The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- Current Contents/Social and Behavioral Sciences®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)
- Scopus™
- Social Sciences Citation Index®

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, Mata, NetCourse, and Stata Press are registered trademarks of StataCorp LP.

The Stata Journal (2010)
10, Number 3, pp. 423–457

Estimation of quantile treatment effects with Stata

Markus Frölich
Universität Mannheim and
Institute for the Study of Labor
Bonn, Germany
froelich@uni-mannheim.de

Blaise Melly
Department of Economics
Brown University
Providence, RI
blaise_melly@brown.edu

Abstract. In this article, we discuss the implementation of various estimators proposed to estimate quantile treatment effects. We distinguish four cases involving conditional and unconditional quantile treatment effects with either exogenous or endogenous treatment variables. The introduced `ivqte` command covers four different estimators: the classical quantile regression estimator of Koenker and Bassett (1978, *Econometrica* 46: 33–50) extended to heteroskedasticity consistent standard errors; the instrumental-variable quantile regression estimator of Abadie, Angrist, and Imbens (2002, *Econometrica* 70: 91–117); the estimator for unconditional quantile treatment effects proposed by Firpo (2007, *Econometrica* 75: 259–276); and the instrumental-variable estimator for unconditional quantile treatment effects proposed by Frölich and Melly (2008, IZA discussion paper 3288). The implemented instrumental-variable procedures estimate the causal effects for the subpopulation of compliers and are only well suited for binary instruments. `ivqte` also provides analytical standard errors and various options for nonparametric estimation. As a by-product, the `locreg` command implements local linear and local logit estimators for mixed data (continuous, ordered discrete, unordered discrete, and binary regressors).

Keywords: st0203, ivqte, locreg, quantile treatment effects, nonparametric regression, instrumental variables

1 Introduction

Ninety-five percent of applied econometrics is concerned with mean effects, yet distributional effects are no less important. The distribution of the dependent variable may change in many ways that are not revealed or are only incompletely revealed by an examination of averages. For example, the wage distribution can become more compressed or the upper-tail inequality may increase while the lower-tail inequality decreases. Therefore, applied economists and policy makers are increasingly interested in distributional effects. The estimation of quantile treatment effects (QTEs) is a powerful and intuitive tool that allows us to discover the effects on the entire distribution. As an alternative motivation, median regression is often preferred to mean regression to reduce susceptibility to outliers. Hence, the estimators presented below may thus be particularly appealing with noisy data such as wages or earnings. In this article, we provide a brief survey over recent developments in this literature and a description of the new `ivqte` command, which implements these estimators.

Depending on the type of endogeneity of the treatment and the definition of the estimand, we can define four different cases. We distinguish between conditional and unconditional effects and whether selection is on observables or on unobservables. Conditional QTEs are defined conditionally on the value of the regressors, whereas unconditional effects summarize the causal effect of a treatment for the entire population. Selection on observables is often referred to as a matching assumption or as exogenous treatment choice (that is, exogenous conditional on X). In contrast, we refer to selection on unobservables as endogenous treatment choice.

First, if we are interested in *conditional* QTEs and we assume that the treatment is exogenous (conditional on X), we can use the quantile regression estimators proposed by Koenker and Bassett (1978). Second, if we are interested in conditional QTEs but the treatment is endogenous, the instrumental-variable (IV) estimator of Abadie, Angrist, and Imbens (2002) may be applied. Third, for estimating *unconditional* QTEs with exogenous treatment, various approaches have been suggested, for example, Firpo (2007), Frölich (2007a), and Melly (2006). Currently, the weighting estimator of Firpo (2007) is implemented. Finally, unconditional QTE in the presence of an endogenous treatment can be estimated with the technique of Frölich and Melly (2008). The estimators for the unconditional treatment effects do not rely on any (parametric) functional forms assumptions. On the other hand, for the conditional treatment effects, $\sqrt{n}$ convergence rate can only be obtained with a parametric restriction. Because estimators affected by the curse of dimensionality are of less interest to the applied economist, we will discuss only parametric (linear) estimators for estimating conditional QTEs.

The implementation of most of these estimators requires the preliminary nonparametric estimation of some kind of (instrument) propensity scores. We use nonparametric linear and logistic regressions to estimate these propensity scores. As a by-product, we also offer the `locreg` command for researchers interested only in these nonparametric regression estimators. We allow for different types of regressors, including continuous, ordered discrete, unordered discrete, and binary variables. A cross-validation routine is implemented for choosing the smoothing parameters.

This article only discusses the implementation of the proposed estimators and the syntax of the commands. It draws heavily on the more technical discussion in the original articles, and the reader is referred to those articles for more background on, and formal derivations of, some of the properties of the estimators described here.

The contributions to this article and the related commands are manyfold. We provide new standardized commands for the estimators proposed in Abadie, Angrist, and Imbens (2002);[1] Firpo (2007); and Frölich and Melly (2008); and estimators of their analytical standard errors. For the conditional exogenous case, we provide heteroskedasticity consistent standard errors. The estimator of Koenker and Bassett (1978) has already been implemented in Stata with the `qreg` command, but its estimated standard errors

[1] Joshua Angrist provides codes in Matlab to replicate the empirical results of Abadie, Angrist, and Imbens (2002). Our codes for this estimator partially build on his codes.

are not consistent in the presence of heteroskedasticity. The ivqte command thus extends upon qreg in providing analytical standard errors for heteroskedastic errors.

At a higher level, locreg implements nonparametric estimation with both categorical and continuous regressors as suggested by Racine and Li (2004). Finally, we incorporate cross-validation procedures to choose the smoothing parameters.

The next section outlines the definition of the estimands, the possible identification approaches, and the estimators. Section 3 describes the ivqte command and its various options, and contains simple applications to illustrate how ivqte can be used. Appendix A describes somewhat more technical aspects for the estimation of the asymptotic variance matrices. Appendix B describes the nonparametric estimators used internally by ivqte and the additional locreg command.

2 Framework, assumptions, and estimators

We consider the effect of a binary treatment variable D on a continuous outcome variable Y. Let Y_i^1 and Y_i^0 be the potential outcomes of individual i. Hence, Y_i^1 would be realized if individual i were to receive treatment 1, and Y_i^0 would be realized otherwise. Y_i is the observed outcome, which is $Y_i \equiv Y_i^1 D_i + Y_i^0(1-D_i)$.

In this article, we identify and estimate the entire distribution functions of Y^1 and Y^0.[2] Because QTEs are an intuitive way to summarize the distributional impact of a treatment, we focus our attention especially on them.

We often observe not only the outcome and the treatment variables but also some characteristics X.[3] We can therefore either define the QTEs conditionally on the covariates or unconditionally. In addition, we have to deal with endogenous treatment choice. We distinguish between the case where selection is only on observables and the case where selection is also on unobservables.

2.1 Conditional exogenous QTEs

We start with the standard model for linear quantile regression, which is a model for conditional effects and where one assumes selection on observables. We assume that Y is a linear function in X and D.

Assumption 1. *Linear model for potential outcomes*

$$Y_i^d = X_i\beta^\tau + d\delta^\tau + \varepsilon_i \text{ and } Q_{\varepsilon_i}^\tau = 0$$

for $i = 1, \ldots, n$ and $d \in (0,1)$. $Q_{\varepsilon_i}^\tau$ refers to the τth quantile of the unobserved random variable ε_i. β^τ and δ^τ are the unknown parameters of the model. Here δ^τ represents the conditional QTEs at quantile τ.

[2] In the case with endogenous treatment, we identify the potential outcomes only for compliers, as defined later.

[3] If we do not observe covariates, then conditional and unconditional QTEs are identical and the estimators simplify accordingly.

Clearly, this linearity assumption is not sufficient for identification of QTEs because the observed D_i may be correlated with the error term ε_i. We assume that both D and X are exogenous.

Assumption 2. *Selection on observables with exogenous X*

$$\varepsilon \perp\!\!\!\perp (D, X)$$

Assumptions 1 and 2 together imply that $Q^{\tau}_{Y|X,D} = X\beta^{\tau} + D\delta^{\tau}$, such that we can recover the unknown parameters of the potential outcomes from the joint distribution of the observed variables Y, X, and D. The unknown coefficients can thus be estimated by the classical quantile regression estimator suggested by Koenker and Bassett (1978). This estimator is defined by

$$(\widehat{\beta}^{\tau}, \widehat{\delta}^{\tau}) = \arg\min_{\beta,\delta} \sum \rho_{\tau}\left(Y_i - X_i\beta - D_i\delta\right) \tag{1}$$

where $\rho_{\tau}(u) = u \times \{\tau - 1\,(u < 0)\}$. This is a convex linear programming problem and is solved rather efficiently by the built-in `qreg` command in Stata. The `ivqte` command produces exactly the same point estimates as does `qreg`. In contrast to `qreg`, however, `ivqte` produces analytical standard errors that are consistent also in the case of heteroskedasticity.

To illustrate the similarity to all the following estimators, we could also write the previous expression as

$$(\widehat{\beta}^{\tau}, \widehat{\delta}^{\tau}) = \arg\min_{\beta,\delta} \sum W_i^{KB} \times \rho_{\tau}\left(Y_i - X_i\beta - D_i\delta\right)$$

where the weights W_i^{KB} are all equal to one.

2.2 Conditional endogenous QTEs

In many applications, the treatment D is self selected and potentially endogenous. We may not be able to observe all covariates to make assumption 2 valid. In this case, the traditional quantile regression estimator will be biased, and we need to use an IV identification strategy to recover the true effects. We assume that we observe a binary instrument Z and can therefore define two potential treatments denoted by D_z.[4] We use the following IV assumption as in Abadie, Angrist, and Imbens (2002).[5]

[4] If the instrument is nonbinary, it must be transformed into a binary variable. See Frölich and Melly (2008).

[5] An alternative approach is given in Chernozhukov and Hansen (2005), who rely on a monotonicity/rank invariance assumption in the outcome equation.

Assumption 3. *IV*

For almost all values of X

$$\begin{array}{l} \left(Y^0, Y^1, D_0, D_1\right) \perp\!\!\!\perp Z \,|X \\ 0 < \Pr\left(Z = 1 \,|X\,\right) < 1 \\ E\left(D_1 \,|X\,\right) \neq E\left(D_0 \,|X\,\right) \\ \Pr\left(D_1 \geq D_0 \,|X\,\right) = 1 \end{array}$$

This assumption is well known and requires monotonicity (that is, the nonexistence of defiers) in addition to a conditional independence assumption on the IV. Individuals with $D_1 > D_0$ are referred to as *compliers*, and treatment effects can be identified only for this group because the always- and never-participants cannot be induced to change treatment status by hypothetical movements of the instrument.

