Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

ivcrc: An Instrumental Variables Estimator for the Correlated Random Coefficients Model

David Benson, Matthew A. Masten, Alexander Torgovitsky

2020-046

Please cite this paper as:
Benson, David, Matthew A. Masten, and Alexander Torgovitsky (2020). "ivcrc: An Instrumental Variables Estimator for the Correlated Random Coefficients Model," Finance and Economics Discussion Series 2020-046. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.046.

`ivcrc`: An Instrumental Variables Estimator for the Correlated Random Coefficients Model

David Benson* Matthew A. Masten† Alexander Torgovitsky‡

June 10, 2020

Abstract

We present the `ivcrc` command, which implements an instrumental variables (IV) estimator for the linear correlated random coefficients (CRC) model. This model is a natural generalization of the standard linear IV model that allows for endogenous, multivalued treatments and unobserved heterogeneity in treatment effects. The proposed estimator uses recent semiparametric identification results that allow for flexible functional forms and permit instruments that may be binary, discrete, or continuous. The command also allows for the estimation of varying coefficients regressions, which are closely related in structure to the proposed IV estimator. We illustrate this IV estimator and the `ivcrc` command by estimating the returns to education in the National Longitudinal Survey of Young Men.

Keywords: `ivregress`, Instrumental Variables, Correlated Random Coefficients, Heterogeneous Treatment Effects, Varying Coefficient Models, Returns to Schooling

*Division of Research and Statistics, Federal Reserve Board of Governors.

†Department of Economics, Duke University.

‡Kenneth C. Griffin Department of Economics, University of Chicago. Research supported in part by National Science Foundation grant SES-1846832.

1 Introduction

In this paper, we describe the `ivcrc` module for Stata, which implements a linear instrumental variables (IV) estimator that is easy to interpret in the presence of heterogeneous treatment effects. The estimator is based on recent identification arguments for a correlated random coefficients (CRC) model. These arguments are described in Masten and Torgovitsky (2016), while details on computation, estimation, and some asymptotic theory for the estimator are developed in Masten and Torgovitsky (2014).

The motivation for the module is causal inference, which is a frequent goal in empirical work. Suppose that we want to estimate the causal effect of a treatment variable X on an outcome Y. A common challenge with observational data is that it may be implausible to view X as if it were randomly assigned, even after conditioning on other observable variables. This is especially important in economic applications, where it is common for X to be a choice variable over which an economic agent has some control. If there are common unobserved variables that affect both the agent's choice of X and their realization of Y, then the relationship between X and Y in the data will reflect both the causal effect of X on Y and the confounding effects of these latent variables. Economists often describe this situation by saying that X is endogenous.

For example, let X be a measure of educational attainment, such as years of completed schooling, and Y a labor market outcome, such as wages. Suppose that we want to estimate the causal effect of X on Y. The magnitude of this effect is one measure of the returns to human capital, which plays a key role in many areas of economics. While X and Y tend to be strongly correlated, knowing the extent to which this correlation reflects an actual causal effect is crucial both for understanding the returns to human capital, and for evaluating the impact of policy counterfactuals, such as an expansion of subsidized tuition loans.

There are several reasons to expect confounding factors that make the direct relationship between X and Y a poor indicator of the causal effect of X on Y. Many of these confounding factors, like family background characteristics, can often be observed in data and controlled for. However, the fact that individuals have some choice over their attainment of education also suggests important confounding factors that are inherently unobservable. Key among these is unobserved heterogeneity in the private costs and benefits of schooling.

For example, suppose as a stylized exercise that the world consists of two types of individuals: Those who are good at mental abstraction, and those who are good at working with their hands. Individuals in the first group find school more enjoyable, and so are prone to obtain more education than individuals in the second group. Suppose in addition that for a given level of education, the labor market rewards to mental skills are greater than those for physical skills. Then individuals who obtain higher levels of education will also be more likely to have better labor market outcomes, even if education itself has no effect on these outcomes. The correlation between X and Y could be positive in this scenario solely because the individuals who choose to obtain more schooling tend to have latent traits that would be more richly rewarded in the labor market anyway.

Instrumental variable (IV) strategies are commonly used to tackle this type of selection bias. The idea of an IV strategy is to use variation in a third variable, Z (the instrument), that is exogenous with respect to the confounding variables, but correlated with X. Instruments that have been used to study the returns to schooling include compulsory schooling laws (Angrist and Krueger, 1991; Oreopoulos, 2006), the distance a teenager lives from a college (Card, 1993; Mountjoy, 2019), and local labor market conditions (Cameron and Heckman, 1998).[1] The argument underlying these strategies is that the proposed instrument affects an individual's educational attainment by shifting the costs and/or benefits involved, but does not itself directly affect labor market outcomes and is uncorrelated with any other factors that do.

Stata already has a built-in command called `ivregress` that can be used to estimate standard linear IV models. However, the estimator computed by this command is generally difficult to interpret unless one assumes that the causal effect of X on Y is unrelated to an individual's choice of X (Angrist and Imbens, 1995; Angrist, Graddy, and Imbens, 2000; Heckman and Vytlacil, 1998, 2005).[2] Assuming away of this type of selection on the gain (a form of heterogeneous treatment effects) is unattractive when X is a choice variable, because it means that an economic agent chooses X without knowing or considering the effect that it will have on Y. In the education example, this would require the unpalatable assumption that individuals decide on their schooling without considering the effects it will have on their future earnings. The CRC model addresses this criticism by allowing the causal effect of X on Y to be an unobservable random variable, one which is potentially correlated with X itself.

In Section 2, we briefly describe the IV CRC model implemented by the `ivcrc` module, as well as the identification results and estimation approach developed in Masten and Torgovitsky (2014, 2016). The structure of this IV estimator turns out to be quite similar to a common estimator for the varying coefficient models (e.g., Fan and Zhang 2008; Park, Mammen, Lee, and Lee 2015). We have written `ivcrc` to be able to treat a standard estimator for these models as a special case.[3] We briefly describe varying coefficient models in Section 3. In Section 4, we discuss syntax and options for the `ivcrc` module. In Section 5, we illustrate the module by estimating the return to schooling with a widely used extract from the National Longitudinal Survey of Young Men. For further usage examples, see Gollin and Udry (2020), who have used the `ivcrc` module to estimate agricultural production functions, and Masten and Torgovitsky (2014), who used the procedure to revisit Chay and Greenstone's (2005) analysis of the effect of air pollution on housing prices.

