ivcrc: An instrumental-variables estimator for the correlated random-coefficients model

David Benson
Division of Research and Statistics
Federal Reserve Board of Governors
Washington, DC
david.a.benson@frb.gov

Matthew A. Masten
Department of Economics
Duke University
Durham, NC
matt.masten@duke.edu

Alexander Torgovitsky
Kenneth C. Griffin Department of Economics
University of Chicago
Chicago, IL
torgovitsky@uchicago.edu

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

Keywords: st0680, ivcrc, ivregress, instrumental variables, correlated random coefficients, heterogeneous treatment effects, varying-coefficient models, returns to schooling

1 Introduction

In this article, we describe the `ivcrc` command for Stata, which implements a linear instrumental-variables (IV) estimator for the correlated random-coefficients (CRC) model. The CRC model relaxes the constant-effects assumption of the standard linear IV model by allowing for random coefficients, which capture unobserved heterogeneity in the causal effect of the endogenous variables, x, on the outcome variable, y. Heckman and Vytlacil (1998) and Wooldridge (1997, 2003, 2008) showed that if there is no unobserved heterogeneity in the way the instrument, z, affects the endogenous variables, x, then the usual linear IV estimator (for example, `ivregress`) will estimate the mean of the random coefficients. However, this assumption is uncomfortably asymmetric: treatment effects can be unobservably heterogeneous, but instrument effects cannot.

Masten and Torgovitsky (2016) address this drawback by showing how to allow for heterogeneity in the relationship between z and x. The identification arguments developed there allow for the instrument to be binary, discrete, or continuous and depend only on familiar, low-level assumptions about exclusion, exogeneity, and instrument relevance. The arguments also lead to a semiparametric estimator that does not suffer from the curse of dimensionality. In section 2, we briefly review the identification results and estimation approach developed in Masten and Torgovitsky (2014, 2016). The `ivcrc` command implements their proposed estimator.

The structure of the estimator turns out to be quite similar to a common estimator for the varying-coefficient model (for example, Fan and Zhang 2008; Park et al. 2015). We have written `ivcrc` to include a standard estimator for this model as a special case. We briefly describe varying-coefficient models in section 3.

In section 4, we discuss syntax and options for the `ivcrc` command. In section 5, we illustrate the command by estimating the return to schooling with a widely used extract from the National Longitudinal Survey of Young Men. We provide some brief concluding remarks in section 6. For further examples, see Gollin and Udry (2021), who used the `ivcrc` command to estimate agricultural production functions; Morales-Mosquera (2020), who used it to estimate the effects of police infrastructure in Colombia; and Masten and Torgovitsky (2014), who used the estimator to revisit Chay and Greenstone's (2005) analysis of the effect of air pollution on housing prices.

2   The correlated random-coefficients model

2.1   Model and motivation

The simplest form of the model fit 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 because of 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.[1]

---

1. 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.

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 0. 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, because 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), β, is deterministic, whereas the coefficient on x in (1), 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 Heckman, Urzua, and Vytlacil (2006). 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 $\mathrm{Cov}(u, z) = 0$ and $\mathrm{Cov}(x, z) \neq 0$. The corresponding IV estimator can be implemented in Stata with the `ivregress` command. However, if the data are in fact generated by (1), then this estimator converges to

$$\frac{\mathrm{Cov}(y, z)}{\mathrm{Cov}(x, z)} = \mathbb{E}\left(b_1 \times \frac{x\{z - \mathbb{E}(z)\}}{\mathbb{E}\left[x\{z - \mathbb{E}(z)\}\right]}\right) \tag{3}$$

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 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 described in Heckman, Urzua, and Vytlacil (2006).

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 homogeneous. While convenient, this type of homogeneity assumption is uncomfortably asymmetric. It enables the additional heterogeneity in (1) relative to (2) only by assuming away the same type of heterogeneity in the analogous relationship between z and x.[2]

---

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

2.2 Identification and estimation by conditional linear regression

Given these negative results, it is worthwhile to consider estimators other than `ivregress`. The `ivcrc` command provides such an estimator. This estimator is based on the following control function argument, which is developed formally in Masten and Torgovitsky (2016).[3]

Suppose that there exists an observable variable r such that $x \perp\!\!\!\perp (b_0, b_1) \mid r$. 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 $\boldsymbol{\beta}(\widetilde{r}) \equiv \mathbb{E}(\boldsymbol{b} \mid r = \widetilde{r})$, where $\boldsymbol{b} \equiv [b_0, b_1]'$ by a linear regression of y on x conditional on $r = \widetilde{r}$. Letting $\boldsymbol{w} \equiv [1, x]'$ so that $y = \boldsymbol{w}'\boldsymbol{b}$, one has

$$\mathbb{E}\left(\boldsymbol{ww}' \mid r = \widetilde{r}\right)^{-1} \mathbb{E}\left(\boldsymbol{wy} \mid r = \widetilde{r}\right) = \mathbb{E}\left(\boldsymbol{ww}' \mid r = \widetilde{r}\right)^{-1} \mathbb{E}\left(\boldsymbol{ww}'\boldsymbol{b} \mid r = \widetilde{r}\right)$$
$$= \boldsymbol{\beta}\left(\widetilde{r}\right) \tag{4}$$

where the second equality uses the assumption that $\boldsymbol{b}$ is independent of x (and hence $\boldsymbol{w}$), conditional on r. For this argument to work, $\mathbb{E}(\boldsymbol{ww}' \mid r = \widetilde{r})$ must be invertible, which is the usual condition of no perfect multicollinearity, but now conditional on $r = \widetilde{r}$. Intuitively, some variation must still be left in x after conditioning on $r = \widetilde{r}$. Assuming that this is the case for all $\widetilde{r}$ in the support of r, one can average up the linear regression estimands on the right-hand side of (4) to obtain $\mathbb{E}(\boldsymbol{b}) \equiv \mathbb{E}\{\boldsymbol{\beta}(r)\}$, and hence the average partial effect of x on y, $\mathbb{E}(b_1)$.

