A rank similarity test for quantile treatment effects in conjunction with propensity score matching: An application to crop yield impacts of agricultural credit

Sumedha Shukla

Indraprastha Institute of Information Technology, Delhi, New Delhi, India

Email: sumedhas@iiitd.ac.in

Gaurav Arora

Indraprastha Institute of Information Technology, Delhi, New Delhi, India

Email: gaurav@iiitd.ac.in

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31 - August 2

A rank similarity test for quantile treatment effects in conjunction with propensity score matching: An application to crop yield impacts of agricultural credit

Abstract:

Distributional heterogeneity of treatment effects is an important consideration for impact evaluation. Quantile regression methods are employed to analyse this heterogeneity by comparing the distance between quantiles of an outcome variable across treated and control groups. However, by virtue of the non-linearity of the quantile function the difference-in-quantiles is *not equal to* the quantiles-of-difference in the outcome variable unless we assume that a representative subject's relative outcome-value or rank is same across counterfactual treatment and control states. This identification condition is termed as *rank invariance or rank preservation*. Rank invariance is statistically untestable and can be relaxed to allow for random deviations of each subject's rank away from her counterfactual self so long as the ranks distribution (conditional on factors influencing treatment status) is identical across treatment states - termed as *rank similarity*. Testing for rank similarity is tricky because the joint distribution of potential outcomes is unknown. Recently, Frandsen and Lefgren (2018) and Dong and Shen (2018) proposed tests for rank similarity where the endogenous treatment variable was modeled using auxiliary instrumental variable regressions. Here, we build on these studies to propose a method for testing rank similarity when treatment endogeneity is resolved by employing a propensity score matching estimator. Specifically, we employ the Kolmogorov-Smirnov distribution test and showcase the proposed rank similarity test through an evaluation of the impact of agricultural credit on crop yield distribution.

JEL Codes: C180, C310, Q140

Keywords: Quantile Treatment Effect, Rank Similarity Condition, Kolmogorov-Smirnov Distribution Test

1. Introduction:

Distributional heterogeneity of treatment effects is an important consideration for impact evaluation (Heckman et al. 1997; Angrist 2004; Bitler et al. 2008; Dammert 2008; Eren and Ozbeklik 2014; Wenz 2019). Traditionally, examination of mean impact has been the natural focal point of treatment effect evaluation (Carneiro et al. 2002; Angrist and Pischke 2009). But an exclusive focus on the mean impact assumes that the distributional aspects of the program are either unimportant (because the impact is identical across units) or can be offset by transfers – neither of which are "attractive" assumptions (Heckman et al. 1997, p. 520). In fact, there is a large and growing body of empirical evidence which suggests heterogeneity in subjects' responses to treatment highlighting the insufficiency of the average-representative-subject paradigm in its ability to approximate reality (Heckman et al. 1997; Carneiro et al. 2003; Bitler et al. 2006; Bitler et al. 2008; Cunha et al. 2010; Bedoya et al. 2018). Quantile regression methods are employed to analyze the potential heterogeneity in treatment effects by comparing the distance between quantiles of an outcome variable across treated and control groups (Koenecker and Basset 1978; Hao and Naiman 2007). However, by virtue of the non-linearity of the quantile function, the difference-in-quantiles is not equal to the quantile-of-difference in the outcome variable unless we assume that a representative subject's relative outcome-value or rank is the same across counterfactual treatment and control states (Imbens and Woolridge 2009). This identification condition is termed *rank invariance* or *rank preservation* (Heckman et al. 1997) and is a statistically "strong" and untestable assumption (Firpo 2007; Chernozhukov & Hansen 2005).

Formally, let Y denote the outcome of interest and F_Y be its unconditional cumulative density function. Our interest lies in exploring the effect of a binary treatment $D \in \{0,1\}$ on Y. Let $Y = h(D, \vec{X}, \vec{V})$ where $\vec{X}$ is the vector of observable determinants of the outcome and $\vec{V}$ is the vector of unobserved factors influencing the outcome. The conditional outcome

distribution is then given by $F_{Y|D,\vec{X},\vec{V}}$. Further, $Q(\tau) = \inf\{y : F_Y(y) \geq \tau\}$ and $Q(\tau \mid D, \vec{X}, \vec{V}) = \inf\{y : F_{Y|D,\vec{X},\vec{V}}(y \mid D, \vec{X}, \vec{V}) \geq \tau\}$ denote the unconditional and conditional quantiles functions of the outcome and $\tau \in (0,1)$. The unconditional quantile treatment effect (UQTE) for the τ^{th} quantile is defined as (Doksum 1974):

$$UQTE(\tau) = Q(\tau \mid D = 1) - Q(\tau \mid D = 0) \quad (1)$$

And the conditional quantile treatment effect (CQTE) at the τ^{th} quantile is given by:

$$CQTE(\tau \mid \vec{X}, \vec{V}) = Q(\tau \mid \vec{X}, \vec{V}, D = 1) - Q(\tau \mid \vec{X}, \vec{V}, D = 0) \quad (2)$$

Unlike average treatment effects where the conditional and unconditional treatment effect coincide as a consequence of the law of iterated expectations (Firpo 2007), CQTE and UQTE are "distinctly different" by virtue of the non-linearity of the quantile function (Firpo 2007; Angrist and Pishke 2009; Liao and Zhao 2019). To see this, suppose the potential outcomes of the treated unit i, $Y_{1i} = Y_i \mid D_i = 1$ and $Y_{0i} = Y_i \mid D_i = 0$ under treatment and control respectively, correspond to quantiles τ_{1i} and τ_{0i} of Y_1 and Y_0 respectively. The treatment effect for subject i could then be written as in Equation 3 where Y_1 and Y_0 have cumulative density functions F_1 and F_0 such that $F_d = \Pr(Y_d \leq y)$ and the quantile functions of Y_1 and Y_0 are given by Q_1 and Q_0 where $Q_d(\tau) = \inf\{y : F_d(y) \geq \tau\}$ and $\tau \in (0,1)$.

$$Y_{1i} - Y_{0i} = Q_1(\tau_{1i}) - Q_0(\tau_{0i}) = \underbrace{[Q_1(\tau_{1i}) - Q_0(\tau_{1i})]}_{\text{Quantile Treatment Effect for } \tau_{1i}} + \underbrace{[Q_0(\tau_{1i}) - Q_0(\tau_{0i})]}_{\text{Mobility Effect: Represents shift in rank of } i \text{ owing to treatment.}} \quad (3)$$

Here CQTE can only be representative of the effect for a subject i belonging to the τ^{th} quantile (i.e., the UQTE) if both Y_{1i} and Y_{0i} belonged to the τ^{th} quantile of Y, i.e., to interpret the difference in ${\tau_{1i}}^{th}$ quantiles of potential outcomes (in equation 3) as the treatment effect for subjects belonging to the ${\tau_{1i}}^{th}$ quantile of the unconditional outcome distribution, we must assume that the *"mobility effect"* (Bedoya, et al. 2018, pp. 7) for each subject is zero, i.e., an

observed subject maintains their rank in the outcome distribution regardless of treatment status. A simple example clarifies this point further (Bedoya et al. 2018). Consider a hypothetical dataset of five subjects who have received fertilizer subsidy from the government. We want to estimate the effect of fertilizer subsidy on yields on average as well as yields on the median. Table 1 presents the data on the subjects' yields when they receive subsidy (column 1) and what their yields would have been had they not received subsidy (column 2)[1]. The average treatment effect is given by the difference in means of the potential outcomes (460 - 280) which is equal to the average of the difference in potential outcomes, i.e., 180[2]. Now, consider the treatment effect at the median which would be given by the difference in the medians of the potential outcomes (500 – 300 = 200) which is higher than the median of the subject-level differences in potential outcomes (i.e., 100). This is because of the *mobility effect* whereby subject I who would have had the lowest yield (among the subjects) without subsidy, has the highest yield among them when they do receive subsidy. The notion of *mobility effect* captures *shifts in subjects' ranking with and without treatment* and is precisely the source of dissimilarity between the difference-in-quantiles of potential outcomes and the quantile-of-difference in potential outcomes[3]. To reconcile this difference and obtain the unconditional quantile treatment effects, rank invariance assumes that the mobility effect is zero for each subject. Stated mathematically, let $U_d = F_d(Y_d)$ be potential ranks by treatment status, $U_d \sim U(0,1)$ by construction[4]. Rank invariance holds if and only if:

[1] These counterfactual outcomes are obviously unobservable in reality as no subject can be observed in both treatment and control state.

