

NBER WORKING PAPER SERIES

TESTS OF HYPOTHESES ARISING IN THE CORRELATED RANDOM COEFFICIENT
MODEL

James J. Heckman
Daniel A. Schmieder

Working Paper 16421
<http://www.nber.org/papers/w16421>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2010

This research was supported by NIH R01-HD043411, NSF SES-024158, the American Bar Foundation and the Geary Institute, University College Dublin, Ireland. The views expressed in this paper are those of the authors and not necessarily those of the funders listed here, nor of the National Bureau of Economic Research. We have received helpful comments from Pedro Carneiro, Jeremy Fox, Joel Horowitz, Benjamin Moll, Azeem Shaikh, Christopher Taber, Edward Vytlačil, and Sergio Urzua.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by James J. Heckman and Daniel A. Schmieder. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Tests of Hypotheses Arising in the Correlated Random Coefficient Model
James J. Heckman and Daniel A. Schmierer
NBER Working Paper No. 16421
September 2010
JEL No. C31

ABSTRACT

This paper examines the correlated random coefficient model. It extends the analysis of Swamy (1971, 1974), who pioneered the uncorrelated random coefficient model in economics. We develop the properties of the correlated random coefficient model and derive a new representation of the variance of the instrumental variable estimator for that model. We develop tests of the validity of the correlated random coefficient model against the null hypothesis of the uncorrelated random coefficient model.

James J. Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and University College Dublin and IZA
and also NBER
jjh@uchicago.edu

Daniel A. Schmierer
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago IL 60637
dschmier@uchicago.edu

1 Introduction

The correlated random coefficient model is the new centerpiece of a large literature in microeconometrics. It extends the classical uncorrelated random coefficient model of Swamy (1971, 1974). For person i , outcome Y_i in terms of choice indicator D_i is written as

$$Y_i = \alpha_i + \beta_i D_i \tag{1}$$

where $D_i = 1$ if a choice is made; $D_i = 0$ if not and both the intercept, α_i , and the slope, β_i , vary among persons. In this expression both the α_i and β_i may depend on regressors X_i which we keep implicit.

β_i is the causal effect of D_i on Y_i holding α_i fixed. If agents make their choices to take treatment based on components of β_i that depend on variables not available to the observing economist, D_i is correlated with β_i even after conditioning on X_i . Most recent studies focus on estimating means or quantiles of the distribution of β_i .¹

The model that motivated the research of a previous generation (see, e.g., Griliches, 1977) assumes no response heterogeneity ($\beta_i = \beta$) or else an uncorrelated random coefficient model as in Swamy (1971, 1974) or Mincer (1974), so β_i is independent of D_i . The correlated random coefficient model assumes that β_i varies in the population and in addition that

$$\text{Cov}(D_i, \beta_i) \neq 0. \tag{C-1}$$

The model also accounts for selection on intercepts, *i.e.* selection on pretreatment unobservables:

$$\text{Cov}(D_i, \alpha_i) \neq 0. \tag{C-2}$$

When (C-1) holds, marginal returns to an activity in general differ from average returns. When assumption (C-2) holds but D_i is independent of β_i , standard IV identifies the mean

¹Abbring and Heckman (2007) discuss methods for estimating the distribution of β_i .

of β_i , which we denote by $\bar{\beta}$. This configuration of assumptions includes the case when β_i is random but independent of D_i and the case when β_i is the same for everyone.^{2,3}

As first noted by Heckman and Robb (1985, reprinted in 2008), instrumental variables (IV) applied to (1) when (C-1) holds produces an instrument-dependent parameter that, in general, is not $\bar{\beta}$.⁴ In general, different instruments identify different parameters. Under conditions specified in Yitzhaki (1989),⁵ Imbens and Angrist (1994), Heckman and Vytlacil (1999), and Heckman, Urzua, and Vytlacil (2006), IV estimates weighted averages of marginal effects. Heckman and Vytlacil (1999, 2001, 2005, 2007a) generalize the marginal treatment effect (MTE) introduced by Björklund and Moffitt (1987) and show that the MTE plays the role of a functional that is invariant to the choice of instrument. The MTE can be used to unify the literature on treatment effects.⁶

Heckman and Vytlacil (2001, 2005, 2007b) derive testable implications of the hypothesis that β_i is statistically independent of D_i given X_i :

$$H_0 : \beta_i \perp\!\!\!\perp D_i \mid X_i,$$

where $A \perp\!\!\!\perp B \mid C$ means A is independent of B given C . In this paper, we examine tests of this hypothesis, drawing heavily on our previous work (Heckman, Schmierer, and Urzua, 2010). We also consider tests of the hypothesis that $E(\beta_i) = \bar{\beta} = 0$ and that the IV estimand of (1) is zero.

The paper proceeds as follows. Section 2 establishes the equivalence of the correlated random coefficient model with the Generalized Roy model. We examine two testable implications of it. One test exploits the insight that, in general, in the case when H_0 is false,

²See Heckman and Vytlacil (1998), Heckman and Vytlacil (2007a,b). The standard “ability bias” problem (Griliches, 1977) assumes that $\beta_i = \beta$, a constant for all i , and that $\text{Cov}(D_i, \alpha_i) \neq 0$.

³Evidence from parametric models on the empirical relevance of (C-1) in a variety of areas of economics is presented in Heckman (2001, Table 3).

⁴See the discussion of the ensuing literature in Heckman, Urzua, and Vytlacil (2006) or Heckman and Vytlacil (2007a,b).

⁵Posted at website for Heckman, Urzua, and Vytlacil (2006), see <http://jenni.uchicago.edu/underiv/>.

⁶See Heckman and Vytlacil (2005, 2007a,b).

different instruments identify different parameters. Section 3 presents some new results on the sampling distribution of the instrumental variable estimator and develops tests for $\bar{\beta} = 0$ and for the IV estimand to be zero. Section 4 develops tests of the correlated random coefficient model. Section 5 develops other tests of H_0 and places all tests in the conditional moment testing framework (Ai and Chen, 2003). Section 6 concludes.

2 Equivalence with the Generalized Roy Model and Two Testable Implications of H_0

An alternative way to represent equation (1) makes the link to economic choice theory more explicit. Individual i experiences outcome $Y_{1,i}$ if $D_i = 1$ and outcome $Y_{0,i}$ if $D_i = 0$, $i = 1, \dots, I$. The observed outcome is $Y_i = D_i Y_{1,i} + (1 - D_i) Y_{0,i}$.⁷ Let $\mu_j(X_i) = E(Y_{j,i} | X_i)$, $j \in \{0, 1\}$. One can write the model for potential outcomes conditional on X_i as $Y_{1,i} = \mu_1(X_i) + U_{1,i}$ and $Y_{0,i} = \mu_0(X_i) + U_{0,i}$ where $E(U_{j,i} | X_i) = 0$, $j \in \{0, 1\}$. In this notation, the observed outcome is

$$Y_i = \mu_0(X_i) + [\mu_1(X_i) - \mu_0(X_i) + U_{1,i} - U_{0,i}] D_i + U_{0,i}.$$

This is the correlated random coefficient model of equation (1) where the baseline outcome is $\alpha_i = \mu_0(X_i) + U_{0,i}$ and the gain is $\beta_i = \mu_1(X_i) - \mu_0(X_i) + U_{1,i} - U_{0,i}$ where, for notational simplicity, we suppress the dependence of α_i and β_i on X_i . To simplify the expressions, we drop the i subscripts throughout the rest of the paper unless their use clarifies the discussion. We define $\alpha = \bar{\alpha} + U_\alpha$ and $\beta = \bar{\beta} + U_\beta$ where $E(U_\alpha | X) = 0$ and $E(U_\beta | X) = 0$. Table 1 shows the equivalent parameters for the two models.

Whether the null hypothesis H_0 is true or not depends on the underlying choice model. We postulate a threshold crossing model which assumes separability between observables

⁷This is the Quandt (1958) switching regression model.

Table 1: Equivalence of Notation Between the Correlated Random Coefficient Model and the Generalized Roy Model. All parameters are defined conditional on X_i which is left implicit.

	Generalized Roy model	Correlated random coefficient model
Baseline outcome	$Y_{0,i} = \mu_0 + U_{0,i}$	α_i
Outcome in treated state	$Y_{1,i} = \mu_1 + U_{1,i}$	$\beta_i + \alpha_i$
Gain to treatment (Individual causal effect)	$Y_{1,i} - Y_{0,i} = \mu_1 - \mu_0 + U_{1,i} - U_{0,i}$	β_i
Outcome	$Y_i = Y_{0,i} + D_i (Y_{1,i} - Y_{0,i})$ $= \mu_{0,i} + (\mu_{1,i} - \mu_{0,i} + U_{1,i} - U_{0,i}) D_i + U_{0,i}$	$Y_i = \alpha_i + \beta_i D_i$

Z that affect choice and an unobservable V : $D = \mathbf{1}(\mu_D(Z) - V > 0)$, where $\mathbf{1}(\cdot)$ is an indicator function that takes the value 1 if its argument is true and is 0 otherwise, and μ_D is a deterministic function of Z .⁸ Z can include components of X . Letting F_V be the distribution of V conditional on X , and assuming that $Z \perp\!\!\!\perp V \mid X$, the choice probability or “propensity score” is

$$P(z) = \Pr(D = 1 \mid Z = z) = F_V(\mu_D(z)),$$

where to simplify the notation, we keep the conditioning on X implicit. The choice equation can be written in several alternative and equivalent ways:

$$D = \mathbf{1}(\mu_D(Z) - V > 0) = \mathbf{1}(F_V(\mu_D(Z)) > F_V(V)) = \mathbf{1}(P(Z) > U_D)$$

where $U_D = F_V(V)$ so $U_D \sim \text{Uniform}[0, 1]$.

⁸See, e.g., Thurstone (1927) and McFadden (1974, 1981). We do not strictly require separability, but we do require that the choice equation has one representation in separable form. See Heckman and Vytlacil (2007b).

We invoke the assumptions of Heckman and Vytlacil (2005, 2007b).⁹ A fundamental treatment parameter introduced by Björklund and Moffitt (1987) is the marginal treatment effect (MTE). The MTE for a given value of $X = x$ is

$$MTE(x, u_D) = E(Y_1 - Y_0 \mid X = x, U_D = u_D) = E(\beta \mid X = x, U_D = u_D).$$

It is the mean effect of treatment when the observables X are fixed at a value x and the unobservable in the choice equation U_D is fixed at a value u_D . Heckman and Vytlacil (1999, 2001, 2005, 2007b) use the MTE to develop implications of the model to test H_0 .

In the general case, the conditional expectation of Y given X and Z is

$$\begin{aligned} E(Y \mid X = x, Z = z) &= E(Y \mid X = x, P(Z) = p) \\ &= E(\alpha \mid X = x) + E(\beta D \mid X = x, P(Z) = p) \\ &= E(\alpha \mid X = x) + E(\beta \mid X = x, D = 1)p \\ &= E(\alpha \mid X = x) + \int_0^p E(\beta \mid X = x, U_D = u_D) du_D, \end{aligned} \quad (2)$$

where the integrand in the final expression is the $MTE(x, u_D)$.¹⁰ Under H_0 ,

$$E(\beta \mid X = x, U_D = u_D) = E(\beta \mid X = x),$$

⁹Their conditions are:

(A-1) $(U_0, U_1, V) \perp\!\!\!\perp Z \mid X$. Alternatively, $(\alpha, \beta, V) \perp\!\!\!\perp Z \mid X$.

