Econometric Theory, **33**, 2017, 69–104.
doi:10.1017/S0266466615000390

INSTRUMENTAL VARIABLES METHODS WITH HETEROGENEITY AND MISMEASURED INSTRUMENTS

KARIM CHALAK
University of Virginia

We study the consequences of substituting an error-laden proxy W for an instrument Z on the interpretation of Wald, local instrumental variable (LIV), and instrumental variable (IV) estimands in an ordered discrete choice structural system with heterogeneity. A proxy W need only satisfy an exclusion restriction and that the treatment and outcome are mean independent from W given Z. Unlike Z, W need not satisfy monotonicity and may, under particular specifications, fail exogeneity. For example, W could code Z with error, with missing observations, or coarsely. We show that Wald, LIV, and IV estimands using W identify weighted averages of local or marginal treatment effects (LATEs or MTEs). We study a necessary and sufficient condition for nonnegative weights. Further, we study a condition under which the Wald or LIV estimand using W identifies the same LATE or MTE that would have been recovered had Z been observed. For example, this holds for binary Z and therefore the Wald estimand using W identifies the same "average causal response," or LATE for binary treatment, that would have been recovered using Z. Also, under this condition, LIV using W can be used to identify MTE and average treatment effects for e.g., the population, treated, and untreated.

1. INTRODUCTION

Data on useful instrumental variables are often error-laden or unavailable. As a result, researchers sometimes employ "proxies" for instruments. For example, Acemoglu, Johnson, and Robinson (2001) use early settlers' mortality rate as an instrument in their study of the effects of institutions on income per capita. To proceed, Acemoglu et al. (2001, pp. 1382–1383) combine data on mortality rates for soldiers, bishops, and sailors from $17^{th} - 19^{th}$ century as "reasonable estimates for settler mortality" and state that "combining data from a variety of

An earlier version of this paper was circulated under the title "Identification of Local Treatment Effects Using a Proxy for an Instrument." I thank the co-editor and three referees for helpful feedback. I also thank the participants in the California Econometrics Conference 2010, the 2011 North American Winter Meeting of the Econometric Society, the Boston College labor lunch seminar, and the econometrics seminars at UCSD, Brown, UC Riverside, Ohio State University, Yale, UCL, Oxford, UBC, Northwestern, Harvard/MIT, UC Davis, Stanford, and the Federal Reserve Bank of Cleveland as well as Gary Chamberlain, Xavier d'Haultfoeuille, Scott Fulford, Peter Gottschalk, Guido Imbens, Stefan Hoderlein, Ivana Komunjer, Tobias Klein, Arthur Lewbel, Ke-Li Xu, Mathis Wagner, and Halbert White for helpful comments and suggestions. I thank Mehmet Onur Ezer for excellent research assistance. Any errors are the author's responsibility. Address correspondence to Karim Calak, Department of Economics, University of Virginia, P.O. Box 400182, Charlottesville, VA 22904-4182, email: chalak@virginia.edu.

© Cambridge University Press 2016

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

sources will introduce measurement error in our estimates of settler mortality." As another example, in studying the returns to education, family background variables such as parental education are often used as instruments (see e.g., references in Card, 1999, pp. 1843–1844). However, reported parental schooling levels are often error-laden (see e.g., Ashenfelter and Krueger, 1994, p. 1161 and the discussion in Card, 1999, p. 1826). As a third example, Hotz, McElroy, and Sanders (2005), use miscarriage as an instrumental variable in estimating the effect of teenage childbearing on socioeconomic attainment of teen mothers. Hotz, Mullin, and Sanders (1997) refer to miscarriage as a "contaminated instrument" because the group of women who miscarry may be an unknown "mixture of women who experience random and non-random miscarriages." Nevertheless, building on the results in Horowitz and Manski (1995), they provide bounds, under restrictions on the extent of contamination, on effects of teenage childbearing that are consistent with the instrumental variable estimates in Hotz et al. (2005).

In the linear constant effect case, an instrument or an error-laden proxy for it (e.g., with error at random) can be used to identify the constant causal effect. What are the implications of substituting a proxy for an instrument on the interpretation of the Wald (1940), local instrumental variable (LIV), and instrumental variable (IV) (e.g., two stage least squares; TSLS) estimands in systems with heterogeneous effects? This paper studies this question in the context of an ordered discrete choice structural system with heterogeneity thereby generalizing the identification results from the homogenous constant effect case and clarifying the scope of the methods of Wald, LIV, and IV using mismeasured instruments. Specifically, this paper gives constructive conditions under which these methods using a proxy identify the local or marginal treatment effects (LATEs or MTEs), as well as average treatment effects e.g., for the population (ATE), treated (ATT), and untreated (ATU), that would have been recovered had the instrument been observed. Moreover, this paper gives weaker conditions under which these methods identify weighted averages of these LATEs or MTEs with nonnegative weights.

To study the Wald and IV estimands using mismeasured instruments, we begin by providing notions for an instrument vector Z and a proxy vector W for it. In this paper, an instrument Z satisfies the assumptions imposed in Imbens and Angrist (1994; thereafter IA), Angrist and Imbens (1995; thereafter AI), Angrist, Imbens, and Rubin (1996, thereafter AIR), Heckman and Vytlacil (2005; thereafter HV), Heckman, Urzua, and Vytlacil (2006; thereafter HUV), and Heckman (2010) for example. As these papers show, in systems with heterogeneous effects, an instrument must satisfy structural and exogeneity assumptions stronger than those required in the linear constant effect case for the methods of Wald and IV to ensure full (point) identification of LATEs, MTEs, or a weighted average of LATEs or MTEs (see also Manski and Pepper, 2000). Following the literature (see e.g., definition 3 in AIR), we assume that Z (1) is exogenous, (2) satisfies "monotonicity," i.e., it induces receipt of treatment D in the same direction across individuals, and (3) drives the outcome Y only via the treatment, an exclusion restriction.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

A proxy W for Z is a vector associated with Z, and possibly driven by it, that need only satisfy an exclusion restriction and that D and Y are mean independent of W conditional on Z. Examples of possible proxies include an error-laden measurement of Z, a measurement of Z with missing or imputed observations, or a coarse (e.g., binary) coding for a multivalued Z. Importantly, a proxy W can fail the assumptions imposed on Z and thus need not be an instrument (e.g., according to definition 3 in AIR). In particular, W need not satisfy "monotonicity" nor be a cause of the treatment. Further, W can fail mean ignorability from potential outcomes and it can thus be endogenous, albeit this requires particular specifications.

Using this notion of an instrument's proxy, we build on the work of IA, AI, AIR, HV, and HUV studying the methods of Wald, LIV, and IV using instruments and extend their analysis to accommodate mismeasured instruments. In particular, we show that the methods of Wald and LIV using a proxy W for an instrument Z as well as the method of IV using a function $g(W)$ of W identify weighted averages of LATEs or MTEs. This is particularly troublesome if some of the weights are negative since, unlike in the linear constant effect case, the sign of the identified weighted average of LATEs or MTEs may then be the opposite of the sign of some, or even all, of the underlying LATEs or MTEs. For the methods of Wald and LIV, we give conditions under which the weights are nonnegative if and only if W is *intensity preserving* for Z, a notion of dependence of Z on W that we study. Similarly, for the method of IV (e.g., TSLS), we give conditions under which the weaker property of *mean quadrant dependence* of $g(W)$ on Z is necessary and sufficient for nonnegative weights. For the methods of Wald, LIV, and IV, we study the conditions for nonnegative weights and show, for example, that these can hold in the presence of misclassification or classical measurement error in coding Z. Further, we show that when W is *value preserving* for Z, a restriction on the dependence of Z on W stronger than intensity preservation, the method of Wald or LIV using W identifies LATEs or MTEs that would have been recovered had Z been observed. We study this condition and show that it can hold in the presence of misclassification or measurement error in coding Z. In this case, substituting a proxy for an instrument in the Wald or LIV estimand can be harmless for the identification of LATE or MTE as well as weighted averages of MTE such as ATE, ATT, and ATU. When multiple proxies are available, this provides the foundation for tests for proxy validity.

The case of binary Z receives significant attention in the literature. For example, in the "compliance problem" studied in AIR, the treatment D and the treatment assignment Z are binary, and LATE is the average effect of the treatment for the subpopulation of "compliers" who always comply with the treatment assignment. We show that for binary Z, value preservation trivially holds, and thus the methods of Wald and IV using any proxy for Z identify the same "average causal response" discussed in AI, or LATE if D is binary, that would have been recovered using Z. Thus, in this case, identification of LATE via the

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

methods of Wald or IV is robust to certain types of instrument misclassification and missing data.

On the one hand, this paper's results suggest that researchers should cautiously interpret estimates of causal effects obtained via the methods of Wald, LIV, or IV (e.g., TSLS) using a proxy for an instrument. On the other hand, if economic theory suggests an unobserved instrument Z, a researcher may make use of a suitable proxy for it, that need not satisfy the instrument assumptions, to identify the LATEs or MTEs that would have been identified had Z been observed, as well as average treatment effects such as ATE, ATT, and ATU. Alternatively, under weaker assumptions, a researcher may employ a proxy to recover an informative weighted average of LATEs or MTEs. For example, a nonzero weighted average of local effects provides evidence against the hypothesis that the treatment effect is zero. Further, the sign of this weighted average is informative when the local or marginal effects have common sign and the weights are nonnegative.

This paper is organized as follows. Section 2 states the data generation assumptions. Section 3 studies the methods of Wald and LIV using a proxy for an instrument. Section 4 studies the IV method using a proxy. Section 5 gives examples and reports Monte Carlo simulations and Section 6 concludes. Mathematical proofs are gathered in Appendix B.

2. DATA GENERATION

This section states the main assumptions imposed in this paper. Following the results in Vytlacil (2002, 2006), we work simultaneously with a structural system and the Rubin Causal Model (Holland, 1986; AIR). In particular, the first assumption defines the ordered discrete choice structural system we are concerned with in this paper.

Assumption 1 (A.1). Data Generating Structural System: (i) Let Z, W, D, and Y be random vectors or variables taking value respectively in $S_Z \subseteq \mathbb{R}^\kappa$, $S_W \subseteq \mathbb{R}^\ell$, $S_D \equiv \{0, 1, ..., J\}$ with J finite, and $S_Y \subseteq \mathbb{R}$. (ii) Let a triangular structural system $\mathcal{S}$ generate the random vectors U_Y of countable dimension and $U_D \equiv (V_1, ..., V_J)$ such that $-\infty \equiv V_0 \leq V_1 \leq V_2 \leq \cdots \leq V_J \leq V_{J+1} \equiv \infty$ with the distribution of V_j, $j = 1, ..., J$, absolutely continuous. Let $\mathcal{S}$ generate the instrument Z, proxy W, treatment D, and outcome Y such that

$$Y = r(D, U_Y),$$

$$D = q(Z, U_D) = \sum_{j=0}^{J} j \mathbf{1}\{V_j \leq \nu(Z) < V_{j+1}\},$$

where ν, q, and r are unknown measurable functions[1] mapping respectively to $S_\nu \subseteq \mathbb{R}$, S_D, and S_Y. Let $E(|r(j, U_Y)|) < \infty$ for all $j \in S_D$. Realizations of W, D, and Y are observed, those of U_D and U_Y are not, and realizations of Z may be unobserved.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

A.1(ii) introduces the random variables. A.1(ii) imposes structure on the data generation. In particular, A.1(ii) imposes exclusion restrictions ensuring that the treatment D is structurally driven by Z but is not otherwise directly affected by the proxy W. Similarly, the outcome Y is driven by D but is not otherwise directly affected by Z or W. Given these restrictions, we define the potential treatment $D(z)$, $z \in S_Z$, and potential outcome $Y(d)$, $d \in S_D$, by

$$D(z) \equiv q(z, U_D) \quad \text{and} \quad Y(d) \equiv r(d, U_Y).$$

The random vectors U_D and U_Y reflect unobserved heterogeneity in the population (e.g., across units or individuals). The assumption $E(|Y(j)|) < \infty$ for all $j \in S_D$ ensures that average effects are meaningful and enables applying integration theorems.

Further, A.1(ii) restricts q by imposing a threshold crossing form. To illustrate, consider the case of binary treatment D (e.g., IA, AIR, HV) such as in the generalized Roy model (HV; Heckman, 2010). In this case, $J = 1$ and q reduces to

$$D = q(Z, U_D) = \mathbf{1}\{V_1 \leq \nu(Z)\} = \mathbf{1}\{F_{V_1}(V_1) \leq F_{V_1}[\nu(Z)]\} \equiv \mathbf{1}\{U \leq P(Z)\},$$

where[2] $U \backsim Unif[0, 1]$ and $P(Z) \equiv \Pr(V_1 \leq \nu(Z))$. The assumption that V_1 is absolutely continuous simplifies the exposition and enables employing the probability transform above. For binary D, we often find it convenient to employ the representation $D = \mathbf{1}\{U \leq P\}$ with scalar instrument $P \equiv P(Z)$.

