T(t|r(t, X), Y(t)) = \int f_T(t|x, r(t, X), Y(t))\, \mathrm{d}F_X(x|Y(t), r(t, X)).$$

By weak unconfoundedness, $f_T(t|x, r(t, X), Y(t)) = f_T(t|x)$, so

$$\begin{aligned} f_T(t|r(t, X), Y(t)) &= \int r(t, x)\, \mathrm{d}F_X(x|Y(t), r(t, X)) \\ &= r(t, X). \end{aligned}$$

Therefore, for each t, $f_T(t|r(t, X), Y(t)) = f_T(t|r(t, X))$. □

When we consider the conditional density of the treatment level at t, we evaluate the generalized propensity score at the corresponding level of the treatment. In that sense, we use as many propensity scores as there are levels of the treatment. Nevertheless, we never use more than a single score at one time.

7.3 Bias removal using the GPS

In this section, we show that the GPS can be used to eliminate any biases associated with differences in the covariates. The approach consists of two steps. First, we estimate the conditional expectation of the outcome as a function of two scalar variables, the treatment level T and the GPS R, $\beta(t, r) = \mathrm{E}[Y|T = t, R = r]$. Second, to estimate the dose–response function at a particular level of the treatment we average this conditional expectation over the GPS at that particular level of the treatment, $\mu(t) = \mathrm{E}[\beta(t, r(t, X))]$. We do not average over the GPS $R = r(T, X)$; rather we average over the score evaluated at the treatment level of interest, $r(t, X)$.

Theorem 2 (BIAS REMOVAL WITH GENERALIZED PROPENSITY SCORE) *Suppose that assignment to the treatment is weakly unconfounded given pretreatment variables X. Then*
(i) $\beta(t, r) = E[Y(t)|r(t, X) = r] = E[Y|T = t, R = r]$.
(ii) $\mu(t) = E[\beta(t, r(t, X)]$.

Proof. Let $f_{Y(t)|T,r(t,X)}(\cdot|t, r)$ denote the conditional density (with respect to some measure) of $Y(t)$ given $T = t$ and $r(t, X) = r$. Then, using Bayes rule and Theorem 1,

$$f_{Y(t)|T,r(t,X)}(y|t, r) = \frac{f_T(t|Y(t) = y, r(t, X) = r) f_{Y(t)|r(t,X)}(y|r)}{f_T(t|r(t, X) = r)} = f_{Y(t)|r(t,X)}(y|r)$$

Hence,

$$\mathrm{E}[Y(t)|T = t, r(t, X) = r] = \mathrm{E}[Y(t)|r(t, X) = r].$$

But we also have

$$\begin{aligned}\mathrm{E}[Y(t)|T = t, R = r] &= \mathrm{E}[Y(t)|T = t, r(T, X) = r] \\ &= \mathrm{E}[Y(t)|T = t, r(t, X) = r] \\ &= \mathrm{E}[Y(t)|r(t, X) = r] = \beta(t, r)\end{aligned}$$

Hence, $\mathrm{E}[Y(t)|r(t, X) = r] = \beta(t, r)$, which proves part (i). For the second part, by iterated expectations, $\mathrm{E}[\beta(t, r(t, X))] = \mathrm{E}[\mathrm{E}[Y(t)|r(t, X)]] = \mathrm{E}[Y(t)]$. □

It should be stressed that the regression function $\beta(t, r)$ does not have a causal interpretation. In particular, the derivative with respect to the treatment level t does not represent an average effect of changing the level of the treatment for any particular subpopulation.

Robins (1998, 1999) and Robins, Hernan, and Brumback (2000) use a related approach. They parameterize or restrict the form of the $Y(t)$ process (and hence the form of $\mu(t)$), and call this a marginal structural model (MSM). The parameters of the MSM are estimated using a weighting scheme based on the GPS. When

the treatment is continuous these weights must be "stabilized" by the marginal probabilities of treatment. In the approach we take here, we would typically employ parametric assumptions about the form of $\beta(t, r)$ instead of $\mu(t)$, and do not need to reweight the observations.