(A-2) The distribution of $\mu_D(Z)$ conditional on X is nondegenerate. Thus the distribution of $P(Z)$ is nondegenerate conditional on X .

(A-3) The distribution of V is continuous (i.e., absolutely continuous with respect to Lebesgue measure). Thus $U_D = F_V(V)$ is uniform.

(A-4) $E|Y_1| < \infty$ and $E|Y_0| < \infty$, so defining $E(\beta) = \bar{\beta}, |\bar{\beta}| < \infty$.

(A-5) $1 > \Pr(D = 1 \mid X) > 0$.

Vytlacil (2002) shows that under mild regularity conditions, assumptions (A-1)-(A-5) are equivalent to the IV conditions of Imbens and Angrist (1994) used to define the local average treatment effect (LATE).

¹⁰The first line follows from (A-1). The rest of the derivation comes from (1) and the law of iterated expectations.

so

$$E(Y | X = x, P(Z) = p) = E(\alpha | X = x) + E(\beta | X = x)p.^{11} \quad (3)$$

Thus the function $E(Y|X = x, P(Z) = p)$ is linear in p , conditional on $X = x$, which is a testable hypothesis.

A second implication of H_0 is that any standard instrument identifies $\bar{\beta} = E(\beta)$.¹² Thus under H_0 all valid instruments have the same estimand. Under conditions presented in this paper, comparing the estimates produced by different instruments tests the weaker hypothesis $H'_0 : \text{Cov}(\beta, D | X) = 0$, which is an implication of the stronger hypothesis H_0 . The analysis in this paper thus provides an alternative interpretation of standard tests of over-identification. A rejection of the null hypothesis that two instrumental variable estimands are different is not necessarily a rejection of the validity of one instrument. It could be interpreted as evidence in support of a correlated random coefficient model.

3 General Properties of the IV Estimator for the Correlated Random Coefficient Model and Tests of the Hypotheses $\bar{\beta} = 0$ and that the IV Estimand Is Zero

We present a new representation of the sampling distribution of the IV estimator. We consider the problem of constructing the power of tests of several hypotheses using the sampling distribution of the IV estimator for the correlated random coefficient model.

¹¹To see this, notice that $\beta \perp\!\!\!\perp D | X \iff \beta \perp\!\!\!\perp \mathbf{1}(P(Z) > U_D) | X \iff \beta \perp\!\!\!\perp U_D | X$ given (A-1).

¹²In the notation of equation (1), but dropping subscripts i , a standard instrument J has the two properties: (i) $\text{Cov}(J, D | X) \neq 0$ and (ii) $\text{Cov}((\alpha, \beta), J | X) = 0$. Note that J is shorthand for $J(Z)$. Note further that the condition $\text{Cov}(\beta, J | X) = 0$ only emerges as an interesting condition in a random coefficient model.

3.1 IV in the Correlated Random Coefficient Model

Consider an instrument $J(Z)$. Denote $J(Z)$ by J and define $\tilde{J} = J - \bar{J}$ where \bar{J} is the sample mean of $J(Z)$. $E(J)$ is assumed to be finite. The IV estimator is

$$\hat{\beta}_{IV,J} = \frac{\sum Y_i \tilde{J}_i}{\sum D_i \tilde{J}_i}.$$

Define $\text{Cov}(J, D) = \omega_J$ and let I denote the sample size. Under a weak law of large numbers, $\frac{1}{I} \sum D_i \tilde{J}_i \xrightarrow{p} \omega_J$ and $\bar{J} \xrightarrow{p} E(J)$. As shown in Heckman and Vytlacil (2005, 2007b), under the conditions (A-1)–(A-5) stated in Section 2,

$$\hat{\beta}_{IV,J} \xrightarrow{p} \beta_{IV,J} = \int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \quad (4)$$

where

$$h_J(u_D) = \frac{E[(J - E(J)) | P(Z) > u_D] \Pr(P(Z) > u_D)}{\omega_J}, \quad (5)$$

and we keep the conditioning on X implicit. Heckman and Vytlacil (2005) show that $\int_0^1 h_J(t) dt = 1$. Thus we can write

$$\beta_{IV,J} = \bar{\beta} + \int_0^1 E(U_\beta | U_D = u_D) h_J(u_D) du_D. \quad (6)$$

For later use we break out the component of $\beta_{IV,J}$ that depends on the instrument J :

$$\int_0^1 E(U_\beta | U_D = u_D) h_J(u_D) du_D = \Upsilon_J,$$

so $\beta_{IV,J} = \bar{\beta} + \Upsilon_J$. By definition, conditional on X , $\bar{\beta}$ does not depend on J .

Under independent sampling,

$$\sqrt{I} \left(\hat{\beta}_{IV,J} - \beta_{IV,J} \right) \xrightarrow{d} N(0, \Omega_J)$$

where

$$\Omega_J = E[\alpha^2] \frac{\text{Var}(J)}{\omega_J^2} + \int_0^1 [2E(\alpha\beta | U_D = u_D) + E(\beta^2 | U_D = u_D)] h_{\Omega_J}(u_D) du_D \quad (7)$$

$$- \left(\int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \right)^2$$

and

$$h_{\Omega_J}(u_D) = \frac{1}{\omega_J^2} \int_{-\infty}^{\infty} (j - E(J))^2 \int_{u_D}^1 f_{P,J}(P(z), j) dP(z) dj \quad (8)$$

$$= \frac{E[(J - E(J))^2 | P(Z) > u_D] \Pr(P(Z) > u_D)}{\omega_J^2}.$$

The weight $h_{\Omega_J}(u_D)$ does not necessarily integrate to 1:

$$\int_0^1 h_{\Omega_J}(t) dt = \frac{\text{Cov}(\tilde{J}^2, D)}{[\text{Cov}(\tilde{J}, D)]^2}.$$

Appendix A presents the full derivation. The weight $h_{\Omega_J}(u_D)$ plays a role in determining the variance of the IV estimator that is analogous to the role of $h_J(u_D)$ in generating the probability limit of the IV estimator. $2E[\alpha\beta | U_D = u_D] + E[\beta^2 | U_D = u_D]$ plays a role in generating the variance of the IV estimator analogous to the role of the MTE in generating the probability limit of the IV estimator. We use this representation to facilitate comparison of the power of the tests under alternative data generating processes and to consider the problem of the optimal choice of instruments.

These formulae hold for general functions $J(\cdot)$ of instruments Z that satisfy assumptions (A-1)–(A-5) given in Section 2. For example, suppose that $J(Z)$ has discrete support on points j_1, \dots, j_K with corresponding values of the propensity score p_1, \dots, p_L with L possibly not equal to K . Let $p_0 = 0$. In this case, for $u_D \in (p_l, p_{l+1})$ both h_J and h_{Ω_J} are constant

¹³ $f_{P,J}(P(z), j)$ is the density of $P(Z)$ and $J(Z)$ evaluated at $P(Z) = P(z)$ and $J(Z) = j$.

so we can write

$$\Omega_J = E[\alpha^2] \frac{\text{Var}(J)}{\omega_J^2} + \sum_{l=0}^{L-1} \lambda_{\Omega_l} \int_{p_l}^{p_{l+1}} [2E(\alpha\beta | U_D = u_D) + E(\beta^2 | U_D = u_D)] \frac{1}{p_{l+1} - p_l} du_D - \left(\sum_{l=1}^{L-1} \lambda_l \int_{p_l}^{p_{l+1}} E(\beta | U_D = u_D) \frac{1}{p_{l+1} - p_l} du_D \right)^2.$$

The weights λ_{Ω_l} and λ_l are defined in the following way. Let j_i be the i th smallest value in the support of $J(Z)$, then

$$\lambda_{\Omega_l} = \frac{\sum_{i=1}^K [j_i - E(J)]^2 \sum_{t>l}^L f_{P,J}(p_t, j_i)}{\text{Cov}(\tilde{J}(Z), D)^2} (p_{l+1} - p_l)$$

$$\lambda_l = \frac{\sum_{i=1}^K [j_i - E(J)] \sum_{t>l}^L f_{P,J}(p_t, j_i)}{\text{Cov}(\tilde{J}(Z), D)} (p_{l+1} - p_l).$$

For the special case of a binary instrument $J(Z)$ has two points of support, j_1 and j_2 , corresponding to the points p_1 and p_2 in the propensity score distribution. Let $\Pr(J(Z) = j_1) = \Pr(P(Z) = p_1) = q$ and $\Pr(J(Z) = j_2) = \Pr(P(Z) = p_2) = 1 - q$. The λ_l are $\lambda_1 = 1$ and $\lambda_l = 0, l > 1$.¹⁴ The weights for the variance simplify to

$$\lambda_{\Omega_0} = \frac{[j_1 - E(J)]^2 q + [j_2 - E(J)]^2 (1 - q)}{\text{Cov}(\tilde{J}(Z), D)^2} (p_1) \text{ and } \lambda_{\Omega_1} = \frac{[j_2 - E(J)]^2 (1 - q)}{\text{Cov}(\tilde{J}(Z), D)^2} (p_2 - p_1),$$

and

$$\lambda_{\Omega_0} + \lambda_{\Omega_1} = \frac{(j_1 - E(J))^2 q p_1 + (j_2 - E(J))^2 (1 - q) p_2}{\text{Cov}(\tilde{J}, D)^2} = \frac{\text{Cov}(\tilde{J}^2, D)}{\text{Cov}(\tilde{J}, D)^2}.$$

Formula (4) extends the representation of IV as weighted averages of slopes of the underlying function, due to Yitzhaki (1989). It allows the instrument $J(Z)$ be different from

14

$$\lambda_1 = \frac{[j_2 - E(J)](1 - q)}{\text{Cov}(\tilde{J}(Z), D)} (p_2 - p_1) = \frac{(j_2 - j_1)(p_2 - p_1)q(1 - q)}{\text{Cov}(\tilde{J}(Z), D)} = \frac{\text{Cov}(\tilde{J}(Z), P(Z))}{\text{Cov}(\tilde{J}(Z), D)} = 1.$$

the propensity score $P(Z)$ or a monotonic function of it. It reveals that, in general, different instruments identify different parameters. Thus, in general, $\beta_{IV,J} \neq \beta_{IV,J'}$ if J and J' apply different weights (5) to a common MTE.

As noted by Heckman and Vytlacil (2005, 2007b), while the weight in (5) integrates to 1, it is not necessarily non-negative for all values of u_D so the interpretation of the weighted average produced by IV is obscure. Even though the MTE is positive everywhere, the IV estimate may be negative.¹⁵

Some applied economists report tests based on IV sampling distributions as if they are testing the null hypothesis that $\bar{\beta} = 0$. Under H_0 , i.e., the absence of a correlated random coefficient model, the sampling distribution of the standard IV estimator, $\hat{\beta}_{IV,J}$, can be used to consistently test the null hypothesis that $\bar{\beta} = 0$. However, when H_0 is false, a test of $\bar{\beta} = 0$ based on the sampling distribution of the IV estimator is, in general, inconsistent and biased because by (6), IV does not, in general, converge to $\bar{\beta}$.

