Searching for Rehabilitation in Nonparametric Regression Models with Exogenous Treatment Assignment

Daniel J. Henderson
Esfandiar Maasoumi

DISCUSSION PAPER SERIES

I Z A

Searching for Rehabilitation in Nonparametric Regression Models with Exogenous Treatment Assignment

Daniel J. Henderson
University of Alabama
and IZA

Esfandiar Maasoumi
Emory University

ABSTRACT

Searching for Rehabilitation in Nonparametric Regression Models with Exogenous Treatment Assignment[*]

This paper offers some new directions in the analysis of nonparamertric models with exogenous treatment assignment. The nonparametric approach opens the door to the examination of potentially different distributed outcomes. When combined with cross-validation, it also identifies potentially irrelevant variables and linear versus nonlinear effects. Examination of the distribution of effects requires distribution metrics, such as stochastic dominance tests for ranking based on a wide range of criterion functions, including dollar valuations. We can identify subgroups with different treatment outcomes. We offer an empirical demonstration based on the GAIN data. In the case of one covariate (English as the primary language), there is support for a statistical inference of uniform first order dominant treatment effects. We also find several others that indicate second and higher order dominance rankings to a statistical degree of confidence.

JEL Classification: C14

Keywords: bootstrap, GAIN, nonparametric, rehabilitation, stochastic dominance, treatment effects

Corresponding author:

Esfandiar Maasoumi
Department of Economics
Emory University
Atlanta, GA 30322-2240
USA
E-mail: emaasou@emory.edu

1 Introduction

As Gary Becker (1981) has argued, "education" is not just an investment good, but also a consumption stream. An educated person will likely drive a larger stream of consumption from reading a book, or a page of NewsWeek, than an uninformed person. This greater benefit is likely related, nonlinearly, to many attributes of the individual and the characteristics of the "goods."

In a parametric model for treatment effects, fixed coefficients for the treatment variable, as well as other variables and attributes, imposes severe, and possibly inadvertent restrictions which may exclude the possibility of observing some aspects of the program or treatment. These may include different distributions of any treatment effect, different distributions of behavioral or response changes due to treatment, and others.

Consider the stylized linear in parameters model:

$$y = \beta X + Z'\gamma + u \tag{1}$$

where X is an exogenous treatment, and Z is a vector of other variables and attributes. A constant coefficient β has several implications, two of which are of immediate concern in this paper. The first is that constant coefficients force a large degree of homogeneity on the individuals irrespective of treatment level (if X is not a binary variable). Everyone in each group (treated and non-treated) has the same response. The second is that changes in X (and/or Z) have no impact on "β" or "γ." We call this a no rehabilitation assumption since it will not allow any behavior modification, either as a direct response to changes in X, or as modifications in γ arising from treatment, or different amounts of treatment. To impose these restrictions, a priori, will exclude the possibility of learning from the observed data whether there are changes in both the outcome variable, y, and perhaps sustainable changes in behavior. In other words, a model of this kind is biased toward observing temporal/local responses, at most, to treatment programs and policies. There is no redemption!

Some aspects of these restrictions may be dealt with by considering variable parameter models when data allow it (e.g., panels are available). Indeed, parameters may be allowed to depend on the observed variables in a parametric way, effectively permitting a priori specific forms of nonlinearity in treatment responses, and changes

in other coefficients. Alternatively, we could estimate nonparametric regressions, letting the data settle the degree of nonlinearity, and the form of dependence of the responses as well as attribute effects (analogous forms of β and γ).

We attempt to address some of these problems in the present study in which we consider the impact of a well known (exogenously assigned treatment) program, Greater Avenues for Independence (GAIN), on labor market outcomes. Our aim is to examine changes in the gradients which will vary continuously with the values of the variables in the model. The traditional approach in this literature focuses on the average treatment effect in the conditional distribution of y. More recent work removes this "veil of ignorance" by looking at the distribution of the treatment effects on various individuals/households, etc.

The heterogenous estimates allowed by the nonparametric approach pose new challenges. In effect, we now have a *distribution* of responses that need to be examined. We could of course report several aspects of this latter distribution, such as the mean value and quantiles of (the analogous version of β), say, for a range of observed variables and characteristics. Alternatively, dominance criteria may be used, as we intend in this work. Suppose that a job or drug treatment program is intended to be life enhancing, so that one values higher responses ("β"). This will be compatible with the class of increasing utility/valuation functions, including "dollar valuations." This is all that is needed for first order stochastic dominance (FSD) rankings. Failing to find it is as informative as when it is found to a statistical degree of confidence, using tests such as in Linton, Maasoumi and Whang (2005). When FSD is found, we do not necessarily need a cardinal valuation function to inform the decision maker that the program is effective, or failed, whatever the criterion function. Only a decision needing to quantify the impact of the treatment by a scalar value will need to select a function for that purpose. Dollar values are useful and sensible, so long as we acknowledge they reflect only one particular valuation function that gives one particular "complete" ranking. On the other hand, failing to find FSD makes it clear that decisions based on any cardinal valuation function, including dollars and averages, will inevitably assign different weights to different members of the population, and is completely subjective. Different people will legitimately differ on both the usefulness of the treatment as well as the magnitude of the effects.

2

When FSD does not hold, it may be the case that second order stochastic dominance (SSD) or higher orders hold to a statistical degree of confidence. Higher order dominance rankings are interesting when **concave** valuation functions are justified, reflecting aversion to too much "dispersion," or "inequality" in treatment outcomes, or aversion to the risk of leaving behind some seriously at risk groups, or overly benefitting the less needy.