Two artificial examples

EXAMPLE 1: Suppose that the conditional distribution of $Y(t)$ given X is

$$Y(t)|X \sim \mathrm{N}(t + X\exp(-tX), 1).$$

The conditional mean of $Y(t)$ given X is $t + X\exp(-tX)$. Suppose also that the marginal distribution of X is unit exponential. The marginal mean of $Y(t)$ is obtained by integrating out the covariate to get

$$\mu(t) = \mathrm{E}[t + X\exp(-tX)] = t + \frac{1}{(t+1)^2}.$$

Now consider estimating the dose–response function using the GPS approach. We assume that the assignment to treatment is weakly unconfounded. For illustrative purposes, we also assume that the conditional distribution of the treatment T given X is exponential with hazard rate X. This implies that the conditional density of T given X is

$$f_{T|X}(t, x) = x\exp(-tx).$$

Hence the generalized propensity score is $R = X\exp(-TX)$.

Next, we consider the conditional expectation of Y given the treatment T and the score R. By weak unconfoundedness, the conditional expectation of Y given T and X is

$$\mathrm{E}[Y|T = t, X = x] = \mathrm{E}[Y(t)|X = x].$$

Then by iterated expectations

$$\begin{aligned}\mathrm{E}[Y|T = t, R = r] &= \mathrm{E}\left[\mathrm{E}[Y|T = t, X]|\, T = t, R = r\right] \\ &= \mathrm{E}[\mathrm{E}[Y(t)|X]|T = t, R = r] \\ &= \mathrm{E}[t + X\exp(-tX)|T = t, R = r] = t + r.\end{aligned}$$

As stressed before, this conditional expectation does not have a causal interpretation as a function of t. For the final step, we average this conditional expectation over the marginal distribution of $r(t, X)$:

$$\mathrm{E}[Y(t)] = \mathrm{E}[t + r(t, X)] = t + \frac{1}{(1+t)^2} = \mu(t).$$

This gives us the dose–response function at treatment level t.

EXAMPLE 2: Suppose that the dose–response function is $\mathrm{E}[Y(t)] = \mu(t)$. Also suppose that X is independent of the level of the treatment so that we do not actually need to adjust for the covariates. Independence of the covariates and the treatment implies that the GPS $r(t, x) = f_{T|X}(t|x) = f_T(t)$ is a function only of t. This creates a lack of uniqueness in the regression of the outcome on the level of the treatment and the GPS. Formally, there is no unique function $\beta(t, r)$ such that $\mathrm{E}[Y|T = t, R = r] = \beta(t, r)$ for all (t, r) in the support of $(T, r(T))$. In practice, this means that the GPS will not be a statistically significant determinant of the average value of the outcome, and in the limit we will have perfect collinearity in the regression of the outcome on the treatment level and the GPS. However, this does not create problems for estimating the dose–response function. To see this, note that any solution $\beta(t, r)$ must satisfy

$$\beta(t, r(t)) = \mathrm{E}[Y|T = t, r(T) = r(t)] = \mathrm{E}[Y|T = t] = \mu(t).$$

Hence, the implied estimate of the dose–response function is

$$\int_x \beta(t, r(t, x)) f_X(x)\, \mathrm{d}x = \beta(t, r(t)) = \mu(t),$$

equal to the dose–response function.

7.4 Estimation and inference

In this section, we consider the practical implementation of the generalized propensity score methodology outlined in the previous section. We use a flexible parametric approach. In the first stage, we use a normal distribution for the treatment given the covariates:

$$T_i|X_i \sim \mathrm{N}(\beta_0 + \beta_1' X_i, \sigma^2).$$

We may consider more general models such as mixtures of normals, or heteroskedastic normal distributions with the variance being a parametric function of the covariates. In the simple normal model, we can estimate β_0, β_1, and σ^2 by maximum likelihood. The estimated GPS is

$$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(T_i - \hat{\beta}_0 - \hat{\beta}_1' X_i)^2\right).$$

In the second stage, we model the conditional expectation of Y_i given T_i and R_i as a flexible function of its two arguments. In the application in the next section, we use a quadratic approximation:

$$\mathrm{E}[Y_i|T_i, R_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 R_i + \alpha_4 R_i^2 + \alpha_5 T_i R_i.$$

We estimate these parameters by ordinary least squares using the estimated GPS $\hat{R}_i$.

Given the estimated parameter in the second stage, we estimate the average potential outcome at treatment level t as

$$\widehat{\mathrm{E}[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 \hat{r}(t, X_i) + \hat{\alpha}_4 \hat{r}(t, X_i)^2 + \hat{\alpha}_5 t \hat{r}(t, X_i)\right).$$

We do this for each level of the treatment we are interested in, to obtain an estimate of the entire dose–response function.