Consider the following example based on the normal generalized Roy Model,

$$\begin{pmatrix} U_1 \\ U_0 \\ V \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{10} & \sigma_{1V} \\ \sigma_{10} & \sigma_0^2 & \sigma_{0V} \\ \sigma_{1V} & \sigma_{0V} & \sigma_V^2 \end{pmatrix} \right), \quad (9)$$

and assume $X = 1$. Recalling that $u_D = F_V(v)$, when V is a normal random variable, the marginal treatment effect is

$$MTE(U_D = u_D) = \bar{\beta} + \left(\frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V} \right) \Phi^{-1}(u_D) \quad (10)$$

where $\Phi^{-1}(\cdot)$ is the inverse of a standard normal CDF (hence $\Phi^{-1}(u_D) = v$). Alternatively, in terms of v ,

$$MTE(V = v) = \bar{\beta} + \frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V} v.$$

¹⁵See the examples in Heckman, Urzua, and Vytlacil (2006).

Let $\tau = \frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V}$. A value of $\tau \neq 0$ produces a correlated random coefficient model. For such values $\text{plim } \hat{\beta}_{IV,J} \neq \bar{\beta}$. The choice equation is assumed to be $D = \mathbf{1}(Z > V)$ where both Z (a single instrument) and V are normally distributed and $Z \perp\!\!\!\perp V$. Additionally, assume that $\sigma_1^2 = \sigma_0^2 = \sigma_U^2$, $\sigma_{10} = 0.5 \times \sigma_1 \times \sigma_0$ and $\sigma_V^2 = 1$.

Figure 1 plots the power of a Wald test of the hypothesis that $\bar{\beta} = 0$ based on $\hat{\beta}_{IV,J}$. We compute the power function for different values of $\bar{\beta}$. Recall from (6) that this is the component of $\beta_{IV,J}$ that does not depend on J . In Panel A, $\hat{\beta}_{IV,J}$ is a consistent estimator for $\bar{\beta}$. In the other two panels it is not. Thus in the top panel of the figures, when $\tau = 0$, and hence H_0 is true, the test of the hypothesis $\bar{\beta} = 0$ is unbiased and consistent and the size of the test is controlled.¹⁶ As expected, smaller values of σ_U^2 produce higher power, and larger values of σ_Z^2 produce higher power. The bottom two panels plot the power of the test that $\bar{\beta} = 0$ when $\tau = -1$ and $\tau = 0.6$, respectively. In these two latter cases, $\text{plim } \hat{\beta}_{IV,J} = \beta_{IV,J} \neq \bar{\beta}$. Hence the tests are biased and inconsistent. The power and size of the test for the existence of an “effect” (i.e., whether $\bar{\beta} = 0$) can be badly distorted. Thus even if $\bar{\beta} = 0$, an “effect” can be detected, and if $\bar{\beta} \neq 0$, no “effect” can be detected.

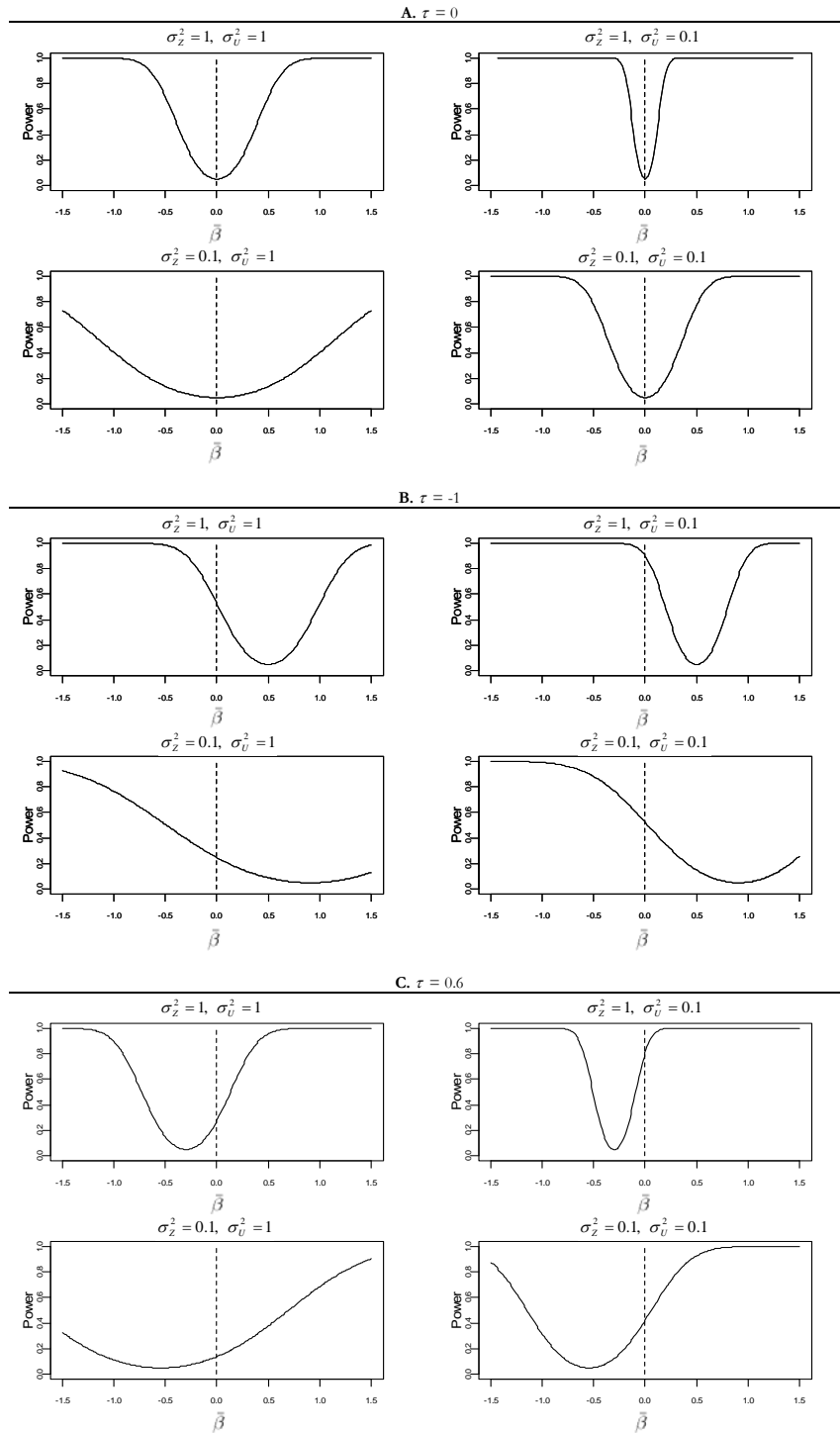
3.2 Testing Hypotheses About Instrument-Dependent Parameters

More recently, many applied economists, following Imbens and Angrist (1994), interpret IV as a weighted average of “LATEs,” or in our framework, a weighted average of MTEs, as in equation (3). It is understood that $\hat{\beta}_{IV,J}$ is not, in general, consistent for the true $\bar{\beta}$. Within this framework, economists often report tests of the hypothesis that $\beta_{IV,J} = 0$.

To calculate the power of such tests, consider alternative values of $\beta_{IV,J} (= \bar{\beta} + \Upsilon_J$ from equation (6)) obtained by varying $\bar{\beta}$ holding Υ_J fixed. Notice that unlike the analysis in the preceding section, in this section we are not testing the hypothesis that $\bar{\beta} = 0$. Instead

¹⁶Although Figure 1 shows the power function only for one sample size, the consistency of the test is readily verified.

Figure 1: Power function for a Wald test of $\bar{\beta} = 0$ based on the sampling distribution of $\hat{\beta}_{IV,J}$.



Note: Each plot shows the power with a hypothetical sample size of 500. The size of the test is 0.05. The model is the normal generalized Roy model with the unobservables jointly normal with variance σ_v^2 and correlation 0.5. The choice equation is $D = \mathbf{1}(Z > V)$ where $V \sim N(0, 1)$ and $Z \sim N(1, \sigma_Z^2)$. The power functions plot the power of the Wald test of $\beta_{IV,J} = 0$ for alternative values of $\bar{\beta}$. The vertical dashed lines denote the null hypothesis $\bar{\beta} = 0$. Each panel fixes $\tau = \text{Cov}(\beta, V) / \text{Var}(V)$ at a different level. When $\tau = 0$, $\text{plim} \hat{\beta}_{IV} = \beta_0$, which in these figures is zero, and hence the test is consistent. For all nonzero values of τ , the test is inconsistent.

we are testing the hypothesis that $\beta_{IV,J} = 0$ (or some other specified value). We vary $\bar{\beta}$ to calculate the power of the test for alternative values of $\beta_{IV,J}$. This is a sensible way to proceed because $\bar{\beta}$ is instrument invariant. Investigating the power of the test in this fashion allows us to construct power functions for instrument-invariant alternatives.

Figure 2 plots the power function for the Wald test of the hypothesis $\beta_{IV,J} = 0$ as a function of $\beta_{IV,J}$ holding \mathcal{Y}_J fixed at -0.5. Consequently, the $\bar{\beta}$ compatible with the null hypothesis, $\bar{\beta}_0$, is 0.5. For the model of unobservables used in the previous subsection, keeping \mathcal{Y}_J fixed entails, among other things, holding $\tau = \frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V}$ fixed along with the weighting function $h_J(u_D)$. For a given τ and a fixed IV weighting function $h_J(u_D)$, we vary the parameters of covariance matrix (9). These parameters affect the sampling distribution of $\hat{\beta}_{IV,J}$ and hence the power of the test.