More generally, we study the case in which an individual chooses "treatment" j among the ordered alternatives $0, 1, ..., J$ if $\upsilon_j \leq \nu(z) < \upsilon_{j+1}$ realizes. This threshold crossing restriction implies the *monotonicity* assumption imposed in e.g., IA and AI:

$$\text{For all } z, z' \in S_Z \text{ either } \Pr[D(z) \leq D(z')] = 1 \text{ or } \Pr[D(z) \geq D\left(z'\right)] = 1.$$

For example, in the compliance problem with binary Z and D, monotonicity essentially rules out from the population "defiers" who systematically undertake the opposite treatment than that assigned to them (see AIR). Further, Vytlacil (2002, 2006) shows that monotonicity and "ignorability" of Z (discussed below) are equivalent to the threshold crossing restriction on q in A.1(ii) with Z independent of potential outcomes and the random thresholds U_D. Our next assumption ensures this independence. We thus employ interchangeably the threshold crossing structure in A.1(ii) and monotonicity[3]. We let $\perp$ denote independence as in Dawid (1979), with $\not\perp$ denoting dependence.

Assumption 2 (A.2). Exogeneity of Z: $(U_D, U_Y) \perp Z$.

A.2 ensures that Z is "ignorable," $\{D(z), Y(d), z \in S_Z \text{ and } d \in S_D\} \perp Z$, as assumed in e.g., IA and AI[4]. Note that, under A.2, $P(Z)$ is the propensity score, $P(Z) = \Pr(D = 1|Z)$.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

2.1. Proxies for Instruments

We assume that the econometrician observes realizations of a proxy W for Z whereas realizations of Z may be unobserved. A.1(ii) does not specify structural equations for W or Z, permitting several possibilities including $W = s(Z, U_W)$ or $Z = \tilde{s}(W, U_Z)$ with U_W and U_Z random vectors of unobservables and s and $\tilde{s}$ unknown measurable functions mapping to S_W and S_Z respectively[5]. However, we impose the following assumption on the proxy W.

Assumption 3 (A.3). Conditional Mean Independence[6] of D and Y from W Given Z: $D \perp_m W|Z$ and $Y \perp_m W|Z$.

A.3 ensures that W is irrelevant for predicting average D and Y given Z. Similar "nondifferential" measurement error assumptions have been recently employed in different contexts (see e.g. Mahajan, 2006; Lewbel, 2007; Hu, 2008). A.3 may fail if the error in coding Z relates systematically to D or Y. If $W = s(Z, U_W)$ then $(Z, D, Y) \perp U_W$ suffices for A.3. For example, U_W may denote measurement or misclassification error occurring at random when recording Z, a binary indicator for missing at random realizations of Z that W codes e.g. by $L \notin S_Z$ so that $W = ZU_W + L(1 - U_W)$, or a constant such that when W is a coarse (e.g. binary) coding for a multivalued Z. Thus, A.3 allows for several of the leading measurement error assumptions studied in the literature. Note that A.3 implies that a function $g(W)$ of W is also a proxy. For binary $D = \mathbf{1}\{U \leq P\}$, A.1–A.3 give that $D \perp_m W|P$ and $Y \perp_m W|P$, and thus A.3 holds using the scalar instrument $P \equiv P(Z)$.

2.1.1. Comparison Between Instruments and Proxies. It is useful to compare the assumptions imposed on instruments and proxies to clearly distinguish Z and W. First, as discussed above, Z and W play different structural roles under A.1. In particular, W need not structurally determine D or Y whereas Z drives D and, through it, Y. Second, W need not satisfy the monotonicity assumption imposed on Z. For instance, consider a binary instrument Z satisfying monotonicity, $\Pr[q(0, U_D) \leq q(1, U_D)] = 1$, and suppose that the proxy $W = (1 - U_W)Z + U_W(1 - Z)$, with nondegenerate U_W taking value in $\{0, 1\}$, can randomly misclassify Z and satisfies A.3. Then, substituting for $Z = \frac{W - U_W}{1 - 2U_W}$ in q gives the nonstructural equation $D = q\left(\frac{W - U_W}{1 - 2U_W}, U_D\right)$. But W does not satisfy monotonicity in this equation since $\Pr\left[q\left(\frac{w - U_W}{1 - 2U_W}, U_D\right) \leq q\left(\frac{w' - U_W}{1 - 2U_W}, U_D\right)\right] \neq 1$ for $w' \neq w \in S_W = \{0, 1\}$. Last, unlike for Z, we allow but do not require that W is exogenous. In particular, given A.2 $(U_D, U_Y) \perp Z$, joint exogeneity of Z and W, $(U_D, U_Y) \perp (Z, W)$, is equivalent to $(U_D, U_Y) \perp W|Z$ (e.g. Dawid 1979, lemma 4.3 and its converse). The latter condition is sufficient, but not necessary, for A.3 to hold. Specifically, it is possible that A.3 holds and $Y(d) \not\perp_m W$ so that W fails mean ignorability from potential outcomes and is therefore endogenous, $(U_D, U_Y) \not\perp W$. However, this requires particular specifications ensuring that $(U_D, U_Y) \not\perp W|Z$ while A.3 holds, i.e., $q(Z, U_D) \perp_m W|Z$

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

and $r(q(Z, U_D), U_Y) \perp_m W|Z$. To illustrate these conditions, Appendix A studies the linear random coefficients model and gives conditions under which an endogenous proxy W (such that $Y(d) \not\perp_m W$) satisfies A.3.

In a nutshell, a proxy W need not be an instrument. Specifically, W can fail monotonicity. Also, under particular specifications, W can fail exogeneity.

3. THE METHODS OF WALD AND LIV USING MISMEASURED INSTRUMENTS

For any $z, z' \in S_Z$ such that $E(D|Z = z) \neq E(D|Z = z')$, define $\alpha_{z,z'}$, the Wald estimand using Z evaluated at z and z':

$$\alpha_{z,z'} \equiv \frac{E(Y|Z = z') - E(Y|Z = z)}{E(D|Z = z') - E(D|Z = z)}.$$

Similarly, for any $w, w' \in S_W$ such that $E(D|W = w) \neq E(D|W = w')$, define $\beta_{w,w'}$, the Wald estimand using W evaluated at w and w':

$$\beta_{w,w'} \equiv \frac{E(Y|W = w') - E(Y|W = w)}{E(D|W = w') - E(D|W = w)}.$$

Begin by considering the compliance problem (e.g. AIR) with binary treatment D and instrument Z so that $S_D = \{0, 1\}$ and $S_Z = \{0, 1\}$. For instance, $Z = 1$ may denote receiving financial aid and $D = 1$ joining a training program. Let

$$\alpha^*_{0,1} \equiv E[Y(1) - Y(0)|D(1) - D(0) = 1],$$

denote LATE, the average treatment effect for the subpopulation of "compliers" who join the program if, and only if, they receive financial aid. Then, under A.1–A.2, we have $\alpha_{0,1} = \alpha^*_{0,1}$ (see e.g. AIR) and thus the Wald estimand $\alpha_{0,1}$ identifies LATE $\alpha^*_{0,1}$. When Z is unobserved, $\alpha_{0,1}$ is not directly estimable. However, data on a proxy W for Z may be available. For example, suppose that a binary W codes Z, possibly with misclassification. Then provided that $E(D|Z = 1) \neq E(D|Z = 0)$ and $Z \not\perp W$, we have

$$\begin{aligned}
\beta_{0,1} &\equiv \frac{E(Y|W = 1) - E(Y|W = 0)}{E(D|W = 1) - E(D|W = 0)} \\
&= \frac{\sum_{k=0}^{1} E(Y|Z = k, W = 1) f_{Z|W}(k|1) - \sum_{k=0}^{1} E(Y|Z = k, W = 0) f_{Z|W}(k|0)}{\sum_{k=0}^{1} E(D|Z = k, W = 1) f_{Z|W}(k|1) - \sum_{k=0}^{1} E(D|Z = k, W = 0) f_{Z|W}(k|0)} \\
&= \frac{[f_{Z|W}(1|1) - f_{Z|W}(1|0)]\,[E(Y|Z = 1) - E(Y|Z = 0)]}{[f_{Z|W}(1|1) - f_{Z|W}(1|0)]\,[E(D|Z = 1) - E(D|Z = 0)]} = \alpha_{0,1} = \alpha^*_{0,1},
\end{aligned}$$

where the second equality obtains by A.3. Thus, the Wald estimand using the proxy W identifies LATE, $\beta_{0,1} = \alpha^*_{0,1}$. Importantly, W need not satisfy the assumptions required for Z, e.g. monotonicity, as discussed in Section 2.1.1. Instead, it suffices that $D \perp_m W|Z$, $Y \perp_m W|Z$, and $Z \not\perp W$. We do not require

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

other restrictions on the extent of misclassification. This relaxes the conditions in AIR: any suitable proxy for Z can be used to identify LATE.

Another example arises when data on Z is missing. For instance, W may take on three values corresponding to no aid ($W = 0$), being assigned aid ($W = 1$), and missing financial aid information ($W = 2$). Then an analogous argument gives that, under A.3 and provided that $E(D|Z = 1) \neq E(D|Z = 0)$ and $f_{Z|W}(1|w') \neq f_{Z|W}(1|w)$, $\beta_{w,w'}$ evaluated at *any* $w, w' \in S_W$ identifies $\alpha^*_{0,1}$. Interestingly, this includes $w' = 2$ which codes a missing realization of Z.

In showing that $\beta_{0,1}$ identifies $\alpha^*_{0,1}$ in the case of binary Z, we used the equality under A.3 of the Wald estimands $\beta_{0,1}$ and $\alpha_{0,1}$. More generally, in interpreting the Wald and LIV estimands using W in the following subsections, we use Lemma B.1 in Appendix B which relates these estimands to Wald and LIV estimands using Z. In particular, for discrete[7] Z, Lemma B.1 shows that, under A.3, $\beta_{w,w'}$ is a weighted average of α_{z_{k-1},z_k}, for adjacent points $z_{k-1}, z_k \in S_Z$. Similarly, for continuous scalar ($\kappa = 1$) Z, we show that, under A.3, $\beta_{w,w'}$ is a weighted average of $\alpha(z)$ over S_Z, where $\alpha(z)$ denotes the local IV (LIV) estimand

$$\alpha(z) \equiv \lim_{dz \downarrow 0} \alpha_{z,z+dz} = \frac{\frac{\partial}{\partial z} E(Y|Z = z)}{\frac{\partial}{\partial z} E(D|Z = z)},$$

with $E(D|Z = \cdot)$ and $E(Y|Z = \cdot)$ differentiable at $z \in S_Z$ and $\frac{\partial}{\partial z} E(D|Z = z) \neq 0$ (see e.g. Angrist, Graddy, and Imbens, 2000 (thereafter AGI); HV; Heckman, 2010). Further, when W is a continuous scalar and $E(D|W = \cdot)$ and $E(Y|W = \cdot)$ are differentiable at $w \in S_W$ with $\frac{\partial}{\partial w} E(D|W = w) \neq 0$, Lemma B.1 shows that, under A.3, the LIV estimand using W,

$$\beta(w) \equiv \frac{\frac{\partial}{\partial w} E(Y|W = w)}{\frac{\partial}{\partial w} E(D|W = w)},$$

is a weighted average of $\alpha(z)$ over S_Z. Lemma B.1 does not assume A.1(ii) or A.2.

Results interpreting the Wald estimand using a binary proxy W in the case of binary D and either a binary Z or a continuous scalar Z such that $\Pr(D = 1|Z) = Z$ appear in the epidemiology literature (Hernán and Robins, 2006). Our results encompass, but are not limited to, these cases. Here, we demonstrate that W need not satisfy monotonicity and can, in particular structures, be endogenous. Further, we allow for multivalued D throughout and for general vectors Z and W. In addition to studying the Wald estimand, we study the interpretation of the LIV and IV estimands using a proxy in discrete choice structural systems and show that these estimands identify weighted averages of LATEs or MTEs. We study necessary and sufficient conditions for nonnegative weights as well as a sufficient condition for the Wald or LIV estimand using W to identify the LATE or MTE (and to recover e.g. ATE, ATT, ATU) that would have been identified had Z been observed.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

3.1. Interpretation of Wald and LIV Estimands Using a Proxy

We build on results in IA, AI, AGI, HV, and HUV studying identification of LATE and MTE via the methods of Wald and LIV using an instrument Z and extend these results to accommodate mismeasured instruments. First, we state a result interpreting the Wald and LIV estimands using Z. This follows from arguments similar to e.g. AI's theorem 1, AGI's theorem 3, and HUV.

THEOREM 3.1. *Assume A.1–A.2. (a) Let* $z, z' \in S_Z$ *with* $\Pr[D(z) < D(z')] > 0$ *then*

$$\alpha_{z,z'} = \alpha^*_{z,z'} \quad \textit{where} \quad \alpha^*_{z,z'} \equiv \sum_{j=1}^{J} \omega_j^{z,z'} E\left[Y(j) - Y(j-1) | \nu(z) < V_j \leq \nu(z')\right],$$

with $\omega_j^{z,z'} \equiv \frac{\Pr[\nu(z) < V_j \leq \nu(z')]}{\sum_{j=1}^{J} \Pr[\nu(z) < V_j \leq \nu(z')]}$, $0 \leq \omega_j^{z,z'} \leq 1$, *and* $\sum_{j=1}^{J} \omega_j^{z,z'} = 1$.