[2] Follows from the linearity of the expectation operator (Liao and Zhao 2019)

[3] A practical illustration of this effect when the independent variable is continuous, is included in Appendix 3 with data on cotton yields and di-ammonium phosphate (DAP) fertilizer usage.

[4] Define $Z = F_x(X)$ and notice that Z takes values between 0 and 1. Then, $F_Z(x) = \Pr(F_X(X) \le x) = \Pr(X \le F_X^{-1}(x)) = F_X(F_X^{-1}(x)) = x$. On the other hand, if U is a random variable that takes values in $(0,1)$, $F_U(x) = \int f_U(u)du = \int_0^x du = x$. So, since $F_Z(x) = F_U(x)$, Z is also uniformly distributed.

$$(U_0 \mid \overrightarrow{X} = x, \overrightarrow{V} = v) = (U_1 \mid \overrightarrow{X} = x, \overrightarrow{V} = v) \quad (4)$$

Importantly, rank invariance is a strong and untestable assumption (Firpo 2007; Imbens and Woolridge 2009). Firpo (2007) develop a semiparametric method for estimating UQTE based on the restriction that selection to treatment is based on observable characteristics but without assuming rank invariance. Firpo et al. (2009) also present a method for estimating the UQTE without assuming rank invariance but restrict themselves to a setting with only exogenous regressors.

The restrictive nature of this assumption is well reflected in the following thought experiment (adapted from Dong and Shen (2018)). Consider a random sample of rice-growing farmers in a village and their observationally equivalent duplicates, i.e., clones. Let the treatment be a binary indicator for being a clone. Naturally, this treatment should not have any effect on yields. However, due to a random chance induced by idiosyncratic shocks (like farm-specific livestock mortality and consequent shortage of farm labour), a farmer and their clone may not have exactly equal yields in one incidence of the experiment. Nevertheless, in infinite repetitions of the experiment, the farmer and their clone would have the same distribution of yields. Rank similarity captures this notion and relaxes rank invariance by allowing for random deviations of each subject's rank away from her counterfactual self so long as the ranks distribution (conditional on factors influencing treatment status) is identical across treatment states (Chernozhukov & Hansen 2005). While rank invariance requires U_0 and U_1 to be the same random variable, rank similarity only assumes they have the same conditional distribution, conditional on a set of rank-shifting covariates, i.e.,

$$(U_0 \mid \overrightarrow{X} = x, \overrightarrow{V} = v) \sim (U_1 \mid \overrightarrow{X} = x, \overrightarrow{V} = v) \quad (4)$$

Testing for rank similarity is tricky because the joint distribution of potential outcomes is unknown (Heckman et al. 1997; Firpo 2007). Recently, Frandsen and Lefgren (2018) and Dong and Shen (2018) in contemporaneous works proposed tests for rank similarity arriving at testable implications of rank similarity in situations where the endogeneity in treatment variable was modeled using auxiliary instrumental variable regressions. As shown by Dong and Shen (2018), rank similarity implies that at the same rank of the potential outcome distributions, the distribution of all relevant observables and unobservables (i.e., covariates) remains the same (Equation 5). In fact, a popular test for rank invariance is to check covariate similarity at the same quantile of the treatment and control outcome distributions (see for example Bitler et. al. 2006).

$$F_{\vec{X},\vec{V}|U_0}(x,v \mid \tau) = F_{\vec{X},\vec{V}|U_1}(x,v \mid \tau) \forall (x,v) \tag{5}$$

Equation 5 follows directly from the definition of Bayes' rule:

$$f_{\vec{X},\vec{V}|U_d}(x,v \mid \tau) = \frac{f_{U_d|\vec{X},\vec{V}}(\tau \mid x,v)\, f_{\vec{X},\vec{V}}(x,v)}{\iint f_{U_d|\vec{X},\vec{V}}(\tau \mid x,v)\, f_{\vec{X},\vec{V}}(x,v) dx dv} \tag{6}$$

Substituting the definition of rank similarity, i.e., $f_{U_1|\vec{X},\vec{V}}(\tau \mid x,v) = f_{U_0|\vec{X},\vec{V}}(\tau \mid x,v)$ in (6) gives (5). Further, since $f_{U_d|\vec{X}}(\tau \mid x) = \int f_{U_d|\vec{X},\vec{V}}(\tau \mid x,v) dF_{\vec{V}|\vec{X}}(v \mid x)$ and $f_{U_1|\vec{X},\vec{V}}(\tau \mid x,v) = f_{U_0|\vec{X},\vec{V}}(\tau \mid x,v)$, substituting the latter into the former, we can say that:

$$F_{U_1|\vec{X}}(\tau \mid x) = F_{U_0|\vec{X}}(\tau \mid x) \tag{7}$$

Equation 7 is the directly testable implication of rank similarity (Dong and Shen 2018; Frandsen and Lefgren 2018) which states that under rank similarity, the **distribution of potential ranks among observationally equivalent subjects is the same across treatment states.** We borrow this testable implication of rank similarity (equation (7)) and adapt the test to reflect our resolution of treatment endogeneity through propensity score matching. Specifically, we employ the Kolmogorov-Smirnov distribution test (Dodge 2008) and

showcase the proposed rank similarity test through an evaluation of the impact of agricultural credit on crop yield distribution (Shukla & Arora 2021). In the following section we lay out the background for studying the impact of credit on crop yields and describe our quasi-experimental framework. Next, we present an understanding of the consequences of rank similarity or lack thereof and showcase the steps for testing it within this framework. This is followed by the results and conclusion.

2. Background: Heterogenous impacts of agricultural credit on farm yield distribution

Consider a random sample of farming households with yields Y who can choose to take agricultural credit ($D=1$) or not ($D=0$). We aim to evaluate the impact of credit on average yields and the downside yield-risk (i.e., the left tail of the yield distribution relative to the higher yield quantiles). Since credit access and yield levels are endogenous (i.e., better farm outcomes lead to potentially higher credit access and vice versa (Feder et al. 1990; Rapsomanikis 2015)), we propose and implement a quasi-experimental design to evaluate the impact of credit on yield. First, we employ propensity score matching (PSM) to match loanee and non-loanee households on their propensity to access credit through their credit worthiness and then, we compare the average yield differentials across matched loanee and non-loanee households. Further, we evaluate the role of credit in downside yield risk mitigation using the quantile regression (QR) framework and compare the credit impact in the 25^{th} and 50^{th} yield quantiles with its mean impact.

2.1. Data Source:

We utilize a plot-level repeated cross-sectional dataset from the Village Dynamics Studies in South Asia (VDSA) program of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) to study 512 sorghum[5]-growing households across fourteen

[5] We choose sorghum because it is the third most important food grain in India and plays a vital role in food security acting as a staple food for some of the most vulnerable populations in the semi-arid tropics of Africa and Asia. Grown in both rainy (kharif) and postrainy (rabi) seasons, its high photosynthetic ability and nitrogen

semi-arid villages in five Indian states: Andhra Pradesh, Gujarat, Karnataka, Madhya Pradesh and Maharashtra, during 2001-'14.

2.2. Designing the Quasi-Experiment (Shukla and Arora 2021):

To implement a quasi-experimental design, we first need to define our treatment indicator. In defining the treatment, while a household can take credit multiple times during the study period, we assign to treatment only the *first-instance of access for agriculturally-productive credit.* Estimating the impact of multiple credit instances was not possible due to insufficient counterfactual data with about 89% of the households in our sample accessing credit at least once. Moreover, the utilization of loan funds for non-agricultural purposes, i.e., the fungibility of cash complicates the isolation of the impact of credit on yield (David and Meyer 1979). To address this, we filter on the purpose for which credit is taken and arrive at the set of households accessing *agriculturally-productive credit* (ICRISAT 2015) i.e., loan funds spent for crop cultivation only[6]. We stick to a binary definition of credit access because there is high likelihood of recall bias in the credit amount figures given the multiple cropping seasons within a year (Beegle et al. 2011). Further, there is also a possibility of reporting bias given that a majority households (about 94%) grow multiple crops in a year and therefore details of credit amount going towards each crop can be hard to reproduce[7].