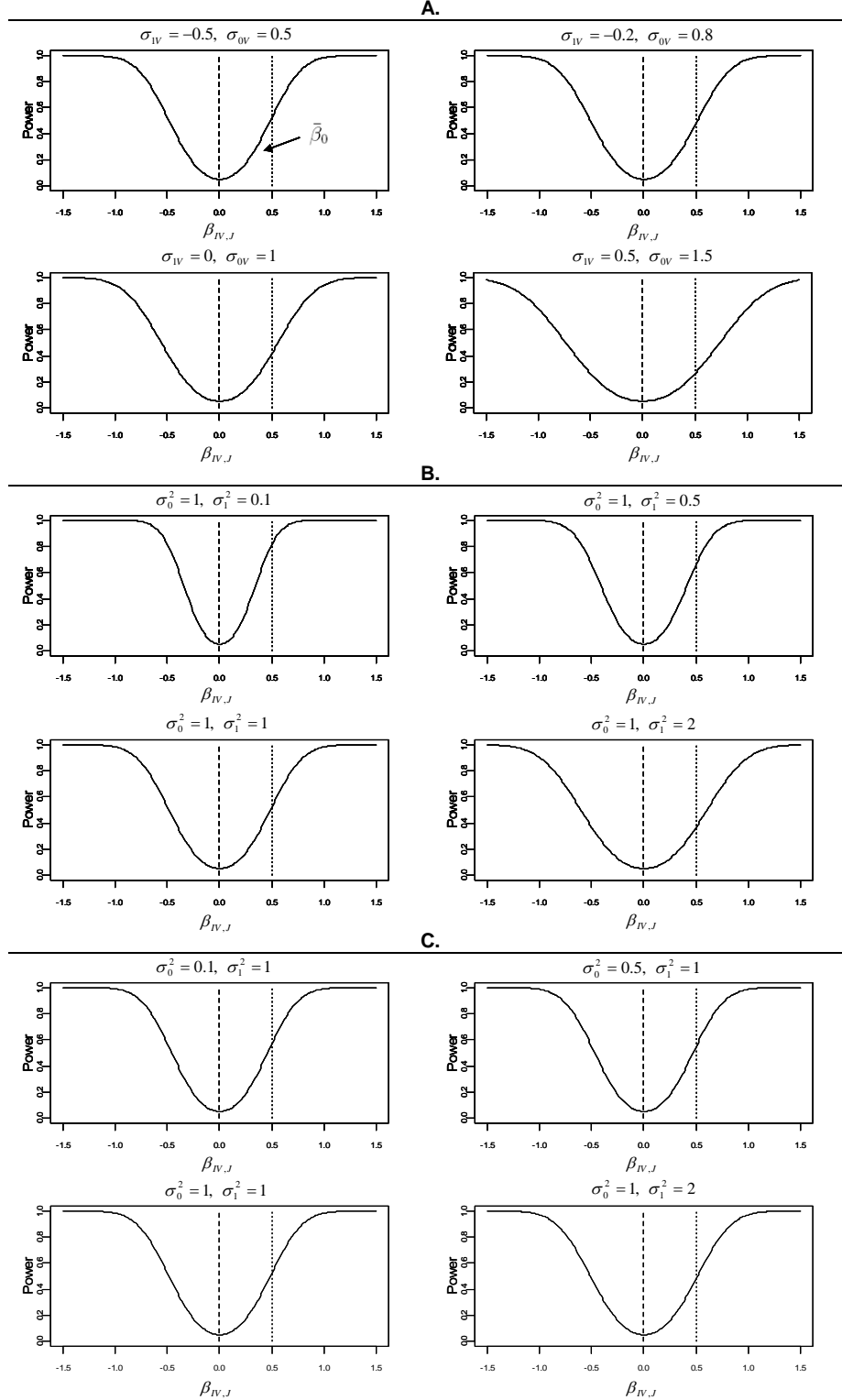
Neither the IV estimand nor the variance of the IV estimator depends on σ_{10} . Therefore, the power of the test of the null hypothesis $\beta_{IV,J} = 0$ does not depend on σ_{10} . The only remaining parameters that can be changed without changing \mathcal{Y}_J are $\sigma_0^2, \sigma_1^2, \sigma_{1V}$ and σ_{0V} . To keep τ fixed, we can only vary σ_{1V} and σ_{0V} subject to a constraint that $\sigma_{1V} - \sigma_{0V}$ is constant.¹⁷ For $\sigma_V = 1$, the four A panels of Figure 2 show the power of the test for different values of $\bar{\beta}$ when we vary σ_{1V} and σ_{0V} such that $\sigma_{1V} - \sigma_{0V} = -1$. The power of the test is highest when σ_{1V} and σ_{0V} are both close to 0 (ie. straddling 0), and lowest when both are far from zero (either positive or negative). The panels in B vary $\beta_{IV,J}$ by varying $\bar{\beta}$ holding \mathcal{Y}_J fixed and hold fixed all of the parameters of (9) except for σ_1^2 , while the panels in C vary $\bar{\beta}$ hold fixed all of the parameters of (9) except σ_0^2 . As expected, power decreases as both variances increase, in general at different rates.

There are other ways to calculate the power of the test that $\beta_{IV,J} = 0$ for alternative values that are obtained by varying $\bar{\beta}$ keeping \mathcal{Y}_J fixed. If the choice equation is

$$D = 1(Z\gamma > V)$$

¹⁷Variations in σ_V^2 affect the denominator of the weights.

Figure 2: Power function for the test of the hypothesis that $\text{plim } \hat{\beta}_{IV,J} = 0$ when $\bar{\beta}_0 = 0.5$. Alternatives are different values of $\beta_{IV,J}$ obtained by fixing Υ_J and varying $\bar{\beta}$.



Note: Each plot shows the power with a hypothetical sample size of 500. The size of the test is 0.05. The instrument is normally distributed, $Z \sim N(1, 1)$; $D = 1(Z > V)$. In panel A, the unobservables are generated with covariances given in the figure and $\sigma_V^2 = 1$, $\sigma_{10} = 0$, $\sigma_1^2 = 1$, $\sigma_0^2 = 1$. In panels B and C the unobservables are generated with variances given in the figure and $\sigma_V^2 = 1$, $\sigma_{10} = 0$, $\sigma_{1V} = -0.5$, $\sigma_{0V} = 0.5$. In all panels, under the null hypothesis $\bar{\beta}_0 = 0.5$, and alternative hypotheses are generated by changing $\bar{\beta}$. The vertical dashed line shows the value of β under the null hypothesis.

and $Z \sim N(\bar{Z}, \Sigma_Z)$ and $V \sim N(0, \sigma_V^2)$, all instruments constructed from linear or affine transformations of Z have the same weight function (5) and hence have the same instrument-dependent value, $\beta_{IV,J}$. For proof of this claim, see Appendix B.¹⁸

This result implies that one can construct power functions for the hypothesis $\beta_{IV,J} = 0$ for different values of $\beta_{IV,J} = \bar{\beta} + \Upsilon_J$ for alternative choices of Σ_Z , holding $\gamma' \Sigma_Z \gamma$, the variance of the choice index, constant. The derivation in Appendix B shows that the IV estimand depends only on the distribution of the index $Z\gamma - V$. From assumption (A-1), $Z\gamma$ and V are statistically independent. σ_V^2 has to be held constant to keep Υ_J fixed. We keep this term fixed by varying components of Z while keeping $\gamma' Z\gamma$ fixed. An instrument with greater variance that obeys this constraint will produce greater power. Figure 3 plots power functions of the test of the hypothesis that $\beta_{IV,J} = 0$ using each component of a two-dimensional instrument $Z = (Z_1, Z_2)$. These plots show that for a given IV estimand $\beta_{IV,J}$, the power of the test is higher when using the instrument that accounts for more of the variance of the index $Z\gamma$. Going from top to bottom, the variance of Z_1 is increasing while the variance of Z_2 is decreasing. Accordingly, from top to bottom the power of the test $\beta_{IV,J} = 0$ using Z_1 as an instrument is increasing while the power of the test using Z_2 as an instrument is decreasing. Each panel shows the fraction of $\gamma' Z\gamma$ accounted for by the variance of the instrument used to construct the power function (either Z_1 or Z_2).¹⁹ We now use the tools developed for IV in a correlated random coefficient model to test H_0 .

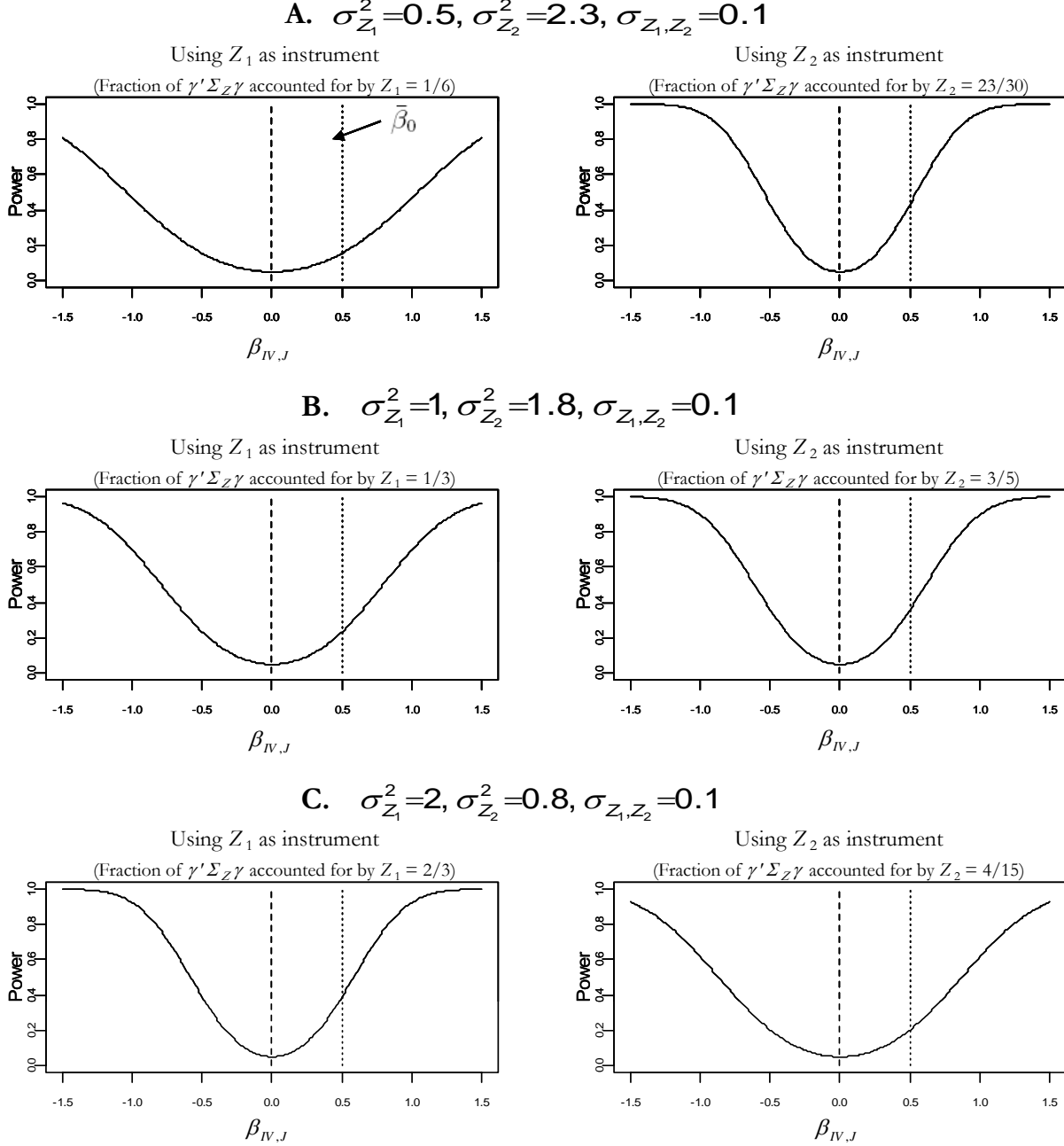
4 Testing H_0 Using Instrumental Variables

Armed with the results of Section 3, we study how to use different IVs to test H_0 . Under H_0 , the probability limits of any two IV estimators are identical, because for any choice of

¹⁸This result is special to the case of $J(Z)$ linear or affine in Z with Z normally distributed, so $J(Z)$ is normally distributed and the further assumption (A-1) that $Z \perp\!\!\!\perp V$, where V is normally distributed. We have not analyzed more general conditions on Z and V under which the invariance holds.