---

[1] See also Carneiro, Heckman, and Vytlacil (2011) for an IV strategy that uses multiple types of instruments.

[2] A notable exception is the case where both X and Z are binary and any additional included covariates are included in a fully saturated way. In this case, the estimator can be interpreted as estimating a weighted average of covariate specific local average treatment effects as long as an additional monotonicity assumption is maintained (Imbens and Angrist, 1994; Angrist and Imbens, 1995; Abadie, 2003). Models that are saturated in covariates quickly succumb to the curse of dimensionality, and so are rarely used in practice.

[3] Also see Rios-Avila (2019).

2 The Correlated Random Coefficients Model

2.1 Model and Motivation

The simplest form of the model estimated by `ivcrc` has the outcome equation

$$Y = B_0 + B_1 X, \tag{1}$$

where Y is an observed outcome, X is an observed explanatory variable, and both B_0 and B_1 are unobserved random variables. The model is described as a random coefficients model due to the treatment of B_1 as an unobserved random variable. Economists have long been interested in such models (Wald, 1947; Hurwicz, 1950; Rubin, 1950; Becker and Chiswick, 1966). To allow for endogeneity, X is permitted to be arbitrarily dependent with both B_0 and B_1. This feature makes the model one of *correlated* random coefficients.[4]

It is helpful to compare (1) with the outcome equation for the textbook linear model:

$$Y = \alpha + \beta X + U, \tag{2}$$

where α and β are fixed (deterministic) parameters, and U is an unobservable random variable with mean zero. This model also allows for endogeneity by permitting X to be dependent with U. The distinction between U in (2) and B_0 in (1) is not important, since one can view B_0 as being equal to $\alpha + U$. Rather, the important difference between (2) and (1) is that the coefficient on X in (2), i.e. β, is deterministic, whereas the coefficient on X in (1), i.e. B_1, is a random variable. The interpretation is that in (2) the causal effect of X on Y is the same for all agents, whereas in (1) it is a random variable that can be dependent with X. This important difference allows for heterogeneous treatment effects and selection on the gain of the sort described in the introduction. One can view (2) as a special case of (1) with a degenerate B_1.

Textbook discussions of (2) show that β is identified if there exists an instrument Z such that $\text{Cov}(U, Z) = 0$ and $\text{Cov}(X, Z) \neq 0$. The corresponding IV estimator can be implemented in Stata with the `ivregress` command. However, if the data is in fact generated by (1), then this estimator converges to

$$\frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)} = \mathbb{E}\left[B_1 \times \frac{X(Z - \mathbb{E}(Z))}{\mathbb{E}[X(Z - \mathbb{E}(Z))]}\right]. \tag{3}$$

This quantity is difficult to interpret in general (Garen, 1984; Wooldridge, 1997; Heckman and Vytlacil, 1998). It is a weighted average of the causal effect of X on Y; that is, a weighted average of B_1. The weights, however, can be both positive and negative. It generally does not equal the

[4] This terminology seems to have been first used by Heckman and Vytlacil (1998). In earlier work, some authors, for example, Conway and Kniesner (1991), had used the adjective "correlated" to describe an unrestricted correlation structure between the random coefficients on different explanatory variables. Our model also allows for this.

unweighted average of B_1 unless B_1 is independent of (X, Z), which would rule out the type of selection on the gain scenario discussed in the introduction.

A natural question is whether there are additional assumptions under which the IV estimator provided by `ivregress` would consistently estimate a parameter that is easier to interpret. For example, are there additional conditions under which this estimator converges to the average partial effect, $\mathbb{E}[B_1]$? Heckman and Vytlacil (1998) and Wooldridge (1997, 2003, 2008) show that there are indeed such conditions, namely, the assumption that the causal effect of Z on X is homogenous. While convenient, this type of homogeneity assumption is uncomfortably asymmetric. It enables the additional heterogeneity in equation (1) relative to equation (2) only by assuming away the same type of heterogeneity in the analogous relationship between Z and X.[5]

2.2 Identification and Estimation by Conditional Linear Regression

Given these negative results, it is worthwhile considering estimators other than `ivregress`. The `ivcrc` module provides such an estimator. This estimator is based on the following intuitive control function argument, which is developed more formally in Masten and Torgovitsky (2016).[6] Suppose that there exists an observable variable R such that $X \perp\!\!\!\perp (B_0, B_1)|R$. The variable R is a "control function" (or sometimes, and more loosely, a "control variable") because it controls for the endogeneity in X. That is, while X is endogenous in the sense of being unconditionally dependent with (B_0, B_1), it is exogenous after conditioning on the control function, R. In practice, R is constructed from the instrument; we explain the derivation and construction of R in more detail in Section 2.3.

Given the availability of a variable R with this property, it is straightforward to see that one could consistently estimate the vector $\beta(r) \equiv \mathbb{E}[B|R = r]$ where $B \equiv [B_0, B_1]'$ by a linear regression of Y on X *conditional* on $R = r$. Letting $W \equiv [1, X]'$ so that $Y = W'B$, one has

$$\mathbb{E}[WW'|R = r]^{-1}\,\mathbb{E}[WY|R = r] = \mathbb{E}[WW'|R = r]^{-1}\,\mathbb{E}[WW'B|R = r] = \beta(r), \qquad (4)$$

where the second equality uses the assumption that B is independent of X (and hence W), conditional on R. In order for this argument to work, it must be the case that $\mathbb{E}[WW'|R = r]$ is invertible, which is the usual condition of no perfect multicollinearity, but now conditional on $R = r$. Intuitively, there must still be some variation left in X after conditioning on $R = r$. Assuming that this is the case for all r in the support of R, one can average up the linear regression

[5] An influential literature started by Imbens and Angrist (1994), Angrist and Imbens (1995), Angrist, Imbens, and Rubin (1996), and Angrist et al. (2000) has provided conditions under which the IV estimator provided by `ivregress` can be interpreted as a local average treatment effect (LATE) or a weighted average of various LATEs. While related, these arguments are nonparametric, and in particular do not use the linearity in X of the CRC model.