This identification argument suggests an estimator given by an average of conditional ordinary least-squares estimators. The conditioning is incorporated by applying kernel weights to each observation, where the weights reflect the distance of r from $\widetilde{r}$. More concretely, given a sample $\{y_i, x_i, r_i\}_{i=1}^{n}$, a conditional regression estimator of y on $\boldsymbol{w}$ near $r = \widetilde{r}$ is given by

$$\widehat{\boldsymbol{\beta}}\left(\widetilde{r}\right) \equiv \left\{\sum_{i=1}^{n} k_i^h\left(\widetilde{r}\right) \boldsymbol{w}_i \boldsymbol{w}_i'\right\}^{-1} \left\{\sum_{i=1}^{n} k_i^h\left(\widetilde{r}\right) \boldsymbol{w}_i y_i\right\} \tag{5}$$

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

The conditional ordinary least-squares estimator (5) displays the same type of bias-variance tradeoff that is familiar from nonparametric kernel regression. As $h \to \infty$,

---

3. This article builds on a large literature on control functions, including Heckman (1979), Heckman and Robb (1985), Smith and Blundell (1986), Blundell and Powell (2004), Florens et al. (2008), and Imbens and Newey (2009), among many others.

$k_i^h(r) \to k(0)$ for all i so that $\widehat{\boldsymbol{\beta}}(\widetilde{r})$ is just the estimator from a usual linear regression of y on $\boldsymbol{w}$. We expect this estimator to be biased for $\mathbb{E}(\boldsymbol{b})$ 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 $\widehat{\boldsymbol{\beta}}(\widetilde{r})$. Balancing these two concerns entails using fewer than n effective observations, and consequently $\widehat{\boldsymbol{\beta}}(\widetilde{r})$ will have a slower-than-parametric rate of convergence for $\boldsymbol{\beta}(\widetilde{r})$.

As a parameter of interest, $\boldsymbol{\beta}(\widetilde{r})$ has a clear interpretation as the average partial effect of x on y, conditional on $r = \widetilde{r}$. Variation in this parameter as a function of $\widetilde{r}$ indicates treatment-effect heterogeneity. We can average $\boldsymbol{\beta}(r)$ 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

$$\widehat{\boldsymbol{\beta}}_{\mathcal{R}} = \frac{\sum_{i=1}^{n} \widehat{\boldsymbol{\beta}}(r_i) \mathbb{1}(r_i \in \mathcal{R})}{\sum_{i=1}^{n} \mathbb{1}(r_i \in \mathcal{R})} \tag{6}$$

where $\mathbb{1}(\cdot)$ is the indicator function that is 1 if $\cdot$ is true and 0 otherwise. At least in principle, $\widehat{\boldsymbol{\beta}}_{\mathcal{R}}$ can be estimated at the parametric $\sqrt{n}$ rate (see Masten and Torgovitsky [2014], or, for a more general discussion, Newey [1994]). If the local Gram matrix is invertible for (almost) every $\widetilde{r}$ 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 $\boldsymbol{b}$.

A more general version of (1) is

$$y = b_0 + \sum_{j=1}^{d_x} b_j x_j + \sum_{j=1}^{d_1} b_{d_x+j} z_{1j} \equiv \boldsymbol{w}'\boldsymbol{b} \tag{7}$$

where $\boldsymbol{x}$ is now a d_x-dimensional vector of potentially endogenous explanatory variables and $\boldsymbol{z}_1 \in \mathbb{R}^{d_1}$ is a vector of exogenous explanatory variables. For notation, we combine these variables and their coefficients together with the constant term as $\boldsymbol{w} \equiv [1, \boldsymbol{x}', \boldsymbol{z}_1']'$ and $\boldsymbol{b}$. We rename the excluded exogenous variable as $\boldsymbol{z}_2$ and combine the exogenous variables (included and excluded instruments) together into a vector $\boldsymbol{z} = [\boldsymbol{z}_1', \boldsymbol{z}_2']'$. The required condition on the control function is now that $\boldsymbol{w} \perp\!\!\!\perp \boldsymbol{b} \mid \boldsymbol{r}$ so that both $\boldsymbol{x}$ and $\boldsymbol{z}_1$ are exogenous after conditioning on $\boldsymbol{r}$. 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 $\boldsymbol{v} \equiv [v_1, \ldots, v_{d_x}]' \in \mathbb{R}^{d_x}$ such that

$$x_j = h_j(\boldsymbol{z}, \boldsymbol{v}_j) \quad \text{for each } j \tag{8}$$

where $h_j(\boldsymbol{z}, \cdot)$ is strictly increasing for each $\boldsymbol{z}$. As shown by Imbens and Newey (2009) and Masten and Torgovitsky (2016), if $(\boldsymbol{b}, \boldsymbol{v}) \perp\!\!\!\perp \boldsymbol{z}$, then $\boldsymbol{r} \equiv [r_1, \ldots, r_{d_x}]'$ is a valid