¹⁹Note that in a given row, the fractions do not sum to 1 because there is a covariance (of 0.1) between Z_1 and Z_2 .

Figure 3: Power functions for the test of the hypothesis that $\beta_{IV,J} = \text{plim } \hat{\beta}_{IV,J} = 0$ for $\bar{\beta}_0 = 0.5$. Alternatives are different values of $\beta_{IV,J}$ obtained by fixing Υ_J and varying $\bar{\beta}$.



Note: Each plot shows the power with a hypothetical sample size of 500 varying $\bar{\beta}$ keeping Υ_J fixed. The size of the test is 0.05. The instruments are distributed normally, $Z_1 \sim N(1, \sigma_{Z_1}^2)$, $Z_2 \sim N(1, \sigma_{Z_2}^2)$ and $\text{Cov}(Z_1, Z_2) = \sigma_{Z_1, Z_2}$, $Z_2 = 0.1$; $D = \mathbf{1}(Z_1 + Z_2 > V)$ so $\gamma = (1, 1)$. The distribution of the index is held fixed and is distributed $N(2, 3)$. The unobservables are jointly normally distributed with zero means and $\sigma_V^2 = 1$, $\sigma_{10} = 0.5$, $\sigma_1^2 = 1$, $\sigma_0^2 = 1$, $\sigma_{1V} = -0.3$, $\sigma_{0V} = 0.3$. In all panels, under the null hypothesis $\bar{\beta}_0 = 0.5$, and alternative hypotheses are generated by changing $\bar{\beta}$. The vertical dashed line shows the value of $\beta_{IV,J}$ under the null hypothesis and the vertical dotted line shows the value of $\bar{\beta} = \bar{\beta}_0$ under the null hypothesis being considered, i.e. that $\beta_{IV,J} = \bar{\beta} + \Upsilon_J$.

J ,

$$\text{plim } \widehat{\beta}_{IV,J} = \beta_{IV,J} = \int_0^1 E(\beta|U_D = u_D)h_J(u_D)du_D = \bar{\beta} \int_0^1 h_J(u_D)du_D = \bar{\beta}.$$

If H_0 is false, in general any two IV estimators will differ. Excluding the case of equal IV weights for the two instruments, our IV test forms two estimators $\widehat{\beta}_{IV,1}$ and $\widehat{\beta}_{IV,2}$, based on $J_1(Z)$ and $J_2(Z)$ respectively, and tests the null hypothesis

$$H_0^{IV} : \beta_{IV,1} - \beta_{IV,2} = 0$$

against the alternative hypothesis

$$H_A^{IV} : \beta_{IV,1} - \beta_{IV,2} \neq 0.$$

This test is identical to a standard test for overidentification. However, within the context of a correlated random coefficient model, we do not interpret rejections of the null hypothesis as evidence of the violation of the assumptions required for the validity of an instrument. Rather, rejections are interpreted as evidence of selection on heterogeneous gains to treatment.

Under the null hypothesis, the Wald test statistic is asymptotically distributed as a χ_1^2 . Under the alternative, in the general case, the Wald statistic converges to a noncentral chi-square distribution. Let $h_1(\cdot)$ and $h_2(\cdot)$ denote the weights (akin to $h_J(\cdot)$ above) corresponding to $J_1(Z)$ and $J_2(Z)$, respectively. To simplify the notation, we suppress the Z argument. Define $\tilde{J}_1 = J_1 - \bar{J}_1$ and $\tilde{J}_2 = J_2 - \bar{J}_2$ as the demeaned values of the instruments. Let $\tilde{\mathbf{J}}_1 = (\tilde{J}_{11}, \dots, \tilde{J}_{1I})'$ and $\tilde{\mathbf{J}}_2 = (\tilde{J}_{21}, \dots, \tilde{J}_{2I})'$ be the matrices of demeaned instruments stacked across individuals. Let $\mathbf{D} = (D_1, \dots, D_I)'$ be the stacked values of the choice variable D_i . Under random sampling, and the assumptions of Section 2, $\frac{\tilde{\mathbf{J}}_1' \mathbf{D}}{I} \xrightarrow{p} \omega_1$ and $\frac{\tilde{\mathbf{J}}_2' \mathbf{D}}{I} \xrightarrow{p} \omega_2$ for some finite constants ω_1 and ω_2 . Under $H_A^{IV} : \beta_{IV,1} - \beta_{IV,2} = \left[\int_0^1 MTE(u_D)(h_1(u_D) - h_2(u_D))du_D \right] / \sqrt{I}$, the noncentrality parameter of the chi-square

distribution of the test statistic is

$$\lambda_{IV,1,2} = \frac{1}{2} \left(\int_0^1 MTE(u_D)(h_1(u_D) - h_2(u_D)) du_D \right)^2 \Psi_{1,2}^{-1} \quad (11)$$

where

$$\begin{aligned} \Psi_{1,2} = E \left(\alpha^2 \right) & \left[\frac{\text{Var}(J_1)}{\omega_1^2} - \frac{2 \text{Cov}(J_1, J_2)}{\omega_1 \omega_2} + \frac{\text{Var}(J_2)}{\omega_2^2} \right] \\ & + \int_0^1 [2E(\alpha\beta \mid U_D = u_D) + E(\beta^2 \mid U_D = u_D)] h_{\Omega, J_1, J_2}(u_D) du_D \\ & - \left[\int_0^1 MTE(u_D)(h_1(u_D) - h_2(u_D)) du_D \right]^2. \end{aligned} \quad (12)$$

Defining $J_1^* = J_1 - E(J_1)$ and $J_2^* = J_2 - E(J_2)$, the weight $h_{\Omega, J_1, J_2}(\cdot)$ is given by

$$\begin{aligned} h_{\Omega, J_1, J_2}(u_D) &= \int_{u_D}^1 \int_{-\infty}^{\infty} \left(\frac{J_1^*}{\omega_1} - \frac{J_2^*}{\omega_2} \right)^2 f_{(J_1 - J_2), P}(j_1 - j_2, P(z)) d(j_1 - j_2) dP(z) \\ &= E \left[\left(\frac{J_1^*}{\omega_1} - \frac{J_2^*}{\omega_2} \right)^2 \mid P(Z) > u_D \right] \Pr(P(Z) > u_D).^{20} \end{aligned}$$

The derivation follows a logic similar to that used to derive (7).²¹ Notice that not only will the difference in the IV estimands depend on the alternative under consideration, but the variance of the difference between the IV estimators will also depend on the alternative under consideration.

We present this characterization of the variance in order to understand the properties of tests of H_0 based on IV estimators. This expression for the variance is not meant as a guide for how to implement such tests. In practice the analyst would form the test statistic using

²⁰ $f_{(J_1 - J_2), P}(j_1 - j_2, P(z))$ is the joint density of $J_1 - J_2$, and $P(Z)$ evaluated at $J_1 - J_2 = j_1 - j_2$ and $P(Z) = P(z)$.

²¹The logic is not, however, identical. Using $(J_1 - J_2)$ as an instrument and testing if $\beta_{IV, J_1 - J_2} = 0$ is *not* equivalent to the test presented in the main text of the paper. The denominators of the IVs differ in the two approaches.

a standard estimator of the variance of the vector of IV estimates.

In general, the weights presented above do not have simple analytical expressions. They do in the case of a model with normal error terms with normally distributed instruments and a linear index structure for the choice equation. However, for this case, the proposed IV test has no power, because, and as previously discussed and as established in Appendix B, $\beta_{IV,J_1} \equiv \beta_{IV,J_2}$ irrespective of the truth or falsity of H_0 . For this case, the noncentrality parameter of the asymptotic chi-square distribution of the test statistic will be zero so the power of the test equals its size. To have a test with any power, we have to rule out instruments with equal weights. Since the weights can be constructed from the data on Z , it is possible to check this condition in any sample.²²

We do not formally analyze conditions that guarantee that the two instruments J_1 and J_2 , constructed from Z , optimize the power function of the test. From the expression for the noncentrality parameter, one can see the ingredients required to construct an asymptotically most powerful test. Let $Z \in \mathbb{R}^k$ be the vector of available instruments and let $\mathcal{J} = \{J \mid J : \mathbb{R}^k \rightarrow \mathbb{R}\}$ be the space of functions which map the vector of instruments to the real line. Then for a given MTE, the optimal choice of J_1 and J_2 solves the problem

$$\max_{J_1 \in \mathcal{J}, J_2 \in \mathcal{J}} \frac{1}{2} \left(\int_0^1 MTE(u_D)(h_{J_1}(u_D) - h_{J_2}(u_D)) du_D \right)^2 \Psi_{1,2}^{-1}.$$

The optimal choice of instruments will generally depend on the shape of the $MTE(u_D)$.²³

We present an example with two non-normal instruments in Figure 4. Specifically, let $D = \mathbf{1}(\gamma_1 Z_1 + \gamma_2 Z_2 > V)$ where the vector $Z = (Z_1, Z_2)$ is distributed as a multivariate *mixture* of normals with the distribution given at the base of the figure. The unobservables are assumed to be generated by a normal generalized Roy model. The test of equality of the IV estimators constructed using these two instruments has power to detect deviations from

²²It would be desirable to develop a formal test for equality of the two IV weights. The required ingredients are in the literature. We leave the formal derivation for another occasion.

²³More generally, one could use multiple instruments and base a test on multiple contrasts of the set of instruments. We do not develop this test in this paper.

H_0 . Figure 4A plots the weights $h_1(\cdot)$ and $h_2(\cdot)$ which the IV estimator places on the MTE, using Z_1 or Z_2 respectively. The weights must differ for the test based on the difference in IV estimators to have power to detect deviations from H_0 . When the mixing proportion in the mixture of normals is 0.45, the instruments are highly non-normal and the IV weights differ substantially. However, when the mixing proportion is 0.75, the instruments become closer to normal, the weights become very similar, and the test of H_0 loses power. This case is discussed further in Heckman, Schmieder, and Urzua (2010).

Another example of a test that has power to detect deviations from H_0 , even with normal instruments, constructs IV estimators using nonlinear functions of Z . We consider a normal generalized Roy model where there is one Z variable in the choice equation that is normally distributed, $D = \mathbf{1}(Z > V)$. We plot the weights of the IV estimators based on Z and Z^2 . Figure 4B plots the weights for these two choices of instruments. The weights differ, and in addition the amount by which they differ generally depends on the distribution of Z . We plot the weights for two choices of the mean of Z presented in the figure. These choices clearly affect the weights and hence will generally affect the power of a test of H_0 based on these IV estimators.