In our work, we examine such rankings for the distribution of the "responses" which are derivatives of the nonparametric version of model (1). Our approach does not compete with quantile techniques, as such. Indeed, SD rankings are equivalent to **joint** testing of ranking "all," or a desired subset of quantiles. The difference is that, a comparison based on individual quantiles may leave one in a quandary when the outcome direction is different for different quantiles. This will be equivalent to not finding FSD on the outcomes. But with SD rankings, comparing quantiles is not the end of the road, as it were. We can look for higher order rankings with meaningful welfare theoretic interpretations that are essential to policy debate and decision making. While the treatment literature has begun to move beyond the "average treatment effect" in the distribution of the outcome variable, to our knowledge, our work is the first exploration of the distributed effects on responses.

In our empirical analysis, we find that future earnings are only (significantly) impacted by a handful of variables. Specifically, we find that enrollment in GAIN as well as higher test scores lead to higher earnings. More importantly, we find that enrollment in GAIN leads to heterogenous impacts across the sample with females having larger returns to GAIN than males, those whose primary language is English over those whose primary language is not English, older individuals over younger, those with no previous earnings over those with previous earnings and those with higher test scores over those with lower test scores on average. However, even though we see higher returns at the quartiles, we find relatively few cases of stochastic dominance. In fact, we only find one case of FSD (English as the primary language versus English not being the primary language). However, we do find SSD for those age 21 and older over those under 21, those with children over those without children, and for those with above median reading skills over those with below median reading skills. From a policy standpoint, this would suggest providing additional training in

3

English reading skills, generally, and prior to enrollment in programs such as GAIN.

The remainder of the paper proceeds as follows: Section 2 describes the stochastic dominance procedure. Section 3 briefly outlines the GAIN program while the fourth section gives the empirical results of our study. The fifth section concludes.

2   Stochastic dominance procedure

In this section we outline our stochastic dominance procedure for gradient estimates. This methodology will also work for nonlinear parametric models, but we discuss a procedure for obtaining the gradient estimates nonparametrically. In our empirical application, we employ local-linear kernel regression for mixed data (Li and Racine, 2004; Racine and Li, 2004) using $AICc$ selected bandwidth vectors (Hurvich, Simonoff and Tsai, 1998), but other regression methods and bandwidth selectors are clearly feasible. We should note here that while we have a relatively large sample of data (6460 observations), we do have a large number of covariates (14) and hence we should keep the curse of dimensionality in mind.

Nonparametric estimation generates unique gradient estimates for each observation (individual) for each variable. This feature of nonparametric estimation enables us to compare (rank) several distributed effects of the exogenous treatment for subgroups and make inferences about who benefits most from the treatment. Here we propose using stochastic dominance tests for empirical examination of such comparisons.[1] The comparison of the effectiveness of a policy on different subpopulations based on a particular index (such as a conditional mean) is highly subjective; different indices may yield substantially different conclusions. Quantile regressions offer a limited solution which can be conclusive only when first order dominance holds. In contrast, finding different orders of stochastic dominance provides uniform ranking regarding the impact of the policy among different groups and offers robust inferences. It is known to be simpler and more powerful than the corresponding tests of joint ranking of simple/marginal quantiles (see Maasoumi, 2001).

To proceed, consider a nonparametric version of the treatment regression

$$y = m(X, Z) + u,$$

_____

[1]For an empirical application of stochastic dominance tests on estimated outcome values obtained via nonparametric regression see Maasoumi, Racine and Stengos (2007).

where $m(\cdot)$ is an unknown smooth function of (the exogenous treatment) X and (covariates) Z. We are particularly interested in the change in the conditional expectation of y with respect to a change in the exogenous treatment variable X. We will denote this change as $\beta(X)$ $(= \nabla_X m(X, Z))$, but wish to emphasize that (as with all nonlinear regression functions with interactions) this gradient will likely depend on the values taken by the control variables Z. While it is possible to fix these control variables at their means (or other values), we prefer to allow them to remain at their individual observed values both because employing fixed values for Z would result in counterfactual estimates not representing any particular individual (see Henderson, Kumbhakar and Parmeter, 2012 for a discussion on the problems of such methods) and because in our case X is binary and thus fixing the Z would lead to scalar estimates and not allow for a distributional analysis.

If distinct and known groups are selected within the sample, we can examine the differences in returns between any two groups, say w and v. Here w and v might refer to males and females, respectively. Denote $\beta_w(X)$ as the effect of the treatment specific to an individual in group w. $\beta_v(X)$ is defined similarly. Again, note that the remaining covariates are not constrained to be equal across or within groups.

In practice, the actual treatment effect is unknown, but the nonparametric regression gives us an estimate of this effect. $\{\widehat{\beta}_{w,i}(X)\}_{i=1}^{N_w}$ is a vector of N_w estimates (one for each individual in group w) of $\beta_w(X)$ and $\{\widehat{\beta}_{v,i}(X)\}_{i=1}^{N_v}$ is an analogous vector of estimates of $\beta_v(X)$. $F[\beta_w(X)]$ and $G[\beta_v(X)]$ represent the cumulative distribution functions of $\beta_w(X)$ and $\beta_v(X)$, respectively.