[6] This paper builds on a large literature on control functions, including Heckman (1979), Heckman and Robb (1985), Smith and Blundell (1986), Blundell and Powell (2004), Florens, Heckman, Meghir, and Vytlacil (2008), and Imbens and Newey (2009), among many others.

estimands on the right-hand side of (4) to obtain $\mathbb{E}[B] \equiv \mathbb{E}[\beta(R)]$, and hence the average partial effect of X on Y, i.e. $\mathbb{E}[B_1]$.

This identification argument suggests an estimator given by an average of conditional ordinary least squares (OLS) estimators. The conditioning is incorporated by applying kernel weights to each observation, where the weights reflect the distance of R from r. More concretely, given a sample $\{Y_i, X_i, R_i\}_{i=1}^n$, a conditional regression estimator of Y on W near $R = r$ is given by

$$\widehat{\beta}(r) \equiv \left(\sum_{i=1}^{n} k_i^h(r) W_i W_i' \right)^{-1} \left(\sum_{i=1}^{n} k_i^h(r) W_i Y_i \right), \tag{5}$$

where $k_i^h(r) \equiv h^{-1} K((R_i - r)/h)$ with K a second-order kernel function and $h > 0$ a bandwidth parameter.

The conditional OLS estimator (5) displays the same type of bias-variance tradeoff that is familiar from nonparametric kernel regression. As $h \to \infty$, $k_i^h(r) \to K(0)$ for all i, so that $\widehat{\beta}(r)$ is just the estimator from a usual linear regression of Y on W. We expect this estimator to be biased for $\mathbb{E}[B]$ if X is endogenous. Given the control function assumption, this bias disappears as $h \to 0$, but at the cost of higher variance in using fewer effective observations in computing $\widehat{\beta}(r)$. Balancing these two concerns leads one to choose a value of h that leads $\widehat{\beta}(r)$ to use fewer than n effective observations, and as a consequence $\widehat{\beta}(r)$ will have a slower-than-parametric rate of convergence for $\beta(r)$.

As a parameter of interest, $\beta(r)$ has a clear interpretation as the average partial effect of X on Y, conditional on $R = r$. Variation in this parameter as a function of r indicates treatment effect heterogeneity. We can average $\beta(R)$ for R in some known set $\mathcal{R}$ to obtain the average partial effect for the subpopulation with $R \in \mathcal{R}$. A natural estimator of this average is given by

$$\widehat{\beta}_{\mathcal{R}} = \frac{\sum_{i=1}^{n} \widehat{\beta}(R_i) \mathbb{1}[R_i \in \mathcal{R}]}{\sum_{i=1}^{n} \mathbb{1}[R_i \in \mathcal{R}]}, \tag{6}$$

where $\mathbb{1}[\cdot]$ is the indicator function that is 1 if $\cdot$ is true and 0 otherwise. At least in principle, $\widehat{\beta}_{\mathcal{R}}$ can be estimated at the parametric $\sqrt{n}$ rate (see Masten and Torgovitsky, 2014, or, for a more general discussion, Newey, 1994). In practice, however, such behavior likely requires R to land in $\mathcal{R}$ with fairly high probability. If the local design matrix in (4) exists for (almost) every r in the support of R, then $\mathcal{R}$ can be taken to be the entire support of R, so that (6) becomes an estimator of the unconditional average of B.

A more general version of (1) is

$$Y = B_0 + \sum_{j=1}^{d_x} B_j X_j + \sum_{j=1}^{d_1} B_{d_x+j} Z_{1j} \equiv W'B, \tag{7}$$

where X is now a d_x–dimensional vector of potentially endogenous explanatory variables and $Z_1 \in \mathbb{R}^{d_1}$ is a vector of exogenous explanatory variables. For notation, we combine these variables and their coefficients together with the constant term as $W \equiv [1, X', Z_1']'$ and B. We rename the excluded exogenous variable as Z_2, and combine the exogenous variables (included and excluded instruments) together into a vector $Z = [Z_1', Z_2']'$. The required condition on the control function is now that $W \perp\!\!\!\perp B|R$, so that both X and Z_1 are exogenous after conditioning on R. Given this condition, the identification argument (4) and the estimators (5) and (6) follow exactly as before.

2.3 Estimation of the Control Function

We have shown how a control function, R, can be used to estimate interesting parameters in a CRC model, but we have not yet explained how one can find or construct such a control function. The most common approach is to assume that for each $j = 1, \ldots, d_x$, there exists a function h_j and unobservables $V \equiv [V_1, \ldots, V_{d_x}]' \in \mathbb{R}^{d_x}$ such that

$$X_j = h_j(Z, V_j) \quad \text{for each } j, \tag{8}$$

where $h_j(z, \cdot)$ is strictly increasing for each z. As shown by Imbens and Newey (2009) and Masten and Torgovitsky (2016), if $(B, V) \perp\!\!\!\perp Z$, then $R \equiv [R_1, \ldots, R_{d_x}]'$ is a valid control function, where $R_j \equiv F_{X_j|Z}(X_j|Z)$ and $F_{X_j|Z}(x_j|z) \equiv \mathbb{P}[X_j \leq x_j|Z = z]$ is the population conditional distribution function of X_j, given Z. The components R_j of this control function can be interpreted as providing the conditional rank (relative position) of X_j given Z. The `ivcrc` module is written primarily with this choice of control function in mind, although the user can provide a different choice if desired. In such cases, the estimator can be viewed as estimating the varying coefficient model discussed in the next section.

We refer to Masten and Torgovitsky (2016) for more theoretical details on the interpretation and restrictiveness of maintaining (8); see also Chernozhukov and Hansen (2005) and Torgovitsky (2015). Here we focus on the implications for implementing (4) and (5) with R as the resulting conditional ranks. The first implication is that it may be useful to make a distinction between different components of the endogenous variables, X. For example, if X_2 is just some deterministic transformation of X_1, say $X_2 = X_1^2$, then X_2 is also fully determined by R_1. As a result, there is no need to separately estimate and condition on R_2. In the terminology of Masten and Torgovitsky (2016), X_1 is a basic endogenous variable, and $X_2 = X_1^2$ is a derived endogenous variable.

Derived endogenous variables require special treatment, since they appear as part of the vector of explanatory variables W, but are not included as part of the conditioning variables Z in the definition of $R_j \equiv F_{X_j|Z}(X_j|Z)$. More formally, a component X_j of X is a derived endogenous variable if it can be written as $X_j = g_j(X_{-j}, Z)$ for some known function g_j. Interaction terms and other nonlinear functions form the primary examples of derived endogenous variables. The `ivcrc` module handles derived endogenous variables using the `dendog` option discussed in Section 4. The empirical illustration in Section 5 provides an example of its use.