control function, where $r_j \equiv F_{x_j|z}(x_j \mid \boldsymbol{z})$ and $F_{x_j|\boldsymbol{z}}(\widetilde{x}_j \mid \widetilde{\boldsymbol{z}}) \equiv \mathbb{P}[x_j \leq \widetilde{x}_j \mid \boldsymbol{z} = \widetilde{\boldsymbol{z}}]$ is the population conditional distribution function of x_j, given $\boldsymbol{z}$. The components r_j of this control function can be interpreted as providing the conditional rank (relative position) of x_j given $\boldsymbol{z}$. The `ivcrc` command is written primarily with this choice of control function in mind, although the user can provide a different choice if desired. In that case, the estimator can be viewed as fitting 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 $\boldsymbol{r}$ 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, $\boldsymbol{x}$. For example, if x_2 is a 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, because they appear as part of the vector of explanatory variables $\boldsymbol{w}$ but are not included as part of the conditioning variables $\boldsymbol{z}$ in the definition of $r_j \equiv F_{x_j|\boldsymbol{z}}(x_j \mid \boldsymbol{z})$. More formally, a component x_j of $\boldsymbol{x}$ is a derived endogenous variable if it can be written as $x_j = g_j(\boldsymbol{x}_{-j}, \boldsymbol{z})$ for some known function g_j. Interaction terms and other nonlinear functions form the primary examples of derived endogenous variables. The `ivcrc` command 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 $\boldsymbol{r}$ is that it is not directly observed in the data. Instead, we need to estimate $r_{ji} = F_{x_j|\boldsymbol{z}}(x_{ji} \mid \boldsymbol{z}_i)$ in a first step for each basic endogenous variable x_j and each observation i. The `ivcrc` command approaches this problem by estimating conditional quantile functions and then inverting them using the prearrangement operator studied by Chernozhukov, Fernández-Val, and Galichon (2010). This operator translates an estimator of a conditional quantile function, say, $\widehat{Q}_{x_j|\boldsymbol{z}}(\cdot \mid \widetilde{\boldsymbol{z}})$, into an estimator of a conditional distribution function through the relationship

$$\widehat{F}_{x_j|\boldsymbol{z}}(\widetilde{x}_j \mid \widetilde{\boldsymbol{z}}) = \int_0^1 \mathbb{1}\left\{\widehat{Q}_{x_j|\boldsymbol{z}}(s \mid \widetilde{\boldsymbol{z}}) \leq \widetilde{x}_j\right\} ds \tag{9}$$

For estimating $\widehat{Q}_{x_j|\boldsymbol{z}}(s \mid \widetilde{\boldsymbol{z}})$, the `ivcrc` command uses linear quantile regression (see, for example, 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 $(x_{ji}, \boldsymbol{z}_i)$ into (9) for every i.

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 $r = r \equiv F_{x|z}(x \mid z)$ is uniformly distributed when $x = x$ is continuous and scalar. As a result, the probability that $r = r$ lands in any region $\mathcal{R}$ is known a priori and does not need to be estimated. In this case, the population average of $\boldsymbol{\beta}(r)$ conditional on $r \in \mathcal{R}$ reduces to

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

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

A fourth point that is worth reemphasizing is that for (4) to exist, the Gram matrix $\mathbb{E}(\boldsymbol{w}\boldsymbol{w}' \mid r = \widetilde{r})$ must be invertible. That is, there must not be perfect multicollinearity among the regressors after conditioning on $r = \widetilde{r}$. When using the conditional rank for r, conditioning on $r = \widetilde{r}$ still leaves variation in the basic endogenous variables as long as the excluded instrument, $\boldsymbol{z}_2$, 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.

2.4   Bandwidth selection

The `ivcrc` command implements automated bandwidth selection for h based on a rule-of-thumb (ROT) method proposed by Fan and Gijbels (1996). We first estimate $\mathbb{E}(y \mid r = \widetilde{r}, \boldsymbol{w} = \widetilde{\boldsymbol{w}})$ as a fourth-order polynomial in $\widetilde{r}$, linearly interacted with $\widetilde{\boldsymbol{w}}$. We use this regression to produce estimates of the second derivative of $\mathbb{E}(y \mid r = \widetilde{r}, \boldsymbol{w} = \widetilde{\boldsymbol{w}})$ with respect to $\widetilde{r}$, which we denote by $\widetilde{\boldsymbol{w}}'\widehat{\boldsymbol{\delta}}(\widetilde{r})$. The ROT bandwidth is then defined as

$$h_{\text{ROT}} \equiv 0.58 \left\{ \frac{\widehat{\sigma}^2}{\sum_{i=1}^{n} \frac{1}{2} \boldsymbol{w}_i' \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{r}}_i)} \right\}^{1/5}$$

where $\widehat{\sigma}^2$ is the homoskedastic error variance from the polynomial regression and the constant 0.58 is derived from the roughness for single-peaked kernels. See the discussion in Hansen (2022, sec. 19.9) for further details.

2.5   Statistical inference

The asymptotic variance of $\widehat{\boldsymbol{\beta}}_{\mathcal{R}}$ needs to account for the statistical error involved in 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, $\widehat{\boldsymbol{\beta}}_{\mathcal{R}}$ is a relatively well-behaved estimator, so the bootstrap should be valid for approximating standard errors and confidence intervals (see, for example, Chen, Linton, and Van Keilegom [2003]). The `ivcrc` command uses Stata's built-in `bootstrap` routine for these purposes.