Consider the null hypotheses of interest as

Equality of distributions :

$$F[\beta(X)] = G[\beta(X)] \quad \forall \beta(X). \tag{2a}$$

First order stochastic dominance : F dominates G if

$$F[\beta(X)] \leq G[\beta(X)] \quad \forall \beta(X), \tag{2b}$$

Second order stochastic dominance : F dominates G if

$$\int_{-\infty}^{\beta(X)} F(t)dt \leq \int_{-\infty}^{\beta(X)} G(t)dt \quad \forall \beta(X), \tag{2c}$$

Third order stochastic dominance : F dominates G if

$$\int_{-\infty}^{\beta(X)} \int_{-\infty}^{s} F(t)dt \ ds \leq \int_{-\infty}^{\beta(X)} \int_{-\infty}^{s} G(t)dt \ ds \quad \forall \beta(X), \tag{2d}$$

and so on. To test the null hypotheses, we define the empirical cumulative distribution function for $\beta_w(X)$ as

$$\widehat{F}[\beta_w(X)] = \frac{1}{N_w} \sum_{i=1}^{N_w} 1\left[\widehat{\beta}_{w,i}(X) \leq \beta_w(X)\right], \tag{3}$$

where $1[\cdot]$ denotes the indicator function and $\widehat{G}[\beta_v(X)]$ is defined similarly. Next, we define the following Kolmogorov-Smirnov statistics ;

$$T_{EQ} = \max \left(\begin{array}{c} \left\{\widehat{F}[\beta(X)] - \widehat{G}[\beta(X)]\right\}, \\ \left\{\widehat{G}[\beta(X)] - \widehat{F}[\beta(X)]\right\} \end{array} \right); \tag{4a}$$

$$T_{FSD} = \min \left(\begin{array}{c} \max\left\{\widehat{F}[\beta(X)] - \widehat{G}[\beta(X)]\right\}, \\ \max\left\{\widehat{G}[\beta(X)] - \widehat{F}[\beta(X)]\right\} \end{array} \right); \tag{4b}$$

$$T_{SSD} = \min \left(\begin{array}{c} \max\left\{\int_{-\infty}^{\beta(X)} \left[\widehat{F}(t) - \widehat{G}(t)\right] dt\right\}, \\ \max\left\{\int_{-\infty}^{\beta(X)} \left[\widehat{G}(t) - \widehat{F}(t)\right] dt\right\} \end{array} \right); \tag{4c}$$

$$T_{TSD} = \min \left(\begin{array}{c} \max\left\{\int_{-\infty}^{\beta(X)} \int_{-\infty}^{s} \left[\widehat{F}(t) - \widehat{G}(t)\right] dt \ ds\right\}, \\ \max\left\{\int_{-\infty}^{\beta(X)} \int_{-\infty}^{s} \left[\widehat{G}(t) - \widehat{F}(t)\right] dt \ ds\right\} \end{array} \right); \tag{4d}$$

for testing the equality, first order stochastic dominance (FSD), second order dominance (SSD) and third order dominance (TSD), respectively.

Consistent estimation of $\beta(X)$ does not require us to split the sample for groups w and v, but our bootstrap procedure does. Specifically, we suggest to split the sample into two distinct groups and run separate nonparametric regressions on each (including estimating bandwidths for each group separately). These estimates of $\beta(X)$ will also be consistent (this is analogous to running separate regressions for a Chow test) and will allow us to compare the distributions of the two groups without the information from one affecting the other. In essence, this is equivalent to setting the bandwidth on the variable we are comparing (say gender) to zero (which will occur asymptotically, in any case).

Based on these estimates, we can construct our test statistics in (4a) – (4d). The asymptotic distributions of these nonparametric statistics are generally unknown because they depend on the underlying distributions of the data. We propose re-sampling approximations for the empirical distributions of these test statistics to overcome this problem. Our bootstrap strategy is as follows:

(i) Using nonparametric regression methods, obtain the estimates of $\beta(X)$ $(\widehat{\beta}(X) = \nabla_X \widehat{m}(X, Z))$ for each group.

(ii) Let T be a generic notation for T_{EQ}, T_{FSD}, T_{SSD} and T_{TSD}. Compute the test statistics T from the original gradient estimates $\{\widehat{\beta}_{w,1}(X), \widehat{\beta}_{w,2}(X), \ldots, \widehat{\beta}_{w,N_w}(X)\}$ and $\{\widehat{\beta}_{v,1}(X), \widehat{\beta}_{v,2}(X), \ldots, \widehat{\beta}_{v,N_v}(X)\}$.

(iii) For each observation in group w, construct the centered bootstrapped residual u^*, where $u^* = \frac{1-\sqrt{5}}{2}\left(\widehat{u} - \overline{\widehat{u}}\right)$ with probability $\frac{1+\sqrt{5}}{2\sqrt{5}}$ and $u^* = \frac{1+\sqrt{5}}{2}\left(\widehat{u} - \overline{\widehat{u}}\right)$ with probability $1 - \frac{1+\sqrt{5}}{2\sqrt{5}}$. Then construct the bootstrapped left-hand-variable as $y^* = \widehat{m}(X, Z) + u^*$ for each observation in group w. Call $\{y_i^*, X_i, Z_i\}_{i=1}^{N_w}$ the bootstrap sample. Repeat this process for group v.

(iv) Re-estimate $\beta(X)$ for each group using the same nonparametric procedure and bandwidths in (i), but replace the data with the bootstrap data obtained in (iii). Call these estimates $\widehat{\beta}^*(X)$.