Abadie, Angrist, and Imbens (2002) (AAI) impose assumption 3. Furthermore, they require assumption 1 to hold for the compliers (that is, those observations with $D_1 > D_0$). They show that the conditional QTE, δ^τ, for the compliers can be estimated consistently by the weighted quantile regression:

$$\begin{aligned} (\widehat{\beta}^\tau_{\text{IV}}, \widehat{\delta}^\tau_{\text{IV}}) &= \arg\min_{\beta,\delta} \sum W_i^{\text{AAI}} \times \rho_\tau \left(Y_i - X_i\beta - D_i\delta\right) \\ W_i^{\text{AAI}} &= 1 - \frac{D_i\left(1 - Z_i\right)}{1 - \Pr\left(Z = 1 \,|X_i\,\right)} - \frac{\left(1 - D_i\right) Z_i}{\Pr\left(Z = 1 \,|X_i\,\right)} \end{aligned} \tag{2}$$

The intuition for these weights can be given in two steps. First, by assumption 3,[6]

$$\begin{aligned} & \left(Y^0, Y^1, D_0, D_1\right) \perp\!\!\!\perp Z \,|X \\ \Longrightarrow\ & \left(Y^0, Y^1\right) \perp\!\!\!\perp Z \,|X, D_1 > D_0 \\ \Longrightarrow\ & \left(Y^0, Y^1\right) \perp\!\!\!\perp D \,|X, D_1 > D_0 \end{aligned}$$

This means that any observed relationship between D and Y has a causal interpretation for compliers. To use this result, we have to find compliers in the population. This is done in the following average sense by the weights W_i^{AAI}:[7]

$$E\left\{W_i^{\text{AAI}} \rho_\tau\left(Y_i - X_i\beta - D_i\delta\right)\right\} = \Pr\left(D_1 > D_0\right) E\left\{\rho_\tau\left(Y_i - X_i\beta - D_i\delta\right) |D_1 > D_0\right\}$$

Intuitively, this result holds because $W_i^{\text{AAI}} = 1$ for the compliers and because $E\left(W_i^{\text{AAI}} | D_{i,1} = D_{i,0} = 0\right) = E\left(W_i^{\text{AAI}} | D_{i,1} = D_{i,0} = 1\right) = 0$.

A preliminary estimator for $\Pr\left(Z = 1 \,|X_i\,\right)$ is needed to implement this estimator. `ivqte` uses the local logit estimator described in appendix B.[8] A problem with

[6] This is the result of lemma 2.1 in Abadie, Angrist, and Imbens (2002).

[7] This is a special case of theorem 3.1.a in Abadie (2003).

[8] In their original article, Abadie, Angrist, and Imbens (2002) use a series estimator instead of a local estimator as in `ivqte`. Nevertheless, one can also use series estimation or, in fact, any other method to estimate the propensity score by first generating a variable containing the estimated propensity score and informing `ivqte` via the `phat()` option that the propensity-score estimate is supplied by the user.

estimator (2) is that the optimization problem is not convex because some of the weights are negative while others are positive. Therefore, this estimator has not been implemented. Instead, `ivqte` implements the AAI estimator with positive weights. Abadie, Angrist, and Imbens (2002) have shown that as an alternative to W_i^{AAI}, one can use the weights

$$W_i^{\mathrm{AAI}+} = E\left(W^{\mathrm{AAI}}\,|Y_i, D_i, X_i\right) \tag{3}$$

instead, which are always positive. Because these weights are unknown, `ivqte` uses local linear regression to estimate $W_i^{\mathrm{AAI}+}$; see appendix B. Some of these estimated weights might be negative in finite samples, which are then set to zero.[9]

2.3 Unconditional QTEs

The two estimators presented above focused on conditional treatment effects, that is, conditional on a set of variables X. We will now consider unconditional QTEs, which have some advantages over the conditional effects. The unconditional QTE (for quantile τ) is given by

$$\Delta^\tau = Q_{Y^1}^\tau - Q_{Y^0}^\tau$$

First, the definition of the unconditional QTE does not change when we change the set of covariates X. Although we aim to estimate the unconditional effect, we still use the covariates X for two reasons. On the one hand, we often need covariates to make the identification assumptions more plausible. On the other hand, covariates can increase efficiency. Therefore, covariates X are included in the first-step regression and then integrated out. However, the definition of the effects is not a function of the covariates. This is an advantage over the conditional QTE, which changes with the set of conditioning variables even if the covariates are not needed to satisfy the selection on observables or the IV assumptions.

A very simple example illustrates this advantage. Assume that the treatment D has been completely randomized and is therefore independent both from the potential outcomes as well as from the covariates. A simple comparison of the distribution of Y in the treated and nontreated populations has a causal interpretation in such a situation. For efficiency reasons, however, we may wish to include covariates in the estimation. If we are interested in mean effects, it is well known that including in a linear regression covariates that are independent from the treatment leaves the estimated treatment effect asymptotically unchanged. This property is lost for QTEs! Including covariates that are independent from the treatment can change the limit of the estimated conditional QTEs. On the other hand, it does not change the unconditional treatment effects if the assumptions of the model are satisfied for both sets of covariates, which is trivially satisfied in our randomized example.

A second advantage of unconditional effects is that they can be estimated consistently at the $\sqrt{n}$ rate without any parametric restrictions, which is not possible for conditional effects. For the conditional QTE, we therefore only implemented estimators

[9] Again, other estimators may be used with `ivqte`. The weights are first estimated by the user and then supplied via the `what()` option.

with a parametric restriction. The following estimators of the unconditional QTE are entirely nonparametric, and we will no longer invoke assumption 1. This is an important advantage because parametric restrictions are often difficult to justify from a theoretical point of view. In addition, assumption 1 restricts the QTE to be the same independently from the value of X. Obviously, interaction terms may be included, but the effects in the entire population are often more interesting than many effects for different covariate combinations.

The interpretation of the unconditional effects is slightly different from the interpretation of the conditional effects, even if the conditional QTE is independent from the value of X. This is because of the definition of the quantile. For instance, if we are interested in a low quantile, the conditional QTE will summarize the effect for individuals with relatively low Y even if their absolute level of Y is high. The unconditional QTE, on the other hand, will summarize the effect with a relatively low absolute Y.

Finally, the conditional and unconditional QTEs are trivially the same in the absence of covariates. They are also the same if the effect is the same independent of the value of the covariates and of the value of the quantile τ. This is often called the location shift model because the treatment affects only the location of the distribution of the potential outcomes.

2.4 Unconditional endogenous QTEs

We consider first the case of an endogenous treatment with a binary IV Z. This includes the situation with exogenous treatment as a special case when we use $Z = D$.

Frölich and Melly (2008) showed that Δ^τ for the compliers is identified under a somewhat weaker version of assumption 3, and they proposed the following estimator:

$$\begin{aligned}(\widehat{\alpha}_{\mathrm{IV}}, \widehat{\Delta}^\tau_{\mathrm{IV}}) &= \arg\min_{\alpha,\Delta} \sum W_i^{\mathrm{FM}} \times \rho_\tau \left(Y_i - \alpha - D_i \Delta\right) \qquad (4)\\ W_i^{\mathrm{FM}} &= \frac{Z_i - \Pr\left(Z = 1 \,|X_i\right)}{\Pr\left(Z = 1 \,|X_i\right)\left\{1 - \Pr\left(Z = 1 \,|X_i\right)\right\}}\left(2D_i - 1\right)\end{aligned}$$

This is a bivariate quantile regressor estimator with weights. One can easily see that $\alpha_{\mathrm{IV}} + \Delta^\tau_{\mathrm{IV}}$ is identified only from the $D = 1$ observations and that α_{IV} is identified only from the $D = 0$ observations. Therefore, this estimator is equivalent to using two univariate weighted quantile regressions separately for the $D = 1$ and the $D = 0$ observations.[10]

There are two differences between (4) and (2): The covariates are not included in the weighted quantile regression in (4), and the weights are different.[11] One might be think-

[10] The previous expression is numerically identical to $\widehat{\alpha}_{\mathrm{IV}} = \arg\min_{q_0} \sum_{i:D_i=0} W_i^{\mathrm{FM}} \times \rho_\tau(Y_i - q_0)$ and $\widehat{\alpha}_{\mathrm{IV}} + \widehat{\Delta}^\tau_{\mathrm{IV}} = \arg\min_{q_1} \sum_{i:D_i=1} W_i^{\mathrm{FM}} \times \rho_\tau(Y_i - q_1)$, from which we thus obtain $\widehat{\Delta}^\tau_{\mathrm{IV}}$ via two univariate quantile regressions.

[11] The weights W_i^{FM} were suggested in theorem 3.1.b and 3.1.c of Abadie (2003) for a general purpose. Frölich and Melly (2008) used these weights to estimate unconditional QTEs.

ing about running a weighted quantile regression of Y on a constant and D by using the weights W_i^{AAI}. For that purpose, however, the weights of Abadie, Angrist, and Imbens (2002) are not correct as shown in Frölich and Melly (2008). This estimator would estimate the difference between the τ quantile of Y^1 for the treated compliers and the τ quantile of Y^0 for the nontreated compliers, which is not meaningful in general. However, weights W_i^{AAI} could be used to estimate unconditional effects in the special case when the IV is independent of X such that $\Pr(Z=1\,|X)$ is not a function of X.

On the other hand, if one is interested in estimating conditional QTE using a parametric specification, the weights W_i^{FM} could be used, as well. Hence, although not developed for this case, the weights W_i^{FM} can be used to identify conditional QTEs. It is not clear whether W_i^{FM} or W_i^{AAI} will be more efficient. For estimating conditional effects, both are inefficient anyway because they do not incorporate the conditional density function of the error term at the quantile.

Intuitively, the difference between the weights W_i^{AAI} and W_i^{FM} can be explained as follows: They both find the compliers in the average sense discussed above. However, only W_i^{FM} simultaneously balances the distribution of the covariates between treated and nontreated compliers. Therefore, W_i^{AAI} can be used only in combination with a conditional model because there is no need to balance covariates in such a case. It can also be used without a conditional model when the treated and nontreated compliers have the same covariate distribution. W_i^{FM}, on the other hand, can be used with or without a conditional model.

A preliminary estimator for $\Pr(Z=1\,|X_i)$ is needed to implement this estimator. `ivqte` uses the local logit estimator described in appendix B. The optimization problem (4) is neither convex nor smooth. However, only two parameters have to be estimated. In fact, one can easily show that the estimator can be written as two univariate quantile regressions, which can easily be solved despite the nonsmoothness; see the previous footnotes. This is the way `ivqte` proceeds when the `positive` option is not activated.[12]

An alternative to solving this nonconvex problem consists in using the weights

$$W_i^{\mathrm{FM}+} = E\left(W^{\mathrm{FM}}\,|Y_i, D_i\right) \tag{5}$$

which are always positive. `ivqte` estimates these weights by local linear regression if the `positive` option has been activated. Again, estimated negative weights will be set to zero.[13]

[12] More precisely, `ivqte` solves the convex problem for the distribution function, and then monotonizes the estimated distribution function using the method of Chernozhukov, Fernández-Val, and Galichon (2010), and finally inverts it to obtain the quantiles. The parameters chosen in this way solve the first-order conditions of the optimization problem, and therefore, the asymptotic results apply to them.

[13] If one is interested in average treatment effects, Frölich (2007b) has proposed an estimator for average treatment effects based on the same set of assumptions. This estimator has been implemented in Stata in the command `nplate`, which can be downloaded from the websites of the authors of this article.