We evaluate the validity of the bootstrap using a small Monte Carlo simulation. The model is (1) with a single endogenous variable, x. The relationship between x, z, and v in (8) is specified as

$$x = 0.3z + 0.4zv + v$$

where z is binary $\{0, 1\}$ with equal probability and v is normally distributed with mean 0.1 and variance 0.2. The unobservables are related by

$$b_0 = 0.3v + \epsilon_0 \quad \text{and} \quad b_1 = 0.7v + \epsilon_1$$

where ϵ_0 and ϵ_1 are normally distributed, independent of v, with means 0.2 and 0.45 and variances 0.2 and 1. The parameter values imply $\mathbb{E}(b_0) = 0.23$ and $\mathbb{E}(b_1) = 0.52$.

We implement the bootstrap on each of 1,000 Monte Carlo replications with 250 bootstraps per replication while using the ROT bandwidth. We first consider a nominal level 5% test that uses bootstrap standard errors together with a normal approximation. We reject the true null hypotheses $\mathbb{E}(b_0) = 0.23$ and $\mathbb{E}(b_1) = 0.52$ in 4.9% and 6.2% of replications, respectively. Then we consider the coverage rate for a nominal 95% bootstrap percentile confidence interval without exploiting normality. The confidence intervals covered the true values of $\mathbb{E}(b_0)$ and $\mathbb{E}(b_1)$ in 94.5% and 94% of replications, respectively. We interpret these findings as providing evidence that the bootstrap provides inference that is close to size-correct.

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

$$y = \beta_0(\boldsymbol{s}) + \beta_1(\boldsymbol{s})x + u \tag{11}$$

where y is an observed outcome, $\boldsymbol{s}$ 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 $\boldsymbol{s}$, this is a parametric model in x. But conditional on x, it is a nonparametric model in $\boldsymbol{s}$. While it is unclear who first proposed such models (for example, see O'Hagan [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 $\{y_i, x_i, \boldsymbol{s}_i\}_{i=1}^n$, the local regression estimator (5) with $\boldsymbol{r}_i = \boldsymbol{s}_i$ is precisely the Nadaraya–Watson (local constant) varying-coefficient estimator, for example, (2.1) of Park et al. (2015). Cleveland, Grosse, and Shyu (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 $\boldsymbol{s}$ to be a generated regressor and b) by considering the asymptotic distribution of average coefficients, such as $\mathbb{E}\{\beta_1(\boldsymbol{s})\}$. 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 fit 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 that 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 command

The syntax for the `ivcrc` command is

`ivcrc` *depvar* $\big[$ *varlist*$_1$ $\big]$ (*varlist*$_{\mathrm{edg}}$ = *varlist*$_2$) $\big[$ *if* $\big]$ $\big[$ *in* $\big]$ $\big[$, *options* $\big]$

In terms of the IV model discussed in section 2, *depvar* is y, *varlist*$_1$ consists of the components in z_1, *varlist*$_{\mathrm{edg}}$ 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*$_{\mathrm{edg}}$, while the remaining terms in brackets are optional.

The command 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).

Table 1. Options for ivcrc

options	Description
dendog(*varlist*)	list of derived endogenous variables, known functions of *varlist*$_{\text{edg}}$
<u>bootstrap</u>(*options*)	bootstrap confidence intervals and standard errors; default setting is no standard errors; allows passing bootstrap command *options*, for example, reps() or cluster(); access additional postestimation statistics via estat bootstrap
<u>kernel</u>(*string*)	choose alternative kernel functions; default is the Epanechnikov kernel; *string* may be uniform, triangle, biweight, triweight, cosine, or gaussian
<u>bandwidth</u>(*numlist*)	bandwidth of kernel; default is the ROT; if multiple (comma-separated) values are specified, estimates for each bandwidth are reported; suboption: together with varcoef(), specify the bandwidth for a varying-coefficient model
<u>ranks</u>(*integer*)	use $[(1/integer), \ldots, 1 - (1/integer)]$ evenly spaced quantiles for computing the conditional rank statistic; default is ranks(50)
<u>av</u>erage(*numlist*$\big[$, report$\big]$)	options for numerical integration, with generalized *numlist* syntax *lb*(*g*)*ub*; specify average(*lb*(0)*ub*) to use the sample average method; default is average(0(0)1); specify nonzero values of *g* to use the grid method, for example, average(.01(.01).99) to numerically integrate over the grid $(0.01, 0.02, \ldots, 0.99)$; the space of integration may comprise nonoverlapping ascending subsets by specifying comma-separated lists; suboption: specifying average(*lb1*(*g1*)*ub1*,...,*lbN*(*gN*)*ubN*, report) returns estimates for each subset as well as estimates over their union; suboption: together with varcoef(), specify the support for kernel weights in a varying-coefficient model

options	Description
<u>generate</u>(*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*)	create a comma-delimited (.csv) dataset of the local rank-specific coefficient estimates, saved to *filename*
varcoef(*varlist*)	fit a varying-coefficient 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*$_{\text{edg}}$; options average() and bandwidth() are required with varcoef()

The average() option determines the set $\mathcal{R}$ over which the local estimates $\widehat{\boldsymbol{\beta}}(\boldsymbol{r})$ are averaged and controls how this averaging is implemented. For example, average(.1 (0).3) sets $\mathcal{R} = [0.1, 0.3]$ and uses the empirical mean to evaluate the integral in (10). The command interprets a grid step of 0 as a request for computing $\widehat{\boldsymbol{\beta}}_{\mathcal{R}}$ using the sample averaging formula (6), which does not use knowledge of the distribution of $\boldsymbol{r}$. Alternatively, specifying average(.1(.01).3) sets $\mathcal{R} = [0.1, 0.3]$ and uses grid steps of 0.01 to numerically evaluate the integral. Multiple nonoverlapping 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 $\boldsymbol{\beta}_{\mathcal{R}}$ just discussed, along with another empirical average estimate for $\mathcal{R} = [0.5, 0.8]$. The grid method supports the report suboption as well.