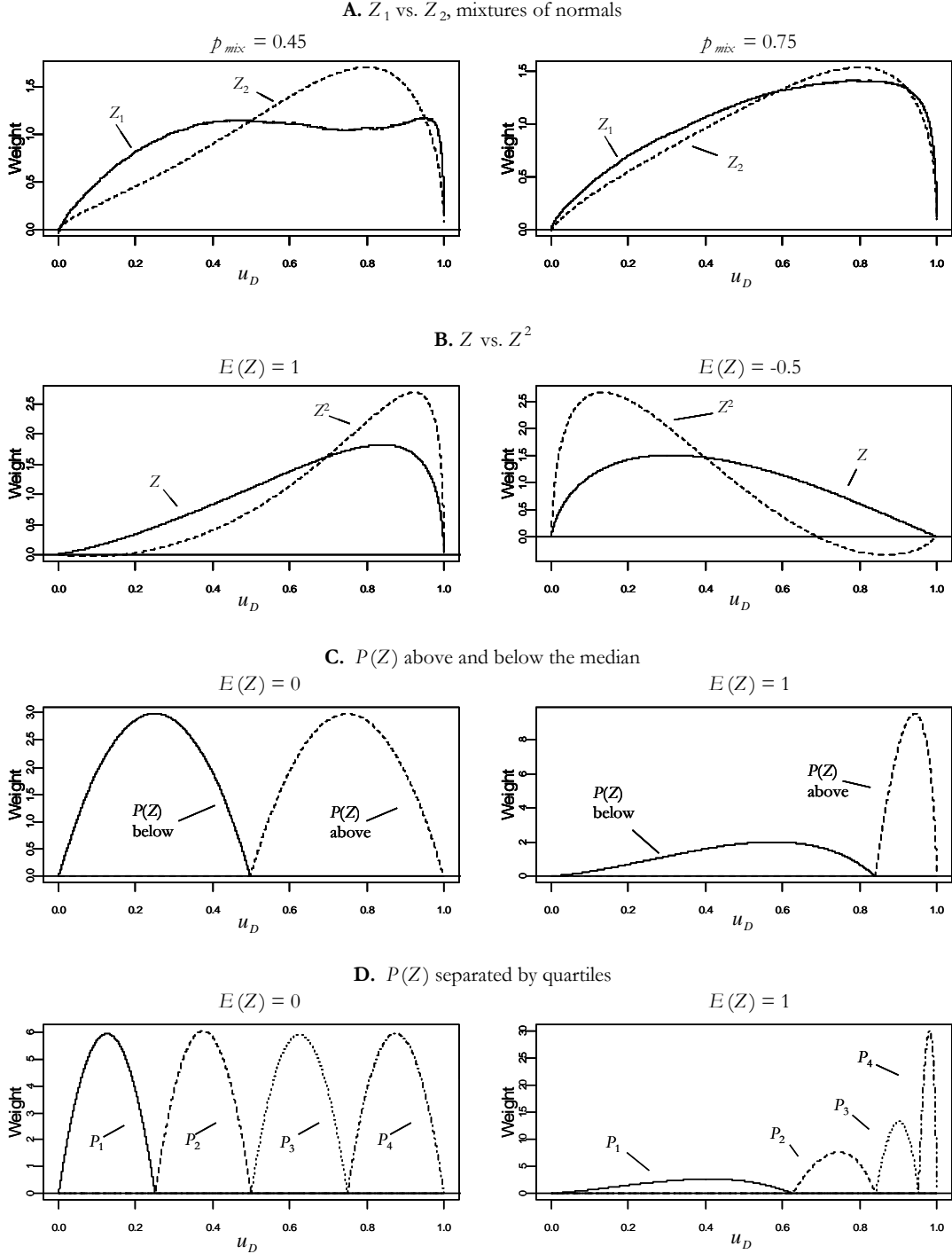
Another choice of instruments uses $P(Z)$ on disjoint intervals of the support of $P(Z)$ as two instruments. Form two disjoint intervals $[\underline{p}_1, \bar{p}_1]$ and $[\underline{p}_2, \bar{p}_2]$, and construct IV estimators over these intervals as sample analogs to

$$\beta_{IV, [\underline{p}_1, \bar{p}_1]} = \frac{\text{Cov}\left(Y, P(Z) \mid P(Z) \in [\underline{p}_1, \bar{p}_1]\right)}{\text{Var}\left(P(Z) \mid P(Z) \in [\underline{p}_1, \bar{p}_1]\right)}$$

and

$$\beta_{IV, [\underline{p}_2, \bar{p}_2]} = \frac{\text{Cov}\left(Y, P(Z) \mid P(Z) \in [\underline{p}_2, \bar{p}_2]\right)}{\text{Var}\left(P(Z) \mid P(Z) \in [\underline{p}_2, \bar{p}_2]\right)}$$

Figure 4: IV weights for alternative choices of the instrument.



Note: Panel A plots the weights of IV estimates constructed using either Z_1 or Z_2 as an instrument where (Z_1, Z_2) is distributed as a multivariate mixture of normals, with $D = \mathbf{1}(\gamma_1 Z_1 + \gamma_2 Z_2 > V)$. To construct these results, we assume

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim p_{mix} \times N \left(\begin{pmatrix} -0.8 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.4 & 0.5 \\ 0.5 & 1.4 \end{pmatrix} \right) + (1 - p_{mix}) \times N \left(\begin{pmatrix} -0.8 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.6 & -0.3 \\ -0.3 & 0.6 \end{pmatrix} \right)$$

and the coefficients in the choice equation are $\gamma_1 = 0.2, \gamma_2 = 1$. In the left plot of Panel A we let $p_{mix} = 0.45$ and in the right plot $p_{mix} = 0.75$. Panel B plots the weights of IV estimates constructed using either Z or Z^2 as an instrument where $Z \sim N(\mu_Z, 1)$, $\mu_Z = 1$ or $\mu_Z = -0.5$, and $D = \mathbf{1}(Z > V)$. Panel C plots the weights of IV estimates constructed using either $P(Z)$ below the median or $P(Z)$ above the median as instruments. Panel D plots the weights of IV estimates constructed using $P(Z)$ in different quartiles of its distribution as instruments. In Panels C and D, $Z \sim N(\mu_Z, 1)$, $\mu_Z = 0$ or $\mu_Z = 1$, and $D = \mathbf{1}(Z > V)$. In all of the plots, we set $\sigma_V^2 = 1$.

and test

$$\begin{aligned}
 H_0^{IV} & : \beta_{IV, [\underline{p}_1, \bar{p}_1]} = \beta_{IV, [\underline{p}_2, \bar{p}_2]} \\
 H_A^{IV} & : \beta_{IV, [\underline{p}_1, \bar{p}_1]} \neq \beta_{IV, [\underline{p}_2, \bar{p}_2]}.
 \end{aligned}$$

There is no *a priori* guidance on which intervals to use so we consider two ways to construct intervals over which to form IV estimates: (1) use the intervals $[0, p_{med})$ and $[p_{med}, 1]$ where p_{med} is the sample median of $P(Z)$, and (2) use the intervals $[0, p_{q1})$, $[p_{q1}, p_{q2})$, $[p_{q2}, p_{q3})$ and $[p_{q3}, 1]$, where p_{qj} is the j th sample quartile of the distribution of $P(Z)$ and form all pairwise contrasts between these estimates. Note that even though we split the propensity score into four intervals, we are still conducting pairwise tests. However, because there is a multiplicity of pairwise tests, we must control the size of the test. We do this by using the stepdown procedure of Romano and Wolf (2005). Figures 4C and 4D plot the weights for the instruments constructed in this manner. These weights are nonoverlapping by construction and will also depend on the distribution of the instrument Z .

The power of the test of H_0 based on IV estimators also depends on the variance (12), which determines the denominator of the noncentrality parameter. The important terms which are affected by the choice of instruments are the variance of the difference in the instruments $\left[\frac{\text{Var}(J_1)}{\omega_1^2} - \frac{2 \text{Cov}(J_1, J_2)}{\omega_1 \omega_2} + \frac{\text{Var}(J_2)}{\omega_2^2} \right]$ and the variance weight $h_{\Omega, J_1, J_2}(\cdot)$. The variance of the difference in the instruments is identified from the distribution of Z given X . The weights $h_{\Omega, J_1, J_2}(\cdot)$, can also be estimated from the data but are less transparent. For each of the examples presented in Figure 4, we plot the variance weights $h_{\Omega, J_1, J_2}(\cdot)$. In the case of the normal generalized Roy model, the weights are more intuitive and more easily calculated when conditioning directly on $V = v$ (rather than $U_D = u_D$), so we plot them as a function of v . Figure 5 plots the variance weights. *Ceteris paribus*, the larger the variance weights, the larger is the variance of the difference in the IV estimators and hence the lower the power of a test based on this difference. In Panel A of Figure 5 we see that when the

mixing proportion is 0.45 the variance of the difference in the estimators is higher than when the mixing proportion is 0.75 due to the fact that the IV weights covary highly when the instruments are closer to normal so the variance of their difference is smaller. In Panel B, the variance weights are roughly similar for $E(Z) = 1$ and $E(Z) = -0.5$. Finally, in Panel C the variance weights are much larger when $E(Z) = 1$ than when $E(Z) = 0$. This demonstrates that even when the IV weights are nonoverlapping, as is the case in both examples in Panel C, the variance of the difference in the IV estimators will generally depend on the distribution of Z .

We emphasize that the specific comparisons of IV estimators presented in this section are illustrative examples. Our formal analysis is completely general and allows for any choice of valid instruments which satisfy (A-1)–(A-5).

5 Testing H_0 by Testing for Linearity

We next consider tests of H_0 based on linearity in p . Keeping the conditioning on X implicit, we can write (3) as