(b) Set $\kappa = 1$ *and let* $z \in S_Z$. *Let* ν *be differentiable at* z *with* $\frac{\partial}{\partial z}\nu(z) \neq 0$ *and let* $E[Y(j) - Y(j-1)|V_j = \cdot]$ *and* $f_{V_j}(\cdot)$ *be continuous at* $\nu(z)$ *for all* $j \in S_D \backslash \{0\}$, *with* $f_{V_j}(\nu(z)) > 0$ *for some* $j \in S_D \backslash \{0\}$. *Then*

$$\alpha(z) = \alpha^*(z) \quad \textit{where} \quad \alpha^*(z) \equiv \sum_{j=1}^{J} \omega_j(z) E[Y(j) - Y(j-1)|V_j = \nu(z)],$$

with $\omega_j(z) \equiv \frac{f_{V_j}(\nu(z))}{\sum_{j=1}^{J} f_{V_j}(\nu(z))}$, $0 \leq \omega_j(z) \leq 1$, *and* $\sum_{j=1}^{J} \omega_j(z) = 1$.

Theorem 3.1(a) shows that the Wald estimand $\alpha_{z,z'}$ is equal to a weighted average of LATEs for subpopulations such that $\nu(z) < V_j \leq \nu(z')$ or $D(z) < j \leq D(z')$. For binary Z, AI refers to $\alpha^*_{0,1}$ as the "average causal response." If D is also binary, $\alpha^*_{0,1}$ is LATE. For continuous scalar Z, part (b) shows that the LIV estimand $\alpha(z)$ equals $\alpha^*(z)$, a weighted average of MTEs. In particular, for $J = 1$, $\alpha(z)$ equals the MTE $E[Y(1) - Y(0)|V_1 = \nu(z)]$, the treatment effect for those indifferent toward receiving the treatment at instrument value z. Average treatment effects such as ATE, ATT, and ATU can be expressed as weighted averages of $E[Y(1) - Y(0)|V_1 = \cdot]$ (see e.g. HV).

When Z is unobserved $\alpha_{z,z'}$ and $\alpha(z)$ are not directly estimable. The next Theorem demonstrates that the Wald or LIV estimand, $\beta_{w,w'}$ or $\beta(w)$, using a proxy W for Z identifies a weighted average of LATEs or MTEs. For discrete variables, we often simplify notation whenever multiple sub- or super-scripts are used. For example, we denote by $\alpha_{k-1,k}$ the Wald estimand α_{z_{k-1},z_k} evaluated at $z_{k-1}, z_k \in S_Z$. Similarly, we sometimes write $\alpha^*_{k-1,k}$ instead of $\alpha^*_{z_{k-1},z_k}$. For $z \neq z' \in S_Z$, $\nu(z) \leq \nu(z')$, and $\sum_{j=1}^{J} \Pr[\nu(z) < V_j \leq \nu(z')] = 0$, we let $\alpha^*_{z,z'} = 0$. Similarly, for $\sum_{j=1}^{J} f_{V_j}(\nu(z)) \frac{\partial}{\partial z}\nu(z) = 0$, we let $\alpha^*(z) = 0$. Throughout the paper, for $z \in S_Z$, we write $G_j(\nu(z)) \equiv E\{\mathbf{1}\{V_j \leq \nu(z)\} E[Y(j) - Y(j-1)|V_j]\}$, $j = 1, ..., J$.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

THEOREM 3.2. *Assume A.1–A.3 and let $w, w' \in S_W$ with $E(Y|W=w^*) < \infty$ for $w^* = w, w'$. (a) Suppose that $S_Z = \{z_0, z_1, ..., z_K\}$ with $z_{k-1} < z_k$ for $k = 1, ..., K$. Then, provided $E(D|W=w) \neq E(D|W=w')$:*

$$\beta_{w,w'} = \sum_{k=1}^{K} \mu_k^{w,w'} \alpha_{k-1,k}^*, \quad \textit{with}$$

$$\mu_k^{w,w'} \equiv \frac{[E(D|Z=z_k) - E(D|Z=z_{k-1})][F_{Z|W}(z_{k-1}|w) - F_{Z|W}(z_{k-1}|w')]}{\sum_{s=1}^{K}[E(D|Z=z_s) - E(D|Z=z_{s-1})][F_{Z|W}(z_{s-1}|w) - F_{Z|W}(z_{s-1}|w')]}$$

$$\textit{and} \quad \sum_{k=1}^{K} \mu_k^{w,w'} = 1.$$

(b.i) Set $\kappa = 1$ and suppose that $F_{Z|W}(\cdot|w)$ and $F_{Z|W}(\cdot|w')$ are absolutely continuous with support S_Z, $v(\cdot)$ is differentiable at a.e. (almost every) $z \in S_Z$, and $F_{V_j}(v(\cdot))$ and $G_j(v(\cdot))$, $j = 1, ..., J$, are absolutely continuous on S_Z. Then, provided $E(D|W=w) \neq E(D|W=w')$:

$$\beta_{w,w'} = \int_{S_Z} \alpha^*(z) \mu^{w,w'}(z) dz,$$

$$\textit{with} \quad \mu^{w,w'}(z) \equiv \frac{\left[\frac{\partial}{\partial z} E(D|Z=z)\right]\left[F_{Z|W}(z|w) - F_{Z|W}(z|w')\right]}{\int_{S_Z}\left[\frac{\partial}{\partial z} E(D|Z=s)\right]\left[F_{Z|W}(s|w) - F_{Z|W}(s|w')\right] ds}$$

$$\textit{and} \quad \int_{S_Z} \mu^{w,w'}(z) dz = 1.$$

(b.ii) In addition, set $\ell = 1$ and let $w \in \mathcal{B}(w)$, a nonempty open subset of S_W, and suppose that for all $w^* \in \mathcal{B}(w)$, $E(Y|W=w^*) < \infty$, $F_{Z|W}(\cdot|w^*)$ is absolutely continuous with support S_Z, $\frac{\partial}{\partial w} F_{Z|W}(z|w^*)$ exists for a.e. $z \in S_Z$, and there exist integrable functions $b_x(\cdot)$ on S_Z such that $\left|\left[\frac{\partial}{\partial z} E(X|Z=z)\right]\left[\frac{\partial}{\partial w} F_{Z|W}(z|w^*)\right]\right| < b_x(z)$ for $X = D, Y$ and a.e. $z \in S_Z$. Then, provided $\frac{\partial}{\partial w} E(D|W=w) \neq 0$:

$$\beta(w) = \int_{S_Z} \alpha^*(z) \mu^w(z) dz,$$

$$\textit{with} \quad \mu^w(z) \equiv \frac{\left[\frac{\partial}{\partial z} E(D|Z=z)\right]\left[\frac{\partial}{\partial w} F_{Z|W}(z|w)\right]}{\int_{S_Z}\left[\frac{\partial}{\partial z} E(D|Z=s)\right]\left[\frac{\partial}{\partial w} F_{Z|W}(s|w)\right] ds} \quad \textit{and} \quad \int_{S_Z} \mu^w(z) dz = 1.$$

Under A.1–A.3, Theorem 3.2 shows that the Wald estimand $\beta_{w,w'}$ identifies a weighted average of "average causal responses" $\alpha_{k-1,k}^*$ or $\alpha^*(z)$, which in turn are weighted averages of LATEs or MTEs with nonnegative weights $\omega_j^{k-1,k}$ or $\omega_j(z)$. Thus, $\beta_{w,w'}$ identifies a weighted average of LATEs, with weights

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

$\sum_{j=1}^{J} \mu_k^{w,w'} \omega_j^{k-1,k}$, or of MTEs with weights $\sum_{j=1}^{J} \mu^{w,w'}(z)\omega_j(z)$. Similarly, the LIV estimand $\beta(w)$ identifies a weighted average of MTEs. We comment on the conditions in Theorem 3.2. First, from the proof of Lemma B.1, we have under A.3 that $E(D|W = w') - E(D|W = w)$ equals the denominator of the weights $\mu_k^{w,w'}$ or $\mu^{w,w'}(z)$. Similarly, $-\frac{\partial}{\partial w}E(D|W = w)$ equals the denominator of $\mu^{w}(z)$. Requiring this difference or derivative to be nonzero permits division by it and rules out uninformative proxies by ensuring that $Z \not\perp W$. Second, to simplify the exposition in Theorem 3.2(b), we let $F_{Z|W}(\cdot|w^*)$ for $w^* = w, w'$ in part ($b.i$) and for all $w^* \in \mathcal{B}(w)$ in part ($b.ii$) be absolutely continuous (e.g., enabling integration by parts) with the same support[8] S_Z as that of Z to remove the complication associated with boundary terms. Similarly, having $F_{V_j}(\nu(\cdot))$ and $G_j(\nu(\cdot))$ be absolutely continuous ensures under A.1–A.2 that $E(X|Z = \cdot)$, $X = D, Y$, is absolutely continuous so that $\frac{\partial}{\partial z}E(X|Z = \cdot)$ exists for a.e. $z \in S_Z$ and is integrable. For example, this obtains if $\nu(\cdot)$ is absolutely continuous and monotonic. Last, the regularity conditions in Theorem 3.2($b.ii$) enable interchanging the order of differentiation $\left(\frac{\partial}{\partial w}\right)$ and integration in $\frac{\partial}{\partial w}\int_{S_Z}\left[\frac{\partial}{\partial z}E(X|Z = z)\right]F_{Z|W}(z|w)dz$ (see proof of Lemma B.1).

To illustrate Theorem 3.2, consider a binary $D = \mathbf{1}\{U \leq P\}$ with scalar instrument $P \equiv P(Z)$, so that $E(D|P = p) = p$. Here, $\nu(p) = p$ and $F_{V_j}(\cdot)$ and $G_j(\cdot)$ are absolutely continuous. Applying Theorem 3.2(b) using P, assumed continuous with support S_P, gives

$$\beta_{w,w'} = \int_{S_P} E[Y(1) - Y(0)|U = p] \frac{\left[F_{P|W}(p|w) - F_{P|W}(p|w')\right]}{\int_{S_P}\left[F_{P|W}(s|w) - F_{P|W}(s|w')\right]ds} dp.$$

Under the conditions of Theorem 3.2($b.ii$), $\frac{\partial}{\partial w}F_{P|W}(p|w)$ exists for a.e. $p \in S_P$ and

$$\beta(w) = \int_{S_P} E[Y(1) - Y(0)|U = p] \frac{\frac{\partial}{\partial w}F_{P|W}(p|w)}{\int_{S_P}\frac{\partial}{\partial w}F_{P|W}(s|w)ds} dp.$$

The weights sum up or integrate to 1. For example, with a homogenous effect, $\beta_{w,w'} = E[Y(1) - Y(0)]$. Note that some of the weights can be negative. Thus, with heterogeneous effects, $\beta_{w,w'}$ or $\beta(w)$ may be negative even when all the LATEs or MTEs are positive.

3.2. Nonnegative Weights

Next, we define a property that a proxy for an instrument may have and that can ensure nonnegative weights. As discussed in the Introduction, when the LATEs or MTEs have a common sign, this property ensures that the Wald or LIV estimand using W recovers this sign.

DEFINITION 3.1. *Intensity Preserving Proxy: Let Z and W be as in A.1(i). For $w, w' \in S_W$, we say that the proxy W at (w, w') is intensity preserving for*

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

the instrument Z if $F_{Z|W}(z|w) \geq F_{Z|W}(z|w')$ for all $z \in S_Z$ or $F_{Z|W}(z|w) \leq F_{Z|W}(z|w')$ for all $z \in S_Z$. Further, provided $\frac{\partial}{\partial w} F_{Z|W}(\cdot|\cdot)$ exists at $w \in S_W \subseteq \mathbb{R}$ and a.e. $z \in S_Z$, we say W at w is intensity preserving for Z if $\frac{\partial}{\partial w} F_{Z|W}(z|w) \leq 0$ for a.e. $z \in S_Z$ or $\frac{\partial}{\partial w} F_{Z|W}(z|w) \geq 0$ for a.e. $z \in S_Z$.

Intensity preservation at (w, w') is a stochastic dominance restriction. Similar first-order stochastic dominance conditions have been assumed for a mismeasured dependent variable in Abrevaya and Hausman (1999) and Hausman, Abrevaya, and Scott-Morton (1998). For scalar proxy W and $w < w'$, we say W at (w, w') is positively (negatively) intensity preserving for Z if W being small increases (decreases) the probability of Z being small. Positive (negative) intensity preservation at w is the marginal analogue with $\frac{\partial}{\partial w} F_{Z|W}(z|w) \leq 0$ ($\frac{\partial}{\partial w} F_{Z|W}(z|w) \geq 0$) for a.e. $z \in S_Z$. When positive (negative) intensity preservation holds for all $w, w' \in S_W$, $w < w'$, or for all $w \in S_W$, we drop the local qualifiers "at (w, w')" or "at w."

A proxy W that fails the monotonicity assumption satisfied by Z in A.1 can be intensity preserving for Z. For instance, in the example in Section 2.1.1 where W may misclassify binary Z, $W = (1 - U_W)Z + U_W(1 - Z)$, W fails monotonicity and is trivially intensity preserving for Z. More generally, Appendix A gives an example for discrete (nonbinary) Z and a misclassified proxy W, that may code the reverse or mirror image of Z, where W fails the monotonicity in A.1, and possibly exogeneity, but is locally intensity preserving.

Next, we study sufficient conditions for intensity preservation. First, in the scalar deterministic case, $W = f(Z)$, monotonicity of f suffices for intensity preservation but is not necessary, as shown via counterexample in the proof of Proposition 3.3.