There are two situations in which the command will always use (6) instead of attempting to numerically integrate (10). The first is when there is more than one basic endogenous variable, in which case $\boldsymbol{r}$ 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 command interprets each subset *lbn*(*gn*)*ubn* as belonging to the *n*th endogenous variable in order of appearance in *varlist*$_{\text{edg}}$. Because of 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 $\widehat{\boldsymbol{\beta}}(\boldsymbol{r})$ 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 $\widehat{\boldsymbol{r}}_i$ and uses the variables in *varlist* in its place. Because 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 the (*varlist*$_{\mathrm{edg}}$ = *varlist*$_2$) syntax in combination with the `varcoef()` option will generate an error.

5   Using ivcrc to estimate the returns to schooling

In this section, we use the `ivcrc` command to estimate the returns to schooling. Our discussion builds off of Card (1995, 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). In the notation of (7), y would be a labor market outcome (for example, wages) and x would be a measure of educational attainment (for example, years of completed schooling). There are several reasons to expect confounding factors that make the direct relationship between x and y a poor indicator of the causal effect. Some of these determinants, like family background characteristics, can be observed in data and controlled for. However, the fact that individuals have some choice over their education also suggests that some important confounding factors are inherently unobservable.

Instrumental-variables strategies have been widely used to tackle this self-selection problem. Some instruments that have been used are compulsory schooling laws (Angrist and Krueger 1991; Oreopoulos 2006), the distance a teenager lives from a college (Card 1993; Mountjoy 2021), and local labor market conditions (Cameron and Heckman 1998).[4] The argument underlying these strategies is that the proposed instrument affects an individual's educational attainment by shifting the costs or benefits involved but does not itself directly affect labor market outcomes and is uncorrelated with any other factors that do. The CRC model discussed in section 2 allows for the effect that an instrument has on an individual's educational choice to be correlated with their returns to schooling. So, for example, those whose education choice is affected by distance might have systematically different returns to schooling than those whose choice is not.

Our analysis uses the same data and regression specification as Card (1993) and Kling (2001), which are available as part of Cameron and Trivedi's (2010) textbook on Stata for microeconometrics. The data are an extract from the National Longitudinal Survey of Young Men 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, are available at https://github.com/a-torgovitsky/ivcrc.

---

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

The outcome variable is `wage76`, the individual's log hourly wage in 1976. The primary endogenous variable of interest is `grade76`, the individual's highest level of schooling measured in years. The regression includes two additional endogenous variables: years of potential work experience in 1976, `exp76`, and squared experience, `expsq76`. Potential work experience is defined as $exp76 = age76 - grade76 - 6$, following the standard convention for Mincer equations (Mincer 1958, 1974). To avoid perfect collinearity, we follow the literature by including experience in the outcome equation and age in the first-stage equation. Because potential experience is a derived endogenous variable, this convention also allows us to illustrate the command's `dendog()` option.

We include a set of sociodemographic controls for race (`black`), parents' education (`daded`, `momed`, `famed1`–`famed8`), family structure at age 14 (`momdad14`, `sinmom14`), and geographic region (`smsa66`, `smsa76`, `reg1`–`reg8`).[5] While not essential to demonstrating use of the `ivcrc` command, the inclusion of these controls shows that the semiparametric estimator does not suffer from the curse of dimensionality. Including these controls also replicates the regression specification in Card (1993) and Kling (2001), which allows us to compare the `ivcrc` command with popular alternative estimators, like two-stage least squares (for example, `ivregress`).

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 equation; see Heckman, Lochner, and Todd (2006) for an in-depth discussion. The estimates indicate that an additional year of schooling is associated with an approximate 7.25% increase in 1976 wages:

```
. use http://www.stata-press.com/data/mus2/mus207klingdata
. local ControlVars black south76 smsa66 smsa76 reg1-reg8
> momdad14 sinmom14 daded momed famed1-famed8
. local TableOptions ci noobs nostar nomtitles nonumbers nolines nonotes nogaps
. regress 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
```

wage76	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]
grade76	.0725423	.0038685	18.75	0.000	.0649572 .0801275

(*output omitted*)

As discussed above, 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.

---

5. For readability, we collect these into a local variable, `ControlVars`, for this exercise.

```
. regress 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
```

grade76	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
col4	.3669905	.1023706	3.58	0.000	.1662663	.5677147

(*output omitted*)

For college proximity to be a valid instrument, it should have no direct effect on wages in 1976 and also be uncorrelated with other factors that are correlated with wages or schooling decisions after conditioning on observables. There are several reasons to be suspect of this requirement; see, for example, Kling (2001), and see Mountjoy (2021) 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 two-stage least-squares estimator suggests that an additional year of schooling causes about a 13.33% 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	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
grade76	.1333034	.0493359	2.70	0.007	.0366068	.2300001

(*output omitted*)

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 is that this arises from a failure to account for heterogeneity in the causal effect of schooling on wages.