(v) Compute (centered[2]) bootstrapped test statistics T_b from the bootstrapped estimates, where (for FSD, the others follow similarly)

$$
T_b = \min \left[\begin{array}{c} \max\left(\begin{array}{c} \left\{\widehat{F}^*[\beta(X)] - \widehat{G}^*[\beta(X)]\right\} \\ -\left\{\widehat{F}[\beta(X)] - \widehat{G}[\beta(X)]\right\} \end{array} \right), \\ \max\left(\begin{array}{c} \left\{\widehat{G}^*[\beta(X)] - \widehat{F}^*[\beta(X)]\right\} \\ -\left\{\widehat{G}[\beta(X)] - \widehat{F}[\beta(X)]\right\} \end{array} \right) \end{array} \right],
$$

[2]The centering of the bootstrap test statistic is performed by subtracting the initial sample estimates of the empirical CDF differences. We do not impose the null hypothesis (specifically, we do not impose the least-favorable case) in step (iii). In this way we obtain consistent estimates of the sampling distributions and coverage probabilities, with i.i.d samples. Standard results for centered bootstrap validity apply here. We have also conducted extensive experiments when the null of the least favorable case is imposed, in addition to centering on the initial test statistics themselves. Our empirical findings are generally the same.

where $\widehat{F}^{*}\left[\beta\left(X\right)\right]$ is the analogous estimate of (3) for the bootstrap estimates.

(vi) Repeat steps (iii-v) B times.

(vii) Calculate the "p-values" of the tests based on the percentage of times the centered bootstrapped test statistic is negative. Reject the null hypotheses if the p-value is smaller than some desired level α, where $\alpha \in (0, 1/2)$.

The careful reader will notice that the main departure from typical SD tests is that the data in question $(\beta(X))$ is unknown and thus must be estimated. Therefore, instead of bootstrapping from $\widehat{\beta}(X)$, it is important to bootstrap from the data and re-estimate $\beta(X)$ in each replication.[3] This allows us to to approximate the distribution of the derivatives. By re-sampling, we take into account the fact that we are dealing with the estimates of the gradients and not the actual gradients.

The most important steps above are the third through fifth. In (iii), we emphasize that we do not impose the least favorable case. Instead we separate the groups and re-sample from each separately. This can be achieved several ways (which we have done), but our preferred procedure is to use a wild bootstrap (to avoid issues with respect to potential heteroskedasticity). Then proceeding to step (iv), we re-estimate each model (using the same bandwidths as in step (i)). Note that we evaluate the bootstrapped gradient estimates at the original X and Z values. In the fifth step, we calculate the bootstrapped based test statistic by evaluating over the same grid we did in step (ii).[4]

We wish to note here that in our empirical example, the gradient in question comes from a binary regressor. Hence, we only achieve a gradient estimate for those observations for which the dummy variable is equal to unity. Therefore, we construct our empirical CDF's with fewer observations than if we had a continuous regressor, but the basic methodology remains the same.

---

[3]Eren and Henderson (2008) and Henderson (2010) simply resample the gradient estimates. If the distribution functions are sufficiently well separated, this should lead to the same conclusions, but we recommend re-estimating the gradients in practice.

[4]We performed simulations to determine the size and power of our bootstrap based test and found that it did well in relatively small samples. These results are available from the author upon request.

3 Greater Avenues for Independence

The Greater Avenues for Independence (GAIN) program was started in California in 1986 in order to help long term welfare recipients "find employment, stay employed, and move on to higher paying jobs, which will lead to self-sufficiency and independence." It is a mandatory (excluding female heads of households with children under age six) program for adults receiving Aid to Families with Dependent Children (AFDC).

The program initially administers screening tests to determine basic math and reading skills. Those deemed to be below a given level are targeted to receive basic education. Those above a given level are moved into either a job search assistance program or a vocational training program. This decision largely falls on the county with some counties preferring one over another.

Starting in 1988, a randomly assigned subset of GAIN registrants in six California counties (Alameda, Butte, Los Angeles, Riverside, San Diego and Tulare) were assigned to a treatment group and the remaining were selected into a control group. Those in the treatment group were allowed to participate in the GAIN program and the remaining were not, but were still allowed to receive standard AFDC benefits. Those in the control group were allowed, but not required, after two years, to join the GAIN program.

From the econometrician's standpoint, this data set is ideal because the participants were randomly assigned to either the treatment or the control group. Table 1 shows that for Riverside County, nearly all means are the same between the two groups, perhaps with the exception of females in the control group. The results are similar for the other counties.

We choose Riverside County for several reasons. It has been highlighted by many as the best performing county. In fact, it has often been referred to as the "Riverside Miracle" (e.g., see Nelson, 1997). This result has led many to study this case (e.g., see Dehejia, 2003) and thus our findings can be compared to past studies. Finally, the sample is relatively large, and given the large number of covariates, our estimation procedure benefits greatly from the relatively large sample size.

Although this data has been studied using rigorous econometric techniques before (e.g., Dehejia, 2003; Hotz, Imbens and Klerman, 2006), to our knowledge, no one has

used nonparametric methods. The need for these methods with this particular data set have been hinted at before. Dehejia (2003, pp. 9) mentions that "an estimator or a functional form that is more flexible in terms of pretreatment covariates should yield a more reliable prediction of the treatment impact."

In addition to having a more flexible approach, we are also able to get a treatment effect for each GAIN recipient in the program. This allows us to look at heterogeneity both across and within groups. Further, it allows us to use the stochastic dominance methods discussed earlier to look for relationships amongst the returns for pre-specified groups in order to better inform policy decisions.

4   Empirical results

We begin by looking at the cross-validated bandwidths from the regression of earnings on pre-treatment attributes (Table 2). These bandwidths can lead to knowledge about whether or not variables are relevant and whether or not they enter the model linearly. We then turn our attention to the gradient estimates (Table 3). Although our primary concern is with respect to the GAIN participation variable, we will also analyze other gradients. We then turn our focus to our primary interest. We split the sample amongst the pre-specified groups and look at their returns distributions to the GAIN program (Table 4). Finally, we perform stochastic dominance tests to determine whether or not we have first or higher-order dominance relationships (Tables 5-6).