2.5 Unconditional exogenous QTEs

Finally, we consider the case where the treatment is exogenous, conditional on X. We assume that X contains all confounding variables, which we denote as the selection on observables assumption. We also have to assume that the support of the covariates is the same independent of the treatment, because in a nonparametric model, we cannot extrapolate the conditional distribution outside the support of the covariates.

Assumption 4. *Selection on observables and common support*

$$(Y^0, Y^1) \perp\!\!\!\perp D | X$$

$$0 < \Pr(D = 1 | X) < 1$$

Assumption 4 identifies the unconditional QTE, as shown in Firpo (2007), Frölich (2007a), and Melly (2006). The estimator of Firpo (2007) is a special case of (4), when D is used as its own instrument. The weighting estimator for Δ^τ therefore is

$$\begin{aligned}(\widehat{\alpha}, \widehat{\Delta}^\tau) &= \arg\min_{\alpha,\Delta} \sum W_i^F \times \rho_\tau (Y_i - \alpha - D_i \Delta) \qquad (6)\\ W_i^F &= \frac{D_i}{\Pr(D = 1 | X_i)} + \frac{1 - D_i}{1 - \Pr(D = 1 | X_i)}\end{aligned}$$

This is a traditional propensity-score weighting estimator, also known as inverse probability weighting. A preliminary estimator for $\Pr(D = 1 | X_i)$ is needed to implement this estimator. `ivqte` uses the local logit estimator described in appendix B.

3 The ivqte command

3.1 Syntax

The syntax of `ivqte` is as follows:

```
ivqte depvar [indepvars] (treatment [= instrument]) [if] [in] [,
   quantiles(numlist) continuous(varlist) dummy(varlist) unordered(varlist)
   aai linear mata_opt kernel(kernel) bandwidth(#) lambda(#) trim(#)
   positive pbandwidth(#) plambda(#) pkernel(kernel) variance
   vbandwidth(#) vlambda(#) vkernel(kernel) level(#)
   generate_p(newvarname [, replace]) generate_w(newvarname [,
   replace]) phat(varname) what(varname) ]
```

3.2 Description

ivqte computes the QTEs of a binary variable using a weighting strategy. This command can estimate both conditional and unconditional QTEs under either exogeneity or endogeneity. The estimator proposed by Frölich and Melly (2008) is used if unconditional QTEs under endogeneity are estimated. The estimator proposed by Abadie, Angrist, and Imbens (2002) is used if conditional QTEs under endogeneity are estimated. The estimator proposed by Firpo (2007) is used if unconditional QTEs under exogeneity are estimated. The estimator proposed by Koenker and Bassett (1978) is used if conditional QTEs under exogeneity are estimated.

The estimator used by ivqte is determined as follows:

- If an *instrument* is provided and aai is not activated, the estimator proposed by Frölich and Melly (2008) is used.
- If an *instrument* is provided and aai is activated, the estimator proposed by Abadie, Angrist, and Imbens (2002) is used.
- If there is no *instrument* and *indepvars* is empty, the estimator proposed by Firpo (2007) is used.
- If there is no *instrument* and *indepvars* contains variables, the estimator proposed by Koenker and Bassett (1978) is used.

indepvars contains the list of X variables for the Koenker and Bassett (1978) estimator for the estimation of exogenous conditional QTEs.[14] For all other estimators, *indepvars* must remain empty, and the control variables X are to be given in continuous(), unordered(), and dummy(). The instrument or the treatment variable is assumed to satisfy the exclusion restriction conditionally on these variables.

The IV has to be provided as a binary variable, taking only the values 0 and 1. If the original IV takes different values, it first has to be transformed to a binary variable. If the original IV is one-dimensional, one may use the endpoints of its support and discard the other observations. If one has several discrete IVs, one would use only those two combinations that maximize and minimize the treatment probability $\Pr(D=1|Z=z)$ and code these two values as 0 and 1. For more details on how to transform several nonbinary IVs to this binary case, see Frölich and Melly (2008).

The estimation of all nonparametric functions is described in detail in appendix B. A mixed kernel as suggested by Racine and Li (2004) is used to smooth over the continuous and categorical data. The more conventional approach of estimating the regression plane inside each cell defined by the discrete variables can be followed by setting lambda() to 0. The propensity score is estimated by default by a local logit estimator. A local linear estimator is used if the linear option is selected. Two algorithms

[14] For the Koenker and Bassett (1978) estimator, the options continuous(), unordered(), and dummy() are not permitted.

are available to maximize the local logistic likelihood function. The default is a simple Gauss–Newton algorithm written for this purpose. If you select the `mata_opt` option, the official Stata 10 optimizer is used. We expect the official estimator to be more stable in difficult situations. However, it can be used only if you have Stata 10 or more recent versions.

The `ivqte` command also requires the packages `moremata` (Jann 2005b) and `kdens` (Jann 2005a).

3.3 Options

Model

`quantiles(`*numlist*`)` specifies the quantiles at which the effects are estimated and should contain numbers between 0 and 1. The computational time needed to estimate an additional quantile is very short compared with the time needed to estimate the preliminary nonparametric regressions. When conditional QTEs are estimated, only one quantile may be specified. If one is interested in several QTEs, then one can save the estimated weights for later use by using the `generate_w()` option. By default, `quantiles()` is set to 0.5 when conditional QTEs are estimated, and `quantiles()` contains the nine deciles from 0.1 to 0.9 when unconditional QTEs are estimated.

`continuous(`*varlist*`)`, `dummy(`*varlist*`)`, and `unordered(`*varlist*`)` specify the names of the covariates depending on their type. Ordered discrete variables should be treated as continuous. For all estimators except Koenker and Bassett (1978), the X variables should be given here and not in *indepvars*. For the Koenker and Bassett (1978) estimator, on the other hand, these options are not permitted and the X variables must be given in *indepvars*.

`aai` selects the Abadie, Angrist, and Imbens (2002) estimator.

With the exception of the Koenker and Bassett (1978) estimator, several further options are needed to control the estimation of the nonparametric components. First, we need to estimate some kind of propensity score. For the Firpo (2007) estimator, we need to estimate $\Pr(D=1|X)$. For the Abadie, Angrist, and Imbens (2002) and Frölich and Melly (2008) estimators, we need to estimate $\Pr(Z=1|X)$, which we also call a propensity score in the following discussion. These propensity scores are then used to calculate the weights W_i^F, W_i^{AAI}, and W_i^{FM}, respectively, as defined in section 2.[15] The QTEs are estimated using these weights after applying some trimming to eliminate observations with very large weights. The amount of trimming is controlled by `trim()`, as explained below. This is the way the Firpo (2007) estimator is implemented.

For the Abadie, Angrist, and Imbens (2002) and Frölich and Melly (2008) estimators, more comments are required. First, the Abadie, Angrist, and Imbens (2002) estimator is only implemented with the positive weights $W_i^{\mathrm{AAI+}}$. Hence, when the Abadie, Angrist, and Imbens (2002) estimator is activated via the `aai` option, first

[15] The weights W_i^{KB} used in the Koenker and Bassett (1978) estimator are always equal to one.

the propensity score is estimated to calculate the weights W_i^{AAI}, which are then automatically projected via nonparametric regression to obtain $W_i^{\mathrm{AAI}+}$. This last nonparametric regression to obtain the positive weights is controlled by the options `pkernel()`, `pbandwidth()`, and `plambda()`, which are explained below. The letter p in front of these options stresses that these are used to obtain the positive weights.

Finally, the Frölich and Melly (2008) estimator is implemented in two ways. We can either use the weights W_i^{FM} after having trimmed very large weights, or alternatively, we could also project these weights and then use $W_i^{\mathrm{FM}+}$ to estimate the QTEs. If one wants to pursue this second implementation, one has to activate the `positive` option and specify `pkernel()` and `pbandwidth()` to control the projection of W_i^{FM} to obtain the positive weights $W_i^{\mathrm{FM}+}$.

Estimation of the propensity score

`linear` selects the method used to estimate the instrument propensity score. If this option is not activated, the local logit estimator is used. If `linear` is activated, the local linear estimator is used.

`mata_opt` selects the official optimizer introduced in Stata 10 to estimate the local logit, Mata's `optimize()`. The default is a simple Gauss–Newton algorithm written for this purpose. This option is only relevant when the `linear` option has not been selected.

`kernel(`*kernel*`)` specifies the kernel function used to estimate the propensity score. *kernel* may be any of the following second-order kernels: `epan2` (Epanechnikov kernel function, the default), `biweight` (biweight kernel function), `triweight` (triweight kernel function), `cosine` (cosine trace), `gaussian` (Gaussian kernel function), `parzen` (Parzen kernel function), `rectangle` (rectangle kernel function), or `triangle` (triangle kernel function). In addition to these second-order kernels, there are also several higher-order kernels: `epanechnikov_o4` (Epanechnikov order 4), `epanechnikov_o6` (order 6), `gaussian_o4` (Gaussian order 4), `gaussian_o6` (order 6), `gaussian_o8` (order 8). By default, `epan2`, which specifies the Epanechnikov kernel, is used.[16]

[16] Here are the formulas for these kernel functions for Epanechnikov of order 4 and 6, respectively:

$$\begin{aligned} K(z) &= \left(\frac{15}{8} - \frac{35}{8}z^2\right)\frac{3}{4}\left(1-z^2\right)1(z<1) \\ K(z) &= \left(\frac{175}{64} - \frac{525}{32}z^2 + \frac{5775}{320}z^4\right)\frac{3}{4}\left(1-z^2\right)1(z<1) \end{aligned}$$

And here are the formulas for Gaussian of order 4, 6, and 8, respectively:

$$\begin{aligned} K(z) &= \frac{1}{2}\left(3-z^2\right)\phi(z) \\ K(z) &= \frac{1}{8}\left(15-10z^2+z^4\right)\phi(z) \\ K(z) &= \frac{1}{48}\left(105-105z^2+21z^4-z^6\right)\phi(z) \end{aligned}$$

`bandwidth(#)` sets the bandwidth h used to smooth over the continuous variables in the estimation of the propensity score. The continuous regressors are first orthogonalized such that their covariance matrix is the identity matrix. The bandwidth must be strictly positive. If the bandwidth h is missing, an infinite bandwidth $h = \infty$ is used. The default value is infinity. If the bandwidth h is infinity and the parameter λ is one, a global model (linear or logit) is estimated without any local smoothing.

The cross-validation procedure implemented in `locreg` can be used to guide the choice of the bandwidth. Because the optimal bandwidth converges at a faster rate than the cross-validated bandwidth, the robustness of the results with respect to a smaller bandwidth should be examined.

`lambda(#)` sets the λ used to smooth over the dummy and unordered discrete variables in the estimation of the propensity score. It must be between 0 and 1. A value of 0 implies that only observations within the cell defined by all discrete regressors are used. The default is `lambda(1)`, which corresponds to global smoothing. If the bandwidth h is infinity and $\lambda = 1$, a global model (linear or logit) is estimated without any local smoothing. The cross-validation procedure implemented in `locreg` can be used to guide the choice of lambda. Again the robustness of the results with respect to a smaller bandwidth should be examined.

Estimation of the weights

`trim(#)` controls the amount of trimming. All observations with an estimated propensity score less than `trim()` or greater than $1 -$ `trim()` are trimmed and not used further by the estimation procedure. This prevents giving very high weights to single observations. The default is `trim(0.001)`. This option is not useful for the Koenker and Bassett (1978) estimator, where no propensity score is estimated.