A second issue raised by this choice of R is that it is not directly observed in the data. Instead, we need to estimate $R_{ji} = F_{X_j|Z}(X_{ji}|Z_i)$ in a first step for each basic endogenous variable X_j and each observation i. The `ivcrc` module approaches this problem by estimating conditional quantile functions and then inverting them using the pre-rearrangement operator studied by Chernozhukov, Fernandez-Val, and Galichon (2010). This operator translates an estimator of a conditional quantile function, say $\widehat{Q}_{X_j|Z}(\cdot|z)$, into an estimator of a conditional distribution function through the relationship

$$\widehat{F}_{X_j|Z}(x_j|z) = \int_0^1 \mathbb{1}\left[\widehat{Q}_{X_j|Z}(s|z) \leq x_j\right] ds. \tag{9}$$

For estimating $\widehat{Q}_{X_j|Z}(s|z)$, the `ivcrc` module uses linear quantile regression (see e.g. Koenker, 2005) as implemented by Stata's built-in `qreg` command. The generated regressors $\{\widehat{R}_{ji}\}_{i=1}^n$ are then constructed by substituting (X_{ji}, Z_i) into (9) for every i.

An expression for the asymptotic variance of $\widehat{\beta}_{\mathcal{R}}$ needs to account for the statistical error involved in this first step estimation of the control function R. Masten and Torgovitsky (2014) report this calculation, but the form of the asymptotic variance is complicated and does not facilitate direct estimation. Fortunately, $\widehat{\beta}_{\mathcal{R}}$ is a relatively well-behaved estimator, so the bootstrap should be valid for approximating standard errors and confidence intervals (see e.g. Chen, Linton, and van Keilegom, 2003). The `ivcrc` module uses Stata's built-in `bootstrap` routine for these purposes.

A third point that arises when using this choice of R is that (6) can be simplified when there is only one basic endogenous variable. This is because $R \equiv F_{X|Z}(X|Z)$ is uniformly distributed when X is continuous. As a result, the probability that R lands in any region $\mathcal{R}$ is known a priori and does not need to be estimated. The population average of $\beta(R)$, conditional on $R \in \mathcal{R}$ in this case reduces to

$$\beta_{\mathcal{R}} = \lambda(\mathcal{R})^{-1} \int_{\mathcal{R}} \beta(r)\, dr, \tag{10}$$

where $\lambda(\mathcal{R})$ is the Lebesgue measure of the set $\mathcal{R}$. When equation (10) holds, `ivcrc` estimates it by substituting the (known) value of $\lambda(\mathcal{R})$ and numerically approximating the integral $\int_{\mathcal{R}} \widehat{\beta}(r)\, dr$ that replaces $\beta(r)$ with $\widehat{\beta}(r)$.

A fourth point that is worth reemphasizing is that in order for (4) to exist, the design matrix $\mathbb{E}[WW'|R = r]$ must be invertible. That is, there must not be perfect multicollinearity among the regressors after conditioning on $R = r$. When using the conditional rank for R, conditioning on $R = r$ still leaves variation in the basic endogenous variables as long as the excluded instrument, Z_2, is appropriately dependent with X near its rth quantile. See Masten and Torgovitsky (2014, 2016) for a more detailed discussion of this point. A consequence for implementation is that it is necessary to exclude from $\mathcal{R}$ regions over which this instrument relevance condition fails.

3 Varying Coefficient Models

The CRC model can be viewed as a special case of a larger class of models called *varying coefficient models*. A simple example of this model is

$$Y = \beta_0(S) + \beta_1(S)X + U, \tag{11}$$

where Y is an observed outcome, S are observed covariates (sometimes called "effect modifiers"), X is our primary observed covariate of interest, and U is an unobserved variable. Both $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are unknown, nonparametrically specified functions. Conditional on S, this is a parametric model in X. But conditional on X, it is a nonparametric model in S. While it is unclear who first proposed such models (e.g., see O'Hagan and Kingman, 1978, for an early citation), their in-depth study began with Cleveland, Grosse, and Shyu (1991) and Hastie and Tibshirani (1993). Fan and Zhang (2008) and Park et al. (2015) provide recent reviews of this literature.

Given a sample $\{Y_i, X_i, S_i\}_{i=1}^n$, the local regression estimator (5) with $R_i = S_i$ is precisely the Nadaraya-Watson (local constant) varying coefficient estimator; e.g., equation (2.1) of Park et al. (2015). Cleveland et al. (1991) proposed a local linear estimator. Fan and Zhang (1999) study these and other alternative estimators in detail. The asymptotic theory in Masten and Torgovitsky (2014) extends that of the varying coefficient literature in two directions: (a) by allowing for S to be a generated regressor and (b) by considering the asymptotic distribution of average coefficients, such as $\mathbb{E}[\beta_1(S)]$. While the literature on varying coefficient models focuses on the functions $\beta_0(\cdot)$ and $\beta_1(\cdot)$ themselves, the econometric models we consider motivate interest in these average coefficients as well.

The `ivcrc` command can estimate varying coefficient models like (11) via the `varcoef` option. See section 4 for details. This estimator allows all components of S to enter all coefficients. Park et al. (2015) discuss estimators which allow one to impose the assumption that some components of S enter some coefficients, but not others.

We conclude this section by briefly showing how the linear CRC model can be seen as a varying coefficient model. For simplicity, we only consider the simple model (1). Write

$$\begin{aligned} Y &= B_0 + B_1 X \\ &= \mathbb{E}(B_0 \mid R) + \mathbb{E}(B_1 \mid R)X + [(B_0 - \mathbb{E}(B_0 \mid R)) + (B_1 - \mathbb{E}(B_1 \mid R))X] \\ &\equiv \beta_0(R) + \beta_1(R)X + U. \end{aligned}$$

By $X \perp\!\!\!\perp (B_0, B_1) \mid R$ and the definition of U, $\mathbb{E}(U \mid R, X) = 0$. Thus the linear CRC model is a varying coefficient model with effect modifier R.