2.2.1. Designing the Propensity Score Matching (PSM) Protocol:

We obtain propensity score estimates $P(\vec{Z}) = \Pr(D = 1 | \vec{Z})$ using a logistic regression with $\vec{Z}$ containing measures of credit-worthiness such as land ownership, soil quality,

and water-use efficiency (Reddy, Kumar and Reddy 2008) make it genetically well-suited for production in the semi-arid (hot and dry) regions with our sample states accounting for over 70 percent of India's sorghum production and close to 80 percent of the cultivated area during 2001-14 (Ministry of Agriculture & Farmers' Welfare (MoAFW) 2019).

[6] This definition does not account for the indirect impact of consumption credit which can increase agricultural labour productivity through improved food choices. But, since consumption credit can be used for several other non-agricultural purposes, its inclusion would complicate the isolation of credit impact on yields and yield risk. Hence, it is kept out of the scope of this study.

[7] For this reason, we also repeat our analysis for a restricted sample of households which grow only sorghum. All our results hold true.

demographics, asset income, ease of institutional credit access, and financial literacy (Shukla and Arora 2021; see Tables 2 and 3(A-D) for detailed variable description and summaries). Then, we construct counterfactuals by matching $\hat{P}(\vec{Z})$ across treated and untreated households. By matching households on their credit worthiness, we are attempting to generate for each loanee household a control group of non-loanee households which (due to the similarity in propensity of credit access) would have similar yields on-average as the loanee household had it not taken credit. Therefore, we argue that any remaining difference in the yields of loanee and non-loanee households can then be attributed to credit access (Shukla and Arora 2021). In this manner, propensity score matching resolves the problem of selection on *observable* factors ($\vec{Z}$) that determine credit access. But *unobservable* factors that influence credit access and crop yields could be different across the treated and control groups and hence bias the treatment effect estimation. Panel data methods (such as difference-in-differences) could potentially remove such a bias arising due to (a) factors that are invariant across households and might vary through time (by incorporating time fixed effects) and (b) factors that are invariant over time and might vary across households (by incorporating household fixed effects) (Angrist and Pischke 2009). However, our data structure does not allow us to exploit such panel data strategies without losing about 84% of the data (see Tables 4 & 5). Moreover, forcing a panel structure by restricting the sample to the 16% of households for which we have at least one year of data prior-to and after treatment (i.e., credit access), would systematically select more households with lower yields thereby introducing a selection bias in the treatment effect estimates (see Figure 1). Therefore, we rely on PSM to obtain the counterfactuals and attempt to reduce the bias arising due to unobservable factors by refining the matching protocol in the following ways:

2.2.1.1. Year-by-Year Matching (Arora et al. 2016):

Potential matches for loanee households in each year are restricted to non-loanee households in the same year according to the rule $|\hat{\Pr}(D=1|\vec{Z}) - \hat{\Pr}(D=0|\vec{Z})| \leq 0.05$. Since our panel data involves a long-time span with over a decade of data, it is plausible that the economic and social conditions faced by a household such as access to roads, ease of electricity access and type of informal credit institutions could be completely dissimilar between the early and late periods. Therefore, a household in 2001 may not serve as a good counterfactual for itself in 2014 (Arora et al. 2016). Hence, we impose restrictions on matching to be within the same year to minimize the potential bias from time-varying unobservable factors (Liu & Lynch 2011).

2.2.1.2. Within Village-Year Matching (Shukla and Arora 2021):

We further refine our matching criterion such that loanee households in each year are only matched to non-loanee households in the same year *and* within the same village. This serves to control for village specific unobservable characteristics which can influence credit access. For example, caste dynamics and its influences on institutions supplying credit (particularly the informal sector) are likely to be different across villages (Karthick and Madheswaran 2018). So, a household in rural Maharashtra may not serve as a good match for a household in rural Andhra Pradesh. Hence, we restrict the matching to the same village (and consequently the same state) in each year to minimize the potential bias from time-invariant village-specific unobservable factors.

2.2.1.3. Inclusion of Household-Specific, Time-Invariant Covariates (Shukla and Arora 2021):

We further minimize the bias arising from time-invariant household specific unobservable variables such as (a) soil nutrient levels (Tesfay & Moral 2021) by controlling for soil quality; (b) social capital (Bastelaer 2000) by differentiating credit access from institutional and informal sources and controlling for household caste as well as household-

specific access to credit prior to first instance of access for sorghum; (c) access to market information and farm management ability (Nuthall 2009) by including proxies for farming ability through age and education of the household head[8] (Shukla and Arora 2021).

2.2.2. Defining the Treatment and Control Groups:

Finally, based on the above, we can formally define our sets of treatment and control households as follows. For each household i, define B_i^F and B_i^I as the sets of all the years in which household i takes credit for agriculturally productive purposes from formal and informal sources of credit respectively. Further, define S_i to be the set of all years in which the household grows sorghum.[9] A household belongs to the treated group only if $B_i^F \cap S_i \neq \varnothing \vee B_i^I \cap S_i \neq \varnothing$.[10] Meanwhile, the control group (for the first strategy, i.e., 2.2.1.1) consists of households for which, $B_i^F \cap S_i = \varnothing \wedge B_i^I \cap S_i = \varnothing$ i.e., households which do not take credit in the years in which they grow sorghum.[11] If a sorghum-growing household gets credit from both formal and informal sources, there can be three cases:

1) $t_i^{FS} = \min_t\{t : t \in B_i^F \cap S_i\} > t_i^{IS} = \min_t\{t : t \in B_i^I \cap S_i\}$, i.e., informal credit is accessed prior to formal credit.
2) $t_i^{FS} < t_i^{IS}$, i.e., informal credit is accessed after formal credit.
3) $t_i^{FS} = t_i^{IS}$, i.e., the year of first-time access to formal and informal credit coincide.

[8] Typically, the household head is also the head of farming operations.

[9] Obviously, since we consider the set of sorghum-growing households in our sample, $S_i \neq \varnothing$.

[10] Note that if for a household the year of first instance of access to agriculturally productive credit *for sorghum* falls *after* the first instance of access to agriculturally productive credit in general then the household is considered treated only after it accesses credit for growing sorghum and not before that.

[11] Importantly, this set will also include households for which the year in which Sorghum is first grown, i.e., $t_i^S = \min_t\{t : t \in S_i\}$ occurs after the first instance formal ($t_i^{FS} = \min_t\{t : t \in B_i^F \cap S_i\}$), and informal credit ($t_i^{IS} = \min_t\{t : t \in B_i^I \cap S_i\}$), obviously taken for some other crop, i.e., $t_i^S > t_i^{FS} \wedge t_i^S > t_i^{IS}$ because the credit (whenever taken), did not go towards sorghum production at all.

In each of the first two cases, the household can only be a part of the control group $\forall t < \min\{t_i^{FS}, t_i^{IS}\}$ and will be part of the respective treated groups post the first instance of access. In the last case, the household can only be a part of the control group $\forall t < t_i^{FS} = t_i^{IS}$ but will not be considered as part of the treated group at all since our interest is in estimating the differential impact of credit from formal and informal sources on sorghum yields.[12] Finally, after matching, the treatment effect of credit (on average and at the quantiles) are obtained by comparing the conditional yields among *matched* treated and untreated households.

2.3. Estimating the impact of credit on downside yield risk: Quantile Regressions (QR) in conjunction with Propensity Score Matching (PSM):

Following the formulation in Section 1, the potential yields are defined as $Y_d = h(d, \vec{X}, \vec{V})$ where $\vec{X}$ is the vector of observable determinants of yield including biological inputs (e.g., weather), physical inputs (e.g., fertilizer), land characteristics (e.g., soil quality), and ease of farming operations and $\vec{V}$ consists of the unobservable factors affecting crop yields. The conditional quantile function for yield is given by:

$$Q(\tau \mid D, \vec{X}, \vec{V}) = \inf\{y : F_{Y|D,\vec{X},\vec{V}}(y \mid D, \vec{X}, \vec{V}) \geq \tau\} \quad (8)$$

and the conditional quantile treatment effect CQTE for $\tau \in \{0.25, 0.5\}$ is given by:

$$\Gamma(\tau \mid \vec{X}, \vec{V}, \vec{Z}) = \left\{Q(\tau \mid \vec{X}, \vec{V}, D=1) - Q(\tau \mid \vec{X}, \vec{V}, D=0)\right\} \mid (\hat{P}(\vec{Z}) \mid D=1) - (\hat{P}(\vec{Z}) \mid D=0) \leq 0.05 \quad (9)$$

measured by $\beta_{1,\tau}$ in the following yield model for the *matched* set of households:

$Q_\tau(Y_{it} \mid \vec{X}_{it}, D_{it}) = \beta_{0,\tau} + \beta_{1,\tau} D_{it} + \beta_{2,\tau} \vec{X}_{it}^{in} + \beta_{3,\tau} \vec{X}_{it}^{l} + \beta_{4,\tau} \vec{X}_{it}^{w} + \beta_{5,\tau} \vec{X}_{it}^{d} + \beta_{6,\tau}(x_t) + \eta_{it,\tau}$ where D_{it} is treatment (i.e., credit access); $\vec{X}_{it}^{w}$ is weather (i.e., seasonal rainfall (mm) and temperature

[12] After accounting for these nuances, we arrive at a sample of 458 loanee and non-loanee, sorghum-growing households.