PROPOSITION 3.3. *Let Z and W be as in A.1(i) with $\kappa = \ell = 1$. Let $f : S_Z \to S_W$ and $W = f(Z)$. Then f being a nondecreasing (nonincreasing) monotonic function is sufficient but not necessary for W to be positively (negatively) intensity preserving for Z.*

Second, "positive likelihood ratio dependence" suffices for W to be positively intensity preserving for Z:

$$f_{Z,W}\left(z, w'\right) f_{Z,W}\left(z', w\right) \leq f_{Z,W}(z, w) f_{Z,W}\left(z', w'\right)$$
for all $z < z' \in S_Z$, $w < w' \in S_W$.

Using this property, Proposition 3.4 gives conditions under which classical measurement error yields an intensity preserving proxy. This follows from arguments in Lehmann[9] (1966).

PROPOSITION 3.4. *Let Z and W be as in A.1(i) with $\kappa = \ell = 1$. Let U_W be a random variable with logarithmically concave density function such that $U_W \perp Z$ and let $W = Z + U_W$. Then W is positively intensity preserving for Z.*

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Examples of log-concave or "strongly unimodal" density functions include the normal, uniform, logistic, and exponential densities (see e.g. Bagnoli and Bergstrom, 2005).

Last, if $\tilde{f}(\cdot)$ is a nondecreasing function on S_Z and W at (w, w'), or at w, is positively (negatively) intensity preserving for Z then W at (w, w'), or at w, is positively (negatively) intensity preserving for $\tilde{f}(Z)$. If $\tilde{f}(\cdot)$ is nonincreasing, this implication holds with opposite signs (e.g. Shaked and Shanthikumar 1994, Thm. 1.A.3(a)). For example, for nondecreasing $\tilde{f}$, W in Proposition 3.4 is also positively intensity preserving for $\tilde{f}(Z)$.

To state necessary and sufficient conditions for nonnegative weights, we employ the following condition, which imposes further structure on the effect of scalar Z on D. A similar condition is employed e.g. in some of the results in IA and AI.

Assumption 4 (A.4). For $S_Z \subseteq \mathbb{R}$ either $\Pr\left[D(z) \leq D\left(z'\right)\right] = 1$ for all $z < z'$ in S_Z or $\Pr\left[D(z) \geq D\left(z'\right)\right] = 1$ for all $z < z'$ in S_Z. If, in addition, $\Pr\left[D(z) \neq D\left(z'\right)\right] > 0$ for all $z < z'$ in S_Z, we say that A.4 holds strictly.

A.4 holds when $\nu(\cdot)$ is monotonic. Under A.4, Z affects D in the same direction at any values $z < z'$ of Z, as may be the case in a training program with Z denoting financial aid. We employ (strict) A.4 only to state sufficient (and necessary) conditions for nonnegative weights. In particular, for scalar Z, we make use of the fact that A.1, A.2, and (strict) A.4 ensure that $E(D|Z = \cdot)$ is (strictly) monotonic over[10] S_Z. Note that strict A.4 is immediately satisfied for binary $D = \mathbf{1}\{U \leq P\}$ with instrument $P \equiv P(Z)$ and $E(D|P = p) = p$ is clearly strictly monotonic in p.

The following corollary to Theorem 3.2 shows that W being locally intensity preserving for Z is sufficient (and necessary under strict A.4) for nonnegative weights.

COROLLARY 3.5. *(a) Under the conditions of Theorem 3.2(a) or (b.i) and (strict) A.4,* $0 \leq \mu_k^{w,w'}$ *for* $k = 1, ..., K$ *or* $0 \leq \mu^{w,w'}(z)$ *for a.e.* $z \in S_Z$ *if (and only if) W at* (w, w') *is intensity preserving for Z. (b) Under the conditions of Theorem 3.2(b.ii) and (strict) A.4,* $0 \leq \mu^{w}(z)$ *for a.e.* $z \in S_Z$ *if (and only if) W at* w *is intensity preserving for Z.*

For example, for binary D with continuous instrument $P \equiv P(Z)$, Corollary 3.5 gives that the weights $\mu^{w,w'}(p)$ (or $\mu^{w}(p)$) are nonnegative if and only if W at (w, w') (or at w) is intensity preserving for P. In this case, the weights $\sum_{j=1}^{J} \mu^{w,w'}(p)\omega_j(p)$ or $\sum_{j=1}^{J} \mu^{w}(p)\omega_j(p)$ on the MTEs are all nonnegative.

3.3. Identification of Treatment Effects Using a Proxy

This section studies conditions under which the Wald or LIV estimand using W identifies the LATE or MTE that would have been identified had Z been observed.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

The next definition gives a property that a proxy may have and that suffices for this identification result.

DEFINITION 3.2. *Value Preserving Proxy: Let Z and W be as in A.1(i). For $z \in S_Z$ and $w, w' \in S_W$, we say that the proxy W at (w, w') is value preserving for the instrument Z at z if $F_{Z|W}(s|w) = F_{Z|W}(s|w')$ for every $s \in S_Z \backslash \{z\}$ and $F_{Z|W}(z|w) \neq F_{Z|W}(z|w')$. Further, for $\kappa = \ell = 1$, $z \in S_Z$ and $w \in S_W$, we say that W at w is value preserving for Z at z if $\frac{\partial}{\partial w} F_{Z|W}(s|w) = \gamma\, \delta(s - z)$ for all $s \in S_Z$ where γ is a nonzero constant and $\delta(\cdot)$ is the Dirac delta function.*

A discrete W at (w, w') is value preserving for discrete Z at z_{k-1} if W being equal to w as opposed to w' is uninformative about the likelihood of Z being equal to $s \in S_Z$ except for when Z equals z_{k-1} or z_k, i.e., $f_{Z|W}(s|w) = f_{Z|W}(s|w')$ for every $s \in S_Z \backslash \{z_{k-1}, z_k\}$ and $f_{Z|W}(s|w) \neq f_{Z|W}(s|w')$ for $s \in \{z_{k-1}, z_k\}$. Thus, $F_{Z|W}(\cdot|w)$ can exceed $F_{Z|W}(\cdot|w')$ only at z_{k-1}, since $F_{Z|W}(\cdot|w')$ catches up with $F_{Z|W}(\cdot|w)$ at z_k. This trivially holds for binary Z with $F_{Z|W}(z_0|w) \neq F_{Z|W}(z_0|w')$. W at w being value preserving for Z at z is the marginal analogue with $\frac{\partial}{\partial w} F_{Z|W}(s|w) = 0$ for $s \neq z$ and having an impulse at $s = z$. We give an example shortly. Note that if W at (w, w') (or at w) is value preserving for Z at z then W at (w, w') (or at w) is intensity preserving for Z. However, a positively intensity preserving proxy W at w may be *value disrupting* for Z at z, for all $w \in S_W$ and $z \in S_Z$ (e.g., let Z and W be as in Proposition 3.4 with classical measurement error and (Z, U_W) jointly normal). The next proposition gives an example of an error-laden proxy W for possibly unobserved Z with $S_W = S_Z$ such that for all $z \in S_Z$, W at z is value preserving for Z at z.

PROPOSITION 3.6. *Let Z and W be as in A.1(i) with $\kappa = \ell = 1$. Let Z and V be identically distributed random variables with a continuous density on the support S_Z and let ϵ be a random variable taking value in $\{0, 1\}$ with $\Pr(\epsilon = 0) > 0$. Let $Z \perp (V, \epsilon)$ and $V \perp \epsilon$ and let $W = (1 - \epsilon)Z + \epsilon V$. Then, for all $z \in S_Z$, W at z is value preserving for Z at z.*

In Proposition 3.6, W codes Z correctly with a nonzero probability or as the realization of an independent and identically distributed V. Note that this allows for $\Pr(\epsilon = 0) = 1$ and $W = Z$. An and Hu (2012) study the related assumption that the probability of truthful reporting in self-reported survey data is nonzero and discuss examples. Appendix A gives another example in which a discrete proxy W, that may code the reverse of Z, fails monotonicity (and can be endogenous) but is intensity and value preserving at certain points.

Corollary 3.7 gives conditions under which the Wald or LIV estimand using W identifies the LATE and MTE that would have been recovered had Z been observed.

COROLLARY 3.7. *(a) Under the conditions of Theorem 3.2(a), if W at (w, w') is value preserving for Z at z_{k-1} then $\beta_{w,w'} = \alpha^*_{k-1,k}$. (b) Assume the conditions of Theorem 3.1(b) with $\nu(\cdot)$ continuously differentiable at z and that the*

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

conditions of Theorem 3.2(b) hold for each element of a sequence $\{F^{\tau}_{Z|W}(\cdot|\cdot)\}_{\tau>0}$ *such that* $\lim_{\tau\to 0}\frac{\partial}{\partial w}F^{\tau}_{Z|W}(s|w)=\frac{\partial}{\partial w}F_{Z|W}(s|w)$ *for all* $s\in S_Z$. *If* W *at* w *is value preserving for* Z *at* z *then* $\beta(w)=\alpha^*(z)$.

To illustrate Corollary 3.7, let the instrument $P\equiv P(Z)$ and its error-laden proxy W be as in Proposition 3.6. Then, for all $p\in S_P$, W at p is value preserving[11] for P at p. Corollary 3.7 gives that for binary treatment, the LIV estimand using W, $\beta(\cdot)$, identifies MTE, i.e., $\beta(p)=E[Y(1)-Y(0)|U=p]$. Building on the results of HV, we can identify various average effects using $\beta(p)$. In particular, when S_W is the unit interval, integrating $\beta(p)$ identifies the average treatment effect:

$$ATE\equiv E[Y(1)-Y(0)]=\int_0^1\beta(p)dp.$$

Similarly, since P and W are identically distributed (see proof of Proposition 3.6), the average treatment effects for the treated and untreated are also identified by:

$$ATT\equiv E[Y(1)-Y(0)|D=1]=\int_0^1\beta(p)(1-F_W(p))\frac{1}{E(W)}dp,\text{ and}$$
$$ATU\equiv E[Y(1)-Y(0)|D=0]=\int_0^1\beta(p)F_W(p)\frac{1}{E(1-W)}dp.$$

Thus, in this case, several average effects can be recovered using the proxy W for $P(Z)$.

4. THE METHOD OF IV USING MISMEASURED INSTRUMENTS

Researchers often employ IV methods such as TSLS to identify causal effects. Consider a function $\tilde{g}(Z)$ of the instrument, with $\tilde{g}:S_Z\to S_{\tilde{g}}\subseteq\mathbb{R}$. The IV estimand $\beta^{\tilde{g}}_Z$ is given by:

$$\beta^{\tilde{g}}_Z\equiv\frac{Cov(Y,\tilde{g}(Z))}{Cov(D,\tilde{g}(Z))}.$$

The choice $\tilde{g}(Z)=E(D|Z)$ corresponds to TSLS. Theorem 2 of IA demonstrates that for binary D, discrete Z, and $\tilde{g}(\cdot)$ monotonic, $\beta^{\tilde{g}}_Z$ identifies under A.1, A.2, and A.4 a weighted average of $\alpha^*_{k-1,k}$, $k=1,...,K$, with nonnegative weights. Theorem 2 of AI generalizes this result to multivalued D in the case of TSLS. HV shows that for binary D, $\beta^{\tilde{g}}_Z$ is a weighted average of marginal treatment effects and note that $E(\tilde{g}(Z)|E(D|Z)\geq m)$ being monotonic in m suffices for nonnegative weights. HUV extends the analysis in HV to accommodate multivalued D in ordered and unordered choice models.

We build on, and extend, these results to interpret the IV estimand β^{g}_W using a function $g(W)$ of a proxy W. We show that, under our assumptions, β^{g}_W identifies

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

a weighted average of LATEs or MTEs and study necessary and sufficient conditions for nonnegative weights. The results obtain by noting that A.3 and the law of iterated expectations give

$$\beta_W^g \equiv \frac{Cov(Y, g(W))}{Cov(D, g(W))} = \frac{Cov(Y, E(g(W)|Z))}{Cov(D, E(g(W)|Z))} \equiv \beta_Z^{E(g(W)|\cdot)},$$

and, under A.1–A.2, $\beta_Z^{E(g(W)|\cdot)}$ is a weighted average of LATEs or MTEs.

4.1. Interpretation of the IV Estimand Using a Proxy

The next result extends the results in e.g., IA and AI to study the method of IV using a function of a proxy W for Z. Throughout this subsection, we let $g : S_W \to S_g \subseteq \mathbb{R}$ and write $\tilde{g}(Z) \equiv E[g(W)|Z]$ and $\ddot{g}(Z) \equiv \tilde{g}(Z) - E(\tilde{g}(Z))$.