$$E(Y \mid P(Z) = p) = \mu + g(p) \tag{13}$$

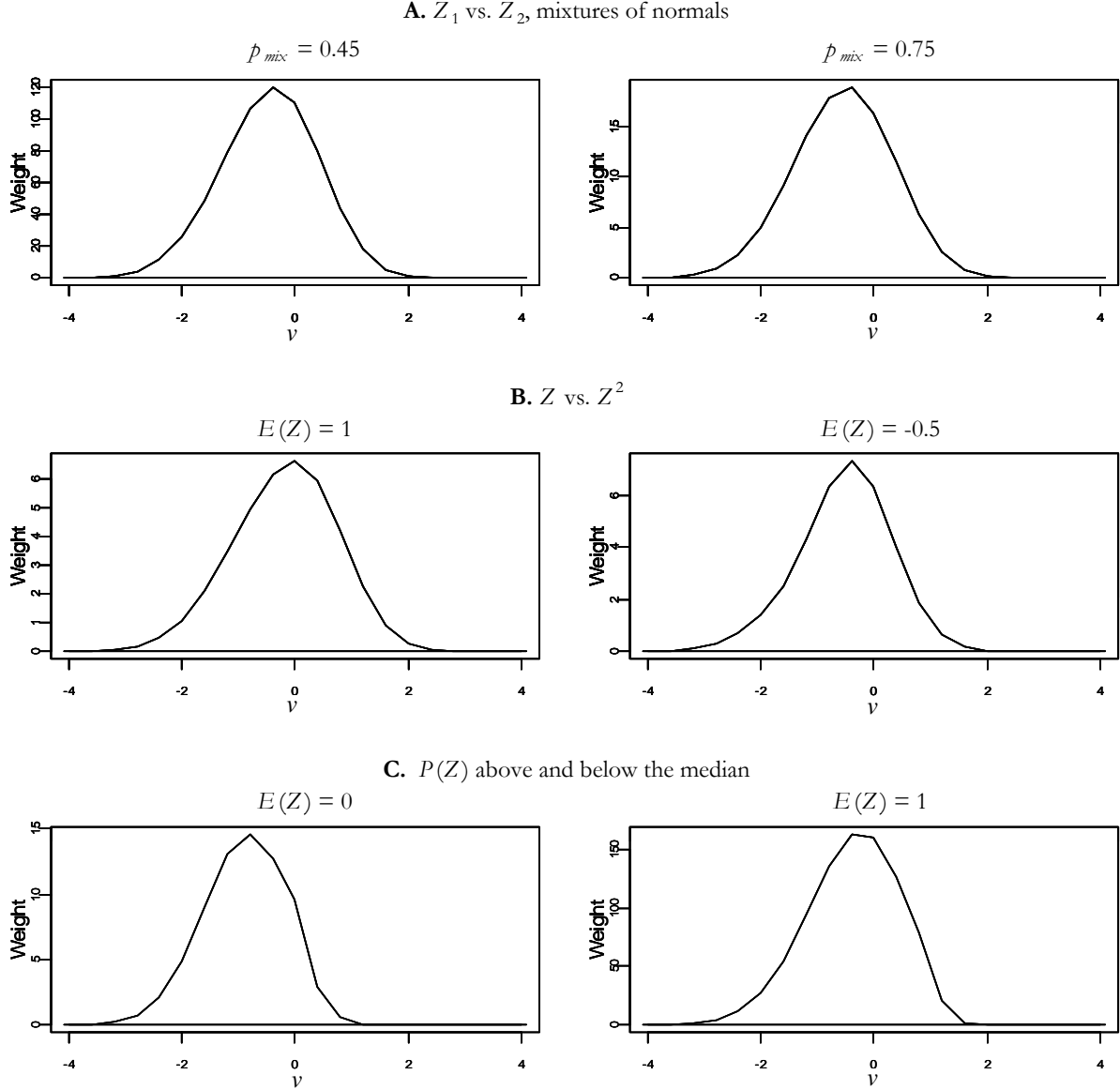
for some general nonlinear function $g(\cdot)$ where μ and g may depend on X . Our test for the absence of selection on the gain to treatment is a test of whether the function $g(\cdot)$ belongs to the linear parametric family $\mathcal{F} = \{a + bp, (a, b) \in \mathbb{R}^2\}$. Let \mathcal{P} be the support of $P(Z)$, with typical element $p \in \mathcal{P}$. The null hypothesis of linearity can be written as

$$H_0^L : \text{There exists some } (a, b) \in \mathbb{R}^2 \text{ such that } g(p) = a + bp \text{ for almost all } p \in \mathcal{P},$$

while the alternative is

$$H_A^L : \text{There exists no } (a, b) \in \mathbb{R}^2 \text{ such that } g(p) = a + bp \text{ for almost all } p \in \mathcal{P}.$$

Figure 5: IV variance weights ($h_{\Omega_{J_1, J_2}}(\cdot)$) as a function of $V = v$ for alternative choices of instruments.



Note: Panel A plots the variance weights of the difference in the IV estimates constructed using either Z_1 or Z_2 as an instrument where (Z_1, Z_2) is distributed as a multivariate mixture of normals, with $D = \mathbf{1}(\gamma_1 Z_1 + \gamma_2 Z_2 > V)$. To construct these results, we assume

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim p_{mix} \times N \left(\begin{pmatrix} -0.8 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.4 & 0.5 \\ 0.5 & 1.4 \end{pmatrix} \right) + (1 - p_{mix}) \times N \left(\begin{pmatrix} -0.8 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.6 & -0.3 \\ -0.3 & 0.6 \end{pmatrix} \right)$$

and the coefficients in the choice equation are $\gamma_1 = 0.2, \gamma_2 = 1$. In the left plot of Panel A we let $p_{mix} = 0.45$ and in the right plot $p_{mix} = 0.75$. Panel B plots the variance weights of the difference in the IV estimates constructed using either Z or Z^2 as an instrument where $Z \sim N(\mu_Z, 1)$, $\mu_Z = 1$ or $\mu_Z = -0.5$, and $D = \mathbf{1}(Z > V)$. Panel C plots the variance weights of the difference in the IV estimates constructed using either $P(Z)$ below the median or $P(Z)$ above the median as instruments. In Panel C, $Z \sim N(\mu_Z, 1)$, $\mu_Z = 0$ or $\mu_Z = 1$, and $D = \mathbf{1}(Z > V)$. In all of the plots, we set $\sigma_V^2 = 1$.

There is a large and still unsettled literature in econometrics and statistics dealing with specification tests of this type.²⁴ These tests proceed in one of two ways: (i) testing orthogonality restrictions implied by the parametric model, or (ii) comparing a nonparametric estimate of $g(p)$ with a parametric estimate, $\hat{a} + \hat{b}p$. We discuss both types of tests. Heckman, Schmieder, and Urzua (2010) present Monte Carlo analyses of the power of these tests. We briefly discuss a third test due to Li and Nie (2007).

Linearity Test 1: Wald Test Based on Series

The first test of linearity of $E(Y|P(Z) = p)$ in p determines whether terms in addition to p are required to fit the data. It is instructive to consider the case of the normal selection model as a baseline. When the data are generated from the normal generalized Roy model, we can characterize $E(Y|P(Z) = p)$ by

$$E(Y|P(Z) = p) = \bar{\alpha} + \bar{\beta}p + \tau \int_0^p \Phi^{-1}(u_D) du_D.$$

Heckman, Schmieder, and Urzua (2010) examine this case in depth.

In the general non-normal case, polynomials can approximate classes of smooth alternatives for the function $g(\cdot)$. One can estimate $E(Y|P(Z) = p)$ using polynomials of degree 2 or higher. Polynomials approximate well a broad class of functions. Exploring power in this class gives us an indication of the power of our procedures against such alternatives.²⁵ One specification of a more general, but still parametric, alternative model is

$$g(P(Z)) = \sum_{l=0}^L \phi_l(P(Z))^l,$$

²⁴See, e.g., Horowitz and Spokoiny (2001) and the references therein. The properties of particular tests depend on the specification of alternatives.

²⁵Ichimura and Todd (2007) discuss the properties of series estimators. Newey (1997) establishes convergence rates and proves asymptotic normality of such estimators.

where L is assumed to be known.²⁶ The proposed test for linearity is

$$H_0 : \phi_l = 0 \text{ for } l = 2, \dots, L$$

$$H_A : \phi_l \neq 0 \text{ for some (or all) } l = 2, \dots, L.$$

Heckman, Schmierer, and Urzua (2010) develop properties of this test for linearity.

Linearity Test 2: Bierens Conditional Moment Test

One can also test the validity of representation (3) using orthogonality restrictions implied by the parametric model. One approach is the conditional moment (CM) test of Bierens (1990).²⁷ This test uses the fact that under the null hypothesis the following moment condition must be satisfied

$$E[Y - a_0 - b_0P(Z) \mid P(Z)] = 0$$

for the true parameter vector $(a_0, b_0) \in \mathbb{R}^2$. This conditional moment restriction implies the set of unconditional moment restrictions

$$E[(Y - a_0 - b_0P(Z)) \exp(t' \Lambda(P(Z)))] = 0 \tag{14}$$

for all $t \in \mathbb{R}$, for some bounded one-to-one, mapping Λ from \mathbb{R} into \mathbb{R} . A test can be constructed using the sample analog of the left-hand side of (14). Bierens (1990) shows how one can use sample analogs to construct a test statistic which, under the null hypothesis, converges in distribution to a χ_1^2 and under the alternative diverges to infinity. Heckman, Schmierer, and Urzua (2010) discuss the properties of this test.

²⁶Below, we discuss a procedure when L is unknown.

²⁷See also Bierens (1982) and Bierens and Ploberger (1997) for related tests. Newey (1985) discusses conditional moment tests more generally.

The Preceding Tests are Conditional Moment Tests²⁸

All of the tests discussed so far test if

$$E(Y | P(Z)) = a + bP(Z)$$

which is equivalent to

$$E [h(P(Z)) [Y - a - bP(Z)]] = 0$$

for any function h . Conditional moment tests would typically use a vector of functions $h(P(z))$ to construct tests.

The tests previously discussed use different choices of $h(P)$. For the Bierens test, one would use $h(P(Z)) = \exp(t'\Lambda(P(Z)))$. The test of linearity based on polynomials takes $h(P(Z)) = [1, P(Z), (P(Z))^2, \dots, (P(Z))^L]$. The IV test can also be cast in this framework, as the following argument shows.

The plim of the IV estimator obtained using $J_k(Z)$, $k = 1, \dots, K$, as an instrument are the values of (a_k, b_k) that solve

$$E [J_k(Z) [Y - a_k - b_k D]] = 0$$

and

$$E [Y - a_k - b_k D] = 0, \quad k = 1, \dots, K.$$

By the law of iterated expectations, this is equivalent to solving

$$E [J_k(Z) [E(Y | Z) - a_k - b_k P(Z)]] = 0$$

$$E [E(Y | Z) - a_k - b_k P(Z)] = 0,$$

²⁸We thank Edward Vytlačil for suggesting this unifying approach.

which is equivalent to solving

$$E [J_k(Z) [Y - a_k - b_k P(Z)]] = 0$$

$$E [Y - a_k - b_k P(Z)] = 0.$$

For one instrument there is no test, but for two or more ($K \geq 2$), one can test if a common pair of (a, b) satisfies all of the moment conditions produced from using different instrumental variables. This is the classical test of overidentification. Thus, all of the tests previously discussed can be viewed as conditional moment tests.

Linearity Test 3: A Semiparametric Test Based on Local Linear Regression²⁹

A potential problem with the test based on series estimators (Linearity Test 1) is that it assumes that the degree of the highest order polynomial in $P(Z)$ is finite and known. A semiparametric approach that did not rely on strong functional form assumptions about the generator model is more desirable.

Li and Nie (2007) use local linear regression methods to develop a test for linearity of an unknown parametric function in a semiparametric model. They develop a test of linearity of the unknown nonparametric component (linearity in $P(Z)$ in our setup) that can be applied to the problem analyzed in this paper if it is adapted to the case of an estimated $P(Z)$. If $P(Z)$ is parametric and its coefficients are \sqrt{N} estimable, their analysis can be applied directly. The case where $P(Z)$ is estimated nonparametrically is left for another occasion.

Li and Nie (2007) conduct a Monte Carlo study of their approach. They show good size and power properties for their test statistic. Their test can be interpreted as a local conditional moment test.

²⁹We thank Xiaohong Chen for directing us to this paper and clarifying our thinking about semiparametric approaches to testing for linearity.

Conditioning on X

Throughout, we have conditioned on X . An important practical problem not addressed in this paper but common to all empirical models is picking the appropriate conditioning set, and determining how to explicitly model the dependence of Y on X . Heckman, Schmieder, and Urzua (2010) discuss the power of these tests and conduct extensive Monte Carlo studies.