(°C)); $\vec{X}_{it}^{in}$ are inputs such as fertilizer, pesticides, seeds; $\vec{X}_{it}^{l}$ represents land (soil) quality; $\vec{X}_{it}^{d}$ represents ease and quality of farming operations measured by the distance between house and plot and farmer experience measured by age and education status and x_t is trend (see Tables 6 and 7(A-D) for detailed variable description and summaries).

3. Rank Similarity in the QR+PSM Framework:

Rank similarity implies that the conditional distribution of potential ranks ($U_d = F_d(Y_d)$) among *observationally equivalent* households is the same across treatment states, i.e.,

$$\left\{U_0 \mid (\vec{X} = x, \vec{V} = v) \sim U_1 \mid (\vec{X} = x, \vec{V} = v)\right\} \mid (P(\vec{Z}) \mid D = 1) - (P(\vec{Z}) \mid D = 0) \leq 0.05 \quad (10)$$

Importantly, by imposing rank similarity on the *matched* set of households, we require *observational equivalence* in terms of (a) yield outcomes conditional on the determinants of yield; and (b) credit worthiness of a household as measured by the propensity of credit access. Assuming distributional equality in the unobservable factors affecting yield, equation (10) can be rewritten as follows:

$$\{F_{U_1|\vec{X}}(\tau \mid \vec{X} = x) = F_{U_0|\vec{X}}(\tau \mid \vec{X} = x)\} \mid (P(\vec{Z}) \mid D = 1) - (P(\vec{Z}) \mid D = 0) \leq 0.05 \quad (11)$$

Equation (11) is the directly testable implication of rank similarity which states that for the matched set of households, the relative yield of a household conditional on the determinants of yield is the same irrespective of whether the household accesses credit or not.

3.1. Interpretation of QR coefficients when rank similarity holds:

Generally rank similarity is implausible when treatment effects differ across observationally identical subjects (Liao and Zhao 2018; Wenz 2019). Nonetheless, if it holds then the CQTE and UQTE coincide and the quantile regression estimates can be interpreted at the 'individual' level, i.e., for individual subjects belonging to a particular quantile in the unconditional outcome distribution (Imbens and Woolridge 2009; Bedoya, et al. 2018). For example, under rank similarity, $\hat{\beta}_{1,\tau=0.5}$ represents the difference in the median yields of loanee

vs. non-loanee households *and equivalently* the impact of credit for households having median yields in the unconditional yield distribution.

3.2. Interpretation of QR coefficients when rank similarity fails:

On the contrary, if rank similarity does not hold, the UQTE and CQTE interpretations are not equivalent and quantile regression estimates no longer provide an interpretation for the quantiles of the unconditional outcome distribution. In this case, $\hat{\beta}_{1,\tau=0.5}$ still represents the difference in the median yields of loanee vs. non-loanee households but it **cannot** be interpreted as the effect of credit for households having median yields at the population level because, simply put, we do not know which households are still in the 50th quantile after accessing credit.

3.3. Steps for Testing Rank Similarity in the QR+PSM Framework:

Step 1: Obtain Y_d for the treated and control households conditional on the relevant covariates ($\vec{X}_{it}$).

Step 2: Compute the empirical cumulative distribution functions (CDFs) for Y_d , which provide the rank distributions, i.e., U_d , for the treated and control groups.

Step 3: Obtain the set of matched treated and control household-pairs.

Step 4: Run the two-sample Kolmogorov-Smirnov (K-S) distributional similarity test to compare U_0 and U_1 . The null hypothesis is that the distribution of ranks across treated and control groups is the same (Dodge 2008).

H_0 : Distribution of ranks across the treated and control groups is same, i.e., $\forall u, F_{U_0}(u) = F_{U_1}(u)$
H_1 : Distribution of ranks across the treated and control groups is different, i.e., $\exists u : F_{U_0}(u) \neq F_{U_1}(u)$

The K-S statistic is given by:

$$D_{mn} = \left[\frac{mn}{(m+n)} \right] \sup_y |U_1(y) - U_0(y)| \qquad (12)$$

where *m* and *n* are sample sizes for the treated and control groups respectively.

Step 5: Interpret the results.

4. Results:

Table 8 summarizes the quantile regression results and Table 9 summarize the results from the Kolmogorov Smirnov Distribution test. Figure 2 below displays the empirical CDF of ranks for the treated and control groups respectively. We reject the null hypothesis that the ranks for treated and control groups are similar. Overall, as is evident in the figure below, the ranks for the treated group are significantly higher than the ranks in the control group. To see this, fix a level of rank (say 0.6) and note that the cumulative density values for the (informal) treated group is close to 0.45 whereas for the (informal) control group it is about 0.65. This implies that roughly 45% of households in the treated group have yield ranks lower than 0.6 whereas the same proportion of households for the control group is higher (i.e., about 65%). In this manner, the control group displays a higher concentration of low ranks as compared to the treated group and therefore rank similarity does not hold. Hence, while the QR estimates can be interpreted as the difference in conditional yield quantiles (25^{th} and 50^{th}) of the loanee vs. non-loanee households, we cannot comment on the magnitude of credit impact on the left tail of the unconditional yield distribution, i.e., the downside yield risk.

5. Conclusion and Future Work:

In conclusion, we propose a general test for rank similarity in a setting where treatment endogeneity is resolved by employing a propensity score matching estimator. The example on risk-mitigative capacity of agricultural credit highlights that this assumption deserves examination for questions pertaining to agricultural risk management. In the future, we hope to tackle this issue by adapting the strategy proposed by Firpo (2007) for our study.

Appendix 1: Figures:

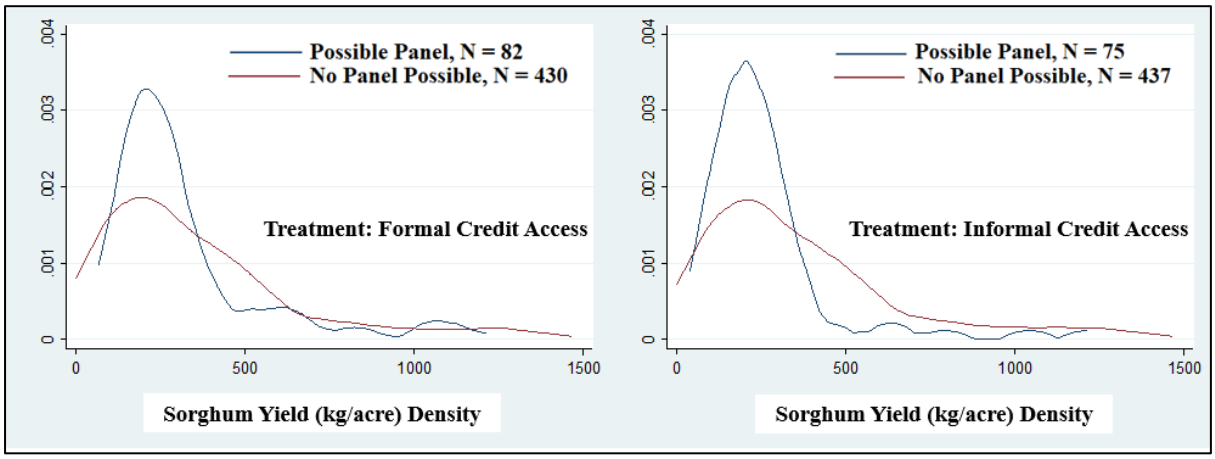

Figure 1: Sorghum Yield Density by Data Availability for Panel Strategies

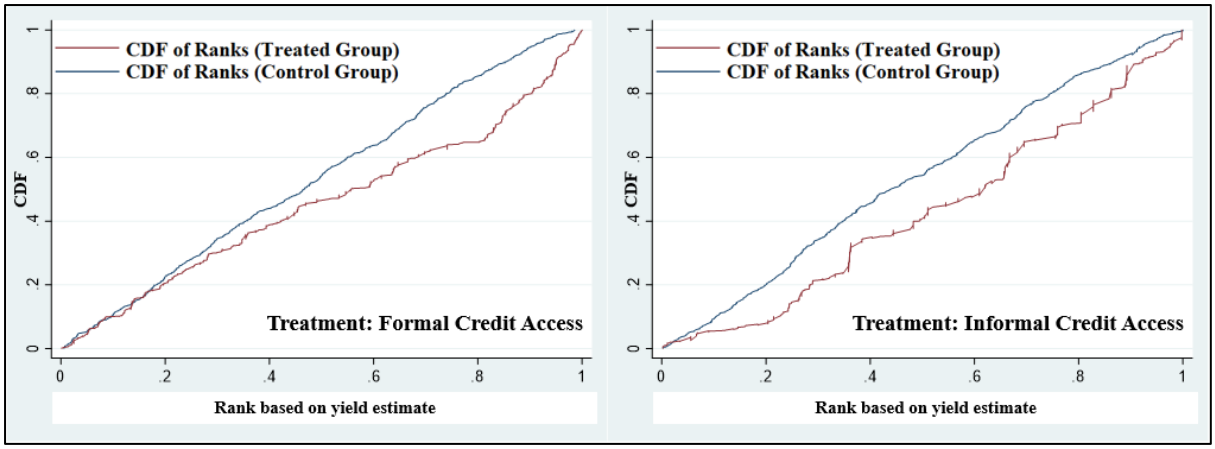

Figure 2: Empirical Rank CDFs for Formal and Informal Credit Access

Appendix 2: Tables:

Subject	Yield with Subsidy	Yield without Subsidy	Difference in Potential Outcomes (i.e., yields with and without subsidy)
I	600	100	500
II	500	300	200
III	500	400	100
IV	500	500	0
V	200	100	100
Average	460	280	180
Median	500	300	100

Table 1: Hypothetical data on potential outcomes (yield) for five subjects.

Dependent Variable: Credit Access	**Indicator for credit access, if credit is accessed, takes value 1 and 0 otherwise.**
Independent Variables	
Backward Caste	Time-invariant indicator for household caste, takes value 1 if caste is backward and 0 otherwise.
Landholding Class	Time-invariant indicator for landholding class, takes value 1 if no land is owned and 0 otherwise.
Operational Holding	Total area (in acres) available to the household for farming.
Ratio of Owned Land to Total Operational Holding	Ratio of owned land (in acres) to operational holding (in acres)
Soil Quality	Time-invariant indicator for soil quality, takes value 1 for erosive or saline soils and 0 otherwise.
Number of Productive Family Members	Number of family member above the age of 18 years who are fit for work, i.e., do not have any disability.
Years of Education of Household Head	Time-invariant variable indicating number of years of education for household head.
Age of Household Head	Age of household head (in years)
Total Durable Value	Total value of household durable assets (such as cycle, TV etc.) in rupees adjusted for inflation at 1986 base year prices.
Quantum of Livestock (Cattle)	Number of livestock (bull, buffaloes and cows) owned by the household.
Non-Farm Income	Total value of household non-farm income (such as wages from teaching at local school) in rupees adjusted for inflation at 1986 base year prices.
Instances of Formal Credit Access prior to accessing credit for Sorghum	Number of times a HH has accessed credit (from formal or informal sources) prior to accessing it for Sorghum for the first time, reflective of the financial literacy and credit experience of the household.
Instances of Informal Credit Access prior to accessing credit for Sorghum	

Institutional Credit Disbursed to Agriculture	Household invariant, district-level variable reflecting ease of credit access measured as the percentage change in total direct (cash) *plus* indirect (e.g., input subsidies) financial support to agriculture from institutional sources such as scheduled commercial banks.
Village Fixed Effects	Village-level indicator variables to control for unobserved spatial variation in credit access.

Table 2: Description of Determinants of Credit Access (Adapted from (Shukla and Arora 2021))

Variable	Mean	SD	Min	P25	P50	P75	Max
Operational holding (Non-Fallow, in acres)	6.69	6.51	0.1	2.75	5	8	40
Ratio of Owned Land to Total Operational Holding (Non-Fallow)	0.94	0.22	0	0.95	1	1	1
Years of Education of Household Head	4.29	4.55	0	0	4	7	18
Age of Household Head	49.62	13.22	18	39	48	60	84
Value of Total Durables (INR, Constant Prices)	2425.22	2614.72	0	835.25	1582.04	3012.65	20019.5
Quantum of Livestock (Cattle)	2.45	2.66	0	0	2	4	16
Non-Farm Income (INR, Constant Prices)	9464.08	11562.54	0	2434.05	6514.66	12210.3	65000
Size of the HH	5.41	2.20	2	4	5	6	19
Instances of Formal Credit Access (prior to first instance for Sorghum)	0.63	1.31	0	0	0	1	11
Instances of Informal Credit Access (prior to first instance for Sorghum)	0.53	1.34	0	0	0	0	11
Percentage Change in Institutional Credit Disbursed to Agriculture [t-(t-1)]/(t-1)	0.17	0.12	-0.05	0.082	0.15	0.24	0.41

Table 3A: Summary statistics for Access Regressions (Formal, Continuous) (Adapted from (Shukla and Arora 2021))

Variable	Mean	SD	Min	P25	P50	P75	Max
Operational Holding (Non-Fallow, in acres)	6.81	5.92	0.25	3	5	8.5	46
Ratio of Owned Land to Total Operational Holding (Non-Fallow)	0.91	0.26	0	0.88	1	1	1
Years of Education of Household Head	3.76	4.21	0	0	2	7	17
Age of Household Head	49.47	13.25	18	39	47	60	85
Value of Total Durables (INR, Constant Prices)	2059.07	2327.23	0	749.19	1312.70	2437.14	20639.36
Quantum of Livestock (Cattle)	2.30	2.59	0	0	2	4	15

Non-Farm Income	8430.94	10214.50	0	2018.98	6168.57	11068.94	65000
Size of the HH	5.16	1.92	1	4	5	6	12
Instances of Formal Credit Access (prior to first instance for Sorghum)	0.55	1.30	0	0	0	0	11
Instances of Informal Credit Access (prior to first instance for Sorghum)	0.70	1.47	0	0	0	1	11

Table 3B: Summary statistics for Access Regressions (Informal, Continuous), (Adapted from (Shukla and Arora 2021))

Variable	Yes = 1	No = 0
Formal Credit	276	642
Backward Caste	179	739
Landholding Class = Labour	30	888
Erosive Soil	98	819

Table 3C: Summary statistics for Access Regressions (Formal, Categorical), (Adapted from (Shukla and Arora 2021))

Variable	Yes = 1	No = 0
Formal Credit	181	600
Backward Caste	171	610
Landholding Class = Labour	32	749
Erosive Soil	102	678

Table 3D: Summary statistics for Access Regressions (Informal, Categorical), (Adapted from (Shukla and Arora 2021))

Treatment = First time access to FORMAL Credit for Agriculturally Productive Purposes	Number of HHs
At least 1 year of Pre-Treatment & Post*-Treatment Data	82
At least 2 years of Pre-Treatment & Post-Treatment Data	47
At least 3 years of Pre-Treatment & Post-Treatment Data	21
At least 4 years of Pre-Treatment & Post-Treatment Data	12
At least 5 years of Pre-Treatment & Post-Treatment Data	6
At least 6 years of Pre-Treatment & Post-Treatment Data	1
At least 7 years of Pre-Treatment & Post-Treatment Data	0
***Post-Treatment refers to periods <u>after (and not including)</u> the first-time access to credit.**	

Table 4: Availability of Pre and Post Treatment (Formal Credit Access) Data for Sorghum-Growing Households