4 The `ivcrc` Module

The `ivcrc` module is available on the Statistical Software Components (SSC) archive and can be installed directly in Stata with the command `ssc install ivcrc`. Alternatively, the latest version of the module can be downloaded from the GitHub repository https://github.com/a-torgovitsky/ivcrc. The code (`ivcrc.ado`) and the help file (`ivcrc.sthlp`) can be downloaded from the repository and placed in the personal ado directory, as described in the Stata FAQ: https://www.stata.com/support/faqs/programming/personal-ado-directory/.

The syntax for the `ivcrc` module is

```
ivcrc depvar [varlist_1] (varlist_edg = varlist_2) [if] [in] [, options]
```

In terms of the IV model discussed in Section 2, *depvar* is Y, $varlist_1$ consists of the components in Z_1, $varlist_{edg}$ are the basic endogenous variable components of X, and $varlist_2$ are the components in Z_2. The required components of the syntax are *depvar*, $varlist_2$, and $varlist_{edg}$, while the remaining terms in brackets are optional.

The module allows for the options shown in Table 1. The `dendog` option allows the user to specify a list of endogenous variables that should be treated as derived (rather than basic), with the implications for implementation discussed in Section 2. The `bootstrap` option controls the calculation of standard errors and confidence intervals. Note that `ivcrc` does not compute these by default, because the bootstrap procedure can be computationally intensive. The `kernel` and `bandwidth` options allow the user to change the kernel function K and bandwidth h used to compute the weights in (5). If the input for `bandwidth` is a list of numbers (separated by commas), then `ivcrc` will compute different estimates for each bandwidth. The computational efficiency of specifying several bandwidths at once is especially useful when calling `bootstrap` for standard errors and confidence intervals. The `ranks` option controls the degree of accuracy for approximating the integral in (9).

The `average` option determines the set $\mathcal{R}$ over which the local estimates $\widehat{\beta}(r)$ are averaged and controls how this averaging is implemented. For example, `average(.1(0).3)` sets $\mathcal{R} = [.1, .3]$ and uses the empirical mean to evaluate the integral in (10). The module interprets a grid step of `0` as a request for computing $\widehat{\beta}_{\mathcal{R}}$ using the sample averaging formula (6) that does not use knowledge of the distribution of R. Alternatively, specifying `average(.1(.01).3)` sets $\mathcal{R} = [.1, .3]$ and uses grid steps of .01 to numerically evaluate the integral. Multiple non-overlapping sets can be specified by adding commas. If the `report` suboption is given, then estimates on each set will be reported separately together with the overall estimate. For example, `average(.1(0).3, .5(0).8, report)` would report the estimate of $\beta_{\mathcal{R}}$ just discussed, along with another empirical average estimate for $\mathcal{R} = [.5, .8]$. The grid method supports the `report` suboption as well.

There are two situations in which the module will always use (6) instead of attempting to

Table 1: Options for `ivcrc`

Option	Description
`dendog(`*varlist*`)`	Specify derived endogenous variables.
`bootstrap()`	Bootstrap confidence intervals and standard errors; default setting is no standard errors. Specify typical bootstrap options in (), e.g. `reps(#)` or `cluster(`*varlist*`)`. Access additional bootstrap statistics via `estat bootstrap`.
`kernel(`*string*`)`	Choose alternative kernel functions; default is the Epanechnikov kernel. Other options: *uniform*, *triangle*, *biweight*, *triweight*, *cosine*, or *gaussian*.
`bandwidth(`*numlist*`)`	Bandwidth of kernel; default is 0.05. If multiple (comma separated) values are specified, estimates for each bandwidth are reported. Sub-option: together with `varcoef`, specify the bandwidth for a varying coefficients model.
`ranks(`*integer*`)`	Use $(\frac{1}{integer}, ..., 1 - \frac{1}{integer})$ evenly spaced quantiles for computing the conditional rank statistic; default is 50.
`average(`*numlist* `[, report])`	Options for numerical integration, with number list syntax: `lb(g)ub`. Specify `average(lb(0)ub)` to use the sample average method; default is `average(0(0)1)`. Specify non-zero values of `g` to use the grid method, e.g. `average(.01(.01).99)` to numerically integrate over the grid $(.01, .02, ..., .99)$. The space of integration may be comprised of non-overlapping ascending subsets by specifying comma separated lists. Sub-option: specifying `average(lb1(g1)ub1,..., lbN(gN)ubN, report)` returns estimates for each subset as well as estimates over their union. Sub-option: together with `varcoef`, specify the support for kernel weights in a varying coefficients model.
`generate(`*varname* `[, replace])`	Save the conditional rank estimates to *varname* in the working dataset; this option is ignored when bootstrapping.
`userank(`*varname*`)`	Use *varname* as the conditional rank statistic, bypassing rank estimation.
`savecoef(`*filename*`)`	Creates a comma delimited (csv) dataset of the local rank-specific coefficient estimates, saved to *filename*.
`varcoef(`*varlist*`)`	Estimate a varying coefficients model, in which coefficients are conditioned on covariates specified in *varlist* as an alternative to conditioning on the ranks of the basic endogenous variables $varlist_{edg}$. Options `average` and `bandwidth` are required with `varcoef`
`noconstant`	Suppress the constant term of the model.

numerically integrate (10). The first is when there is more than one basic endogenous variable, in which case R is a vector with a joint distribution that is not known a priori and (10) is not valid. If a user specifies a list of subsets `average(lb1(g1)ub1,..., lbN(gN)ubN)` when there are multiple basic endogenous variables, the module interprets each subset `lbn(gn)ubn` as belonging to the `n`th endogenous variable in order of appearance in *varlist*$_{edg}$. Due to the difficulty of specifying sets in higher dimensions, more general multidimensional subset estimates may be obtained either by permuting this syntax, or by storing the local estimates $\widehat{\beta}(r)$ using the `savecoef` option and subsequently computing any desired subset average. This is not essential to the method, but allowing for more general specifications would complicate the syntax significantly without providing much in the way of useful flexibility.

The second case in which `ivcrc` only uses the empirical average (6) is when the `varcoef` option is called. Passing `varcoef(`*varlist*`)` skips the estimation of $\widehat{R}_i$ and uses the variables in *varlist* in its place. Since the density of these variables is generally not known a priori, (10) may not be true, so (6) is used. The average in (6) can still be taken over some specified subset $\mathcal{R}$, and such a set is still specified using the `average(lb(0)ub)` syntax. Note that using both the ($varlist_{edg} = varlist_2$) syntax and passing `varcoef` as an option will generate an error.