`positive` is used only with the Frölich and Melly (2008) estimator. If it is activated, the positive weights $W_i^{\mathrm{FM}+}$ defined in (5) are estimated by the projection of the weights W^{FM} on the dependent and the treatment variable. Weights $W^{\mathrm{FM}+}$ are estimated by nonparametric regression on Y, separately for the $D = 1$ and the $D = 0$ samples. After the estimation, negative estimated weights in $\widehat{W}_i^{\mathrm{FM}+}$ are set to zero.

`pbandwidth(#)`, `plambda(#)`, and `pkernel(`*kernel*`)` are used to calculate the positive weights. These options are useful only for the Abadie, Angrist, and Imbens (2002) estimator, which can be activated via the `aai` option, and for the Frölich and Melly (2008) estimator, but only when the `positive` option has been activated to estimate $W^{\mathrm{FM}+}$. `pkernel()` and `pbandwidth()` are used to calculate the positive weights if the `positive` option has been selected. They are defined similarly to `kernel()`, `bandwidth()`, and `lambda()`. When `pkernel()`, `pbandwidth()`, and `plambda()` are not specified, the values given in `kernel()`, `bandwidth()`, and `lambda()` are taken as default.

The positive weights are always estimated by local linear regression. After estimation, negative estimated weights are set to zero. The smoothing parameters `pbandwidth()` and `plambda()` are in principle as important as the other smoothing parameters `bandwidth()` and `lambda()`, and it is worth inspecting the robustness of the results with respect to these parameters. Cross-validation can also be used to guide these choices.

Inference

`variance` activates the estimation of the variance. By default, no standard errors are estimated because the estimation of the variance can be computationally demanding. Except for the classical linear quantile regression estimator, it requires the estimation of many nonparametric functions. This option should not be activated if you bootstrap the results unless you bootstrap t-values to exploit possible asymptotic refinements.

`vbandwidth(`#`)`, `vlambda(`#`)`, and `vkernel(`*kernel*`)` are used to calculate the variance if the `variance` option has been selected. They are defined similarly to `bandwidth()`, `lambda()`, and `kernel()`. They are used only to estimate the variance. A "quick and dirty" estimate of the variance can be obtained by setting `vbandwidth()` to infinity and `vlambda()` to 1, which is much faster than any other choice. When `vkernel()`, `vbandwidth()`, or `vlambda()` is not specified, the values given in `kernel()`, `bandwidth()`, and `lambda()` are taken as default.

`level(`#`)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`.

Saved propensity scores and weights

`generate_p(`*newvarname* [`, replace`]`)` generates *newvarname* containing the estimated propensity score. Remember that the propensity score is $\Pr(Z=1|X)$ for the Abadie, Angrist, and Imbens (2002) and Frölich and Melly (2008) estimators and is $\Pr(D=1|X)$ for the Firpo (2007) estimator. This may be useful if one wants to compare the results with and without the projection of the weights or to compare the conditional and unconditional QTEs under endogeneity. One can first estimate the QTEs using one method and save the propensity score in the variable *newvarname*. In the second step, one can use the already estimated propensity score as an input in the `phat()` option. The `replace` option allows `ivqte` to overwrite an existing variable or to create a new one where none exists.

`generate_w(`*newvarname* [`, replace`]`)` generates *newvarname* containing the estimated weights. This may be useful if you want to estimate several conditional QTEs. The weights must be estimated only once and then be given as an input in the `what()` option. The `replace` option allows `ivqte` to overwrite an existing variable or to create a new one where none exists.

`phat(`*varname*`)` gives the name of an existing variable containing the estimated instrument propensity score. The propensity score may have been estimated using `ivqte` or with any other command such as a series estimator.

`what(`*varname*`)` gives the name of an existing variable containing the estimated weights. The weights may have been estimated using `ivqte` or with any other command such as a series estimator.

3.4 Saved results

`ivqte` saves the following in `e()`:

Scalars

`e(N)`	number of observations
`e(bandwidth)`	bandwidth
`e(lambda)`	lambda
`e(pbandwidth)`	pbandwidth
`e(plambda)`	plambda
`e(vbandwidth)`	vbandwidth
`e(vlambda)`	vlambda
`e(pseudo_r2)`	pseudo-R^2 of the quantile regression
`e(compliers)`	proportion of compliers
`e(trimmed)`	number of observations trimmed

Macros

`e(command)`	`ivqte`
`e(depvar)`	name of dependent variable
`e(treatment)`	name of treatment variable
`e(instrument)`	name of IV
`e(continuous)`	name of continuous covariates
`e(dummy)`	name of binary covariates
`e(regressors)`	name of regressors (conditional QTEs)
`e(unordered)`	name of unordered covariates
`e(estimator)`	name of estimator
`e(ps_method)`	linear or logistic model
`e(optimization)`	algorithm used
`e(kernel)`	kernel function
`e(pkernel)`	kernel function for positive weights
`e(vkernel)`	kernel function for variance estimation

Matrices

`e(b)`	row vector containing the QTEs
`e(quantiles)`	row vector containing the quantiles at which the QTEs have been estimated
`e(V)`	matrix containing the variances of the estimated QTEs in the diagonal and 0 otherwise

Functions

`e(sample)`	marks the estimation sample

3.5 Simple examples (without local smoothing)

Having given the syntax for `ivqte` in a previous subsection, we now illustrate how the command can be used with some very simple examples. In particular, we defer the use of smoothing parameters (h, λ) to the next subsection to keep things simple here. This means that all regressions will be estimated parametrically because the default values are $h = \infty$ and $\lambda = 1$.

We use the "distance to college" dataset of `card.dta`.[17] The aim is to estimate the effect of having a college degree (`college`) on log wages (`lwage`), controlling for parental education, experience, race, and region. A potential instrument is living near a four-year college (`nearc4`). The control variables are experience, `exper` (continuous variable); mother's education, `motheduc` (ordered discrete); `region` (unordered discrete); and `black` (dummy).

We first consider the quantile regression estimator with exogenous regressors for the first decile. As mentioned, this estimator is already implemented in Stata with the `qreg` command:

```
. use card
. qreg lwage college exper black motheduc reg662 reg663 reg664 reg665 reg666
> reg667 reg668 reg669, quantile(0.1)
  (output omitted)
```

The syntax of the `ivqte` command is similar with the exception that the treatment variable has to be included in parentheses after all other regressors:[18]

```
. ivqte lwage exper black motheduc reg662 reg663 reg664 reg665 reg666 reg667
> reg668 reg669 (college), quantiles(0.1) variance
  (output omitted)
```

The point estimates are exactly identical because `ivqte` calls `qreg`, but the standard errors differ. We recommend using the standard errors of `ivqte` because they are robust against heteroskedasticity and other forms of dependence between the residuals and the regressors.

We may be concerned that having a college degree might be endogenous and consider using the "proximity of a four-year college" as an instrument. The proximity of a four-year college is a binary variable, taking the value 1 if a college was close by. If we are interested in the conditional QTE, we can apply the estimator suggested by Abadie, Angrist, and Imbens (2002), as follows:

```
. ivqte lwage (college=nearc4), quantiles(0.1) variance dummy(black)
> continuous(exper motheduc) unordered(region) aai
  (output omitted)
```

There are three differences compared with the previous syntax: First, the instrument has to be given in parentheses after the treatment variable and the equal sign, that is, `(college=nearc4)`. Second, the control variables X are to be given in the corresponding options—`dummy()`, `continuous()`, and `unordered()`—because they are used not only to define the conditional QTE but also in the nonparametric estimation of the weights. Third, the `aai` option must be activated. `region` enters here as a single unordered

[17] This dataset is available for download from Stata together with the programs described in the present article. The description of the variables can be found in Card (1995).

[18] For this case of exogenous conditional QTEs, it is in principle arbitrary which variable is defined as the treatment variable because the coefficients are estimated for all regressors. In addition, nonbinary treatments are permitted here.

discrete variable, which is expanded by `ivqte` to eight regional dummy variables in the parametric model.

The two examples discussed so far refer to the conditional treatment effect of college degree. We might be more interested in the unconditional QTE, which we examine in the following example. Consider first the case where college degree is exogenous conditional on X. The weighting estimator of Firpo (2007) is implemented by `ivqte`. We are interested in the nine decile treatment effects with this estimator:

```
. ivqte lwage (college), variance dummy(black) continuous(exper motheduc)
> unordered(region)
  (output omitted)
```

Only the treatment is given in parentheses, and the `aai` option is no longer activated.

Finally, to estimate unconditional QTEs with an endogenous treatment, the estimator of Frölich and Melly (2008) is implemented in `ivqte`. The only difference from the previous syntax is that the instrument (`nearc4`) now has to be given after the treatment variable:

```
. ivqte lwage (college = nearc4), variance dummy(black)
> continuous(exper motheduc) unordered(region)
  (output omitted)
```

By default, the weights defined in (4) are used. If the `positive` option is activated, the positive weights (5) are estimated and used:

```
. ivqte lwage (college = nearc4), variance dummy(black)
> continuous(exper motheduc) unordered(region) positive
  (output omitted)
```

If no control variables are included, then `ivqte lwage (college = nearc4), aai` and `ivqte lwage (college = nearc4), positive` produce the same results.

3.6 Advanced examples (with local smoothing)

In the examples given above, we have not used the smoothing options. Therefore, by default, parametric regressions have been used to estimate all functions. In an application, we should use smoothing parameters converging to 0, unless we have strong reasons to believe that we do know the true functional forms. Appendix B contains many details about the nonparametric estimation of functions. We illustrate here the use of these techniques for `ivqte`.

(*Continued on next page*)

We use card.dta and keep only 500 randomly sampled observations from card.dta to reduce computation time. Because of missing values on covariates, eventually only 394 observations are retained in the estimation. The aim is to estimate the effect of having a college degree (college) on log wages (lwage). A potential instrument is living near a four-year college (nearc4). For ease of presentation, we use only experience (exper) as a continuous control variable here. The other control variables are region (unordered()) and black (dummy()).

Depending on the estimator, up to three functions have to be estimated nonparametrically. Three sets of options correspond to these three functions. The options kernel(), bandwidth(), and lambda() determine the kernel and the parameters h and λ used for the estimation of the propensity score. It corresponds to $\Pr(Z=1|X)$ for the Abadie, Angrist, and Imbens (2002) and Frölich and Melly (2008) estimators and to $\Pr(D=1|X)$ for the Firpo (2007) estimator.

The options pkernel(), pbandwidth(), and plambda() determine the kernel and smoothing parameters used for the estimation of the positive weights defined in (3) for the estimator of Abadie, Angrist, and Imbens (2002). If the Frölich and Melly (2008) estimator is to be used and the positive option has been activated, pkernel() and pbandwidth() are used to estimate the positive weights (5).

Finally, the options vkernel(), vbandwidth(), and vlambda() are used for the estimation of the variances of the estimators of Abadie, Angrist, and Imbens (2002), Firpo (2007), and Frölich and Melly (2008).