Treatment = First time access to INFORMAL Credit for Agriculturally Productive Purposes	Number of HHs
At least 1 year of Pre-Treatment & Post*-Treatment Data	75
At least 2 years of Pre-Treatment & Post-Treatment Data	41

At least 3 years of Pre-Treatment & Post-Treatment Data	23
At least 4 years of Pre-Treatment & Post-Treatment Data	17
At least 5 years of Pre-Treatment & Post-Treatment Data	8
At least 6 years of Pre-Treatment & Post-Treatment Data	2
At least 7 years of Pre-Treatment & Post-Treatment Data	0
***Post-Treatment refers to periods <u>after (and not including)</u> the first-time access to credit.**	

Table 5: Availability of Pre and Post Treatment (Informal Credit Access) Data for Sorghum-Growing Households

Dependent Variable: Sorghum Yield	**Total Output (in kg) / Total Crop Area (in acres) for Sorghum**
Independent Variables	
Credit	Indicator for credit access, if credit is accessed, takes value 1 and 0 otherwise.
Rainfall	Household invariant, village level rainfall in millimetre.
Temperature	Household invariant, village level temperature in degree Celsius.
Soil Quality	Time invariant indicator for soil quality, takes value 1 for erosive or saline soils and 0 otherwise.
Seed Type	Indicator for seed type, takes value 1 for locally produced, non- high yielding variety (non-HYV) seeds and 0 otherwise.
Irrigation	Irrigation input measured in motor hours per acre.
Fertilizer	Nitrogen (quantity of nitrogen in fertilizers like urea) input measured in kilograms per acre
Distance between House and Plot	Distance between house and plot reflective of ease of farming management measured in kilometres.
Soil Depth	Indicator for soil depth. If deep soils, i.e., greater than 1.5 metres, takes value 1 and 0 otherwise.
Intercropping Instances	Indicator for whether Sorghum is intercropped with some other crop or not.
Age of HH Head	Age of household head (in years)
Education of HH Head	Time-invariant variable indicating number of years of education for household head.

Table 6: Description of Yield Determinants (Adapted from Shukla and Arora (2021))

Variable	Mean	SD	Minimum	P25	P50	P75	Maximum
Sorghum Yield (kg/acre)	318.89	339.63	0	106.67	225.01	400	2380.95
Rainfall (in mm.)	399.51	262.27	101.33	203.77	292.9	500.33	1417.47
Temperature (in °C)	25.11	0.56	24.26	24.66	24.785	25.65	26.245
Electric Irrigation (hours/acre)	0.10	0.36	0	0	0	0	3.07
Nitrogen (kg/acre))	0.15	0.34	0	0	0	0.16	3.83
Distance between House and Plot	1.61	0.89	0	1	1.5	2.07	5.2
Age of HH Head	48.88	12.52	25	39	47	59	84
Years of Education of HH Head	4.51	4.57	0	0	4	8	18

Table 7A: Summary Statistics Yield Regressions (Formal Credit, Continuous), (Adapted from (Shukla and Arora 2021))

Credit Access	
Yes	199
No	428
Soil Quality	
Erosive	60
Non-Erosive	567
Soil Depth	
Deep (>=1.5 metres)	430
Shallow (<1.5 metres)	197
Intercropping	
Yes	270
No	357
Seed Type	
Local	162
High Yielding Variety	452

Table 7B: Summary Statistics Yield Regressions (Formal Credit, Categorical), (Adapted from (Shukla and Arora 2021))

Variable	Mean	SD	Minimum	P25	P50	P75	Maximum
Sorghum Yield (kg/acre)	285.88	332.30	0	88	190.16	360	2801.12
Rainfall (in mm.)	403.52	250.53	101.33	203.63	316.94	531.7	1417.47
Temperature (in °C)	25.19	0.54	24.5	24.7	24.96	25.68	26.205
Electric Irrigation (hours/acre)	0.13	0.47	0	0	0	0	5.23
Nitrogen (kg/acre))	0.18	0.37	0	0	0	0.191667	2.28
Distance between House and Plot	1.72	0.90	0	1.045455	1.5	2.35	7
Age of HH Head	48.10	12.67	22	39	45	57	85
Years of Education of HH Head	3.34	4.15	0	0	1.5	6	16

Table 7C: Summary Statistics Yield Regressions (Informal Credit, Continuous), (Adapted from (Shukla and Arora 2021))

Credit Access	
Yes	101
No	272
Soil Quality	
Erosive	50
Non-Erosive	324
Soil Depth	
Deep (>=1.5 metres)	231
Shallow (<1.5 metres)	142
Intercropping	
Yes	179
No	194
Seed Type	
Local	76
High Yielding Variety	297

Table 7D: Summary Statistics Yield Regressions (Informal Credit, Categorical), (Adapted from (Shukla and Arora 2021))

Quantile Regressions				
Dependent Variable: Yield (kg/acre)	**Formal Credit**		**Informal Credit**	
	Q (0.25)	**Q (0.5)**	**Q (0.25)**	**Q (0.5)**
Credit (Yes = 1)	24.43* (13.65)	23.16 (22.91)	-3.76 (14.68)	46.82 (31.14)
Trend	3.06 (2.77)	2.14 (3.44)	-1.97 (2.90)	3.23 (3.76)
Rainfall (in mm.)	0.09* (0.05)	1.07 (11.35)	0.10* (0.05)	0.31** (0.09)
Temperature (in °C)	13.25 (20.67)	-22.24* (12.67)	30.51 (31.11)	-10.36 (37.13)
Soil Quality (Erosive/Saline = 1, 0 otherwise)	-39.04* (23.01)	-53.81 (43.11)	-24.74 (17.41)	-47.41 (43.60)
Seed Type (Local = 1)	-30.55 (21.80)	-65.48*** (22.60)	-53.43* (30.82)	-129.19*** (40.94)
Irrigation (hours/acre)	113.41*** (23.62)	103.15*** (36.47)	111.38*** (23.58)	109.64*** (29.70)
Fertilizer Quantity (Nitrogen (kg/acre))	112.98*** (42.47)	124.07*** (51.80)	136.54*** (35.96)	200.98** (91.21)
Distance between House and Plot (in km)	3.31 (7.18)	-8.9 (10.06)	1.03 (8.47)	-1.89 (12.71)
Deep Soils (>=1.5 metres, Yes =1)	40.32* (23.83)	51.83 (39.00)	1.66 (22.03)	11.15 (30.49)
Intercropping with Nitrogen Regulating Crops (Yes = 1)	27.38* (19.11)	68.16*** (21.69)	-16.61 (19.56)	24.74 (23.88)
Years of Education of Household Head	3.18 (2.22)	4.13 (2.70)	-2.36 (1.55)	-2.34 (3.53)
Age of Household Head	-0.18	-0.63	-1.29*	-1.24

	(0.54)	(0.64)	(0.69)	(0.87)
Constant	-328.62 (510.46)	293.69 (789.37)	-632.55 (768.22)	361.47 (927.58)

Table 8: Yield Quantile Regression Results: Formal and Informal Credit | *p<0.10, **p<0.05, ***p<0.01, (Adapted from (Shukla and Arora 2021))

Group	m	n	D-Statistic	p-value
Treatment: Informal Credit Access	897	931	0.18	0
Treatment: Formal Credit Access	1919	1968	0.21	0

Table 9: Results for the Rank Similarity Test using the Two Sample Kilmogorov-Smirnov Distribution Test

Appendix 3: Visualizing the Need for Rank Invariance in a Quantile Regression Model:

Suppose you have a simple regression of the form: $Y_i = \alpha + \beta F_i + \varepsilon_i$ where Y_i is the yield, F_i is the fertilizer application (Diammonium Phosphate (DAP)) per acre for the i^{th} household. First, let's plot the unconditional density of yields.