THEOREM 4.1. *Assume A.1–A.3 and that $Cov(D, g(W))$ is finite and nonzero and $Cov(Y, g(W))$ is finite. (a) Suppose that $S_Z = \{z_0, z_1, ..., z_K\}$ with $z_{k-1} < z_k$ for $k = 1, .., K$. Then*

$$\beta_W^g = \sum_{k=1}^{K} \phi_k^g \alpha_{k-1,k}^*,$$

with

$$\phi_k^g = \frac{[E(D|Z = z_k) - E(D|Z = z_{k-1})]E(\ddot{g}(Z)|Z \geq z_k)\Pr(Z \geq z_k)}{\sum_{m=1}^{K}[E(D|Z = z_m) - E(D|Z = z_{m-1})]E(\ddot{g}(Z)|Z \geq z_m)\Pr(Z \geq z_m)}$$

and $\sum_{k=1}^{K} \phi_k^g = 1.$

(b) Set $\kappa = 1$ and suppose that $v(\cdot)$ is differentiable at a.e. $z \in S_Z$ and that $F_Z(\cdot)$, $F_{V_j}(v(\cdot))$, and $G_j(v(\cdot))$, $j = 1, ..., J$, are absolutely continuous on S_Z. Then

$$\beta_W^g = \int_{S_Z} \alpha^*(z)\phi^g(z)dz,$$

with

$$\phi^g(z) = \frac{[\frac{\partial}{\partial z}E(D|Z = z)]E(\ddot{g}(Z)|Z \geq z)\Pr(Z \geq z)}{\int_{S_Z}[\frac{\partial}{\partial z}E(D|Z = m)]E(\ddot{g}(Z)|Z \geq m)\Pr(Z \geq m)dm}$$

and $\int_{S_Z} \phi^g(z)dz = 1$

Theorem 4.1 shows that β_W^g is a weighted average of Wald estimands $\alpha_{k-1,k}$ for discrete Z or LIV estimands $\alpha(z)$ for scalar continuous Z, which in turn

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

are weighted averages $\alpha^*_{k-1,k}$ or $\alpha^*(z)$ of LATEs or MTEs with nonnegative weights $\omega_j^{k-1,k}$ or $\omega_j(z)$ respectively. For example, for binary $D = \mathbf{1}\{U \leq P\}$, $E(D|P = p) = p$ and Theorem 4.1(*b*) using a continuous instrument P gives

$$\beta^g_W = \int_{S_P} E[Y(1) - Y(0)|U = p] \frac{E(\ddot{g}(Z)|P \geq p)\Pr(P \geq p)}{\int_{S_P} E(\ddot{g}(Z)|P \geq m)\Pr(P \geq m)dm} dp.$$

The weights ϕ^g_k or $\phi^g(z)$ sum or integrate to one but can be negative. Thus, β^g_W may be negative even when all the underlying LATEs or MTEs are positive. We now study a property that is useful in ensuring nonnegative weights.

DEFINITION 4.1. *Mean Quadrant Dependence: Let Z and W be as in A.1(i). We say that W is positively (negatively) mean quadrant dependent on Z if $E(W|Z \geq z) \geq (\leq) E(W)$ for all $z \in S_Z$.*

We say that W is mean quadrant dependent on Z if W is either positively or negatively mean quadrant dependent on Z. A special case of Proposition 4.2 gives that if W is positively (negatively) intensity preserving for Z then W is positively (negatively) mean quadrant dependent on Z.

PROPOSITION 4.2. *Let Z and W be as in A.1(i) with $\kappa = \ell = 1$ and $g : S_W \to S_g \subseteq \mathbb{R}$ with $E(g(W)) < \infty$. If $g(\cdot)$ is nondecreasing and W is positively (negatively) intensity preserving for Z then $g(W)$ is positively (negatively) mean quadrant dependent on Z. If $g(\cdot)$ is nonincreasing, this implication holds with opposite signs.*

The next result relates mean quadrant dependence to monotonic functions and classical measurement error.

COROLLARY 4.3. *Let Z and W be as in A.1(i) with $\kappa = \ell = 1$ and U_W be a random variable such that $U_W \perp_m Z$. For $f : S_Z \to S_f \subseteq \mathbb{R}$ let $W = f(Z) + U_W$. Then f being a nondecreasing (nonincreasing) monotonic function suffices but is not necessary for W to be positively (negatively) mean quadrant dependent on Z.*

It follows that for classical measurement error $U_W \perp Z$, $W = Z + U_W$ is positively mean quadrant dependent on Z. Monotonicity of f is sufficient for mean quadrant dependence but is not necessary, as shown via counterexample in the proof of Corollary 4.3.

A corollary to Theorem 4.1 gives a necessary and sufficient condition for nonnegative weights.

COROLLARY 4.4. *Under the conditions of Theorem 4.1(a) or (b) and (strict) A.4, $0 \leq \phi^g_k$ for $k = 1, ..., K$ or $0 \leq \phi^g(z)$ for a.e. $z \in S_Z$ if (and only if) $g(W)$ is mean quadrant dependent on Z.*

Thus, under A.1–A.4, $g(W)$ being mean quadrant dependent on Z ensures that ϕ^g_k or $\phi^g(z)$ are all positive and thus that the weights $\sum_{j=1}^J \phi^g_k \omega_j^{k-1,k}$ or

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

$\sum_{j=1}^{J} \phi^g(z)\omega_j(z)$ on the underlying LATEs or MTEs are all positive[12]. For example, for binary D and instrument $P \equiv P(Z)$, the weights ϕ_k^g or $\phi^g(p)$ are all nonnegative if and only if $g(W)$ is mean quadrant dependent on P.

Theorem 4.1 and Corollary 4.4 encompass previous results for the special case β_Z^g in which $W = Z$ and $\tilde{g}(Z) = g(Z)$. For example, theorem 2 of IA with binary D supposes that $g(\cdot)$ is monotonic in discrete Z and theorem 2 of AI with multivalued D considers the TSLS case $g(Z) = E(D|Z)$ with discrete Z, which, under A.1, A.2, and A.4, immediately gives that $g(\cdot)$ is monotonic. In these cases, Corollary 4.3 gives that $g(Z)$ is mean quadrant dependent on Z, ensuring nonnegative weights. Also, note that for binary Z, β_W^g is equal to the "average causal response" $\alpha_{0,1}^*$, or LATE for binary D.

HUV provides another decomposition for the IV estimand using a function of instruments in an ordered discrete choice model and show that it identifies a weighted average of marginal treatment effects. The next result builds on HUV's analysis to provide an interpretation for IV estimands using a function of proxies for instruments. Here, we consider general vectors Z and W. We let S_{V_j} denote the support of V_j for $j = 1, ..., J$.

THEOREM 4.5. *Assume A.1–A.3 and suppose* $Cov(D, g(W))$ *is finite and nonzero and* $Cov(Y, g(W))$ *is finite then*

$$\beta_W^g = \sum_{j=1}^{J} E\left[\psi_j^g(V_j)\, E\left[Y(j) - Y(j-1)|V_j\right]\right]$$
$$= \sum_{j=1}^{J} \int_{S_{V_j}} \psi_j^g(v)\, E\left[Y(j) - Y(j-1)|V_j = v\right]\, f_{V_j}(v)dv,$$

with

$$\psi_j^g(v) \equiv \frac{\vartheta_j^g(v)}{\sum_{j=1}^{J} \int_{S_{V_j}} \vartheta_j^g(v) f_{V_j}(v)dv} \quad \textit{and}$$
$$\vartheta_j^g(v) \equiv E[\ddot{g}(Z)|v \leq \nu(Z)]\Pr(v \leq \nu(Z)),$$

and $\sum_{j=1}^{J} \int_{S_{V_j}} \psi_j^g(v) f_{V_j}(v)dv = 1$. *Further,* $0 \leq \psi_j^g(v)$ *for all* $v \in S_{V_j}$, $j = 1, ..., J$, *if and only if* $E[\ddot{g}(Z)|v \leq \nu(Z)]$ *does not change sign for all* $v \in S_{V_j}$, $j = 1, ..., J$.

Theorem 4.5 shows that the IV estimand using a function of a proxy is a weighted average of marginal treatment effects for individuals indifferent between treatments $j-1$ and j. Note that under A.1–A.2, we may impose the normalization $\nu(z) = E(D|Z = z)$ without loss of generality by corollary 1 of Vytlacil (2006). Thus, under strict A.4, $\nu(z) = E(D|Z = z)$ is strictly monotonic over $S_Z \subseteq \mathbb{R}$, and it suffices for nonnegative weights that $E[\ddot{g}(Z)|z \leq Z]$ does not

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

change sign for all $z \in S_Z$ or, by the law of iterated expectations, that $g(W)$ is mean quadrant dependent on Z. When $W = Z$, Theorem 4.5 reduces to, or complements, previous results in the literature, e.g., in HV with binary D and HUV with multivalued D and thresholds possibly depending on exogenous covariates.

5. EXAMPLES AND SIMULATIONS

To illustrate this paper's results, consider the case of a binary treatment D (e.g., training program) and outcome Y (e.g., wage) with randomized Z (e.g., financial aid). In particular, consider a version of the generalized Roy model (see e.g., HUV) where

$$Y(0) = \theta + U_0, \quad Y(1) = \theta + \delta + U_1, \quad \text{and} \quad D = \mathbf{1}\{-[Y(1) - Y(0)] \leq -\lambda + \gamma Z\},$$

with coefficients $\theta = 1$, $\delta = 0.2$, $\lambda = 2$, and $\gamma = 1$. Here, an individual selects into treatment if his/her payoff $Y(1) - Y(0)$ exceeds the cost λ net of financial aid γZ. We let the joint distribution of (U_0, U_1) be normal with zero means, unit variances, and covariance $\rho = 0.5$.

We begin by considering the case of a discrete exogenous instrument Z that takes value in $\{0, 1, 2\}$ with probability c_0, c_1, and c_2 respectively (Z is binary if $c_2 = 0$). For instance, Z denotes zero financial aid ($Z = 0$), partial financial aid covering tuition ($Z = 1$), or full financial aid covering tuition and living expenses ($Z = 2$). Here, the LATE $\alpha^*_{z,z'}$ is given by

$$\begin{aligned}\alpha^*_{z,z'} &= E[Y(1) - Y(0) \mid -\lambda + \gamma z < -[Y(1) - Y(0)] \leq -\lambda + \gamma z'] \\ &= \delta + \frac{\phi(-\lambda + \gamma z' + \delta) - \phi(-\lambda + \gamma z + \delta)}{\Phi(-\lambda + \gamma z' + \delta) - \Phi(-\lambda + \gamma z + \delta)},\end{aligned}$$

where ϕ and Φ are the standard normal pdf and cdf respectively and where we make use of $Y(1) - Y(0) \sim N(\delta, 1)$ along with the properties of the truncated normal distribution. The econometrician observes realizations of D and Y and those of an error-laden proxy that codes the information in Z coarsely. First, we consider the proxy W_1 that records only whether individuals are offered any financial aid ($W_1 = 1$) or none ($W_1 = 0$) but that misclassifies this information when the binary indicator U_W equals 1, with $\Pr(U_W = 1) = m$:

$$W_1 = [1 - \mathbf{1}\{Z = 0\}](1 - U_W) + \mathbf{1}\{Z = 0\}U_W.$$

Second, we consider the proxy W_2 that records only whether individuals are offered partial aid ($W_2 = 1$) or not ($W_2 = 0$) but that is misclassified when $U_W = 1$, with probability m:

$$W_2 = \mathbf{1}\{Z = 1\}(1 - U_W) + [1 - \mathbf{1}\{Z = 1\}]U_W.$$

We maintain that $(U_0, U_1) \perp (Z, U_W)$ and $U_W \perp Z$. Applying Theorem 3.2 with $W = W_h$, $h = 1, 2$, gives[13] that the Wald estimand $\beta^h_{0,1}$ is a weighted average of

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

LATEs $\alpha^*_{0,1}$ and $\alpha^*_{1,2}$:

$$\beta^h_{0,1} \equiv \frac{E(Y|W_h=1)-E(Y|W_h=0)}{E(D|W_h=1)-E(D|W_h=0)} = \mu^h_1\alpha^*_{0,1}+\mu^h_2\alpha^*_{1,2}.$$

For binary Z ($c_2=0$), $W_1=W_2$ which we denote by W, and the closer m is to 0 or 1 (the further m is from 0.5) the closer is W to coding the financial aid information Z perfectly. Nevertheless, for $m\neq 0.5$ (or $W \not\perp Z$), $\beta_{0,1}=\alpha^*_{0,1}$, i.e., the Wald estimand using W identifies the same LATE that would have been recovered using Z. Table 1 reports the median bias (MBias) and interquartile range (IQR) (statistics are computed over 1000 replications in all the simulations) of Wald plug-in estimators[14] $\hat{\alpha}_{0,1}$ and $\hat{\beta}_{0,1}$ with $c_0=0.45$ and $c_1=0.55$. In (small) finite samples, IV-based estimators may fail to have moments and thus MBias and IQR may be more informative than the average bias (ABias) and standard deviation (SD). Although our focus here is on identification rather than estimation or finite sample behavior of estimators, the simulations in this section help shed light on features of the Wald and LIV estimators. For instance, the IQR of $\hat{\alpha}_{0,1}$ and $\hat{\beta}_{0,1}$ decrease as the sample size n increases. Further, the IQR of $\hat{\beta}_{0,1}$ increases as m approaches 0.5. In general, ABias and SD behave relatively similarly to MBias and IQR respectively as the sample size increases in our simulations[15].

When Z takes value in $\{0,1,2\}$, calculating $F_{Z|W_h}(z|w)$ for $z=0,1$ and $w=0,1$ shows that W_1 is intensity preserving for Z (note that, in the special case of $m=0$, W_1 is a monotonic function of Z and therefore W_1 is intensity preserving for Z by Proposition 3.3) whereas W_2 is *intensity disrupting* for Z and thus either μ^2_1 or μ^2_2 is negative. In particular, unlike in the constant effect case, $\beta^2_{0,1}$ in this

TABLE 1. Identification of LATE in a Generalized Roy Model with Misclassified Binary IV.