A general finding in the literature is that the choice of the kernel functions—kernel(), pkernel(), and vkernel()—is rarely crucial. The options vbandwidth() and vlambda() are used only for the estimation of the variance. Therefore, during the exploratory analysis, it may make sense to reduce the computational time by setting vbandwidth() to infinity and vlambda() to one, that is, a parametric model. For the final set of estimates, it often makes sense to set vbandwidth() equal to bandwidth() and vlambda() equal to lambda(). This is done by default. In the following illustration, we show how the cross-validation procedure implemented in locreg can be used to guide the choice of the important smoothing parameters.

We start with the estimator proposed by Firpo (2007). We do not need to use the options pkernel(), pbandwidth(), and plambda() because the weights are always positive by definition. We choose $h=2$ and $\lambda=0.8$ for the estimation of the propensity score. In addition, we use $h=\infty$ and $\lambda=1$ for the estimation of the variance. We use the default Epanechnikov kernel in all cases.

```
. use card, clear
. set seed 123
. sample 500, count
(2510 observations deleted)
. ivqte lwage (college), quantiles(0.5) dummy(black) continuous(exper)
> unordered(region) bandwidth(2) lambda(0.8) variance vbandwidth(.) vlambda(1)
  (output omitted)
```

Of course, these choices of the smoothing parameters are arbitrary. One can use the cross-validation option of `locreg` to choose the smoothing parameters. When we use the Firpo (2007) estimator, we know that the options `bandwidth()` and `lambda()` are used to estimate $\Pr(D=1\,|X)$. Therefore, we can select the smoothing parameters from a grid of, say, four possible values, as follows. We use the `logit` option because D is a binary variable.

```
. locreg college, dummy(black) continuous(exper) unordered(region)
> bandwidth(1 2) lambda(0.8 1) logit
  (output omitted)
```

The scalars `r(optb)` and `r(optl)` indicate that the choices of $h=1$ and $\lambda=1$ minimize the cross-validation criterion. We use the 2×2 search grid only for ease of exposition. Usually, one would search within a much larger grid. We can now obtain the point estimate using this choice of the smoothing parameters:

```
. ivqte lwage (college), quantiles(0.5) dummy(black) continuous(exper)
> unordered(region) bandwidth(1) lambda(1)
  (output omitted)
```

In addition to the values suggested by cross-validation, the user is encouraged to also try other—especially smaller—smoothing parameters and examine the robustness of the final results. For instance, we examine the results with $h=0.5$ and $\lambda=0.5$.

```
. ivqte lwage (college), quantiles(0.5) dummy(black) continuous(exper)
> unordered(region) bandwidth(0.5) lambda(0.5)
  (output omitted)
```

In this case, the results are relatively stable.

When we use the estimator of Abadie, Angrist, and Imbens (2002), we have to additionally specify `pbandwidth()` and `plambda()` for estimating the positive weights. We proceed by first choosing values for `bandwidth()` and `lambda()` and thereafter choosing values for `pbandwidth()` and `plambda()`. We know that the options `bandwidth()` and `lambda()` are used to estimate $\Pr(Z=1\,|X)$. Therefore, we can select the smoothing parameters from a grid of four possible values, as follows. Again we use the `logit` option because Z is a binary variable.

```
. locreg nearc4, dummy(black) continuous(exper) unordered(region)
> bandwidth(0.5 0.8) lambda(0.8 1) generate(ps) logit
  (output omitted)
```

The optimal smoothing parameters are $h=0.8$ and $\lambda=0.8$. The `generate(ps)` option implies that the fitted values $E(Z=1\,|X)$ are saved in the variable `ps`. These fitted values are generated using the optimal bandwidth. That is, they are generated after the cross-validation has selected the optimal bandwidth.

In the next step, we need to select bandwidths for `pbandwidth()` and `plambda()`. We know by equations (2) and (3) that `pbandwidth()` and `plambda()` are used to estimate

$$E\left(W_i^{\mathrm{AAI}}\,|Y_i,D_i,X_i\right)=E\left\{1-\frac{D_i\,(1-Z_i)}{1-\Pr\left(Z=1\,|X_i\right)}-\frac{(1-D_i)\,Z_i}{\Pr\left(Z=1\,|X_i\right)}\,|Y_i,D_i,X_i\right\}$$

We first generate W_i^{AAI}:

```
. generate waai=1-college*(1-nearc4)/(1-ps)-(1-college)*nearc4/ps
```

Then we can use `locreg` to find the optimal parameters. The positive weights are obtained by a nonparametric regression of W^{AAI} on X and Y and D. This is implemented in `ivqte` via two separate regressions: one nonparametric regression of W^{AAI} on X and Y for the $D=1$ subsample and one separate nonparametric regression of W^{AAI} on X and Y for the $D=0$ subsample. We proceed in the same way here by adding Y, which in our example above is `lwage`, as a continuous regressor and running separate regressions for the `college==1` and the `college==0` subsamples:

```
. locreg waai if college==1, dummy(black) continuous(exper lwage)
> unordered(region) bandwidth(0.5 0.8) lambda(0.8 1)
  (output omitted)
. locreg waai if college==0, dummy(black) continuous(exper lwage)
> unordered(region) bandwidth(0.5 0.8) lambda(0.8 1)
  (output omitted)
```

In the first case (that is, for the `college==1` subsample), the optimal smoothing parameters are $h=0.8$ and $\lambda=1$. For the `college==0` subsample, the optimal smoothing parameters are $h=0.8$ and $\lambda=0.8$. The current implementation of `ivqte` permits only one value of h and λ in the options `pbandwidth()` and `plambda()` to not overburden the user with choosing nuisance parameters. If the suggested values for h and λ are different for the `college==1` and the `college==0` subsamples, we recommend choosing the smaller of these values but also examining the robustness of the results to the other values. We suggest using the smaller bandwidth because the inference provided by `ivqte` is based on the asymptotic formula given in (7) (see Appendix A.2), which only contains variance but no bias term. To increase the accurateness of the inference based on (7), one would prefer bandwidth choices that lead to smaller biases.

In our example, we choose `pbandwidth(0.8)`, which was suggested by cross-validation in the `college==1` and the `college==0` subsamples, and `plambda(0.8)`, which is the smaller value of λ. With these bandwidth choices, we obtain the final estimates:

```
. ivqte lwage (college=nearc4), aai quantiles(0.5) dummy(black)
> continuous(age fatheduc motheduc) unordered(region) bandwidth(0.8) lambda(0.8)
> pbandwidth(0.8) plambda(0.8) variance
  (output omitted)
```

By using the `variance` option without specifying `vbandwidth()` nor `vlambda()`, the values given for `bandwidth()` and `lambda()` are used as defaults for `vbandwidth()` and `vlambda()`. Alternatively, we could have specified different values for `vbandwidth()` and vlambda(). In an exploratory analysis, one could use, for example, $h = \infty$ and $\lambda = 1$, which are certainly nonoptimal choices but reduce computation time considerably.

```
. ivqte lwage (college=nearc4), aai quantiles(0.5) dummy(black)
> continuous(age fatheduc motheduc) unordered(region) bandwidth(0.8) lambda(0.8)
> pbandwidth(0.8) plambda(0.8) variance vbandwidth(.) vlambda(1)
  (output omitted)
```

3.7 Replication of results of AAI

In the last illustration of the ivqte command, we replicate tables II-A and III-A of Abadie, Angrist, and Imbens (2002). jtpa.dta contains their dataset for males. We can replicate the point estimates of table II A with the official Stata qreg command:

```
. use jtpa, clear
. global reg "highschool black hispanic married part_time classroom
> OJT_JSA age5 age4 age3 age2 age1 second_follow"
. qreg earnings treatment $reg, quantile(0.5)
  (output omitted)
```

The same point estimates can also be obtained using ivqte:

```
. ivqte earnings $reg (treatment), quantiles(0.5) variance
  (output omitted)
```

The standard errors are still different from the standard errors in the published article because Abadie, Angrist, and Imbens (2002) have used a somewhat unconventional bandwidth. We can replicate their standard errors of table II-A by activating the aai option.[19] The following command calculates the results for the median.

```
. ivqte earnings (treatment=treatment), quantiles(0.5) dummy($reg) variance
> aai
  (output omitted)
```

Now we attempt to replicate the results of table III-A. Using a bandwidth of 2,

```
. ivqte earnings (treatment=assignment), quantiles(0.5) dummy($reg) variance
> aai pbandwidth(2)
  (output omitted)
```

[19] In this command, we use the fact that the estimator of Abadie, Angrist, and Imbens (2002) simplifies to the standard quantile regression estimator when the treatment is used as its own instrument. Similar relationships exist for the estimator proposed by Frölich and Melly (2008) and are discussed in their article.

gives results that are slightly different from their table III-A. We cannot exactly replicate their results because they have used series estimators to estimate the nonparametric components of their estimator and because they have exploited the fact that the instrument was completely randomized.[20] In the following commands, we show how `ivqte` with the options `phat(`*varname*`)` and `what(`*varname*`)` can be used to replicate their results. The parameters and standard errors are then almost identical to the original results.[21]

Abadie, Angrist, and Imbens (2002) first note that the instrument `assignment` has been fully randomized. Therefore, they estimate $\Pr(Z = 1 | X)$ by the sample mean of Z. Then the negative and positive weights W_i^{AAI} can be generated:

```
. summarize assignment
  (output omitted)
. generate pi=r(mean)
. generate kappa=1-treatment*(1-assignment)/(1-pi)-(1-treatment)*assignment/pi
```

In a second step, the positive weights $W_i^{\mathrm{AAI}+}$ are estimated by a linear regression of W_i^{AAI} on a polynomial of order 5 in Y and D:

```
. forvalues i=1/5 {
  2. generate e`i´=earnings^`i´
  3. generate de`i´=e`i´*treatment
  4. }
. regress kappa earnings treatment e2 e3 e4 e5 de1 de2 de3 de4 de5
  (output omitted)
. predict kappa_pos
(option xb assumed; fitted values)
. ivqte earnings (treatment=assignment) if kappa_pos>0, dummy($reg)
> quantiles(0.5) variance aai what(kappa_pos) phat(pi)
  (output omitted)
```

which gives almost the same estimates and standard errors as their table III-A.

4 Acknowledgments

We would like to thank Ben Jann, Andreas Landmann, Robert Poppe, and an anonymous referee for helpful comments.

[20] There are no strong reasons to prefer series estimators to local nonparametric estimators. We will use a series estimator here only to show that we can replicate their results. Actually, Frölich and Melly (2008) require, in some sense, weaker regularity assumptions for the local estimator than what was required for the existing series estimators.

[21] A small difference still remains, which is due to differences in the implementation of the estimator of $H(X)$. With slight adaptations, which would restrict the generality of the estimator, we can replicate their results exactly.

5 References

Abadie, A. 2003. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* 113: 231–263.

Abadie, A., J. Angrist, and G. Imbens. 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70: 91–117.

Card, D. E. 1995. Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of Labour Economics: Essays in Honour of John Vanderkamp*, ed. L. Christofides, E. K. Grant, and R. Swindinsky. Toronto, Canada: University of Toronto Press.

Chernozhukov, V., I. Fernández-Val, and A. Galichon. 2010. Quantile and probability curves without crossing. *Econometrica* 78: 1093–1125.