P5	P10	P25	P50	P75	P90	P95	P99
58.67	100.41	195.01	337.12	600	888.29	1050	1425

Table 10: Cotton Yield(kg/acre) Summary

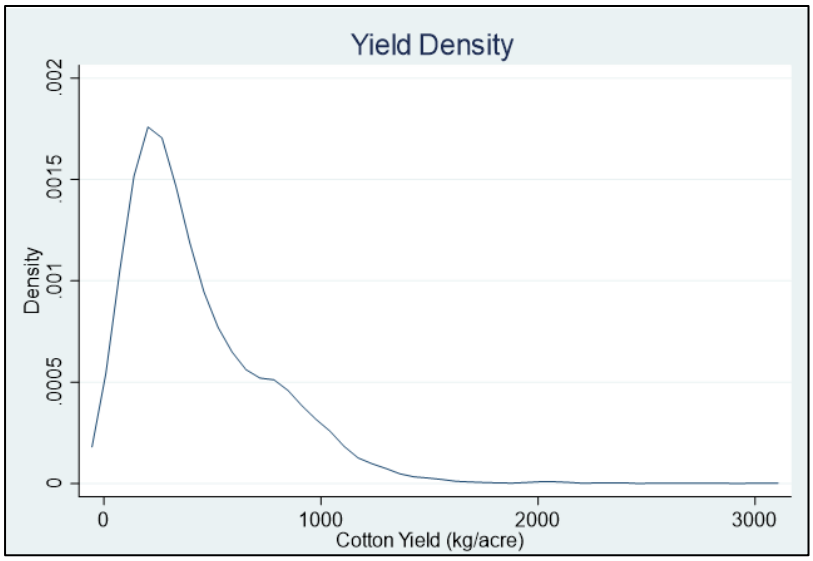

Figure 3: Unconditional Cotton Yield(kg/acre) Density Plot

Next, pick two households B and G where B (Yield = 337 kg/acre) belongs to a relatively lower (50th) quantile of the unconditional yield distribution and G (Yield = 1050 kg/acre) belongs to one of the relatively higher (95th) quantiles.

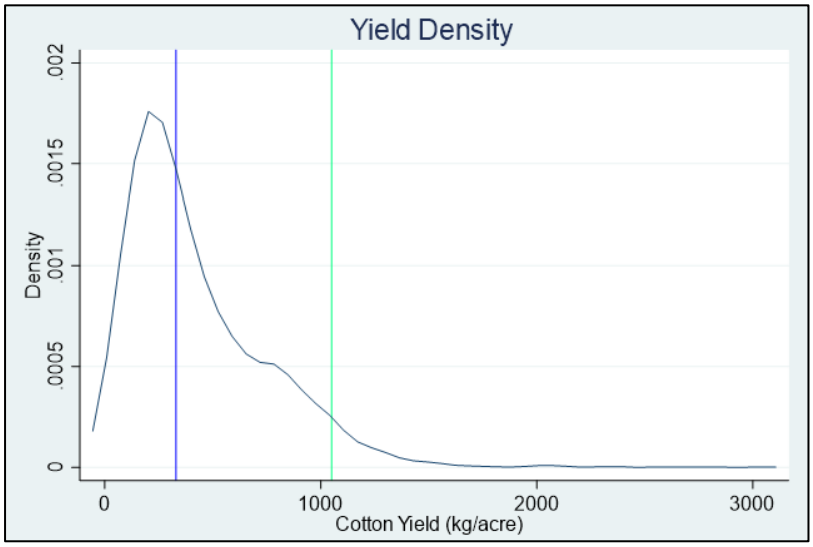

Figure 4: Picking two households B and G

Now we condition our yields on the level of fertilizer application, i.e., for each level of fertilizer application, we obtain get a conditional yield distribution, i.e., we would come up with a density plot as above (Figure 3) but for each level of fertilizer application separately. As we can observe in the graph below, the level of fertilizer application is 5.26 kg/acre for B and 2.38 kg/acre for G (see the blue and green circles respectively):

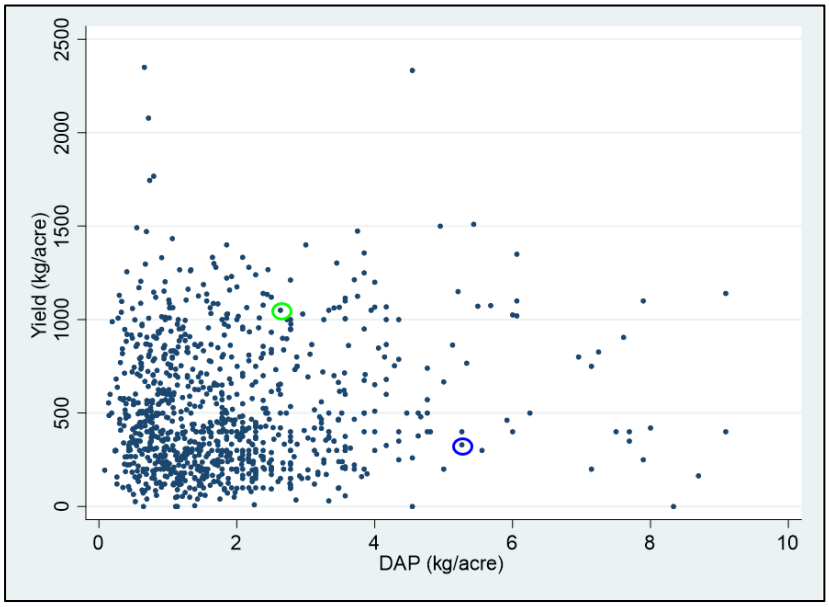

Figure 5: Conditioning Yields on Fertilizer (DAP) Application

Let us now look at the conditional quantile estimates of DAP on cotton yields.

Dependent Variable = Yield (kg/acre)	**Coefficient (SE)**					
	Q5	**Q25**	**Q50**	**Q75**	**Q90**	**Q95**
DAP (kg/acre)	22.99*** (4.04)	30.25*** (3.77)	50.26*** (5.32)	73.5*** (9.89)	76.97*** (10.31)	94.14*** (15.09)
Constant	48.14*** (5.47)	177.9*** (5.11)	306.6*** (7.21)	550*** (13.4)	836.36*** (13.96)	978*** (20.44)

Table 11: Quantile Regression Estimates for Yield as a function of Fertilizer Application

Plotting these conditional quantile estimates we can see that apparently, neither households B or G is doing quite as well among their peers in the 5.26kg/acre and 2.38 kg/acre fertilizer usage brackets and hence are in the 25th and 90th percentile respectively as compared to their ranks in the unconditional yield distribution, i.e., 50th and 95th percentile respectively. So, once you condition on another variable, the ranking of households has shifted in this new distribution. Specifically, both the households considered in this example have a lower place in the conditional distribution as compared to the unconditional distribution. The question is, how doe this change our interpretation of the QR coefficient estimates?

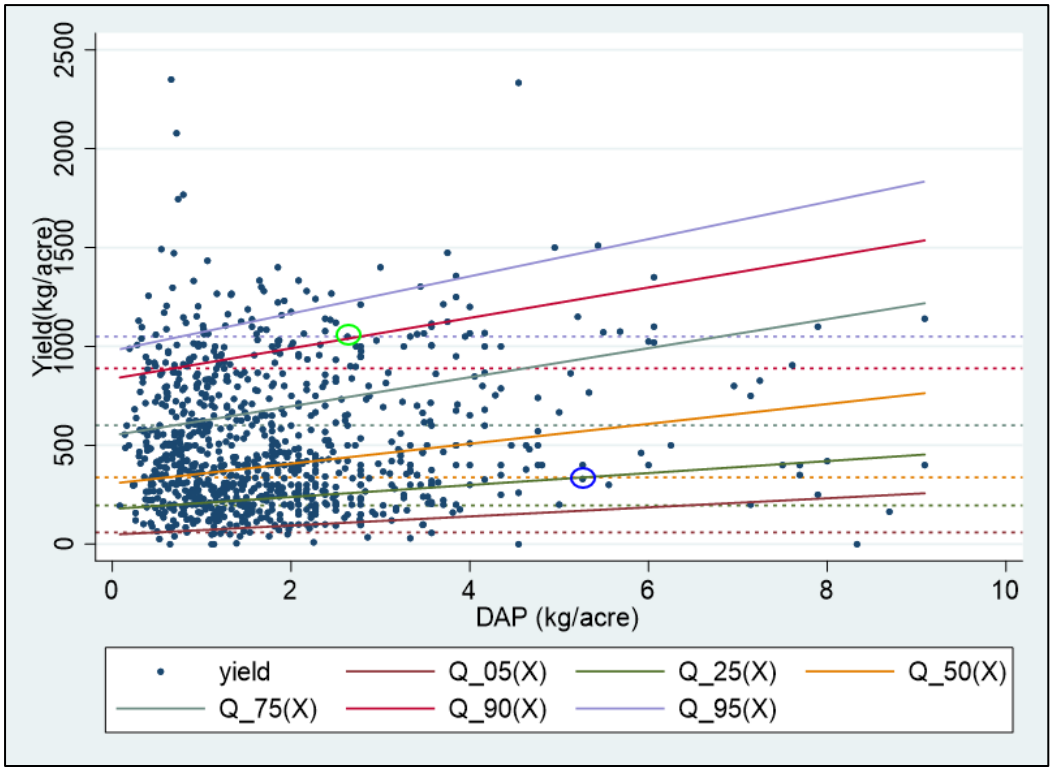

Figure 6: Plotting Quantile Regression (QR) Estimates of Yields Conditional on Fertilizer (DAP) Application