$\alpha^*_{0,1}=1.398$		$\alpha_{0,1}=1.398$		$\beta_{0,1}=1.398$	
		$\hat{\alpha}_{0,1}$		$\hat{\beta}_{0,1}$	
n	m	MBias	IQR	MBias	IQR
500	0.05	0.009	0.688	0.033	0.783
	0.1	0.001	0.676	−0.022	0.860
	0.15	−0.033	0.717	−0.018	1.014
1000	0.05	−0.005	0.485	−0.006	0.582
	0.1	0.003	0.487	0.024	0.605
	0.15	−0.005	0.455	−0.016	0.644
2000	0.05	0.004	0.350	0.008	0.368
	0.1	−0.0002	0.354	−0.002	0.434
	0.15	−0.0001	0.324	0.008	0.484

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

example has the opposite sign than that of both local effects $\alpha^*_{0,1}$ and $\alpha^*_{1,2}$. The simulation results in Table 2 demonstrate these examples, with $c_0 = c_1 = 0.3$ and $c_2 = 0.4$, and reports the median bias (relative to each estimator's probability limit), and IQR of the Wald estimators $\hat{\alpha}_{0,1}$, $\hat{\beta}^1_{0,1}$, and $\hat{\beta}^2_{0,1}$.

Suppose instead that Z is continuous, $Z \sim N(0,1)$ (e.g., the logarithm of continuous financial aid). In this case, MTE is given by

$$\alpha^*(z) = E[Y(1) - Y(0) \mid -[Y(1) - Y(0)] = -\lambda + \gamma z] = \lambda - \gamma z.$$

Consider two proxies W_3 and W_4 for Z

$$W_3 = Z + U_W \quad \text{and} \quad W_4 = (1-\epsilon)Z + \epsilon V.$$

The first proxy W_3 suffers from classical measurement error $U_W \sim N\left(0, \sigma^2\right)$ whereas the second proxy W_4 is as in Proposition 3.6, where $V \sim N(0,1)$ and ϵ is binary with $\Pr(\epsilon = 1) = m$. We maintain that (U_1, U_0), Z, U_W, V, and ϵ are jointly independent. Proposition 3.4 gives that W_3 is intensity preserving for Z and it can be shown (we omit the calculation for brevity) that

$$\beta_3\left(w_3|\sigma^2\right) = \int \mu^{w_3}(z)\alpha^*(z)dz = \lambda - \gamma \frac{\left(-\gamma\,\sigma^2(\delta - \lambda) + w_3\right)}{\left(\sigma^2\left(\gamma^2 + 1\right) + 1\right)},$$

where we sometimes denote the LIV estimand using W_3 by $\beta_3(w_3|\sigma^2)$ to emphasize that it depends on σ^2. Moreover, Proposition 3.6 gives that for all z, W_4 at z is value preserving for Z at z so that the LIV estimands $\alpha(z)$ using Z and $\beta_4(z)$ using W_4 evaluated at z both equal $\alpha^*(z)$. Table 3 demonstrates these results and

TABLE 2. Identification of LATE in a Generalized Roy Model with Misclassified Discrete IV.

$\alpha^*_{0,1} = 1.398$		$\alpha_{0,1} = 1.398$		$\beta^1_{0,1} = 0.896$		$\beta^2_{0,1} = -0.041$	
$\alpha^*_{1,2} = 0.476$		$\hat{\alpha}_{0,1}$		$\hat{\beta}^1_{0,1}$		$\hat{\beta}^2_{0,1}$	
n	m	MBias	IQR	MBias	IQR	MBias	IQR
500	0.05	0.026	0.863	−0.0003	0.370	−0.001	1.138
	0.1	0.025	0.850	0.009	0.417	0.013	1.229
	0.15	−0.018	0.908	−0.023	0.529	0.091	1.484
1000	0.05	−0.049	0.602	−0.006	0.261	0.014	0.865
	0.1	−0.011	0.605	−0.004	0.290	0.028	0.941
	0.15	0.020	0.601	−0.002	0.341	−0.011	1.043
2000	0.05	0.015	0.440	0.00003	0.195	0.006	0.534
	0.1	−0.016	0.429	−0.007	0.224	0.005	0.617
	0.15	−0.006	0.440	−0.014	0.258	0.032	0.730

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

TABLE 3. Identification of MTE in a Generalized Roy Model with Mismeasured Continuous IV.

$\alpha^*(1.5)=0.5$			$\alpha(1.5)=0.5$		$\beta_3(1.5\|0.1)=0.6$ $\beta_3(1.5\|0.5)=0.8$ $\beta_3(1.5\|1)=0.9$		$\beta_4(1.5)=0.5$	
			$\hat{\alpha}(1.5)$		$\hat{\beta}_3(1.5)$		$\hat{\beta}_4(1.5)$	
n	σ^2	m	MBias	IQR	MBias	IQR	MBias	IQR
500	0.1	0.05	−0.016	0.728	−0.044	0.728	−0.026	0.741
	0.5	0.1	0.009	0.775	0.020	0.922	0.027	0.865
	1	0.15	−0.011	0.734	−0.007	1.044	0.013	0.887
1000	0.1	0.05	−0.001	0.494	−0.014	0.507	0.003	0.543
	0.5	0.1	−0.004	0.514	−0.014	0.634	0.003	0.550
	1	0.15	0.001	0.500	−0.031	0.739	−0.012	0.598
2000	0.1	0.05	0.005	0.389	−0.025	0.371	−0.001	0.396
	0.5	0.1	−0.003	0.364	−0.053	0.459	−0.001	0.405
	1	0.15	−0.019	0.376	−0.039	0.516	−0.018	0.431

reports the median bias (relative to each estimator's probability limit) and IQR of the LIV estimators $\hat{\alpha}(z)$, $\hat{\beta}_3(w_3)$ and $\hat{\beta}_4(w_4)$ at $z = w_3 = w_4 = 1.5$. Other values yield qualitatively comparable[16] large sample results. These LIV estimators employ local quadratic kernel estimators using the rule of thumb bandwidth[17] (see e.g., Fan and Gijbels, 1996, p. 111) to estimate the derivatives of the conditional means in the numerators and denominators of $\alpha(z)$, $\beta_3(w_3)$ and $\beta_4(w_4)$. Here, the IQR of $\hat{\alpha}(z)$, $\hat{\beta}_3(w_3)$ and $\hat{\beta}_4(w_4)$ decrease as n increases and the IQR of $\hat{\beta}_4(w)$ generally increases as m increases.

6. CONCLUSION

This paper provides a notion for a proxy vector W for an instrument vector Z in an ordered discrete choice structural system with heterogeneous effects and studies the implications of substituting W for Z on the interpretation of the Wald, LIV, and IV estimands. The proxy need only satisfy an exclusion restriction and that the treatment and outcome are mean independent of the proxy given the instrument. Unlike the instrument, the proxy need not satisfy "monotonicity" nor drive the treatment and it may, under particular specifications, be endogenous. For example, W could code Z with error, with missing observations, or coarsely. We build on, and extend, the analysis in e.g., IA, AI, HV, and HUV to accommodate proxies for instruments. We show that the methods of Wald and LIV using a proxy and the method of IV using a function of a proxy identify weighted averages of LATEs or MTEs. We study necessary and sufficient conditions for nonnegative weights,

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

demonstrating for example that misclassification or classical measurement error in coding Z can yield nonnegative weights. Further, we provide a sufficient condition for the Wald or LIV estimand using W to identify the same local or marginal effects, or weighted averages of these, that would have been recovered using Z. We study this condition and show that it can hold when the proxy codes the instrument values with measurement error or misclassification. Because this condition trivially holds for binary instruments, the Wald or IV estimand using any proxy for a binary instrument identifies the "average causal response," or LATE for binary treatment. Further, this condition permits employing the LIV estimand using W to identify MTE and average effects such as ATE, ATT, and ATU.

The results of this paper complement those in Chalak and White (2011) who distinguish between an instrument and its proxy in linear systems with homogeneous effects (see also Heckman 1996, p. 460) and results in Schennach, White, and Chalak (2012; SWC) who show that, in a triangular system of nonseparable structural equations with continuous Z, D, and Y, the LIV estimand using Z recovers a specific weighted average marginal effect. SWC study estimating this LIV estimand (and weighted averages of it) when Z is observed or using two or more proxies for unobserved[18] Z. Several extensions of this paper's results are of interest including developing a test for the hypotheses of zero or nonnegative causal effect of the treatment when only a proxy for an instrument is available and developing a test for proxy validity based on the result that multiple proxies for an instrument can be used to over-identify the same LATE, MTE, or weighted averages of these. Another extension studies the implications of using proxies on the asymptotic behavior of the Wald, LIV, and IV estimators (see e.g., theorem 3 in IA). Last, it is of interest to build on this paper's results, which relate the Wald, LIV, and IV estimands using proxies to LATEs or MTEs, to study bounding causal effects under restrictions on the measurement error in coding Z. We leave studying these extensions for future research.

NOTES

1. $\mathbf{1}\{A\} = 1$ if A is true and equals 0 otherwise.

2. Throughout, for generic random vectors X_1 and X_2, we denote the cumulative distribution function (cdf) for X_1 and the cdf for X_1 conditional on $X_2 = x_2$ by $F_{X_1}(\cdot)$ and $F_{X_1|X_2}(\cdot|x_2)$ respectively. We denote the corresponding probability density or mass functions (pdf) by $f_{X_1}(\cdot)$ and $f_{X_1|X_2}(\cdot|x_2)$.

3. See also Klein (2010) who approximates the bias of estimators for marginal and related treatment effects resulting from local random departures from monotonicity and Small and Tan (2007) who study a "stochastic monotonicity" condition.

4. In turn, ignorability (and thus A.2) implies that $Y(D(z)) \perp_m Z$ and $D(z) \perp_m Z$ for all $z \in S_Z$. These last conditions permit identifying the average effects of Z on D and Y and suffice for the Wald or LIV estimands to recover LATE or MTE when Z is observed (see the proof of Theorem 3.1 below). For simplicity, we employ the stronger condition A.2.

5. A.1(ii) assumes that treatment receipt and response are *individualistic* since these only depend on one's variables (see e.g., Manski, 2013). A.1(ii) does not restrict an individual's proxy from depending on others' instrument values. Strengthening A.1(ii) to impose individualistic proxies gives the "stable unit treatment value assumption" (see e.g. Rubin, 1986; AIR).

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

6. We let $\perp_m$ denote mean independence and write $X_1 \perp_m X_2|X_3$ if $E(X_1|X_2, X_3) = E(X_1|X_3)$ provided these means exist. We let $\not\perp_m$ denote mean dependence.

7. A discrete scalar Z can be used to code a discrete instrument vector of finite dimension.

8. The analysis may be adapted to let $Z|W = w^*$ for $w^* = w, w'$ in part $(b.i)$ and for all $w^* \in \mathcal{B}(w)$ in part $(b.ii)$ have the same support S_{Z_w}, which need not equal S_Z, and then considering the integrals over S_{Z_w}.

9. Lehmann (1966) studies necessary and sufficient conditions for Z being positively regression dependent on W or, equivalently, W being positively intensity preserving for Z.

10. For discrete Z, we can order the elements of S_Z such that $E(D|Z = \cdot)$ is monotonic over S_Z.

11. From the proof of Proposition 3.6, $\frac{\partial}{\partial w} F_{Z|W}(s|z) = -\Pr(\epsilon = 0)\delta(s - z)$ where it may be convenient to interpret the Dirac delta function $\delta(\cdot)$ such that $\lim_{\tau\to 0} \int \frac{\partial}{\partial w} F^{\tau}_{Z|W}(s|z)ds = \int \frac{\partial}{\partial w} F_{Z|W}(s|z)ds = \int -\Pr(\epsilon = 0)\delta(s - z)ds = -\Pr(\epsilon = 0)$ (see e.g., Bracewell, 1986).

12. In Theorem 4.1(a), we note that if $E(g(W)|Z \geq z_m) = E(g(W))$ for all $z_m \in S_Z\backslash\{z_k\}$ and $E(g(W)|Z \geq z_k) \neq E(g(W))$, then $\beta^g_W = \alpha^*_{k-1,k}$.

13. We slightly adjust the notation for the weights given that the proxies are binary here.

14. The Wald estimator using discrete Z or W replaces the conditional expectations in the numerator and denominator of the Wald estimand with sample averages.

15. Here, the asymptotic variance of the IV estimator using binary W (which equals $\hat{\beta}_{0,1}$) is proportional to $\frac{Var(W)}{Cov(D,W)^2}$ (see e.g., Theorem 3 in IA). Since W is Bernoulli distributed, it can be shown that $Var(W) \to \frac{1}{4}$ and $Cov(D, W) \to 0$ as $m \to \frac{1}{2}$.

16. For instance, at different $z = w_4$ values, the median biases of $\hat{\alpha}(z)$ and $\hat{\beta}_4(w_4)$ vary relatively similarly and, in general, the larger m is the more the IQR of $\hat{\beta}_4(w_4)$ exceeds that of $\hat{\alpha}(z)$.