5 Using `ivcrc` to Estimate the Returns to Schooling

In this section, we apply the `ivcrc` module to the problem discussed in the introduction of estimating the returns to schooling. Our discussion builds off of Card (1994, 2001) and Heckman and Vytlacil (1998), who note that a simple model of optimal schooling decisions (such as Becker, 1975) would generate a CRC model like (1) or (7). Our analysis uses the same data as Card (1993) and Kling (2001), which is available as part of Cameron and Trivedi's (2009) textbook on Stata for Microeconometrics. The data is an extract from the National Longitudinal Survey of Young Men (NLSYM) that consists of 3,010 men who were aged 24–34 in 1966. The extract contains variables from both 1966 and a follow-up survey in 1976. The data, as well as the code for the following analysis, is available at https://github.com/a-torgovitsky/ivcrc.

We begin by estimating a linear regression of log wages on schooling, potential work experience, and demographic control variables. This type of regression is often referred to as a Mincer (1958, 1974) equation; see Heckman, Lochner, and Todd (2006) for an in-depth discussion. The estimates indicate that an additional year of schooling is associated with approximately a 7.25 percent increase in 1976 wages:

```
. reg wage76 grade76 exp76 expsq76 `ControlVars', robust

Linear regression                                   Number of obs     =      3,010
                                                    F(27, 2982)       =      52.45
```

```
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.3040
                                                  Root MSE          =     .37191

------------------------------------------------------------------------------
             |               Robust
      wage76 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |   .0725423   .0038685    18.75   0.000     .0649572    .0801275
...
```

In this regression and throughout the subsequent analysis, we include a set of sociodemographic controls for race (`black`), parent's education (`daded`, `momed`, `famed1-8`), family structure at age 14 (`momdad14`, `sinmom14`), and geographic region (`smsa66`, `smsa76`, `reg1-reg8`).[7] While not essential to demonstrating the usage of the `ivcrc` module, the inclusion of these controls shows that the semiparametric estimator implemented by this model does not suffer from the curse of dimensionality. Also, note that potential work experience, `exp76` is defined as $\texttt{exp76} = \texttt{grade76} - \texttt{age76} - 6$, following the standard convention for Mincer equations.

As discussed in the introduction, education is a choice variable that is likely correlated with latent factors that affect wages, even after controlling for sociodemographic characteristics. Card (1993) used an indicator for living (at age 14) in a county with a four-year college as an instrument for education. Proximity to a four-year college is associated with about a third of a grade higher educational attainment:

```
. reg grade76 col4 `ControlVars', robust

Linear regression                                 Number of obs     =      3,010
                                                  F(25, 2984)       =      53.50
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.2937
                                                  Root MSE          =     2.2591

------------------------------------------------------------------------------
             |               Robust
     grade76 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        col4 |   .3669905   .1023706     3.58   0.000     .1662663    .5677147
```

[7] For readability, we collect these into a local variable `ControlVars` in the do file for this exercise. The local variable `TableOptions` contains a list of formatting and display options for `estout`.

...

In order for college proximity to be a valid instrument, it should, after accounting for control variables, have no direct effect on wages in 1976 and also be uncorrelated with other factors that are correlated with wages or schooling decisions. There are several reasons to be suspect of this requirement; see for example Kling (2001), or Mountjoy (2019) for a modern discussion with richer geographic data. Here, we simply compare estimators and take the validity of the college proximity instrument for granted.

The textbook linear IV estimator suggests that an additional year of schooling causes about a 13.33 percent increase in 1976 wages:

```
. ivregress 2sls wage76 (grade76 exp76 expsq76 = col4 age76 agesq76) ///
>     ‘ControlVars’, perfect

Instrumental variables (2SLS) regression          Number of obs   =      3,010
                                                  Wald chi2(27)   =    1007.25
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2030
                                                  Root MSE        =     .39614

------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |   .1333034   .0493359     2.70   0.007     .0366068    .2300001
...
```

This interpretation presumes that the causal effect of schooling on wages is constant. It yields the potentially puzzling conclusion that the raw association between education and wages actually substantially *understates* the causal effect of education on wages. As Card (2001) documents, this conclusion about the returns to schooling is actually fairly common across diverse studies that use a variety of IV strategies and data sources. One explanation proposed by Card (2001) is that this arises from a failure to account for heterogeneity in the causal effect of schooling on wages.

We can use the `ivcrc` module to assess this explanation. The syntax is similar to that for the IV estimator:

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) ‘ControlVars’, ///
>     dendog(exp76 expsq76)
(default settings do not compute standard errors, see bootstrap() option)
(estimating the conditional rank of grade76)
(estimating beta(r) at each r[i] rank in the sample)
```

```
IVCRC                                             Number of obs     =      3,010
--------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z      P>|z|     [95% Conf. Interval]
-------------+------------------------------------------------------------------
     grade76 |   .0807563          .        .        .             .           .
   ...
--------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .05
```

We treat potential experience, `exp76`, as a derived endogenous variable here because it is defined as a deterministic function of `grade76` and `age76`. Whereas the coefficient on `grade76` reported by the standard linear regression estimator implemented by `ivregress` will estimate a difficult-to-interpret quantity like (3), the coefficient on `grade76` produces an estimator of the average causal effect of a one year increase in `grade76`. The causal effect estimated here of 8.08 percent is significantly lower than the linear IV estimate of 13.33 percent. This supports Card's (2001) reasoning if, as he argues, the usual linear IV estimator places more weight on individuals with higher returns to schooling. The `ivcrc` estimate is also similar to the linear regression coefficient 0.0725.

We now demonstrate some of the options for `ivcrc` by evaluating the statistical significance and robustness of this estimate. First, we compute standard errors, which tends to be time-consuming due to the necessity of using the bootstrap. The syntax and results are:

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) `ControlVars', ///
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020))
(running _ivcrc_estimator on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                             Number of obs     =      3,010
                                                  Replications      =        100
--------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z      P>|z|     [95% Conf. Interval]
-------------+------------------------------------------------------------------
     grade76 |   .0807563   .0188294     4.29    0.000      .0438514    .1176612
...
--------------------------------------------------------------------------------
```

```
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .05
```