4.1   Bandwidth estimates

Table 2 presents the bandwidths for the nonparametric model. The bandwidths reveal three salient points. First, the bandwidths on the CASAS reading and math score variables each exceed 3.94E+06. Since continuous regressors behave linearly as the bandwidths approach infinity, this suggests that a linear approximation for these two variables may be reasonable. The bandwidth on the "previous earnings" in the past 12 quarters is relatively small, indicating nonlinear effects. Employing a model which is linear in this variable would most likely lead to biased estimates. Second, the bandwidths on the treatment, gender, prior employment or training, and

ethnic group are much smaller than their respective upper bounds, implying that these variables are relevant in the model. Finally, the bandwidths on the primary language variables, as well as family status, age, highest school grade completed, number of children and random assignment month are each close to their respective upper bounds; thus, these variables are (likely) statistically irrelevant in explaining treatment effect on earnings .

In sum, examination of the bandwidths suggest that some variables are relevant and some variables are irrelevant. Further, it suggests that some variables enter the model nonlinearly and some variables enter the model linearly. However, this does not mean we should automatically switch to a semiparametric estimation procedure. Linearity is not synonymous with homogeneous effects of the covariates. Consequently, while the assumption of linearity receives the most attention, heterogeneity may be just as problematic. We now turn to the actual results, as well as more formal statistical tests.

4.2   Parameter estimates

4.2.1   All covariates

Table 3 presents the results for the gradient estimates. We present the nonparametric estimates corresponding to the 25th, 50th, and 75th percentiles of the estimated gradient distributions (labelled $Q1$, $Q2$, and $Q3$). Estimates which are statistically significant at the 5% level are listed. To conserve space, we exclude any regressor for which each of the quartiles are insignificant. The full set of estimates with corresponding standard errors is available from the authors upon request.

In terms of the unordered categorical variables, several findings stand out. First, nonparametric estimates of the treatment (enrollment in GAIN) are positive and significant at the median and upper quartile. Perhaps more important for this study is that the third quartile is over three times the value of the second quartile. This shows prevalence of heterogeneity in the effect of the treatment across the sample. Finally, while some of the bandwidths suggest relevance, we did not find significance of any of the other unordered categorical regressors at the quartile values.

Likewise, for the ordered categorical variables, none of the quartile gradient estimates are significant. Again, these results are expected because, as was observed

before, their bandwidths approached their upper bounds of unity. The implication is that they are not important in prediction of earnings. However, this does not mean that they do not play a role in terms of the impact of the treatment, as we will check later.

Finally, for the continuous variables, it is seen that CASAS reading and math scores have effects on earnings. The partial effect at the median for reading scores is 89.6598 (s.e. = 22.4854) and the partial effect at the median for the math score is 37.6296 (s.e. = 12.4125). This result suggests that improving basic reading and math scores would lead to higher earnings (with improvements in reading skills typically being more beneficial than mathematics). While the bandwidths suggest that each of these variables enter linearly, they do not shed light on possible heterogeneity. The results at the quartiles show heterogeneity in the partial effects and re-emphasize the importance of a nonlinear estimation procedure.

4.2.2 Treatment variable

The results across different covariates are interesting, but a main purpose of this study and the GAIN experiment is to determine the effect of the treatment. In most studies, a single coefficient is obtained for the (average) treatment and its magnitude determines whether or not the treatment was successful. Here we obtain a separate estimate for each person receiving the treatment. Thus, we can examine the effect of the treatment among pre-specified groups.

Table 4 gives the nonparametric estimates corresponding to the 25th, 50th, and 75th percentiles of the distribution for the treatment (GAIN) for specific sub-groups. Specifically, we broke the sample across each pre-specified group and ran separate nonparametric regressions on each sub-group (including calculating bandwidths for each). The quartile estimates for those in the GAIN program are given in the table.

The results for the groupings by unordered categorical variables are quite striking. First, the effect of the treatment on women is larger than the effect of the treatment on men at median, the first, and the third quartiles (note that the first quartile estimate for men is negative and insignificant). Although the results at the quartiles are strong, we cannot determine whether or not the effect of the treatment for women dominates the effect of the treatment for men. We will examine this further in the

next sub-section. Second, there is some evidence that individuals who received the treatment and had previous training or work experience in the prior year experienced larger returns to GAIN than treated individuals who did not have employment or training in the prior year. However, these results only hold at the median and at the upper quartile. This would suggest that no dominance relation exists. Third, Asians are the only ethnic group who did not experience significant returns at these quartiles. Finally, although the bandwidth for English or Spanish as a first language was near its upper bound, treated individuals who spoke English as their native language, had positive and significant treatment effects for enrollment in GAIN and those whose primary language was not English did not experience significant returns to GAIN at any of the quartiles. This result may suggest that immigrants are not benefitting from the program. This result is consistent with other inferences below related to spoken or written English.

For the ordered categorical variables, we see that treated individuals aged 21 and over had larger effects than treated individuals under 21. The level of schooling seemed to make little difference on who benefitted the most from the program. Finally, treated individuals who have one or more children have larger treatment effects at each quartile than individuals who did not have any children (perhaps a sign of necessity). Again, these results at these selective quartiles are strong, but it is premature to conclude that any of these groups "dominate" one another in terms of the partial effect of the treatment variable.

Finally, for groupings corresponding to the continuous variables, treated individuals with no earnings in the previous 12 quarters had larger effects of the treatment than treated individuals who had positive earnings in the previous 12 quarters at each quartile. The test scores results are as expected. Treated individuals obtaining scores above the median (either in math or reading) have larger treatment effects as compared to their counterparts who scored below the median. This shows that higher ability individuals are able to benefit more from the treatment. We return to the Gary Becker argument which we paraphrased in the Introduction: education is both an investment and a consumption good. The greater benefit of education is likely related, nonlinearly, to many attributes of the individual and the characteristics of the "goods."