We can use the `ivcrc` command 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 rule-of-thumb bandwidth)
(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 |   .0852767         .         .       .              .            .
  (output omitted)

Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .0370327
```

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 8.53% is substantially lower than the linear IV estimate of 13.33%. 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 somewhat larger than the linear regression coefficient 0.0725.

We now demonstrate some of the options for `ivcrc` by evaluating the statistical significance and robustness of the above average treatment-effect estimate. First, we compute standard errors, which tends to be time consuming because of the necessity of using `bootstrap()`. The syntax and results are

```
. ivcrc wage76 (grade76 = col4 age76 agesq76) `ControlVars',
> dendog(exp76 expsq76) bootstrap(reps(100) seed(5282020))
(estimating rule-of-thumb bandwidth)
(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 |   .0852767   .0267166    3.19   0.001     .0329131    .1376403
  (output omitted)

Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .0370327
```

The confidence interval here is a bit wider than for the linear regression estimator, although it is substantially narrower than for the usual linear IV estimator. The textbook IV estimator and the `ivcrc` estimates are constructed under nonnested assumptions, so this by itself is not unexpected. However, because the bandwidth controls a bias-variance tradeoff in the `ivcrc` estimator, it does suggest that we may want to explore bandwidths other than the default ROT ($\widehat{h}_{\text{ROT}} = 0.037$) to guard against potential bias due to oversmoothing. So we next evaluate the point estimates at several specified 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	.	.	.	.	.

(output omitted)

```
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025
```

grade76	.0807563	.	.	.	.	.

(output omitted)

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

grade76	.0779116	.	.	.	.	.

(output omitted)

```
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 $\widehat{\boldsymbol{\beta}}(\widetilde{r})$ are computed using larger neighborhoods of $\widetilde{r}$. The standard errors and confidence intervals using the smallest bandwidth in this list are

```
. 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

(output omitted)

```
Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025
```

We find a slightly larger standard error and a wider confidence interval, as anticipated. Though more comparable with the standard error and confidence interval from the linear IV model, the `ivcrc` standard error remains roughly 1.5 times smaller, even 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 yields

```
. 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	.0320328	2.44	0.015	.0155079 .1410741

(output omitted)

```
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 (9) is approximated.

Next we swap 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	.0302556	2.58	0.010	.0187693	.137369

(output omitted)

```
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 the 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 ($\hat{r}_i$) directly as in (6). Alternatively, if we are concerned about results being driven by outliers in the conditional distribution of education, we can specify $\mathcal{R}$ to be $[0.05, 0.95]$. Trimming the distribution in this way, while maintaining the smaller bandwidth and more accurate rank estimation from above, yields

```
. 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	.0345893	2.13	0.033	.0057681	.1413558

(output omitted)

```
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 that used the full observed conditional 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 0.01 over the outlier-trimmed region [0.05, 0.95] from above gives

```
. 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	.0341265	2.14	0.032	.0062111 .1399845

(output omitted)

```
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	.0531316	1.96	0.051	−.0002521 .2080199

(output omitted)

```
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	.0320328	2.44	0.015	.0155079	.1410741

(output omitted)

Note: Average coefficients over R = [0,1] rank subset; Bandwidth = .025

grade76	.0651947	.0616646	1.06	0.290	−.0556657	.1860551

(output omitted)

Note: Average coefficients over R = [0,.25] rank subset; Bandwidth = .025

grade76	.1418537	.1006071	1.41	0.159	−.0553327	.33904

(output omitted)

Note: Average coefficients over R = [.2501,.5] rank subset; Bandwidth = .025

grade76	.0264401	.066345	0.40	0.690	−.1035938	.1564739

(output omitted)

Note: Average coefficients over R = [.5001,.75] rank subset; Bandwidth = .025

grade76	.0787306	.0512649	1.54	0.125	−.0217468	.1792081

(output omitted)

Note: Average coefficients over R = [.7501,1] rank subset; Bandwidth = .025

The command 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 that we specified in average(). We find that the estimates of the returns to schooling vary across the education distribution, with the second quartile exhibiting large returns that are comparable with 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.

We conclude with an illustration of varying-coefficient estimation with `ivcrc`. Cai (2010) suggests that the relationship between education and wages might be increasing in work experience. We can explore this hypothesis by using the `varcoef()` option, which allows the linear coefficients to be nonparametric functions of experience. To use `varcoef()`, we must also specify the `average()` option. To obtain the support points of experience for `average()`, we first quietly calculate summary statistics. Then we call `varcoef()` using the `savecoef()` option to store the experience-conditioned coefficient estimates. Loading the dataset stored with `savecoef()`, we plot our estimates of the coefficient on education as a function of experience in figure 1.

```
. quietly summarize exp76

. * Returns to education as a function of experience
. ivcrc wage76 grade76 `ControlVars',
> varcoef(exp76) bandwidth(4)
> average(`r(min)'(0)`r(max)') savecoef(varcoef_exp)
(default settings do not compute standard errors, see bootstrap() option)
(estimating beta(exp76) at each exp76[i] in the sample)
(file varcoef_exp.csv not found)
```

IVCRC					Number of obs = 3,010
wage76	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
grade76	.0602117	.	.	.	. .

(output omitted)

```
Note: Average coefficients over  = [0,23]; Bandwidth = 4

. * Load dataset of saved variable coefs, beta(exp)
. import delimited "varcoef_exp.csv", colrange(2:3) clear
(encoding automatically selected: ISO-8859-1)
(2 vars, 3,010 obs)

. quietly rename rank exp76

. * Plot without 1% tails
. local plotopt title("") scheme(sj)
> ylabel(, angle(0) labsize(4)) legend(off)
> xtitle("Years of experience", size(4))
> ytitle("Coefficient on education", size(4))