17. We leave studying LIV estimators and optimal bandwidth choice to other work in order to keep a sharp focus here on identification of MTE using mismeasured instruments.

18. Here we study the consequences of substituting one proxy W for Z on the interpretation of Wald, LIV, and IV estimands in a discrete choice structural system. Because the setup and goals of SWC and this paper are different, the assumptions imposed on proxies differ accordingly. For example, SWC require two proxies for unobserved Z given by $W_h = Z + U_h$, $h = 1, 2$, with unobserved U_1 and U_2 satisfying assumptions including $U_2 \perp Z$ and $E(U_1|Z, U_2) = 0$. We don't require this here; instead it suffices that one proxy W satisfies A.3.

19. If desirable, one can ensure that D takes value in $\{0, ..., J\}$ by reparametrizing this example such that $D = \delta Z + a + \upsilon$ and $Y = \theta(D - a) + \epsilon$ where a is a constant (e.g., set $a = 3$ if $\upsilon = 0$ and δ takes value in $\{-1, 0, 1\}$). The remaining analysis holds true.

20. For example, if δ takes value in $\{1, 2, 3\}$, for $\Pr(U_W = u, \delta = \upsilon) = p_{u\upsilon}$, put $p_{01} = \frac{1}{12}$, $p_{02} = \frac{3}{12}$, $p_{03} = \frac{2}{12}$, $p_{11} = p_{12} = \frac{5}{36}$, and $p_{13} = \frac{2}{9}$. Then $E(\delta|U_W) = E(\delta) = \frac{13}{6}$ and $Cov(\frac{c}{\delta}, U_W) = \frac{c}{108}$.

21. For example, if symmetric $\delta|U_W$ takes value in $\{1, 2, 3\}$ then $E(\delta|U_W) = 2$ and the skewness $E(\theta\delta|U_W) = 0$ whereas $E(\theta|U_W = u) = \frac{-1}{3\sqrt{2p_u}}$ with $p_u = \Pr(\delta = 1|U_W = u) = \Pr(\delta = 3|U_W = u)$.

REFERENCES

Abrevaya, J. & J. Hausman (1999) Semiparametric estimation with mismeasured dependent variables: An application to duration models for unemployment spells. *Annals of Economics and Statistics / Annales d'Économie et de Statistique* 55/56, 243–275.

Acemoglu, D., S. Johnson, & J. Robinson (2001) The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91, 1369–1401.

An, Y. & Y. Hu (2012) Well-posedness of measurement error models for self-reported data. *Journal of Econometrics* 168, 259–269.

Angrist, J., K. Graddy, & G. Imbens (2000) The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies* 67, 499–527.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Angrist, J. & G. 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.
Angrist, J., G. Imbens, & D. Rubin (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–472 (with discussion).
Ashenfelter, O. & A. Krueger (1994) Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review* 84, 1157–1173.
Bagnoli, M. & T. Bergstrom (2005) Log-concave probability and its applications. *Economic Theory* 26, 445–469.
Bartle, R. (1966) *Elements of Integration*. Wiley.
Bracewell, R. (1986) *The Fourier Transform and Its Applications*. McGraw-Hill.
Card, D. (1999) The causal effect of education on earnings. In O. Ashenfelter & D. Card (eds.), *Handbook of Labor Economics*, vol. 3, Part A. Elsevier.
Chalak, K. & White H. (2011) An extended class of instrumental variables for the estimation of causal effects. *Canadian Journal of Economics* 44, 1–51.
Corbae, D., M. Stinchcombe, & J. Zeman (2009) *An Introduction to Mathematical Analysis for Economic Theory and Econometrics*. Princeton University Press.
Dawid, A.P. (1979) Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B*. 41, 1–31 (with discussion).
Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman and Hall.
Hausman J., J. Abrevaya, & F.M. Scott-Morton (1998) Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* 87, 239–269.
Heckman, J. (1996) Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association* 91, 459–462.
Heckman, J. (2010) Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature* 48, 356–98.
Heckman, J. & E. Vytlacil (2005) Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73, 669–738.
Heckman, J., S. Urzua, & E. Vytlacil (2006) Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88, 389–432.
Hernán, M. & J. Robins (2006) Instruments for causal inference: An epidemiologist's dream? *Epidemiology* 17, 360–372.
Holland, P.W. (1986) Statistics and causal inference. *Journal of the American Statistical Association* 81, 945–970 (with discussion).
Horowitz, J. & C. Manski (1995) Identification and robustness with contaminated and corrupted data. *Econometrica* 63, 281–302.
Hotz, V., S. McElroy, & S. Sanders (2005) Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *Journal of Human Resources* 40, 683–715.
Hotz, V., C. Mullin, & S. Sanders (1997) Bounding causal effects using data from a contaminated natural experiment: Analyzing the effects of teenage childbearing. *The Review of Economic Studies* 64, 575–603.
Hu, Y. (2008) Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics* 144, 27–61.
Imbens, G. & J. Angrist (1994) Identification and estimation of local average treatment effects. *Econometrica* 61, 467–476.
Klein, T. (2010) Heterogeneous treatment effects: Instrumental variables without monotonicity? *Journal of Econometrics* 155, 99–116.
Lehmann, E. (1966) Some concepts of dependence. *The Annals of Mathematical Statistics* 37, 1137–1153.
Leoni, G. (2009) *A First Course in Sobolev Spaces*. American Mathematical Society.
Lewbel, A. (2007) Estimation of average treatment effects with misclassification. *Econometrica* 75, 537–551.
Li, Q. & J. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Mahajan, A. (2006) Identification and estimation of regression models with misclassification. *Econometrica* 74, 631–665.
Manski, C. (2013) Identification of treatment response with social interactions. *The Econometrics Journal* 16, S1–S23.
Manski, C. & J. Pepper (2000) Monotone instrumental variables: With an application to the returns to schooling. *Econometrica* 68, 997–1010.
Rubin, D. (1986) Statistics and causal inference: Comment: which ifs have causal answers. *Journal of the American Statistical Association* 81, 961–962.
Schennach, S.M., H. White, & K. Chalak (2012) Local indirect least squares and average marginal effects in nonseparable structural systems. *Journal of Econometrics* 166, 282–302.
Shaked, M. & J.G. Shanthikumar (1994) *Stochastic Orders and Their Applications*. Associated Press.
Small, D. & Z. Tan (2007) A Stochastic Monotonicity Assumption for the Instrumental Variables Method. Working paper, The Wharton School of the University of Pennsylvania Department of Statistics.
Vytlacil, E. (2002) Independence, monotonicity, and latent index models: An equivalence result. *Econometrica* 70, 331–341.
Vytlacil, E. (2006) Ordered discrete-choice selection models: Equivalence, nonequivalence, and representation results. *Review of Economics and Statistics* 88, 578–581.
Wald, A. (1940) The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11, 284–300.

APPENDIX A. A Proxy Violating the Instrument Assumptions

It is useful to compare the assumptions imposed on instruments and proxies to clearly distinguish Z and W. For this, we provide examples in which a locally intensity and value preserving proxy W that satisfies A.3 fails the instrument assumptions satisfied by Z. In particular, W fails monotonicity in A.1. Further, under particular specifications, W fails mean ignorability from potential outcomes and it therefore fails A.2. This demonstrates that W need not be an instrument. Nevertheless, the method of Wald using W can identify a local treatment effect.

To illustrate, consider the training program example with Z, $S_Z = \{0, 1, 2, 3\}$, denoting the financial aid level. The econometrician does not observe realizations of Z but only those of its misclassified proxy $W = (1 - U_W)Z + U_W(3 - Z)$ that satisfies A.3. Here, nondegenerate U_W takes value in $\{0, 1\}$ and W codes Z correctly when $U_W = 0$ and its reverse (or mirror image) $3 - Z$ when $U_W = 1$. Let $D = q(Z, U_D)$ denote joining the training program (or the number of courses completed), and suppose that Z satisfies monotonicity so that for all $z, z' \in S_Z$ either $\Pr[q(z, U_D) \leq q(z', U_D)] = 1$ or $\Pr[q(z, U_D) \geq q(z', U_D)] = 1$. Substituting for Z in q gives the nonstructural equation $D = q\left(\frac{W - 3U_W}{1 - 2U_W}, U_D\right)$ and W fails monotonicity in this equation since for all $w \neq w' \in S_W$, $\Pr\left[q\left(\frac{w - 3U_W}{1 - 2U_W}, U_D\right) \leq q\left(\frac{w' - 3U_W}{1 - 2U_W}, U_D\right)\right] \neq 1$. Further, it can be verified that W at $(0, 3)$ or $(1, 2)$ is intensity preserving for Z whereas W at e.g., $(0, 1)$ or $(0, 2)$ is intensity disrupting for Z. Similarly, W at $(1, 2)$ is value preserving for Z at 1 whereas W at $(0, 1)$ is value disrupting for Z at e.g., 0. It follows from Corollary 3.7 that, under A.1-A.3, the Wald estimand $\beta_{1,2}$ using W identifies the LATE or "average causal response" $\alpha^*_{1,2}$ which would have been recovered using the Wald estimand $\alpha_{1,2}$ had Z been observed. For these results, we do not restrict the structural equations for Y and D apart from exclusion restrictions and monotonicity.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Next, we construct examples in which A.2 and A.3 hold but W is endogenous. For simplicity, we consider the linear random coefficient model for Y :

$$Y = r(D, U_Y) = \theta D + \epsilon \quad \text{with} \quad D = q(Z, U_D) \quad \text{and} \quad W = s(Z, U_W),$$

where $(U_D, U_Y, U_W) \perp Z$ and (θ, ϵ) are functions of U_Y. Given A.2, A.3 then requires that for all (z, w) in the support of (Z, W)

$$E(D|Z = z, W = w) = E[q(z, U_D)] \text{ and}$$
$$E(Y|Z = z, W = w) = E[\theta q(z, U_D)] + E(\epsilon), \tag{1}$$

and we examine cases in which $Y(d) \not\perp_m W$, so that W is endogenous, and for some $w \in S_W$:

$$E[Y(d)|W = w] = E(\theta d + \epsilon | W = w) \neq d E(\theta) + E(\epsilon). \tag{2}$$

For this random coefficient model, a proxy that satisfies conditions (1) and (2) respects A.3 but is nevertheless endogenous. For example, consider the proxy $W = (1 - U_W)Z + U_W(3 - Z)$ for Z described above and suppose that q is also linear with integer random coefficients δ and υ that depend on U_D:

$$D = q(Z, U_D) = \delta Z + \upsilon.$$

Provided that δ does not change sign, Z satisfies monotonicity (and strict A.4 if δ is also nonzero) whereas W fails monotonicity as discussed above[19]. Given $(U_D, U_Y, U_W) \perp Z$ and letting $E(\epsilon|U_W) = 0$ and $E(\upsilon|\theta, U_W) = 0$, we have

$$E(D|Z = z, W = w) = z E(\delta | U_W = \frac{w - z}{(3 - 2z)}),$$
$$E(Y|Z = z, W = w) = z E(\theta\delta | U_W = \frac{w - z}{(3 - 2z)}), \text{ and}$$
$$Cov(Y(d), W) = Cov[\theta d + \epsilon, Z + (3 - 2Z)U_W] = d(3 - 2E(Z))Cov(\theta, U_W).$$

Provided $E(Z) \neq \frac{3}{2}$ (and $d \neq 0$), conditions (1) and (2) then obtain if for all $u_w \in \{0, 1\}$

$$E(\delta|U_W = u_w) = E(\delta), \ E(\theta\delta|U_W = u_w) = E(\theta\delta), \text{ and}$$
$$E(\theta|U_W = 0) \neq E(\theta|U_W = 1). \tag{3}$$

A very simple example in which these restrictions hold arises when $\theta\delta = c$, a nonzero constant, and $E(\delta|U_W) = E(\delta)$ but[20] $Cov(\theta, U_W) \neq 0$. Other simple examples obtain by letting the expectations in (3) denote moments of the distribution of δ. For instance, let the measurement error U_W alter the distribution of δ but preserve its symmetry. For example, if $E(\delta|U_W) = 0$ and $\theta = \delta^2$ then $E(\theta\delta|U_W) = 0$ since the skewness of $\delta|U_W$ is zero, and it suffices that U_W relates to the variance of δ, $E(\theta|U_W = 0) < E(\theta|U_W = 1)$. Alternatively, to ensure monotonicity for Z, let the support of symmetric $\delta|U_W$ be positive with $E(\delta|U_W) = E(\delta)$ and set $\theta = \frac{1}{\delta}\left(\frac{\delta - E(\delta)}{\sqrt{Var(\delta|U_W)}}\right)^3$ so that the skewness $E(\theta\delta|U_W) = 0$ but[21] $E(\theta|U_W = 0) < E(\theta|U_W = 1)$. In these examples, the average treatment effect is higher when Z is misclassified.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

In summary, in these examples the proxy W satisfies A.3, fails monotonicity in A.1, and W at e.g., $(1, 2)$ is intensity preserving for Z and value preserving for Z at 1. Further, W can be endogenous, albeit this requires particular specifications.