Given the parametric model we use for the GPS and the regression function one can demonstrate root-N consistency and asymptotic normality for the estimator. Asymptotic standard errors can be calculated using expansions based on the estimating equations; these should take into account estimation of the GPS as well as the α parameters. In practice, however, it is convenient to use bootstrap methods to form standard errors and confidence intervals.

7.5 Application: the Imbens–Rubin–Sacerdote lottery sample

The data

The data we use to illustrate the methods discussed in the previous section come from the survey of Massachusetts lottery winners, which is described in further detail in the chapter by Sacerdote in this volume, and in Imbens, Rubin, and Sacerdote (2001). Here we analyze the effect of the prize amount on subsequent labor earnings (from social security records), without discretizing the prize variable.

Although the lottery prize is obviously randomly assigned, there is substantial correlation between some of the background variables and the lottery prize in our sample. The main source of potential bias is the unit and item nonresponse. In the survey unit, nonresponse was about 50%. In fact, it was possible to directly demonstrate that this nonresponse was nonrandom, since for all units the lottery prize was observed. It was shown that the higher the lottery prize, the lower the probability of responding to the survey. The missing data imply that the amount of the prize is potentially correlated with background characteristics and potential outcomes. In order to remove such biases, we make the weak unconfoundedness assumption that conditional on the covariates the lottery prize is independent of the potential outcomes.

The sample we use in this analysis is the "winners" sample of 237 individuals who won a major prize in the lottery. In Table 7.1, we present means and standard deviations for this sample. To demonstrate the effects of nonresponse, we also report the correlation coefficients between each of the covariates and the prize, with the t-statistic for the test that the correlation is equal to zero. We see that many of the covariates have substantial and significant correlations with the prize.

Variable	Mean	S.D.	Corr. w/Prize	t-stat	GPS Est.	GPS SE
Intercept					2.32	(0.48)
Age	47.0	13.8	0.2	2.4	0.02	(0.01)
Years high school	3.6	1.1	−0.1	−1.4	0.02	(0.06)
Years college	1.4	1.6	0.0	0.5	0.04	(0.04)
Male	0.6	0.5	0.3	4.1	0.44	(0.14)
Tickets bought	4.6	3.3	0.1	1.6	0.00	(0.02)
Working then	0.8	0.4	0.1	1.4	0.13	(0.17)
Year won	1986.1	1.3	−0.0	−0.4	−0.00	(0.05)
Earnings year−1	14.5	13.6	0.1	1.7	0.01	(0.01)
Earnings year−2	13.5	13.0	0.1	2.1	−0.01	(0.02)
Earnings year−3	12.8	12.7	0.2	2.3	0.01	(0.02)
Earnings year−4	12.0	12.1	0.1	2.0	0.02	(0.02)
Earnings year−5	12.2	12.4	0.1	1.1	−0.02	(0.02)
Earnings year−6	12.1	12.4	0.1	1.1	−0.01	(0.01)

Table 7.1 Summary statistics and parameter estimates of generalized propensity score.

Modeling the conditional distribution of the prize given covariates

The first step is to estimate the conditional distribution of the prize given the covariates. The distribution of the prize is highly skewed, with a skewness of 2.9 and a kurtosis of 15.0. We therefore first transform the prize by taking logarithms. The logarithm of the prize has a skewness of −0.02 and a kurtosis of 3.4. We then use a normal linear model for the logarithm of the prize:

$$\log T_i | X_i \sim \mathrm{N}(\beta_0 + \beta_1' X_i, \sigma^2).$$

The estimated coefficients from this model are presented in Table 7.1.

To see whether this specification of the propensity score is adequate, we investigate how it affects the balance of the covariates. This idea is again borrowed from the analysis of binary treatment cases, in which Rosenbaum and Rubin (1983) stress the balancing properties of the propensity score. We divide the range of prizes into three treatment intervals, [0, 23], [23, 80], and [80, 485], with 79 observations in the first group, 106 in the second, and 52 in the last treatment group. For each of the thirteen covariates, we investigate the balance by testing whether the mean in one of the three treatment groups was different from the mean in the other two treatment groups combined. (Alternatively, we could carry out various joint tests to assess the covariate balance.) In Table 7.2, we report the t-tests for each of the thirteen covariates and each of the three groups. The results show a clear lack of balance, with 14 (17) of 39 t-statistics greater than 1.96 (1.645) in absolute value.