Chernozhukov, V., and C. Hansen. 2005. An IV model of quantile treatment effects. *Econometrica* 73: 245–261.

Firpo, S. 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75: 259–276.

Frölich, M. 2007a. Propensity score matching without conditional independence assumption—with an application to the gender wage gap in the United Kingdom. *Econometrics Journal* 10: 359–407.

———. 2007b. Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics* 139: 35–75.

Frölich, M., and B. Melly. 2008. Unconditional quantile treatment effects under endogeneity. Discussion Paper No. 3288, Institute for the Study of Labor (IZA). http://ideas.repec.org/p/iza/izadps/dp3288.html.

Hall, P., and S. J. Sheather. 1988. On the distribution of a studentized quantile. *Journal of the Royal Statistical Society, Series B* 50: 381–391.

Jann, B. 2005a. kdens: Stata module for univariate kernel density estimation. Statistical Software Components S456410, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s456410.html.

———. 2005b. moremata: Stata module (Mata) to provide various Mata functions. Statistical Software Components S455001, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s455001.html.

Koenker, R. 2005. *Quantile Regression*. New York: Cambridge University Press.

Koenker, R., and G. Bassett Jr. 1978. Regression quantiles. *Econometrica* 46: 33–50.

Melly, B. 2006. Estimation of counterfactual distributions using quantile regression. Discussion paper, Universität St. Gallen. http://www.alexandria.unisg.ch/Publikationen/22644.

Powell, J. L. 1986. Censored regression quantiles. *Journal of Econometrics* 32: 143–155.

Racine, J., and Q. Li. 2004. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* 119: 99–130.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

About the authors

Markus Frölich is a full professor of econometrics at the University of Mannheim and is Program Director for Employment and Development at the Institute for the Study of Labor. His research interests include policy evaluation, microeconometrics, labor economics, and development economics.

Blaise Melly is an assistant professor of economics at Brown University. He specializes in microeconometrics and applied labor economics and has special interests in the effects of policies on the distribution of outcome variables.

A Variance estimation

In this section, we describe the analytical variance estimators implemented in `ivqte`. The bootstrap represents an alternative and can be implemented in Stata using the `bootstrap` prefix. The validity of the bootstrap has been proven for standard quantile regression but not for the other estimators so far, but it seems likely that it is valid.

A.1 Conditional exogenous QTEs

Let $\mathbf{X} = (D, X')'$ and $\gamma^\tau = (\delta^\tau, \beta^{\tau\prime})'$. The asymptotic distribution of the quantile regression estimator defined in (1) is given by[22]

$$\sqrt{n}\left(\widehat{\gamma}^\tau - \gamma^\tau\right) \to N\left(0, J_\tau^{-1}\Sigma_\tau J_\tau^{-1}\right)$$

where $J_\tau = E\left\{f_{Y|\mathbf{X}}\left(\mathbf{X}'\gamma^\tau\right) \times \mathbf{X}\mathbf{X}'\right\}$ and $\Sigma_\tau = \tau\left(1-\tau\right)E\left(\mathbf{X}\mathbf{X}'\right)$. The term Σ_τ is straightforward to estimate by $\tau\left(1-\tau\right)n^{-1}\sum \mathbf{X}_i\mathbf{X}_i'$. We estimate J_τ by the kernel method of Powell (1986),

$$\widehat{J}_\tau = \frac{1}{nh_n}\sum k\left(\frac{Y_i - \mathbf{X}_i'\widehat{\gamma}^\tau}{h_n}\right)\mathbf{X}_i\mathbf{X}_i'$$

[22] See, for example, Koenker (2005).

where k is a univariate kernel function and h_n is a bandwidth sequence. In the actual implementation, we use a normal kernel and the bandwidth suggested by Hall and Sheather (1988),

$$h_n = n^{-1/3}\Phi^{-1}\left(1 - \mathrm{level}/2\right)^{2/3}\left[\frac{1.5\phi\left\{\Phi^{-1}\left(\tau\right)\right\}^2}{2\phi\left\{\Phi^{-1}\left(\tau\right)\right\}^2 + 1}\right]^{1/3}$$

where level is the level for the intended confidence interval, and ϕ and Φ are the normal density and distribution functions, respectively. This estimator of the asymptotic variance is consistent under heteroskedasticity, which is in contrast to the official Stata command for quantile regression, `qreg`. This is important because quantile regression only becomes interesting when the errors are not independent and identically distributed.

A.2 Conditional endogenous QTEs

The asymptotic distribution of the IV quantile regression estimator defined in (2) is given by

$$\sqrt{n}\left(\widehat{\gamma}_{\mathrm{IV}}^{\tau} - \gamma^{\tau}\right) \rightarrow N\left(0, I_\tau^{-1}\Omega_\tau I_\tau^{-1}\right) \tag{7}$$

where $I_\tau = E\left\{f_{Y|\mathbf{X},D_1>D_0}\left(\mathbf{X}'\gamma^\tau\right) \times \mathbf{X}\mathbf{X}'\left|D_1 > D_0\right.\right\}\Pr\left(D_1 > D_0\right)$ and $\Omega_\tau = E\left(\psi\psi'\right)$ with $\psi = W^{\mathrm{AAI}} m_\tau\left(\mathbf{X}, Y\right) + H\left(X\right)\left\{Z - \Pr\left(Z = 1\left|X\right.\right)\right\}$ and
$m_\tau\left(\mathbf{X}, Y\right) = \left\{\tau - 1\left(Y - \mathbf{X}'\gamma^\tau < 0\right)\right\}\mathbf{X}$ and
$H\left(X\right) = E\Big\{m_\tau\left(\mathbf{X}, Y\right)\Big(-\Big[D\left(1 - Z\right)/\left\{1 - \Pr\left(Z = 1\left|X\right.\right)\right\}^2\Big] +$
$\left(1 - D\right)Z/\Pr\left(Z = 1\left|X\right.\right)^2\Big)\left|X\right.\Big\}.$

We estimate these elements as

$$\widehat{I}_\tau = \frac{1}{nh_n}\sum \widehat{W}_i^{\mathrm{AAI+}} \times k\left(\frac{Y_i - \mathbf{X}_i'\widehat{\gamma}_{\mathrm{IV}}^{\tau}}{h_n}\right)\mathbf{X}_i\mathbf{X}_i'$$

where $\widehat{W}_i^{\mathrm{AAI+}}$ are estimates of the projected weights. For the kernel function in the previous regression, we use an Epanechnikov kernel and $h_n = n^{-0.2}\sqrt{\mathrm{Var}\left(\mathrm{Y_i} - \mathbf{X}_\mathrm{i}'\widehat{\gamma}_{\mathrm{IV}}^{\tau}\right)}$ as proposed by Abadie, Angrist, and Imbens (2002).[23]

Furthermore, $\widehat{H\left(X_i\right)}$ is estimated by the local linear regression of

$$\left\{\tau - 1\left(Y_i - \mathbf{X}_i'\widehat{\gamma}_{\mathrm{IV}}^{\tau} < 0\right)\right\}\mathbf{X}_i\left[-\frac{D_i\left(1 - Z_i\right)}{\left\{1 - \widehat{\Pr}\left(Z = 1\left|X_i\right.\right)\right\}^2} + \frac{\left(1 - D_i\right)Z_i}{\widehat{\Pr}\left(Z = 1\left|X\right.\right)^2}\right] \quad \text{on} \quad X_i$$

This nonparametric regression is controlled by the options `vkernel()`, `vbandwidth()`, and `vlambda()` in `ivqte`. With these ingredients, we calculate

[23] In principle, the same kernel and bandwidth as those for quantile regression can be used. These choices were made to replicate the results of Abadie, Angrist, and Imbens (2002).

$$\begin{aligned}
\widehat{\psi}_i &= \widehat{W}_i^{\mathrm{AAI}} \left\{\tau - 1\left(Y_i - \mathbf{X}_i'\widehat{\gamma}_{\mathrm{IV}}^{\tau} < 0\right)\right\} \mathbf{X}_i + \widehat{H\left(X_i\right)} \times \left\{Z_i - \widehat{\Pr}\left(Z = 1\left|X_i\right.\right)\right\} \\
\widehat{\Omega}_\tau &= \frac{1}{n} \sum \widehat{\psi}_i \widehat{\psi}_i'
\end{aligned}$$

where $\widehat{W}_i^{\mathrm{AAI}}$ are estimates of the weights.

A.3 Unconditional exogenous QTEs

The asymptotic distribution of the estimator defined in (6) is given by

$$\sqrt{n}\left(\widehat{\Delta}^\tau - \Delta^\tau\right) \to N\left(0, \mathcal{V}\right)$$

with

$$\begin{aligned}
\mathcal{V} &= \frac{1}{f_{Y^1}^2\left(Q_{Y^1}^\tau\right)} E\left[\frac{F_{Y|D=1,X}\left(Q_{Y^1}^\tau\right)\left\{1 - F_{Y|D=1,X}\left(Q_{Y^1}^\tau\right)\right\}}{\Pr(D=1|X)}\right] \\
&\quad + \frac{1}{f_{Y^0}^2\left(Q_{Y^0}^\tau\right)} E\left[\frac{F_{Y|D=0,X}\left(Q_{Y^0}^\tau\right)\left\{1 - F_{Y|D=0,X}\left(Q_{Y^0}^\tau\right)\right\}}{1 - \Pr(D=1|X)}\right] \\
&\quad + E\left[\left\{\vartheta_1(X) - \vartheta_0(X)\right\}^2\right]
\end{aligned}$$

where $\vartheta_d(x) = \tau - F_{Y|D=d,X}(Q_{Y^d}^\tau)/f_{Y^d}(Q_{Y^d}^\tau)$. $Q_{Y^0}^\tau$ and $Q_{Y^1}^\tau$ have already been estimated by $\widehat{\alpha}$ and $\widehat{\alpha}+\widehat{\Delta}^\tau$, respectively. The densities $f_{Y^d}(Q_{Y^d}^\tau)$ are estimated by weighted kernel estimators

$$\widehat{f}_{Y^d}\left(\widehat{Q}_{Y^d}^\tau\right) = \frac{1}{nh_n} \sum_{D_i = d} \widehat{W}_i^F \times k\left(\frac{Y_i - \widehat{Q}_{Y^d}^\tau}{h_n}\right)$$

with Epanechnikov kernel function and Silverman (1986) bandwidth choice, and where $F_{Y|D=d,X}(Q_{Y^d}^\tau)$ is estimated by the local logit estimator described in appendix B.