The confidence interval here is a bit wider than for the linear regression estimator, although substantially narrower than for the usual linear IV estimator. The textbook IV estimator and the `ivcrc` estimates are constructed under non-nested assumptions, so this by itself is not unexpected. However, since the bandwidth controls a bias-variance trade-off in the `ivcrc` estimator, it does suggest that we may want to explore decreasing the bandwidth in order to guard against potential bias due to oversmoothing. So next we evaluate the point estimates at several bandwidths:

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) ‘ControlVars’, ///
>     dendog(exp76 expsq76) bandwidth(.025, .05, .075)
(default settings do not compute standard errors, see bootstrap() option)
(estimating the conditional rank of grade76)
(estimating beta(r) at each r[i] rank in the sample)

IVCRC                                           Number of obs     =      3,010
------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |   .0869784          .        .        .             .         .
  ...
------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025
------------------------------------------------------------------------------
     grade76 |   .0807563          .        .        .             .         .
...
------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .05
------------------------------------------------------------------------------
     grade76 |   .0779116          .        .        .             .         .
...
------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .075
```

The estimate is relatively stable over different bandwidths, but does decline somewhat as the local estimates $\widehat{\beta}(r)$ are computed using larger neighborhoods of r. Obtaining standard errors and confidence intervals using the smallest bandwidth in this list,

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) ‘ControlVars’, ///
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020)) bandwidth(.025)
```

```
(running _ivcrc_estimator on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                           Number of obs     =      3,010
                                                Replications      =        100
------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |   .0869784    .029612     2.94   0.003      .0289401   .1450168
...
------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025
```

we find a larger standard error and a wider confidence interval, as anticipated. Though more comparable to the standard error and confidence interval from the linear IV model, the `ivcrc` standard error remains roughly 1.5 times smaller at this smaller bandwidth.

The number of quantiles used to approximate the integral in (9) and the functional form of the kernel weights K could in principle also impact the `ivcrc` estimates. Quadrupling the number of quantiles from its default of 50 while carrying forward the smaller bandwidth from above,

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) `ControlVars', ///
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020)) bandwidth(.025) ///
>     ranks(200)
(running _ivcrc_estimator on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                           Number of obs     =      3,010
                                                Replications      =        100
------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
     grade76 |    .078291  .0320301     2.44    0.015      .0155132   .1410688
   ...
   ------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025
```

we find that the results are not very sensitive to how finely the integral in (8) is approximated. Swapping a uniform kernel for the (default) Epanechnikov kernel, while carrying forward a smaller bandwidth and more accurate rank estimation from above,

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) ‘ControlVars’, ///
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020)) bandwidth(.025) ///
>     ranks(200) kernel(uniform)
(running _ivcrc_estimator on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                           Number of obs     =      3,010
                                                Replications      =        100
------------------------------------------------------------------------------
      wage76 |     Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |   .0780691  .0302534     2.58    0.010      .0187736   .1373647
   ...
   ------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025
```

we find that the results are also not sensitive to the functional form of the kernel, in concordance with the usual folklore for nonparametric kernel regression.

One interesting way to explore both the robustness and potential explanations for our finding is to change the set $\mathcal{R}$ over which the average is being taken. By default, `ivcrc` averages over all estimated conditional ranks ($\widehat{R}_i$) directly as in (6). Alternatively, if we are concerned about results being driven by outliers in the education distribution, we can specify $\mathcal{R}$ to be $[.05, .95]$. Trimming the education distribution in this way, while maintaining the smaller bandwidth and more accurate rank estimation from above,

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) ‘ControlVars’, ///
```

```
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020)) bandwidth(.025) ///
>     ranks(200) average(.05(0).95)
(running _ivcrc_estimator on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                           Number of obs     =      3,010
                                                Replications      =        100
------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |   .0735619   .0345887     2.13   0.033      .0057692   .1413546
...
------------------------------------------------------------------------------
Note: Average coefficients over R = [.05,.95] rank subset; Bandwidth = .025
```

we obtain slightly lower estimated returns to education and a slightly larger standard error, but overall similar results to the estimates which used the full observed distribution of education. When there is a single basic endogenous variable, as in the present application, another check on the estimates is to use numerical integration based on (10). Specifying an equally spaced grid with steps of .01 over the outlier-trimmed region $[.05, .95]$ from above,

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) ‘ControlVars’, ///
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020)) bandwidth(.025) ///
>     ranks(200) average(.05(.01).95)
(running _ivcrc_estimator on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                           Number of obs     =      3,010
                                                Replications      =        100
------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
     grade76 |   .0730978  .0341266     2.14    0.032      .0062108   .1399847
 ...
------------------------------------------------------------------------------
Note: Average coefficients over R = [.05,.95] rank subset; Bandwidth = .025
```

we obtain estimates that are nearly identical to those obtained using the default sample average method, (6).

We can also consider smaller sets of R to explore heterogeneity in the return to schooling. For example, an estimate for individuals in the lower half of the education distribution is:

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) `ControlVars', ///
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020)) bandwidth(.025) ///
>     ranks(200) average(0(0).5)
(running _ivcrc_estimator on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                           Number of obs     =      3,010
                                                Replications      =        100
------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |   .1038839   .053131     1.96    0.051      -.000251   .2080187
  ...
------------------------------------------------------------------------------
Note: Average coefficients over R = [0,.5] rank subset; Bandwidth = .025
```

This suggests that individuals with lower schooling have higher returns to schooling. Specifying a set for each quartile of the education distribution reveals a pattern that supports this explanation, while indicating potentially more nuance,

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) `ControlVars', ///
>     dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020)) bandwidth(.025) ///
>     ranks(200) average(0(0).25, .2501(0).5, .5001(0).75, .7501(0)1, report)
(running _ivcrc_estimator on estimation sample)
```

```
Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

IVCRC                                           Number of obs     =      3,010
                                                Replications      =        100
------------------------------------------------------------------------------
      wage76 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     grade76 |    .078291   .0320301     2.44   0.015      .0155132   .1410688
  ...
------------------------------------------------------------------------------
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025
------------------------------------------------------------------------------
     grade76 |   .0651947   .0616853     1.06   0.291     -.0557062   .1860957
 ...
------------------------------------------------------------------------------
Note: Average coefficients over R = [0,.25] rank subset; Bandwidth = .025
------------------------------------------------------------------------------
     grade76 |   .1418537   .1006071     1.41   0.159     -.0553327     .33904
 ...
------------------------------------------------------------------------------
Note: Average coefficients over R = [.2501,.5] rank subset; Bandwidth = .025
------------------------------------------------------------------------------
     grade76 |   .0264401    .066345     0.40   0.690     -.1035938   .1564739
...
------------------------------------------------------------------------------
Note: Average coefficients over R = [.5001,.75] rank subset; Bandwidth = .025
------------------------------------------------------------------------------
     grade76 |   .0787306   .0512649     1.54   0.125     -.0217468   .1792081
   ...
------------------------------------------------------------------------------
Note: Average coefficients over R = [.7501,1] rank subset; Bandwidth = .025
```