Variable	Unadjusted			Adjusted for the GPS		
	[0, 23]	[23, 80]	[80, 485]	[0, 23]	[23, 80]	[80, 485]
Age	−1.7	−0.1	2.0	0.1	0.3	1.7
Years high school	−0.9	1.7	−0.7	−0.5	0.8	−1.0
Years college	−1.2	0.7	0.5	−0.5	0.7	−0.7
Male	−3.6	0.5	4.0	−0.4	0.2	0.1
Tickets bought	−1.1	0.5	0.6	−0.7	0.7	−0.2
Working then	−1.1	−0.3	2.0	−0.0	−0.2	0.3
Year won	−0.6	2.0	−1.6	−0.1	1.1	−1.0
Earnings year−1	−1.8	−0.5	2.3	−0.3	−0.7	0.5
Earnings year−2	−2.3	−0.4	2.6	−1.0	−0.4	0.5
Earnings year−3	−2.7	−0.6	3.1	−1.4	−0.6	1.2
Earnings year−4	−2.7	−0.7	3.1	−0.9	−0.6	1.7
Earnings year−5	−2.2	−0.3	2.4	−1.1	−0.0	2.1
Earnings year−6	−2.1	−0.1	2.3	−1.5	0.4	2.2

Table 7.2 Balance given the generalized propensity score: t-statistics for equality of means.

Next, we report GPS-adjusted versions of these statistics. Take the first covariate (age), and the test whether the adjusted mean in the first group (with prizes less than 23 K) is different from the mean in the other two groups. Recall that we should have

$$X_i \perp 1\{T_i = t\} | r(t, X_i).$$

We implement this by discretizing both the level of the treatment and the GPS. First, we check independence of X_i and the indicator that $0 \leq T_i \leq 23$, conditional on $r(t, X_i)$. To implement this we approximate $r(t, X_i)$ by evaluating the GPS at the median of the prize in this group, which is 14. Thus, we test

$$X_i \perp 1\{0 \leq T_i \leq 23\} \mid r(14, X_i).$$

We test this by blocking on the score $r(14, X_i)$. We use five blocks, defined by quintiles of $r(14, X_i)$ in the group with $1\{0 \leq T_i \leq 23\}$. The five groups are defined by the GPS values for $r(14, X_i)$ in the intervals [0.06, 0.21], [0.21, 0.28], [0.28, 0.34], [0.34, 0.39], and [0.39, 0.45]. (The full range of values for the GPS $r(T, X)$ evaluated at received treatment and covariates is [0.00, 0.45], but the range of $r(14, X)$ is [0.06, 0.45].) For example, the first of these five groups, with $r(14, X_i) \in [0.06, 0.21]$ has a total of 84 observations (16 with $T_i \in [0, 23]$ and 68 with $T_i \notin [0, 23]$). Testing for equality of the average age in the first versus the other two prize groups in this GPS group gives a mean difference of −5.5 with a standard error of 2.2. In the second GPS group, with $r(14, X_i) \in [0.21, 0.28]$ there are 39 observations (16 with $T_i \in [0, 23]$ and 23 with $T_i \notin [0, 23]$), leading to

a mean difference of -3.2 (SE 5.3). In the third GPS group, with $r(14, X_i) \in [0.28, 0.34]$ there are 53 observations (15 with $T_i \in [0, 23]$ and 38 with $T_i \notin [0, 23]$), leading to a mean difference of 8.2 (SE 4.4). In the fourth GPS group, with $r(14, X_i) \in [0.34, 0.39]$ there are 36 observations (16 with $T_i \in [0, 23]$ and 20 with $T_i \notin [0, 23]$), leading to a mean difference of 4.7 (SE 3.0). In the fifth GPS group, with $r(14, X_i) \in [0.39, 0.45]$ there are 25 observations (16 with $T_i \in [0, 23]$ and 9 with $T_i \notin [0, 23]$), leading to a mean difference of 0.4 (SE 4.0). Combining these five differences in means, weighted by the number of observations in each GPS group, leads to a mean difference of 0.1 (SE 0.9), and thus a t-statistic of 0.1, compared to an unadjusted mean of -3.1 (SE 1.8) and t-statistic of -1.7.

The adjustment for the GPS improves the balance. After the adjustment for the GPS, only 2 t-statistics are larger than 1.96 (compared to 16 prior to adjustment) and 4 out of 39 are larger than 1.645 (compared to 17 prior to adjustment). These lower t-statistics are not merely the result of increased variances. For example, for earnings in year -1, the mean difference between treatment group [0, 23] and the other two is -3.1 (SE 1.7). After adjusting for the GPS, this is reduced to -0.3 (SE 0.9).