A.4 Unconditional endogenous QTEs

Finally, the asymptotic variance of the estimator defined in (4) is the most tedious and is given by

$$
\begin{aligned}
\mathcal{V} =& \frac{1}{P_c^2 f^2_{Y^1|c}\left(Q^\tau_{Y^1|c}\right)} E\left[\frac{\pi(X,1)}{p(X)} F_{Y|D=1,Z=1,X}\left(Q^\tau_{Y^1|c}\right)\left\{1 - F_{Y|D=1,Z=1,X}\left(Q^\tau_{Y^1|c}\right)\right\}\right] \\
&+ \frac{1}{P_c^2 f^2_{Y^1|c}\left(Q^\tau_{Y^1|c}\right)} E\left[\frac{\pi(X,0)}{1-p(x)} F_{Y|D=1,Z=0,X}\left(Q^\tau_{Y^1|c}\right)\left\{1 - F_{Y|D=1,Z=0,X}\left(Q^\tau_{Y^1|c}\right)\right\}\right] \\
&+ \frac{1}{P_c^2 f^2_{Y^0|c}\left(Q^\tau_{Y^0|c}\right)} E\left[\frac{1-\pi(X,1)}{p(X)} F_{Y|D=0,Z=1,X}\left(Q^\tau_{Y^0|c}\right)\left\{1 - F_{Y|D=0,Z=1,X}\left(Q^\tau_{Y^0|c}\right)\right\}\right] \\
&+ \frac{1}{P_c^2 f^2_{Y^0|c}\left(Q^\tau_{Y^0|c}\right)} E\left[\frac{1-\pi(X,0)}{1-p(X)} F_{Y|D=0,Z=0,X}\left(Q^\tau_{Y^0|c}\right)\left\{1 - F_{Y|D=0,Z=0,X}\left(Q^\tau_{Y^0|c}\right)\right\}\right] \\
&+ E\left[\frac{\pi(X,1)\vartheta^2_{11}(X) + \{1-\pi(X,1)\}\,\vartheta^2_{01}(X)}{p(X)} + \frac{\pi(X,0)\vartheta^2_{10}(X) + \{1-\pi(X,0)\}\,\vartheta^2_{00}(X)}{1-p(X)}\right] \\
&- E\Bigg(p(X)\{1-p(X)\}\left[\frac{\pi(X,1)\vartheta_{11}(X) + \{1-\pi(X,1)\}\,\vartheta_{01}(X)}{p(X)}\right. \\
&\quad \left. + \frac{\pi(X,0)\vartheta_{10}(X) + \{1-\pi(X,0)\}\,\vartheta_{00}(X)}{1-p(X)}\right]^2\Bigg)
\end{aligned}
$$

where $\vartheta_{dz}(x) = \tau - F_{Y|D=d,Z=z,X}(Q^\tau_{Y^d|c})/P_c \times f_{Y^d|c}(Q^\tau_{Y^d|c})$; $p(x) = \Pr(Z = 1|X = x)$; $\pi(x, z) = \Pr(D = 1|X = x, Z = z)$; and P_c is the fraction of compliers. $Q^\tau_{Y^0|c}$ and $Q^\tau_{Y^1|c}$ have already been estimated by $\widehat{\alpha}_{\mathrm{IV}}$ and $\widehat{\alpha}_{\mathrm{IV}} + \widehat{\Delta}^\tau_{\mathrm{IV}}$, respectively. The terms $F_{Y|D=d,Z=z,X}(Q^\tau_{Y^d})$, $p(X)$, and $\pi(x, z)$ are estimated by the local logit estimator described in appendix B. Finally, P_c is estimated by $\sum \widehat{\pi}(X_i, 1) - \widehat{\pi}(X_i, 0)$.

(*Continued on next page*)

To estimate the densities $f_{Y^d|c}(Q^\tau_{Y^d|c})$, we note that[24]

$$f_{Y^d|c}(Q^\tau_{Y^d|c}) = \lim_{h\to\infty} \frac{1}{h}\int_0^1 k\left(\frac{Q^\theta_{Y^d|c} - Q^\tau_{Y^d|c}}{h}\right) d\theta$$

where k is the Epanechnikov kernel function with Silverman (1986) bandwidth. We therefore estimate $f_{Y^d|c}$ as

$$\widehat{f}_{Y^d|c}(\widehat{Q}^\tau_{Y^d|c}) = \frac{1}{h}\int_0^1 k\left(\frac{\widehat{Q}^\theta_{Y^d|c} - \widehat{Q}^\tau_{Y^d|c}}{h}\right) d\theta$$

where we replace the integral by a sum of n uniformly spaced values of θ between 0 and 1.

B Nonparametric regression with mixed data

B.1 Local parametric regression

A key ingredient for the previously introduced estimators (except for the exogenous conditional quantile regression estimator) is the nonparametric estimation of some weights. Local linear and local logit estimators have been implemented for this purpose. This is fully automated in the `ivqte` command. Nevertheless, some understanding of the nonparametric estimators facilitates the use of the `ivqte` command.

In many instances, we need to estimate conditional expected values like $E\left(Y\,|X = X_i\,\right)$. We use a local parametric approach throughout; that is, we estimate a locally weighted version of the parametric model. A complication is that many econometric applications contain continuous as well as discrete regressors X. Both types of regressors need to be accommodated in the local parametric model and in the kernel function defining the local neighborhood. The traditional nonparametric approach consists of estimating the model within each of the cells defined by the discrete regressors

[24] To see this, note that

$$\frac{1}{h}\int_0^1 k\left(\frac{Q^\theta_{Y^d|c} - Q^\tau_{Y^d|c}}{h}\right) d\theta = \int_{-\infty}^{\infty} k\left(u\right) \times f_{Y^d|c}(uh + Q^\tau_{Y^d|c}) \times du$$

where we used a change in variables $uh = Q^\theta_{Y^d|c} - Q^\tau_{Y^d|c}$, which implies that $\theta = F_{Y^d|c}(uh + Q^\tau_{Y^d|c})$ and $d\theta = f_{Y^d|c}(uh + Q^\tau_{Y^d|c}) \times hdu$. By the mean value theorem, $f_{Y^d|c}(uh + Q^\tau_{Y^d|c}) = f_{Y^d|c}(Q^\tau_{Y^d|c}) + uh \times f'_{Y^d|c}(\overline{Q})$, where $\overline{Q}$ is on the line between $Q^\tau_{Y^d|c}$ and $uh + Q^\tau_{Y^d|c}$. Hence,

$$= f_{Y^d|c}(Q^\tau_{Y^d|c}) \int_{-\infty}^{\infty} k\left(u\right) \times du + O(h) = f_{Y^d|c}(Q^\tau_{Y^d|c}) + O(h)$$

and of smoothing only with respect to the continuous covariates. When the number of cells in a dataset is large, each cell may not have enough observations to nonparametrically estimate the relationship among the remaining continuous variables. For this reason, many applied researchers have treated discrete variables in a parametric way. We follow an intermediate way and use the hybrid product kernel developed by Racine and Li (2004). This estimator covers all cases from the fully parametric model up to the traditional nonparametric estimator.

Overall, we can distinguish four different types of regressors: continuous (for example, age), ordered discrete (for example, family size), unordered discrete (for example, regions), and binary variables (for example, gender). We will treat ordered discrete and continuous variables in the same way and will refer to them as continuous variables in the following discussion.[25]

The unordered discrete and the binary variables are handled differently in the kernel function and in the local parametric model. The binary variables enter into both as single regressors. The unordered discrete variables, however, enter as a single regressor in the kernel function and as a vector of dummy variables in the local model. Consider, for example, a variable called region that takes four different values: north, south, east, and west. This variable enters as a single variable in the kernel function but is included in the local model in the form of three dummies: "south", "east", and "west".

The kernel function is defined in the following paragraph. Suppose that the variables in X are arranged such that the first q_1 regressors are continuous (including the ordered discrete variables) and the remaining $Q - q_1$ regressors are discrete without natural ordering (including binary variables). The kernel weights $K(X_i - x)$ are computed as

$$K_{h,\lambda}(X_i - x) = \prod_{q=1}^{q_1} \kappa\left(\frac{X_{q,i} - x_q}{h}\right) \times \prod_{q=q_1+1}^{Q} \lambda^{1(X_{q,i} \neq x_q)}$$

where $X_{q,i}$ and x_q denote the qth element of X_i and x, respectively; $1(\cdot)$ is the indicator function; κ is a symmetric univariate weighting function; and h and λ are positive bandwidth parameters with $0 \leq \lambda \leq 1$. This kernel function measures the distance between X_i and x through two components: The first term is the standard product kernel for continuous regressors with h defining the size of the local neighborhood. The second term measures the mismatch between the unordered discrete (including binary) regressors. λ defines the penalty for the unordered discrete regressors. For example, the multiplicative weight contribution of the Qth regressor is 1 if the Qth element of X_i and x are identical, and it is λ if they are different. If $h = \infty$ and $\lambda = 1$, then the nonparametric estimator corresponds to the global parametric estimator and no interaction term between the covariates is allowed. On the other hand, if λ is zero and h is small, then smoothing proceeds only within each of the cells defined by the

[25] Racine and Li (2004) suggest using a geometrically declining kernel function for the ordered discrete regressors. There are no reasons, however, against using quadratically declining kernel weights. In other words, we can use the same (for example, Epanechnikov) kernel for ordered discrete as for continuous regressors. We therefore treat ordered discrete regressors in the same way as continuous regressors in the following discussion.

discrete regressors and only observations with similar continuous covariates will be used. Finally, if λ and h are in the intermediate range, observations with similar discrete and continuous covariates will be weighted more but further observations will also be used.

Principally, instead of using only two bandwidth values h, λ for all regressors, a different bandwidth could be employed for each regressor, but doing so would substantially increase the computational burden for bandwidth selection. This approach might lead to additional noise due to estimating these bandwidth parameters. Therefore, we prefer to use only two smoothing parameters. `ivqte` automatically orthogonalizes the data matrix of all continuous regressors to create an identity covariance matrix. This greatly diminishes the appeal of having multiple bandwidths.

This kernel function, combined with a local model, is used to estimate $E(Y|X)$. If Y is a continuous variable, then `ivqte` uses by default a local linear estimator to estimate $E(Y|X=x)$ as $\widehat{\alpha}$ in

$$(\widehat{\alpha}, \widehat{\beta}) = \arg\min_{a,b} \sum_{j=1}^{n} \{Y_j - a - b(X_j - x)\}^2 \times K_{h,\lambda}(X_j - x)$$

If Y is bound from above and below, a local logistic model is usually preferred. We suppose in the following discussion that Y is bound within $[0,1]$.[26] This includes the special case where Y is binary. The local logit estimator guarantees that the fitted values are always between 0 and 1. The local logit estimator can be used by selecting the `logit` option. In this case, $E(Y|X=x)$ is estimated by $\Lambda(\widehat{\alpha})$, where

$$\begin{aligned}(\widehat{\alpha}, \widehat{\beta}) = \arg\min_{a,b} \sum_{j=1}^{n} & \left(Y_j \ln \Lambda\{a + b(X_j - x)\}\right. \\ & \left. + (1 - Y_j) \ln\left[1 - \Lambda\{a + b(X_j - x)\}\right]\right) \times K_{h,\lambda}(X_j - x)\end{aligned}$$

and $\Lambda(x) = 1/1 + e^{-x}$.

As mentioned before, each of the unordered discrete variables enters in the form of a dummy variable for each of its support points except for an arbitrary base category; for example, if the region variable takes four different values, then three dummies are included.

The `ivqte` command requires that the values of the smoothing parameters h and λ are supplied by the user. Before estimating local linear or local logit with these smoothing parameters, `ivqte` (as well as `locreg`) first attempts to estimate the global model (that is, with $h = \infty$ and $\lambda = 1$). If estimation fails due to collinearity or perfect prediction, the regressors which caused these problems are eliminated.[27] Thereafter, the model is estimated locally with the user-supplied smoothing parameters. If estimation

[26] If the lower and upper bounds of Y are different from 0 and 1, Y should be rescaled to the interval $[0,1]$.