13

4.3   Dominance tests

The results of the previous subsection showed that certain groups appeared to have higher returns from the treatment than other groups, at certain quantiles. Here we use formal tests to compare the effect of the treatment between two pre-specified groups across all quantiles. Tables 5-6 break down the results for tests of equality, first order, second order and third order dominance. Table 5 gives the test statistics. A negative sign of a test statistic is a sign of possibly significant dominance relation. The entries in Table 6 are the "p-values" for the corresponding tests.

4.3.1   Test statistics

In Table 5, the entries are the sample value of the test statistics. The left-hand-side of the table gives the pre-specified groups being compared. In each case we are comparing the treated individuals in each group. The first column of numbers gives the test statistic for the equality of the distributions of the gradient of the conditional mean with respect to the treatment (GAIN). The second through fourth columns give the test statistic for first, second and third order dominance, respectively. In order for a dominance relation to exist, the test statistic must be negative. For example, for the first order dominance case, if the test statistic is negative, then first order dominance is observed. If the test statistic is positive, then there is no observed ranking in the first order sense. Similar interpretations are given to higher order dominance relations.

When examining the test statistics for first order dominance, there is only the possibility of FSD for three of the 17 comparisons. The comparisons with negative FSD test statistics are: white versus Asian, primary language English versus primary language not being English, and CHASS reading score above the median versus score below the median. The lack of negative test statistics for the comparison between those with and without previous earnings may be surprising given the results at the quartiles, but these suggest crossing of the distributions closer to the tails.

As expected, more cases of second order dominance are observed. The third column of numbers in Table 5 gives the test statistics for the null of second order dominance (noting that first order dominance implies second order dominance, and so on). Here we also find negative test statistics for each ethnic group versus Asians,

those 21 and over versus those under 21, and those with children over those not having children. For third order dominance, we also find a negative test statistic for white versus black. These higher order dominance rankings imply that policy makers with an aversion, or increasing aversion to earnings poverty, would find the program to be beneficial, whatever cardinal weighting function/utility is adopted.

4.3.2 Probability values

Each value in Table 6 is the p-value associated with a particular test. The first column rejects the null of equality of the distributions of the treatment effects when the p-value is below α. In columns 2-4, the respective order of dominance is rejected (when its associated test statistic is negative) if the p-value is less than 0.400 (see Maasoumi, Millimet, and Sarkar, 2009). Substantial coverage probability for negative values of the statistic support an inference of dominance to a degree of statistical confidence.

In Table 6, we reject each null that the pairs of treatment effect estimates are equal. These results are not surprising given what we have seen thus far. For the dominance tests (in Table 5) with negative sample statistics, there are cases where there is significant evidence of dominance. The strongest ranking is the finding of first order dominance. We find that those whose primary language is English have uniformly higher returns to GAIN than those whose first language is not English. First order dominance implies higher order dominance, and we see that the p-values for second and third order dominance are larger in magnitude than that of the first order test. In two other cases where we found negative test statistics for first order dominance (white vs. Asian and above median reading score versus below median) both have p-values much less than 0.40.

We find three strong cases for second order dominance. In addition to white versus Asian, we also see that those who received the treatment and were 21 years and older gained more than those under 21; similarly, those with children versus those without children. It may be that older individuals and those who have dependents took better advantage of the program. Finally, we have one test statistic with a p-value near the border of 0.40. Reading score above the median versus reading score below the median (p-value = 0.3924) is likely related to the result of language ability. This, along with the previous results, suggest that the program may want to focus

more on basic language and reading skills.

Finally, for third order dominance, in addition to those listed above, we find a further ranking of white versus black treatment outcomes. Those with increasing aversion to inequality of earnings at the lower end of the earnings distribution, would infer a greater benefit to whites vs blacks treated in GAIN.

5   Conclusions

In this paper we outlined a method to compare gradient estimates from a nonparametric regression via stochastic dominance techniques. Our goal here was to look at the impact of an exogenous treatment across different pre-specified groups.

To showcase the methodology, we applied our procedure to the California GAIN program. Here we found that relatively few inputs commonly used in determining labor outcomes are significant. Specifically, we only found significant quartile estimates for improving earnings for enrollment in GAIN and for test scores. Although many results were insignificant, we did find that certain groups had higher returns to GAIN. For example, we found that females, those whose primary language was English, those individuals over the age of 21 and those with higher test scores had higher returns to the treatment. However, we only found one case of first order dominance: English as the primary language versus English not being the primary language. We also found some evidence of second and higher order dominance, for example, for above median versus below median reading scores. From a policy standpoint, this suggests that improving basic reading skills can increase the impact of GAIN.

An interesting extension to our work would be to calculate "collateral effects," which we define as changes to the gradients of the other regressors (Z) arising from the treatment, or different amounts of the treatment (if the treatment were continuous). These can be calculated as the cross-partial derivatives with respect to X and any element in Z. In other words, we would like to allow for the treatment to have effects on other attributes of the individual.

References

[1] Becker, Gary S. 1981. *A Treatise on the Family* (Enlarged edition). Cambridge: Harvard University Press.

[2] Dehejia, Rajeev H. 2003. "Was There a Riverside Miracle? A Hierarchical Framework for Evaluating Programs with Grouped Data." *Journal of Business and Economic Statistics*, 21, pp. 1-11.

[3] Eren, Ozkan, and Daniel J. Henderson. 2008. "The Impact of Homework on Student Achievement." *Econometrics Journal*, 11, pp. 326-348.