. twoway (scatter grade76 exp76, msize(small) mcolor(black))
> (lpoly grade76 exp76, degree(0) bwidth(1) lcolor(black) `plotopt')
```
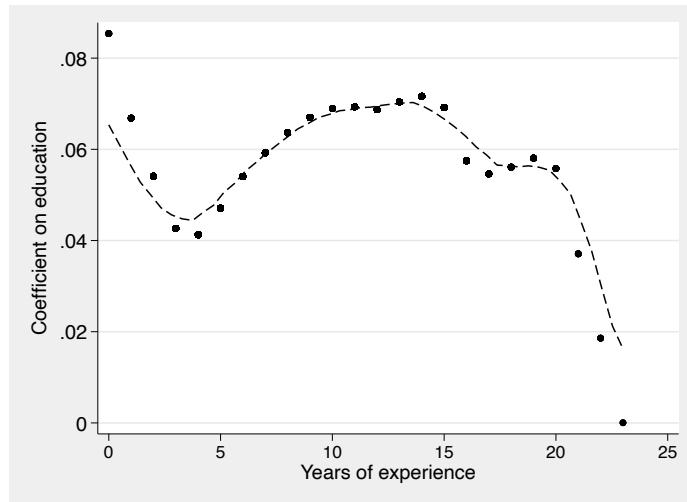
Figure 1. Varying-coefficient estimation: Coefficients as a function of experience

The coefficient on education is positive for all levels of experience and averages about 6% per year of education across experience levels. The coefficient decreases as experience increases for both small and large values of experience. However, as suggested by Cai (2010), the coefficient on education is increasing over the 10th to 90th percentiles (4 to 15 years) of experience.

6   Conclusion

In this article, we discussed the CRC model, which is a parsimonious IV model that explicitly incorporates heterogeneous treatment effects. The `ivcrc` command for Stata implements an estimator for the CRC model. The estimator can be used to carefully analyze heterogeneous treatment effects in a way that the usual linear estimator implemented by `ivregress` cannot. Because the estimator is based on averaging conditional linear regressions, it scales well and incorporates covariates easily, making it attractive from a practical perspective.

A few important limitations to the estimator are worth mentioning both as caveats and as directions for future research. First, the conditions described in section 2.3 to construct the control function, r, from the instrument, z, require the endogenous variable, x, to be continuously distributed. The extent to which these methods fail for discrete endogenous variables is an interesting theoretical question that may have important implications for practice. Second, the `ivcrc` command implements an ROT bandwidth. Other superior methods for automated bandwidth selection may exist. Third, defining, detecting, and correcting for weak instruments is an unexplored topic for the CRC model.

7 Acknowledgment

8 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 22-3
. net install st0680      (to install program files, if available)
. net get st0680          (to install ancillary files, if available)
```

The `ivcrc` command also is available on the Statistical Software Components archive and can be installed directly in Stata with the command

```
. ssc install ivcrc
```

Alternatively, the latest version of the command 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.[6]

9 References

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. *Review of Economic Studies* 67: 499–527. https://doi.org/10.1111/1467-937X.00141.

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. https://doi.org/10.1080/01621459.1995.10476535.

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. https://doi.org/10.2307/2291629.

Angrist, J. D., and A. B. Krueger. 1991. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106: 979–1014. https://doi.org/10.2307/2937954.

---

6. As described in the Stata FAQ: https://www.stata.com/support/faqs/programming/personal-ado-directory/.

Becker, G. 1975. Human capital and the personal distribution of income: An analytical approach. In *Human Capital*, 2nd ed. New York: Columbia University Press.

Becker, G. S., and B. R. Chiswick. 1966. Education and the distribution of earnings. *American Economic Review* 56: 358–369.

Blundell, R. W., and J. L. Powell. 2004. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71: 655–679. https://doi.org/10.1111/j.1467-937X.2004.00299.x.

Cai, Z. 2010. Functional Coefficient Models for Economic and Financial Data. In *The Oxford Handbook of Functional Data Analysis*, ed. F. Ferraty and Y. Romain, 166–188. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199568444.013.6.

Cameron, A. C., and P. K. Trivedi. 2010. *Microeconometrics Using Stata.* Rev. ed. College Station, TX: Stata Press.

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. https://doi.org/10.1086/250010.

Card, D. E. 1993. Using geographic variation in college proximity to estimate the return to schooling. NBER Working Paper No. 4483, The National Bureau of Economic Research. https://doi.org/10.3386/w4483.

———. 1995. Earnings, schooling, and ability revisited. *Research in Labor Economics* 14: 23–48.

———. 2001. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69: 1127–1160. https://doi.org/10.1111/1468-0262.00237.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil. 2011. Estimating marginal returns to education. *American Economic Review* 101: 2754–2781. https://doi.org/10.1257/aer.101.6.2754.

Chay, K. Y., and M. Greenstone. 2005. Does air quality matter? Evidence from the housing market. *Journal of Political Economy* 113: 376–424. https://doi.org/10.1086/427462.

Chen, X., O. Linton, and I. Van Keilegom. 2003. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71: 1591–1608. https://doi.org/10.1111/1468-0262.00461.

Chernozhukov, V., I. Fernández-Val, and A. Galichon. 2010. Quantile and probability curves without crossing. *Econometrica* 78: 1093–1125. https://doi.org/10.3982/ECTA7880.

Chernozhukov, V., and C. Hansen. 2005. An IV model of quantile treatment effects. *Econometrica* 73: 245–261. https://doi.org/10.1111/j.1468-0262.2005.00570.x.