[27] This is done using `_rmcollright`, where `ivqte` first searches for collinearity among the continuous regressors and thereafter among all other regressors.

fails locally because of collinearity or perfect prediction, the bandwidths are increased locally. This is repeated until convergence is achieved.

The `locreg` command also contains a leave-one-out cross-validation procedure to choose the smoothing parameters.[28] The user provides a grid of values for h and λ, and the cross-validation criterion is computed for all possible combinations of these values. The values of the cross-validation criterion are returned in `r(cross_valid)` and the combination that minimizes this criterion is chosen. If only one value is given for h and λ, no grid search is performed.

B.2 The locreg command

Because the codes implementing the nonparametric regressions are likely to be of independent interest in other contexts, we offer a separate command for the local parametric regressions. This `locreg` command implements local linear and local logit regression and chooses the smoothing parameters by leave-one-out cross-validation. The formal syntax of `locreg` is as follows:

```
locreg depvar [if] [in] [weight] [, generate(newvarname [, replace])
   continuous(varlist) dummy(varlist) unordered(varlist) kernel(kernel)
   bandwidth(#[#[# ...]]) lambda(#[#[# ...]]) logit mata_opt
   sample(varname [, replace]) ]
```

`aweights` and `pweights` are allowed. See the [U] **11.1.6 weight** for more information on weights.

B.3 Description

`locreg` computes the nonparametric estimation of the mean of *depvar* conditionally on the regressors given in `continuous()`, `dummy()`, and `unordered()`. A mixed kernel is used to smooth over the continuous and discrete regressors. The fitted values are saved in the variable *newvarname*. If a list of values is given in `bandwidth()` or `lambda()`, the smoothing parameters h and λ are estimated via leave-one-out cross-validation. The values of h and λ minimizing the cross-validation criterion are selected. These values are then used to predict *depvar*, and the fitted values are saved in the variable *newvarname*.

`locreg` can be used in three different ways. First, if only one value is given in `bandwidth()` and one in `lambda()`, `locreg` estimates the nonparametric regression using these values and saves the fitted values in `generate(`*newvarname*`)`. Alternatively,

[28] The cross-validated parameters are optimal to estimate the weights but are not optimal to estimate the unconditional QTE. In the absence of a better method, we offer cross-validation, but the user should keep in mind that the optimal bandwidths for the unconditional QTE converge to zero at a faster rate than the bandwidths delivered by cross-validation. The user is therefore encouraged to also examine the estimated QTE when using some undersmoothing relative to the cross-validation bandwidths.

`locreg` can also be used to estimate the smoothing parameters via leave-one-out cross-validation. If we do not specify the `generate()` option but supply a list of values in the `bandwidth()` or `lambda()` option only the cross-validation is performed. Finally, if several values are specified in `bandwidth()` or `lambda()` when the `generate()` option is also specified, `locreg` estimates the optimal smoothing parameters via cross-validation. Thereafter, it estimates the conditional means with these smoothing parameters and returns the fitted values in the variable `generate(`*newvarname*`)`.

For the nonparametric regression, `locreg` offers two local models: linear and logistic. The logistic model is usually preferred if *depvar* is bound within $[0, 1]$. This includes the case where *depvar* is binary but also incorporates cases where *depvar* is nonbinary but bound from above and below. If the lower and upper bounds of *depvar* are different from 0 and 1, the variable *depvar* should be rescaled to the interval $[0, 1]$ before using this command. If *depvar* is not bound from above and below, the linear model should be used.[29]

B.4 Options

`generate(`*newvarname* [`, replace`]`)` specifies the name of the variable that will contain the fitted values. If this option is not used, only the leave-one-out cross-validation estimation of the smoothing parameters h and λ will be performed. The `replace` option allows `locreg` to overwrite an existing variable or to create a new one where none exists.

`continuous(`*varlist*`)`, `dummy(`*varlist*`)`, and `unordered(`*varlist*`)` specify the names of the covariates depending on their type. Ordered discrete variables should be treated as continuous.

`kernel(`*kernel*`)` specifies the kernel function. *kernel* may be `epan2` (Epanechnikov kernel function; the default), `biweight` (biweight kernel function), `triweight` (triweight kernel function), `cosine` (cosine trace), `gaussian` (Gaussian kernel function), `parzen` (Parzen kernel function), `rectangle` (rectangle kernel function), or `triangle` (triangle kernel function). In addition to these second-order kernels, there are also several higher-order kernels: `epanechnikov_o4` (Epanechnikov order 4), `epanechnikov_o6` (order 6), `gaussian_o4` (Gaussian order 4), `gaussian_o6` (order 6), `gaussian_o8` (order 8).[30]

`bandwidth(`#[#[# ...]]`)` is used to smooth over the continuous variables. The default is $h = \infty$. The continuous regressors are first orthogonalized such that their covariance matrix is the identity matrix. The bandwidth must be strictly positive. If the bandwidth h is missing, an infinite bandwidth $h = \infty$ is used. The default value is infinity.

[29] In the current implementation, there is not yet a local model specifically designed for *depvar* that is bound only from above or only from below. A local tobit or local exponential model may be added in future versions.

[30] The formulas for the higher-order kernel functions are given in footnote 16.

If a list of values is supplied for `bandwidth()`, cross-validation is used with respect to each value in this list to estimate the bandwidth among the proposed values. If a list of values is supplied for `bandwidth()` and for `lambda()`, cross-validation considers all pairwise combinations from these two lists. In case of local multicollinearity, the bandwidth is progressively increased until the multicollinearity problem disappears.[31]

`lambda(`#[#[# ...]]`)` is used to smooth over the dummy and unordered discrete variables. It must be between 0 and 1. A value of 0 implies that only observations within the cell defined by all discrete regressors are used to estimate the conditional mean. The default is `lambda(1)`, which corresponds to global smoothing. If a list of values is supplied for `lambda()`, cross-validation is used with respect to each value in this list to estimate the lambda among the proposed values. If a list of values is supplied for `bandwidth()` and for `lambda()`, cross-validation considers all pairwise combinations from these two lists.

`logit` activates the local logit estimator. If it is not activated, the local linear estimator is used as the default.[32]

`mata_opt` selects the official optimizer introduced in Stata 10, Mata's `optimize()`, to obtain the local logit. The default is a simple Gauss–Newton algorithm written for this purpose. This option is only relevant when the `logit` option has been specified.

`sample(`*varname* [`, replace`]`)` specifies the name of the variable that marks the estimation sample. This is similar to the function `e(sample)` for e-class commands.

[31] In case of multicollinearity, h is increased repeatedly until the problem disappears. If multicollinearity is still present at $h = 100$, then λ is increased repeatedly. Note that `locreg` first examined whether multicollinearity is a problem in the global model ($h = \infty$, $\lambda = 1$) before attempting to estimate locally.

[32] This is different from `ivqte`, where local logit is the default for binary dependent variables.

B.5 Saved results

`locreg` saves the following in `r()`:

Scalars	
`r(N)`	number of observations
`r(optb)`	optimal bandwidth
`r(optl)`	optimal lambda
`r(best_mse)`	smallest value of the cross-validation criterion
Macros	
`r(command)`	`locreg`
`r(depvar)`	name of the dependent variable
`r(continuous)`	name of the continuous covariates
`r(dummy)`	name of the binary covariates
`r(unordered)`	name of the unordered covariates
`r(kernel)`	kernel function
`r(model)`	linear or logistic model used
`r(optimization)`	algorithm used
Matrices	
`r(cross_valid)`	bandwidths, lambda, and resulting values of the cross-validation criterion

B.6 Examples

We briefly illustrate the use of `locreg` with a few examples. We use `card.dta` and keep only 200 observations to keep the computational time reasonable for this illustration. (Because of missing values on covariates, eventually only 184 observations are retained in the estimation.) The aim is to estimate the probability of living near a four-year college (`nearc4`) as a function of experience, `exper` (`continuous()` variable); mother's education, `motheduc` (ordered discrete); `region` (`unordered()`); and `black` (`dummy()`). `locreg` can be used in three different ways. First, if only one value is given in `bandwidth(#)` and one in `lambda(#)`, `locreg` estimates the nonparametric regression using these values h and λ and saves the fitted values in *newvarname*:

```
. use card, clear

. set seed 123

. sample 200, count
(2810 observations deleted)

. locreg nearc4, generate(fitted1) bandwidth(0.5) lambda(0.5)
> continuous(exper motheduc) dummy(black) unordered(region)
  (output omitted)
```

The `fitted1` variable contains the estimated probabilities. Because some of them turn out to be negative and others to be larger than one, we may prefer to fit a local logit regression and add the `logit` option:

```
. locreg nearc4, generate(fitted2) bandwidth(0.5) lambda(0.5)
> continuous(exper motheduc) dummy(black) unordered(region) logit
  (output omitted)
```

locreg can also be used to estimate the smoothing parameters via leave-one-out cross-validation. If we do not specify the generate() option but instead supply a list of values in the bandwidth() or the lambda() option (or both), only the cross-validation is performed:

```
. locreg nearc4, bandwidth(0.2 0.5) lambda(0.5 0.8) continuous(exper motheduc)
> dummy(black) unordered(region)
  (output omitted)
```

In this example, the cross-validation criterion is calculated for each of the four cases: $(h,\lambda) = (0.2, 0.5)$, $(0.2, 0.8)$, $(0.5, 0.5)$, and $(0.5, 0.8)$. The scalars r(optb) and r(optl) indicate the values that minimized the cross-validation criterion. In our example, we obtain $\widehat{h} = 0.2$ and $\widehat{\lambda} = 0.5$. The cross-validation results are saved in the matrix r(cross_valid) for every h and λ combination of the search grid.

If we would like to include in our cross-validation search the values to infinity, that is, the global parametric model, we would supply a missing value for h and a value of 1 for λ. For example, specifying bandwidth(0.2 .) lambda(0.5 1) implies that the cross-validation criterion is calculated for each of the four cases: $(h,\lambda) = (0.2, 0.5)$, $(0.2, 1)$, $(\infty, 0.5)$, and $(\infty, 1)$. Similarly, specifying bandwidth(0.2 0.5 .) lambda(0.5 0.8 1) implies a search grid with nine values: $(h,\lambda) = (0.2, 0.5)$, $(0.2, 0.8)$, $(0.2, 1)$, $(0.5, 0.5)$, $(0.5, 0.8)$, $(0.5, 1)$, $(\infty, 0.5)$, $(\infty, 0.8)$, and $(\infty, 1)$.

Finally, if several values are specified for the smoothing parameters and the generate() option is also activated, then locreg first estimates $\widehat{h}$ and $\widehat{\lambda}$ via cross-validation and thereafter returns the fitted values obtained with $\widehat{h}$ and $\widehat{\lambda}$ in the fitted3 variable.

```
. locreg nearc4, generate(fitted3) bandwidth(0.2 0.5) lambda(0.5 0.8)
> continuous(exper motheduc) dummy(black) unordered(region)
  (output omitted)
```