[4] Henderson, Daniel J. 2010. "A Test for Multimodality of Regression Derivatives with Application to Nonparametric Growth Regressions." *Journal of Applied Econometrics*, 25, pp. 458-480.

[5] Henderson, Daniel .J., Christopher F. Parmeter, and Subal C. Kumbhakar. 2012. "A Simple Method to Visualize Results in Nonlinear Regression Models." *Economics Letters*, 117, pp. 578-581.

[6] Hotz, V. Joseph, Guido W. Imbens, and Jacob A. Klerman. 2006. "Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Reanalysis of the California GAIN Program." *Journal of Labor Economics*, 24, pp. 521-566.

[7] Hurvich, Clifford M., Jeffrey S. Simonoff, and Chih-Ling Tsai. 1998. "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion." *Journal of the Royal Statistical Society, Series B*, 60, pp. 271-293.

[8] Li, Qi, and Jeffrey S. Racine. 2004. "Cross-Validated Local Linear Nonparametric Regression." *Statistica Sinica*, 14, pp. 485-512.

[9] Linton, Oliver, Esfandiar Maasoumi, and Yoon-Jae Whang. 2005. "Consistent Testing for Stochastic Dominance under General Sampling Schemes." *Review of Economic Studies*, 72, pp. 735-765, Also the Corrigendum to the same, (2007).

[10] Maasoumi, Esfandiar. 2001. "Parametric and Nonparametric Tests of Limited Domain and Ordered Hypotheses in Economics." In *A Companion to Econometric Theory* (Chapter 25), ed. Badi Baltagi, Malden: Basil Blackwell Publishers.

[11] Maasoumi, Esfandiar, Jeffrey S. Racine, and Thanasis Stengos. 2007. "Growth and Convergence: A Profile of Distribution Dynamics and Mobility." *Journal of Econometrics*, 136, pp. 483-508.

[12] Maasoumi, Esfandiar, Daniel L. Millimet, and Dipa Sarkar. 2009. "A Distributional Analysis of Marriage Effects." *Oxford Bulletin of Economics and Statistics,* 71, pp. 1-33.

[13] Nelson, Doug. 1997. "Some 'Best Practices' and 'Most Promising Models' for Welfare Reform." memorandum, Baltimore: Annie E. Casey Foundation.

[14] Racine, Jeffrey S., and Qi Li. 2004. "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data." *Journal of Econometrics,* 119, pp. 99-130.

Table 1: Descriptive Statistics

Variables By Type	All	Treatment Group	Control Group
Dependent Variable:			
Earnings	10079.5147	10696.7075	7790.6829
Unordered Categorical Variables:			
Experimental (GAINS)	0.7876	1.0000	0.0000
Female	0.6819	0.7040	0.5999
Employment or Training (Prior Year)	0.2376	0.2390	0.2325
White	0.5500	0.5513	0.5452
Not White	0.4500	0.4487	0.4548
Hispanic	0.2464	0.2435	0.2573
Black	0.1551	0.1584	0.1429
Asian	0.0334	0.0311	0.0423
Primary Language English	0.9610	0.9636	0.9512
Primary Language Spanish	0.0197	0.0181	0.0255
Ordered Categorical Variables:			
Age	32.2918	32.3143	32.2085
Highest School Grade Completed	11.1642	11.1733	11.1305
Number of Children	2.0193	2.0161	2.0313
Continuous Variables:			
Earnings Previous 12 Quarters	2335.7927	2293.0782	2494.1975
CASAS Reading Score	232.6416	232.6085	232.7646
CASAS Math Score	219.5249	219.5871	219.2945
Number of Observations	6460	5088	1372

Notes: Average values are listed. The first column of numbers is for the entire Riverside sample, the second is for the treatement group and the final is for the control group.

Table 2: Bandwidths

Variables By Type	Bandwidth	Upper Bound	Interpretation
Unordered Categorical Variables:			
Experimental (GAINS)	0.2630	0.5000	relevant
Sex	0.2382	0.5000	relevant
Employment or Training (Prior Year)	0.3451	0.5000	relevant
Ethnic Group	0.7721	0.8750	relevant
Primary Language English	0.4993	0.5000	most likely irrelevant
Primary Language Spanish	0.4993	0.5000	most likely irrelevant
Family Status	0.7590	0.8000	most likely irrelevant
Ordered Categorical Variables:			
Age	0.9986	1.0000	most likely irrelevant
Highest School Grade Completed	0.9986	1.0000	most likely irrelevant
Number of Children	0.9986	1.0000	most likely irrelevant
Random Assignment Month	0.9986	1.0000	most likely irrelevant
Continuous Variables:			
Earnings Previous 12 Quarters	1.35E-01	∞	nonlinear
CASAS Reading Score	3.94E+06	∞	most likely linear
CASAS Math Score	5.22E+07	∞	most likely linear

Notes: Bandwidths selected via AICc. Aitchison and Aitken (1976) kernel used for unordered data, Wang and van Ryzin (1981) kernel used for ordered data and second-order Gaussian kernel used for continuous data.

Table 3: Significant nonparametric gradient estimates at the quartiles

Variable	Q1	Q2	Q3
Unordered Categorical Variables:			
Treatment (GAIN)		184.1634	644.8306
Continuous Variables:			
CASAS Reading Score		89.6598	123.8145
CASAS Math Score		37.6296	60.5834

Notes: Significant gradient estimates for the first, second and third quartiles are listed above (standard errors obtained via bootstrapping are available upon request). For those variables with no significant quartiles, the estimates are excluded. For discrete regressors, the lowest value taken by the gradiet is exactly zero by definition.