The module first displays the estimate taken over the union of the given sets, in this case the overall sample average. Then `ivcrc` reports the estimates over each subset. We find that the estimates of the returns to schooling vary across the education distribution, with second quartile exhibiting large returns that are comparable to the linear IV estimate. However, the estimates are less precisely estimated than the average return using the entire sample, which reflects the fact that each subset only uses approximately one fourth of the number of effective observations.

References

Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263. 3

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *The Review of Economic Studies*, 67, 499–527. 3, 5

Angrist, J. D. and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442. 3, 5

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455. 5

Angrist, J. D. and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106, 979–1014. 3

Becker, G. (1975): "Human Capital and The Personal Distribution of Income: An Analytical Approach," in *Human Capital*, New York: Columbia University Press, second ed. 12

Becker, G. S. and B. R. Chiswick (1966): "Education and the Distribution of Earnings," *The American Economic Review*, 56, 358–369. 4

Blundell, R. W. and J. L. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *The Review of Economic Studies*, 71, 655–679. 5

Cameron, A. C. and P. K. Trivedi (2009): *Microeconometrics Using Stata*, Stata Press. 12

Cameron, S. V. and J. J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106, 262–333. 3

Card, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," *NBER Working Paper No. 4483*. 3, 12, 13

——— (1994): "Earnings, Schooling, and Ability Revisited," *NBER Working Paper No. 4832*. 12

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160. 12, 14, 15

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101, 2754–81. 3

Chay, K. Y. and M. Greenstone (2005): "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 113, 376–424. 3

Chen, X., O. Linton, and I. van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591–1608. 8

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125. 8

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. 7

Cleveland, W., E. Grosse, and W. Shyu (1991): "Local Regression Models," in *Statistical Models in S*, ed. by J. Chambers and T. Hastie, Chapman & Hall, London, chap. 8, 309–376. 9

Conway, K. S. and T. J. Kniesner (1991): "The Important Econometric Features of a Linear Regression Model with Cross-Correlated Random Coefficients," *Economics Letters*, 35, 143–147. 4

Fan, J. and W. Zhang (1999): "Statistical Estimation in Varying Coefficient Models," *Annals of Statistics*, 1491–1518. 9

——— (2008): "Statistical Methods with Varying Coefficient Models," *Statistics and its Interface*, 1, 179. 3, 9

Florens, J. P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008): "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 1191–1206. 5

Garen, J. (1984): "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 52, 1199. 4

Gollin, D. and C. Udry (2020): "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture," *Journal of Political Economy (forthcoming)*. 3

Hastie, T. and R. Tibshirani (1993): "Varying-Coefficient Models," *Journal of the Royal Statistical Society. Series B (Methodological)*, 757–796. 9

Heckman, J. and E. Vytlacil (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling," *The Journal of Human Resources*, 33, 974–987. 3, 4, 5, 12

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161. 5

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. Hanushek and F. Welch, Elsevier, vol. 1, chap. 7, 307–458. 12

Heckman, J. J. and R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions: An Overview," *Journal of Econometrics*, 30, 239–267. 5

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. 3

Hurwicz, L. (1950): "Systems with Nonadditive Disturbances," in *Statistical Inference in Dynamic Economic Models*, ed. by T. Koopmans, no. 10 in Cowles Commission Monographs, 410–418. 4

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. 3, 5

Imbens, G. W. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512. 5, 7

Kling, J. R. (2001): "Interpreting Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business & Economic Statistics*, 19, 358–364. 12, 14

Koenker, R. (2005): *Quantile Regression*, Cambridge University Press. 8

Masten, M. A. and A. Torgovitsky (2014): "Instrumental Variables Estimation of a Generalized Correlated Random Coefficients Model," *cemmap working paper 02/14*. 2, 3, 6, 8, 9

——— (2016): "Identification of Instrumental Variable Correlated Random Coefficients Models," *The Review of Economics and Statistics*, 98, 1001–1005. 2, 3, 5, 7, 8

Mincer, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302. 12

——— (1974): *Schooling, Experience, and Earnings*, NBER Press. 12

Mountjoy, J. (2019): "Community Colleges and Upward Mobility," *Working paper*. 3, 14

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382. 6

O'Hagan, A. and J. Kingman (1978): "Curve Fitting and Optimal Design for Prediction," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–42. 9

Oreopoulos, P. (2006): "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter," *The American Economic Review*, 96, 152–175. 3

Park, B. U., E. Mammen, Y. K. Lee, and E. R. Lee (2015): "Varying Coefficient Regression Models: A Review and New Developments," *International Statistical Review*, 83, 36–64. 3, 9

Rios-Avila, F. (2019): "Varying Coefficient Models in Stata," *Stata Conference Chicago 2019, Poster Presentation*. 3

Rubin, H. (1950): "Note on Random Coefficients," in *Statistical Inference in Dynamic Economic Models*, ed. by T. Koopmans, no. 10 in Cowles Commission Monographs, 419–421. 4

Smith, R. J. and R. W. Blundell (1986): "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54, 679–685. 5

Torgovitsky, A. (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185–1197. 7

Wald, A. (1947): "A Note on Regression Analysis," *The Annals of Mathematical Statistics*, 18, 586–589. 4

Wooldridge, J. M. (1997): "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model," *Economics Letters*, 56, 129–133. 4, 5

——— (2003): "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model," *Economics Letters*, 79, 185–191. 5

——— (2008): "Instrumental Variables Estimation of the Average Treatment Effect in Correlated Random Coefficient Models," in *Modeling and Evaluating Treatment Effects in Econometrics*, ed. by D. Millimet, J. Smith, and E. Vytlacil, Elsevier. 5