Cleveland, W. S., E. Grosse, and W. M. Shyu. 1991. Local regression models. In *Statistical Models in S*, ed. J. M. Chambers and T. J. Hastie, 309–376. London: Chapman & Hall/CRC.

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. https://doi.org/10.1016/0165-1765(91)90161-d.

Fan, J., and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall/CRC.

Fan, J., and W. Zhang. 1999. Statistical estimation in varying coefficient models. *Annals of Statistics* 27: 1491–1518. https://doi.org/10.1214/aos/1017939139.

———. 2008. Statistical methods with varying coefficient models. *Statistics and Its Interface* 1: 179–195. https://doi.org/10.4310/SII.2008.v1.n1.a15.

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. https://doi.org/10.3982/ECTA5317.

Garen, J. 1984. The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica* 52: 1199–1218. https://doi.org/10.2307/1910996.

Gollin, D., and C. Udry. 2021. Heterogeneity, measurement error, and misallocation: Evidence from African agriculture. *Journal of Political Economy* 129: 1–80. https://doi.org/10.1086/711369.

Hansen, B. E. 2022. *Econometrics*. Princeton, NJ: Princeton University Press.

Hastie, T. J., and R. J. Tibshirani. 1993. Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B* 55: 757–796. https://doi.org/10.1111/j.2517-6161.1993.tb01939.x.

Heckman, J., and R. Robb, Jr. 1985. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30: 239–267. https://doi.org/10.1016/0304-4076(85)90139-3.

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. *Journal of Human Resources* 33: 974–987. https://doi.org/10.2307/146405.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161. https://doi.org/10.2307/1912352.

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 Vol. 1 of *Handbook of the Economics of Education*, ed. E. Hanushek and F. Welch, 307–458. Amsterdam: North-Holland. https://doi.org/10.1016/S1574-0692(06)01007-5.

Heckman, J. J., S. Urzua, and E. J. Vytlacil. 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88: 389–432. https://doi.org/10.1162/rest.88.3.389.

Hurwicz, L. 1950. Systems with nonadditive disturbances. In *Statistical Inference in Dynamic Economic Models*, ed. T. C. Koopmans, 410–418. No. Monograph No. 10 in Cowles Commission for Research in Economics, New York: Wiley.

Imbens, G. W., and J. D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62: 467–475. https://doi.org/10.2307/2951620.

Imbens, G. W., and W. K. Newey. 2009. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77: 1481–1512. https://doi.org/10.3982/ECTA7108.

Kling, J. R. 2001. Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business and Economic Statistics* 19: 358–364. https://doi.org/10.1198/073500101681020006.

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

Masten, M. A., and A. Torgovitsky. 2014. Instrumental variables estimation of a generalized correlated random coefficients model. Working Paper CWP02/14, Centre for Microdata Methods and Practice, Institute for Fiscal Studies. https://ideas.repec.org/p/ifs/cemmap/02-14.html.

———. 2016. Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics* 98: 1001–1005. https://doi.org/10.1162/rest_a_00603.

Mincer, J. 1958. Investment in human capital and personal income distribution. *Journal of Political Economy* 66: 281–302. https://doi.org/10.1086/258055.

———. 1974. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.

Morales-Mosquera, M. 2020. The economic value of crime control: Evidence from a large investment on police infrastructure. https://www.dropbox.com/s/d08n134b418fp6n/Paper_2_Crime.pdf?dl=0.

Mountjoy, J. 2021. Community colleges and upward mobility. NBER Working Paper No. 29254, The National Bureau of Economic Research. https://doi.org/10.3386/w29254.

Newey, W. K. 1994. The asymptotic variance of semiparametric estimators. *Econometrica* 62: 1349–1382. https://doi.org/10.2307/2951752.

O'Hagan, A. 1978. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society, Series B* 40: 1–24. https://doi.org/10.1111/j.2517-6161.1978.tb01643.x.

Oreopoulos, P. 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* 96: 152–175. https://doi.org/10.1257/000282806776157641.

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. https://doi.org/10.1111/insr.12029.

Rubin, H. 1950. Note on random coefficients. In *Statistical Inference in Dynamic Economic Models*, ed. T. C. Koopmans, 419–421. No. Monograph No. 10 in Cowles Commission for Research in Economics, New York: Wiley.

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. https://doi.org/10.2307/1911314.

Torgovitsky, A. 2015. Identification of nonseparable models using instruments with small support. *Econometrica* 83: 1185–1197. https://doi.org/10.3982/ECTA9984.

Wald, A. 1947. A note on regression analysis. *Annals of Mathematical Statistics* 18: 586–589.

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. https://doi.org/10.1016/S0165-1765(97)81890-3.

———. 2003. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters* 79: 185–191. https://doi.org/10.1016/S0165-1765(02)00318-X.

———. 2008. Instrumental variables estimation of the average treatment effect in the correlated random coefficient model. In *Modeling and Evaluating Treatment Effects in Econometrics*, ed. D. Millimet, J. Smith, and E. Vytlacil, 93–116. Bingley, U.K.: Emerald. https://doi.org/10.1016/S0731-9053(07)00004-7.

About the authors

David Benson is a senior economist in the Division of Research and Statistics at the Federal Reserve Board of Governors.

Matthew A. Masten is an associate professor of economics in the department of economics at Duke University.

Alexander Torgovitsky is a professor of economics in the Kenneth C. Griffin Department of Economics at the University of Chicago.