As evident with the two households taken as an example, since we cannot comment on where a household will be in the outcome distribution before and after a treatment (in our case additional DAP usage), we can only make statements about the distribution as a whole. Hence in the above example, while $\beta_{0.5} = 50.27$ can be interpreted as the difference in the median yields due to an additional kilogram of DAP being used per acre, it **cannot** be interpreted as the effect of DAP for households having median yields at the population level because we don't know which households are still in the 50th quantile after using an additional kilogram of DAP per acre. The only way we can match the two interpretations is if the same households were retained at the median under both the conditional and unconditional distributions. This is exactly what rank invariance assumes. Under rank invariance, the QR coefficient can be interpreted as the coefficient for households at quantile τ.

References:

- Angrist, J. D., 2004. Treatment Effect Heterogeneity in Theory and Practice. The Economic Journal, 114, C52-C83.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. New Jersey: Princeton University Press.
- Arora, G., Wolter, P. T., Feng, H. & Hennessy, D. A., 2016. Role of Ethanol Plants in Dakotas Land Use Change: Incorporating Flexible Trends in the Difference-in-Difference Framework with Remotely-Sensed Data, Iowa: Iowa State University. Accessed here.
- Bastelaer, T., 2000. Does Social Capital Facilitate the Poor's Access to Credit?: A Review of the Microeconomic Literature, Washington, DC: The World Bank. Accessed here.
- Bedoya, Guadalupe, Luca Bittarello, Jonathan Davis, and Nikolas Mittag. 2018. Distributional Impact Analysis: Toolkit and Illustrations of Impacts beyond the Average Treatment Effect. IZA Discussion Papers, No. 11863, Bonn: Institute of Labor Economics (IZA). Accessed here.
- Beegle, K., Carletto, C. & Himelein, K., 2012. Reliability of Recall in Agricultural Data, Journal of Development Economics, 98(1), pp. 34-41.
- Bitler, M. P., Gelbach. J. B. & Hoynes, H. W., 2006. What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. American Economic Review, 96(4), 988-1012.
- Bitler, M. P., Gelbach, . J. B. & Hoynes, H. W., 2008. Distributional Impacts of the Self-Sufficiency Project. Journal of Public Economics, 92, 748-765.
- Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2002. Removing the Veil of Ignorance in Assessing the Distributional Impacts of Social Policies. NBER Working Paper No. 8840, Cambridge: National Bureau of Economic Research (NBER). Accessed here.
- Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." International Economic Review, 44(2) 361-422.
- Chernozhukov, V. & Hansen, C., 2005. An IV model of quantile treatment effects. Econometrica, 73(1), 245–261.
- Cunha, F., Heckman JJ., & Schennach S. 2010: Estimating the technology of cognitive and non-cognitive skill formation, Econometrica 78(3), 883–931.

- Dammert, A. C. (2008). Child labor and schooling response to changes in coca production in rural Peru. Journal of development Economics, 86(1), 164–180.
- David, Cristina, and Meyer, Richard. 1980. “Measuring the Farm Level Impact of Agriculture Loans." Borrowers Lenders: Rural Financial Markets and Institutions in Developing Countries, ed. John Howell. London: Overseas Development Institute. Accessed here.
- Dodge, Yadolah, 2008. Kolmogorov-Smirnov Test. In: The Concise Encyclopaedia of Statistics. Switzerland: Springer Science, pp. 283-287.
- Doksum, K. (1974): “Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case,” The Annals of Statistics, 2, 267–277.
- Eren, O., and S. Ozbeklik (2014): Who Benefits from Job Corps? A Distributional Analysis of An Active Lablor Market Program, Journal of Applied Econometrics, 29. 586–611.
- Feder, G., Lau, L. J. & Lin, J., 1990. The Relationship between Credit and Productivity in Chinese Agriculture: A Microeconomic Model of Disequilibrium. American Journal of Agricultural Economics (Proceedings Issue), 72(5), pp. 1151-57.
- Firpo, S., 2007. Efficient Semiparametric Estimation of Quantile Treatment Effects. Econometrica, 75(1), 259-276.
- Firpo S., Fortin N., Lemieux T. 2009. Unconditional Quantile Regressions. Econometrica, 77(3), 953-973
- Heckman, J.J., Smith, J., & Clements, N. 1997. Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. Review of Economic Studies, 64(4): 487-535
- Imbens Guido W., Woolridge Jeffrey .M., 2009, Recent Developments in the Econometrics of Program Evaluation. Journal of Economic Literature, 47(1): 5-86
- International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), 2015. Lack of formal credit constraining rural transformation: ICRISAT Happenings Newsletter, No. 1701, India: ICRISAT. Accessed here.
- Karlan, D., Osei, R., Osei-Akoto, I. & Udry, C., 2014. Agricultural Decisions after Relaxing Credit and Risk Constraints. The Quarterly Journal of Economics, 129(2), pp. 597-652.
- Karthick V., Madheswaran S., 2018. Access to Formal Credit in the Indian Agriculture: Does Caste matter? Journal of Social Inclusion Studies, 4(2), pp. 1-27.

- Lehmann, E. (1974): Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.
- Liao, W-C., Zhao, D. 2019. The selection and quantile treatment effects on the economic returns of green buildings. Regional Science and Urban Economics, 74: 38-48.
- Liu, X. & Lynch, L., 2011. Do Agricultural Land Preservation Programs Reduce Farmland Loss? Evidence from a Propensity Score Matching Estimator. Land Economics, 87(2), pp. 183-201.
- Ministry of Agriculture & Farmers' Welfare (MoAFW) . 2019. "EPWRF Time Series: Area-Production-Yield Statistics for Sorghum." EPWRF Time Series: Area-Production-Yield Statistics for Sorghum. New Delhi, Delhi: Ministry of Agriculture & Farmers' Welfare (MoAFW), November 7th.
- Nuthall, P., 2009. Modelling the origins of managerial ability in agricultural production. The Australian Journal of Agriculture and Resource Economics, 53(3), pp. 413-436.
- Petrick, M., 2004. A microeconometric analysis of credit rationing in the Polish farm sector. European Review of Agricultural Economics, 31, pp. 23-47.
- Ramaswami, B., 2019. Agricultural Subsidies: Study Prepared for XV Finance Commission, New Delhi: Indian Statistical Institute. Accessed here.
- Rapsomanikis, G., 2015. The economic lives of smallholder farmers: An analysis based on household data from nine countries, Rome: Food and Agriculture Organization of the United Nations. Accessed here.
- Reddy, Belum VS, A Ashok Kumar, and P Sanjana Reddy. 2008. "Genetic improvement of sorghum in the semi-arid tropics." In Sorghum Improvement in the New Millennium. International Crops Research Institute for the Semi-Arid Tropics, by BVS, Ramesh, S, Kumar, A and Gowda, CLL Reddy, 105-123. Andhra Pradesh: International Crops Research Institute for the Semi-Arid Tropics.
- Shukla, S. & Arora, G., 2021. The impact of agricultural credit on farm yield risk: A quantile regression approach in conjunction with propensity score matching: Selected Paper for the 2021 AAEA Annual Meeting. Accessed here.
- Tesfay, M. G. & Moral, M. T., 2021. The impact of participation in rural credit program on adoption of inorganic fertilizer: A panel data evidence from Northern Ethiopia. Cogent Food & Agriculture, 7(1).
- Wenz, Sebastian E. 2019. “What Quantile Regression Does and Doesn’t Do: A Commentary on Petscher and Logan (2014).” Child Development 90(4):1442-52.

- Wilcox, R. R., 1997. Some practical reasons for reconsidering the Kolmogorov-Smirnov test. British Journal of Mathematical and Statistical Psychology, Volume 50, 9-20.