Table 4: Significant returns to GAIN by group at the quartiles

Variable	Q1	Q2	Q3
Unordered Categorical Variables:			
Gender			
Female	691.3410	1125.3420	1652.9889
Male		103.3956	331.2337
Previous Training			
Employment or Training (Prior Year)	338.7188	1122.1333	1974.1487
No Employment or Training (Prior Year)	379.5400	897.4375	1457.1979
Ethnic Group			
White	173.4314	971.4750	1754.1438
Not White	202.2289	897.1133	1368.9425
Hispanic	127.9419	362.3600	504.5695
Black	385.7882	996.9691	1412.9833
Asian			
Language			
Primary Language English	554.1186	1035.6361	1513.4771
Primary Language is Not English			
Primary Language Spanish			
Ordered Categorical Variables:			
Age			
Under 21			386.0289
21 and Over	545.5721	1090.4551	1696.0552
Highest School Grade Completed			
Less Than High School		48.5136	80.2157
High School Diploma and Over		32.7699	46.8052
Number of Children			
Zero			
One or More	344.6863	723.9139	1074.3786
Continuous Variables:			
Previous Earnings			
Positive Earnings Previous 12 Quarters	281.0717	604.8498	790.5349
No Earning in Previous 12 Quarters	716.2675	1133.7201	1580.6813
Test Scores			
CASAS Reading Score Above Median	864.1227	1400.7653	1848.2009
CASAS Reading Score Below Median	225.2025	486.8775	719.7718
CASAS Math Score Above Median	481.5905	1066.8121	1674.2816
CASAS Math Score Below Median		247.4605	740.0365

Notes: Returns to GAIN for the first, second and third quartiles for particular sub-groups are listed above. Only those that are significant are listed (standard errors obtained via bootstrapping are available upon request). Each estimate is obtained by splitting the sample and running a separate regression (including cross-validation routine) on the pre-specified group.

Table 5: Stochastic dominance test statistics

Comparison	EQ	FSD	SSD	TSD
Unordered Categorical Variables:				
Female vs. Male	0.6862	0.0336	3.3447	383.2913
Previous Employment or Training vs. No Previous Employment or Training	0.1520	0.0423	11.0134	3381.1571
White vs. Not White	0.1397	0.0187	0.8180	24.1794
White vs. Black	0.1422	0.0728	3.9184	-0.0219
White vs. Hispanic	0.5074	0.0829	24.1267	7098.3774
White vs. Asian	0.6407	-0.0042	-0.0340	-0.0340
Black vs. Hispanic	0.5790	0.0824	31.0026	11533.7441
Black vs. Asian	0.6989	0.0091	-0.0120	-0.0120
Hispanic vs. Asian	0.6296	0.0275	-0.0435	-0.0435
Primary Language English vs. Primary Language Not English	0.8821	-0.0102	-0.3854	-0.3854
Primary Language Spanish vs. Primary Language Not Spanish	0.7590	0.1264	1.4339	15.6832
Ordered Categorical Variables:				
21 and Over vs. Under 21	0.5492	0.0029	-0.0230	-0.0230
High School Diploma vs. No High School Diploma	0.2761	0.0609	0.5873	6.0183
Children vs. No Children	0.7084	0.0091	-0.1010	-0.1010
Continuous Variables:				
No Earnings in Previous 12 Quarters vs. Earnings in Previous 12 Quarters	0.4951	0.0077	0.2199	14.0987
CHASS Reading Score Above Median vs. CHASS Reading Score Below Median	0.5799	-0.0014	-0.0041	-0.0041
CHASS Math Score Above Median vs. CHASS Math Score Below Median	0.4135	0.0017	0.0182	0.5662

Notes: The number in each cell is the test statistic for the comparison of the returns to enrollment in GAIN between two pre-specified groups for a particular test. The first column is a test for equality. The second through fourth columns are tests for stochastic dominance (first, second and third order, respectively). For the stochastic dominance tests, those test statistics which are negative are possible cases where dominance may exist. For the negative test statistics, the p-values in Table 6 will determine whether or not dominance exists.

Table 6: Stochastic dominance test p-values

Comparison	EQ	FSD	SSD	TSD
Unordered Categorical Variables:				
Female vs. Male	0.0000			
Previous Employment or Training vs. No Previous Employment or Training	0.0000			
White vs. Not White	0.0000			
White vs. Black	0.0000			**0.4810**
White vs. Hispanic	0.0000			
White vs. Asian	0.0000	0.0253	0.0506	0.0506
Black vs. Hispanic	0.0000			
Black vs. Asian	0.0000		0.2025	0.2405
Hispanic vs. Asian	0.0000		0.0886	0.0886
Primary Language English vs. Primary Language Not English	0.0000	**0.8734**	**0.8861**	**0.8861**
Primary Language Spanish vs. Primary Language Not Spanish	0.0000			
Ordered Categorical Variables:				
21 and Over vs. Under 21	0.0000		**0.8228**	**0.8608**
High School Diploma vs. No High School Diploma	0.0000			
Children vs. No Children	0.0000		**0.7722**	**0.8101**
Continuous Variables:				
No Earnings in Previous 12 Quarters vs. Earnings in Previous 12 Quarters	0.0000			
CHASS Reading Score Above Median vs. CHASS Reading Score Below Median	0.0000	0.1013	0.3924	**0.5443**
CHASS Math Score Above Median vs. CHASS Math Score Below Median	0.0000			

Notes: The number in each cell is the p-value for the comparison of the returns to enrollment in GAIN between two pre-specified groups for a particular test. For the stochastic dominance tests (columns 2-4), the p-value is included only if the corresponding test statistic in Table 5 is negative. Cases where we fail to reject the null of dominance are listed in bold. 399 bootstrap replications are performed for each SD test.