Estimating the conditional expectation of outcome given prize and generalized propensity score

Next, we regress the outcome, earnings six years after winning the lottery, on the prize T_i, and the logarithm of the score R_i. We include all second-order moments of prize and log score. The estimated coefficients are presented in Table 7.3. Again, it should be stressed that there is no direct meaning to the estimated coefficients in this model, except that testing whether all coefficients involving the GPS are equal to zero can be interpreted as a test of whether the covariates introduce any bias.

Estimating the dose-response function

The last step consists of averaging the estimated regression function over the score function evaluated at the desired level of the prize. Rather than report the dose–response function, we report the derivative of the dose–response function.

Variable	Est.	SE
Intercept	9.68	3.34
Prize	−0.03	0.03
Prize-squared/1,000	0.40	0.20
Log(score)	−3.33	3.41
Log(score)-squared	−0.28	0.46
Log(score) × prize	0.05	0.02

Table 7.3 Parameter estimates of conditional distribution of prize given covariates.

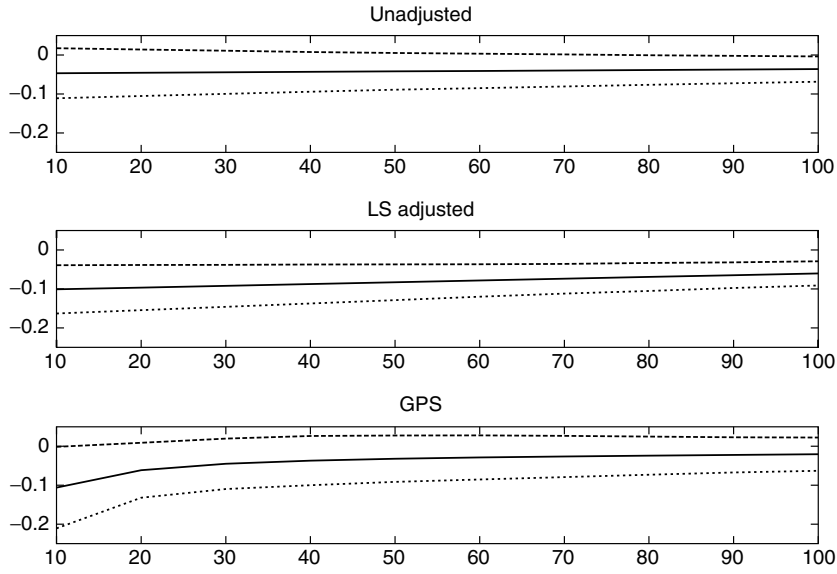

Figure 7.1 Estimated derivatives and 95% confidence bands.

In economic terminology, this is the marginal propensity to earn out of unearned income. (The yearly prize money is viewed as unearned income, and the derivative of average labor income with respect to this is the marginal propensity to earn out of unearned income.) We report the value of the derivative at $10,000 increments for all values of the yearly lottery prize between $10,000 and $100,000. The results are shown in Figure 7.1, along with pointwise 95% confidence bands. The bands are based on 1,000 bootstrap replications, taking into account estimation of the GPS.

The GPS-based estimates are compared to linear regression estimates based on a regression function that is quadratic in the prize, either without additional covariates ("unadjusted") or with the additional covariates included linearly ("LS adjusted").

The GPS estimates imply that the absolute value of the propensity to earn out of unearned income goes down sharply with the level of unearned income, from -0.10 at $10,000 to -0.02 at $100,000, suggesting that those with lower earnings are much more sensitive to income changes than those with higher earnings. The linear regression estimates suggest a much smaller change, with the derivative at a prize of $100,000 equal to -0.04, compared to -0.05 at $10,000.

7.6 Conclusion

Propensity score methods have become one of the most important tools for analyzing causal effects in observational studies. Although the original work of Rosenbaum

and Rubin (1983) considered applications with binary treatments, many of the ideas readily extend to multivalued and continuous treatments. We have discussed some of the issues involved in handling continuous treatments, and emphasized how the propensity score methodology can be extended to this case. We applied these ideas to a data set previously studied by Imbens, Rubin, and Sacerdote (2001). We expect that coming years will see further work applying the Rubin causal model approach to a range of settings.