6 Summary and Conclusion

P.A.V.B Swamy's classic work (1971, 1974) developed estimators for the uncorrelated random coefficient model. This is the case when H_0 is true. In this paper, building on the work of Heckman and Vytlačil (1999, 2005, 2007a,b) and Heckman, Schmieder, and Urzua (2010), we develop tests for the presence of a correlated random coefficient model and related tests on parameters derived from the model. All of the tests we consider can be interpreted as conditional moment tests. We develop instrumental variable tests for the null hypothesis of the absence of a correlated random coefficient model. To implement it, we develop the sampling distribution of the IV estimator using the marginal treatment effect and its extensions to higher moments of the distribution of the heterogeneity on which agents select.

Heckman, Schmieder, and Urzua (2010) conduct a Monte Carlo investigation of the power of these tests. One disturbing finding from their work is that the power of all of the tests we consider is low. They show that among all of the tests considered, the test based on comparing alternative IV estimators above and below the median propensity score has the highest power.

This paper analyzes the case of a binary treatment. Heckman, Urzua, and Vytlačil (2006) and Heckman and Vytlačil (2007b) analyze the cases of a multiple treatment model generated by an ordered choice model with stochastic thresholds and a multiple treatment model generated by an unordered choice model. In all of these cases, IV produces an instrument-dependent parameter so the IV test for selection on unobserved gains based on comparing

the estimands of two different IVs developed in this paper carries over in general to these settings. A test of linearity of the conditional expectation of Y given P is developed for the outcome model for multiple treatments generated by the ordered choice model in Heckman, Urzua, and Vytlačil (2006). It also applies to the unordered multiple choice model that identifies the treatment effect of a gain compared to the next best option which Heckman, Urzua and Vytlačil show is a direct extension of the binary model.

A The Variance of Linear IV in the Correlated Random Coefficient Model

The IV estimator, using instrument $J(Z)$, is

$$\widehat{\beta}_{IV,J} = \frac{\sum Y_i \tilde{J}_i}{\sum D_i \tilde{J}_i}$$

and hence

$$\sqrt{I} \widehat{\beta}_{IV,J} = \frac{\frac{1}{\sqrt{I}} \sum (J_i - \bar{J})(\alpha_i + \beta_i D_i)}{\frac{1}{I} \sum D_i \tilde{J}_i}.$$

Invoking standard central limit theorems,

$$\sqrt{I} \left(\widehat{\beta}_{IV,J} - \beta_{IV,J} \right) \xrightarrow{d} N(0, \Omega_J).$$

Defining $J^* = J - E(J)$, where Ω_J is given by

$$\begin{aligned} \Omega_J &= E \left\{ \left[\frac{YJ^*}{\omega_J} - E \left(\frac{YJ^*}{\omega_J} \right) \right]^2 \right\} \\ &= E \left[\frac{Y^2(J^*)^2}{\omega_J^2} \right] - \left(E \left[\frac{YJ^*}{\omega_J} \right] \right)^2 \\ &= E \left[\frac{Y^2(J^*)^2}{\omega_J^2} \right] - \left(\int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \right)^2 \\ &= \frac{1}{\omega_J^2} E [(\alpha + \beta D)^2 (J^*)^2] - \left(\int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \right)^2 \\ &= \frac{1}{\omega_J^2} E [\alpha^2 (J^*)^2] + \frac{2}{\omega_J^2} E [\alpha \beta D (J^*)^2] + \frac{1}{\omega_J^2} E [\beta^2 D (J^*)^2] \\ &\quad - \left(\int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \right)^2. \end{aligned}$$

Using the law of iterated expectations as well as the assumption that α is independent of Z , this expression can be written as

$$\Omega_J = E[\alpha^2] \frac{\text{Var}(J)}{\omega_J^2} + \frac{1}{\omega_J^2} 2E[\alpha\beta(J^*)^2 | D = 1] \Pr(D = 1) + \frac{1}{\omega_J^2} E[\beta^2(J^*)^2 | D = 1] \Pr(D = 1) - \left(\int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \right)^2$$

where

$$h_J(u_D) = \frac{E[J^* | P(Z) > u_D] \Pr(P(Z) > u_D)}{\omega_J}.$$

Under the conditions of Fubini's Theorem, we can exchange the order of integration and write

$$\begin{aligned} \Omega_J &= E[\alpha^2] \frac{\text{Var}(J^*)}{\omega_J^2} \\ &+ \int_0^1 [2E(\alpha\beta | U_D = u_D) + E(\beta^2 | U_D = u_D)] \frac{E((J^*)^2 | P(Z) > u_D) \Pr(P(Z) > u_D)}{\omega_J^2} du_D \\ &- \left(\int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \right)^2 \\ &= E[\alpha^2] \frac{\text{Var}(J)}{\omega_J^2} + \int_0^1 [2E(\alpha\beta | U_D = u_D) + E(\beta^2 | U_D = u_D)] h_{\Omega_J}(u_D) du_D \\ &- \left(\int_0^1 E(\beta | U_D = u_D) h_J(u_D) du_D \right)^2 \end{aligned}$$

where

$$\begin{aligned} h_{\Omega_J}(u_D) &= \frac{1}{\omega_J^2} \int_{-\infty}^{\infty} (j - E(J))^2 \int_{u_D}^1 f_{P,J}(P(z), j) dP(z) dj \\ &= \frac{E[(J - E(J))^2 | P(Z) > u_D] \Pr(P(Z) > u_D)}{\omega_J^2} \end{aligned}$$

which is the expression in the text.

B Proof of Invariance of the IV Estimand to the Choice of a Linear Instrument under Normality with a Linear Index Choice Equation

Suppose that the choice equation has a linear index structure, so that

$$D = \mathbf{1}(Z\gamma > V)$$

where $Z \sim N(\bar{Z}, \Sigma_Z)$, an L -dimensional multivariate normal random variable, γ an $L \times 1$ vector and $V \sim N(0, \sigma_V^2)$. Consider the instrument $J(Z)$, which is a linear function of Z , say $Z'\eta$. In this case, the IV estimand (written as a weighted average over the support of V) is

$$\beta_{IV,J} = \int_{-\infty}^{\infty} MTE(v) h_J(v) \phi\left(\frac{v}{\sigma_V}\right) dv$$

where $\phi(\cdot)$ is a standard normal pdf and the IV weight is

$$h_J(v) = \frac{E[J(Z) - E(J(Z)) | Z\gamma > v] \Pr(Z\gamma > v)}{\text{Cov}(J(Z), D)}.$$

Under the assumption of multivariate normality for the instruments,

$$h_J(v) = \frac{\frac{\text{Cov}(J(Z), Z\gamma)}{\sqrt{\text{Var}(Z\gamma)}} \phi\left(\frac{v - \bar{Z}\gamma}{\sqrt{\text{Var}(Z\gamma)}}\right)}{\frac{\text{Cov}(J(Z), Z\gamma - V)}{\sqrt{\text{Var}(Z\gamma - V)}} \phi\left(\frac{-\bar{Z}\gamma}{\sqrt{\text{Var}(Z\gamma - V)}}\right)}.$$

Under assumption (A-1) in Section 2, $\text{Cov}(J(Z), Z\gamma) = \text{Cov}(J(Z), Z\gamma - V)$, and we obtain

$$h_J(v) = \frac{\frac{1}{\sqrt{\text{Var}(Z\gamma)}} \phi\left(\frac{v - \bar{Z}\gamma}{\sqrt{\text{Var}(Z\gamma)}}\right)}{\frac{1}{\sqrt{\text{Var}(Z\gamma - V)}} \phi\left(\frac{-\bar{Z}\gamma}{\sqrt{\text{Var}(Z\gamma - V)}}\right)}.$$

That is, the IV weights, and hence the IV estimand, are the *same* for all $J(Z) = Z'\eta$ for any finite η .

Acknowledgments

This research was supported by NIH R01-HD043411, NSF SES-024158, the American Bar Foundation and the Geary Institute, University College Dublin, Ireland. The views expressed in this paper are those of the authors and not necessarily those of the funders listed here. We have received helpful comments from Pedro Carneiro, Jeremy Fox, Joel Horowitz, Benjamin Moll, Azeem Shaikh, Christopher Taber, Edward Vytlačil, and Sergio Urzua.

References

- Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 5145–5303. Amsterdam: Elsevier.
- Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71(6), 1795–1843.
- Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics* 20(1), 105–134.
- Bierens, H. J. (1990, November). A consistent conditional moment test of functional form. *Econometrica* 58(6), 1443–1458.
- Bierens, H. J. and W. Ploberger (1997, September). Asymptotic theory of integrated conditional moment tests. *Econometrica* 65(5), 1129–1151.
- Björklund, A. and R. Moffitt (1987, February). The estimation of wage gains and welfare gains in self-selection. *Review of Economics and Statistics* 69(1), 42–49.
- Griliches, Z. (1977, January). Estimating the returns to schooling: Some econometric problems. *Econometrica* 45(1), 1–22.
- Heckman, J. J. (2001, August). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy* 109(4), 673–748.
- Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions. In J. Heckman and B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, Volume 10, pp. 156–245. New York: Cambridge University Press. Reprinted in 2008, Econometric Society Monograph 10.

- Heckman, J. J., D. Schmieder, and S. Urzua (2010). Testing the correlated random coefficient model. *Journal of Econometrics* 158(2), 177–203.
- Heckman, J. J., S. Urzua, and E. J. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88(3), 389–432.
- Heckman, J. J. and E. J. Vytlacil (1998, Fall). 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(4), 974–987.
- Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* 96(8), 4730–4734.
- Heckman, J. J. and E. J. Vytlacil (2001). Local instrumental variables. In C. Hsiao, K. Morimune, and J. L. Powell (Eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, pp. 1–46. New York: Cambridge University Press.
- Heckman, J. J. and E. J. Vytlacil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73(3), 669–738.
- Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4779–4874. Amsterdam: Elsevier.
- Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4875–5144. Amsterdam: Elsevier.

- Horowitz, J. L. and V. G. Spokoiny (2001, May). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69(3), 599–631.
- Ichimura, H. and P. E. Todd (2007). Implementing nonparametric and semiparametric estimators. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 5369–5468. Amsterdam: Elsevier.
- Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Li, R. and L. Nie (2007, November). Efficient statistical inference procedures for partially nonlinear models and their applications. *Biometrics* 64(3), 904–911.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.
- McFadden, D. (1981). Econometric models of probabilistic choice. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 198–272. Cambridge, MA: MIT Press.
- Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: Columbia University Press for National Bureau of Economic Research.
- Newey, W. K. (1985, September). Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53(5), 1047–1070.
- Newey, W. K. (1997, July). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79(1), 147–168.
- Quandt, R. E. (1958, December). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* 53(284), 873–880.

- Romano, J. P. and M. Wolf (2005, March). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association* 100(469), 94–108.
- Swamy, P. (1971). *Statistical Inference in Random Coefficient Regression Models*. New York: Springer-Verlag.
- Swamy, P. (1974). Linear models with random coefficients. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 143–168. New York: Academic Press.
- Thurstone, L. L. (1927, July). A law of comparative judgement. *Psychological Review* 34(4), 273–286.
- Vytlacil, E. J. (2002, January). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica* 70(1), 331–341.
- Yitzhaki, S. (1989). On using linear regression in welfare economics. Working Paper 217, Department of Economics, Hebrew University.