APPENDIX B. Mathematical Proofs

Proof of Theorem 3.1. (a) Similar to the proof of theorem 1 in AI, let $\rho_j(z) \equiv \mathbf{1}\{V_j \leq \nu(z)\} = \mathbf{1}\{j \leq D(z)\}$ for $z \in S_Z$ and $j \in S_D \cup \{J+1\}$, then $\rho_0(z) = 1$ and $\rho_{J+1}(z) = 0$ for $z \in S_Z$. A.1 and A.2 then give

$$\begin{aligned} E(Y|Z=z) = E[Y(D(z))] &= E\left\{\sum_{j=0}^{J}[\rho_j(z)-\rho_{j+1}(z)]Y(j)\right\} \\ &= E[\rho_0(z)Y(0)]+\sum_{j=1}^{J}E\left\{\rho_j(z)[Y(j)-Y(j-1)]\right\}-E[\rho_{J+1}(z)Y(J)] \\ &= E[Y(0)]+\sum_{j=1}^{J}E\left\{\rho_j(z)[Y(j)-Y(j-1)]\right\}. \end{aligned}$$

A.1 and $\Pr\left[D(z) < D\left(z'\right)\right] > 0$ ensure that $\nu(z) < \nu\left(z'\right)$ and $\rho_j\left(z'\right) - \rho_j(z)$ equals 0 or 1 w.p.1 for all $j \in S_D$, thus

$$\begin{aligned} E\left(Y|Z=z'\right) - E(Y|Z=z) &= \sum_{j=1}^{J}E\left\{\left[\rho_j\left(z'\right)-\rho_j(z)\right][Y(j)-Y(j-1)]\right\} \\ &= \sum_{j=1}^{J}E\left[Y(j)-Y(j-1)|\nu(z) < V_j \leq \nu\left(z'\right)\right]\Pr[\nu(z) < V_j \leq \nu\left(z'\right)]. \end{aligned}$$

Similarly, under A.1, $D(z) = \sum_{j=0}^{J}[\rho_j(z)-\rho_{j+1}(z)]j$ and given A.2 we have

$$E\left(D|Z=z'\right) - E(D|Z=z) = E\left[D\left(z'\right)-D(z)\right] = \sum_{j=1}^{J}\Pr\left[\nu(z) < V_j \leq \nu\left(z'\right)\right].$$

$\Pr\left[D(z) < D\left(z'\right)\right] > 0$ ensures that $\sum_{j=1}^{J}\Pr\left[\nu(z) < V_j \leq \nu\left(z'\right)\right] > 0$ and therefore we have $\alpha_{z,z'} = \alpha^*_{z,z'}$. It is immediate that $0 \leq \omega_j^{z,z'} \leq 1$ and $\sum_{j=1}^{J}\omega_j^{z,z'} = 1$.

(b) A.1-A.2, $\rho_j(z) = \mathbf{1}\left\{V_j \leq \nu(z)\right\}$, and the law of iterated expectations give

$$\begin{aligned} E(Y|Z=z) &= E[Y(0)]+\sum_{j=1}^{J}E\left\{\mathbf{1}\left\{V_j \leq \nu(z)\right\}[Y(j)-Y(j-1)]\right\} \\ &= E[Y(0)]+\sum_{j=1}^{J}\int_{-\infty}^{\nu(z)}E\left[Y(j)-Y(j-1)|V_j=v\right]f_{V_j}(v)dv \\ &\equiv E[Y(0)]+\sum_{j=1}^{J}G_j(\nu(z)). \end{aligned}$$

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Since $\nu(\cdot)$ is differentiable at z and $E\big[Y(j)-Y(j-1)|V_j=\cdot\big]$ and $f_{V_j}(\cdot)$ are continuous at $\nu(z)$, it follows from the Lebesgue differentiation theorem that (see e.g., footnote 53 in Heckman (2010) and apply the chain rule)

$$\frac{\partial}{\partial z}E(Y|Z=z)=\sum_{j=1}^{J}E\big[Y(j)-Y(j-1)|V_j=\nu(z)\big]\,f_{V_j}(\nu(z))\frac{\partial}{\partial z}\nu(z).$$

A similar derivation gives that

$$E(D|Z=z)=\sum_{j=1}^{J}\Pr\big(V_j\le\nu(z)\big)=\sum_{j=1}^{J}\int_{-\infty}^{\nu(z)}f_{V_j}(v)dv\equiv\sum_{j=1}^{J}F_{V_j}(\nu(z))$$

with $\frac{\partial}{\partial z}E(D|Z=z)=\sum_{j=1}^{J}f_{V_j}(\nu(z))\frac{\partial}{\partial z}\nu(z)$. Since $\frac{\partial}{\partial z}\nu(z)\neq 0$ and $f_{V_j}(\nu(z))>0$ for some $j\in S_D\backslash\{0\}$, $\alpha(z)=\alpha^*(z)$. Clearly, $0\le\omega_j(z)\le 1$ and $\sum_{j=1}^{J}\omega_j(z)=1$. ■

In proving Theorem 3.2, we employ the following Lemma which shows that $\beta_{w,w'}$ and $\beta(w)$ are weighted averages of α_{z_{k-1},z_k} or $\alpha(z)$, with necessary and sufficient condition for nonnegative weights. This result does not require A.1(ii), A.2, or A.4. For the remainder of this Appendix, if $E(D|Z=z)=E\big(D|Z=z'\big)$ for $z,z'\in S_Z$ it is convenient to define $\alpha_{z,z'}\equiv c$ for some constant $c<\infty$, and similarly if $\frac{\partial}{\partial z}E(D|Z=z)=0$ we let $\alpha(z)\equiv c'$ for some constant $c'<\infty$. The definitions of $\mu_k^{w,w'}$, $\mu^{w,w'}(z)$, and $\mu^{w}(z)$ are given in Theorem 3.2.

LEMMA B.1. *Assume A.1(i) and A.3 and let $w,w'\in S_W$ with $E(Y|W=w^*)<\infty$ for $w^*=w,w'$. (a) Suppose that $S_Z=\{z_0,z_1,...,z_K\}$ with $z_{k-1}<z_k$ for $k=1,...,K$, $E(Y|Z=z_k)<\infty$ for all $z_k\in S_Z$, and that, for all $z_{k-1},z_k\in S_Z$, if $E(D|Z=z_k)=E(D|Z=z_{k-1})$ then $E(Y|Z=z_k)=E(Y|Z=z_{k-1})$. Then, provided $E(D|W=w)\neq E(D|W=w')$, we have*

$$\beta_{w,w'}=\sum_{k=1}^{K}\mu_k^{w,w'}\alpha_{k-1,k},\quad with\quad\sum_{k=1}^{K}\mu_k^{w,w'}=1.$$

(b.i) Set $\kappa=1$ and suppose that $F_{Z|W}(\cdot|w)$ and $F_{Z|W}(\cdot|w')$ are absolutely continuous with support S_Z, $E(D|Z=\cdot)$ and $E(Y|Z=\cdot)$ are absolutely continuous on S_Z, and that for all $z\in S_Z$, if $\frac{\partial}{\partial z}E(D|Z=z)=0$ then $\frac{\partial}{\partial z}E(Y|Z=z)=0$. Then, provided $E(D|W=w)\neq E(D|W=w')$, we have

$$\beta_{w,w'}=\int_{S_Z}\alpha(z)\mu^{w,w'}(z)dz,\quad with\quad\int_{S_Z}\mu^{w,w'}(z)dz=1.$$

(b.ii) In addition, set $\ell=1$ and assume the regularity conditions in part ($b.ii$) of Theorem 3.2. Then, provided $\frac{\partial}{\partial w}E(D|W=w)\neq 0$, we have

$$\beta(w)=\int_{S_Z}\alpha(z)\mu^{w}(z)dz,\quad with\quad\int_{S_Z}\mu^{w}(z)dz=1.$$

(c) Further, suppose that in (a) and (b) $E(D|Z=z)$ is (strictly) monotonic over S_Z. Then $0\le\mu_k^{w,w'}$ for $k=1,...,K$ or $0\le\mu^{w,w'}(z)$ for a.e. $z\in S_Z$ if (and only if) W at (w,w') is intensity preserving for Z. Also, $0\le\mu^{w}(z)$ for a.e. $z\in S_Z$ if (and only if) W at w is intensity preserving for Z.

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Note that A.1 and A.2 ensure that $E(D|Z=z_k)=E(D|Z=z_{k-1})$ implies $E(Y|Z=z_k)=E(Y|Z=z_{k-1})$ or that $\frac{\partial}{\partial z}E(D|Z=z)=0$ implies $\frac{\partial}{\partial z}E(Y|Z=z)=0$. Further, (strict) A.4 ensures that $E(D|Z=z)$ is (strictly) monotonic over S_Z, in which case W at (w,w') (or at w) being intensity preserving for Z is sufficient (and necessary) for nonnegative weights.

Proof of Lemma B.1. (a) By A.3

$$E(Y|W=w)=\sum_{m=0}^{K}E(Y|Z=z_m)f_{Z|W}(z_m|w).$$

Thus, we have

$$\begin{aligned}
E\left(Y|W=w'\right)-E(Y|W=w)&=\sum_{m=0}^{K}\left[f_{Z|W}\left(z_m|w'\right)-f_{Z|W}(z_m|w)\right]E(Y|Z=z_m)\\
&=\sum_{m=1}^{K}\left[f_{Z|W}\left(z_m|w'\right)-f_{Z|W}(z_m|w)\right]\left[E(Y|Z=z_m)-E(Y|Z=z_0)\right]\\
&=\sum_{m=1}^{K}\left[f_{Z|W}\left(z_m|w'\right)-f_{Z|W}(z_m|w)\right]\sum_{k=1}^{m}\left[E(Y|Z=z_k)-E(Y|Z=z_{k-1})\right]\\
&=\sum_{k=1}^{K}\left[E(Y|Z=z_k)-E(Y|Z=z_{k-1})\right]\left[\sum_{m=k}^{K}\left[f_{Z|W}\left(z_m|w'\right)-f_{Z|W}(z_m|w)\right]\right]\\
&=\sum_{k=1}^{K}\alpha_{k-1,k}\left[E(D|Z=z_k)-E(D|Z=z_{k-1})\right]\left[F_{Z|W}(z_{k-1}|w)-F_{Z|W}\left(z_{k-1}|w'\right)\right].
\end{aligned}$$

where we use $f_{Z|W}(z_0|w)=1-\sum_{m=1}^{K}f_{Z|W}(z_m|w)$ in the second equality and that, for all $z_{k-1},z_k\in S_Z$, $E(D|Z=z_k)=E(D|Z=z_{k-1})$ implies $E(Y|Z=z_k)=E(Y|Z=z_{k-1})$ with $\alpha_{k-1,k}=c$ in the last equality. A similar derivation gives

$$\begin{aligned}
E(D|W=w')-E(D|W=w)=&\sum_{k=1}^{K}\left[E(D|Z=z_k)-E(D|Z=z_{k-1})\right]\\
&\times\left[F_{Z|W}(z_{k-1}|w)-F_{Z|W}\left(z_{k-1}|w'\right)\right].
\end{aligned}$$

$E\left(D|W=w'\right)\neq E(D|W=w)$ gives $\beta_{w,w'}=\sum_{k=1}^{K}\mu_k^{w,w'}\alpha_{k-1,k}$. Clearly, $\sum_{k=1}^{K}\mu_k^{w,w'}=1$.

$(b.i)$ By A.3, we have

$$E(Y|W=w)=\int_{S_Z}E(Y|W=w,Z=z)f_{Z|W}(z|w)dz=\int_{S_Z}E(Y|Z=z)f_{Z|W}(z|w)dz.$$

Integration by parts applies over S_Z since $E(Y|Z=\cdot)$, $F_{Z|W}(\cdot|w)$, and $F_{Z|W}(\cdot|w')$ are absolutely continuous on S_Z and gives

$$\begin{aligned}
E\left(Y|W=w'\right)-E(Y|W=w)&=\int_{S_Z}E(Y|Z=z)\left[f_{Z|W}\left(z|w'\right)-f_{Z|W}(z|w)\right]dz\\
&=E(Y|Z=z)\left[F_{Z|W}\left(z|w'\right)-F_{Z|W}(z|w)\right]\Big|_{\underline{z}}^{\bar{z}}
\end{aligned}$$

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

$$-\int_{S_Z}\left[\frac{\partial}{\partial z}E(Y|Z=z)\right]\left[F_{Z|W}\left(z|w'\right)-F_{Z|W}(z|w)\right]dz$$
$$=\int_{S_Z}\alpha(z)\left[\frac{\partial}{\partial z}E(D|Z=z)\right]\left[F_{Z|W}(z|w)-F_{Z|W}\left(z|w'\right)\right]dz,$$

with $\underline{z}$ and $\bar{z}$ the (possibly infinite) infimum and supremum over S_Z and since for all $z\in S_Z$, $\frac{\partial}{\partial z}E(D|Z=z)=0$ implies $\frac{\partial}{\partial z}E(Y|Z=z)=0$ with $\alpha(z)=c'$. A similar derivation gives

$$E\left(D|W=w'\right)-E(D|W=w)=\int_{S_Z}\left[\frac{\partial}{\partial z}E(D|Z=z)\right]\left[F_{Z|W}(z|w)-F_{Z|W}\left(z|w'\right)\right]dz.$$

Since $E\left(D|W=w'\right)\neq E(D|W=w)$, $\beta_{w,w'}=\int_{S_Z}\alpha(z)\mu^{w,w'}(z)dz$ with $\int_{S_Z}\mu^{w,w'}(z)dz=1$.

$(b.ii)$ Since $\frac{\partial}{\partial z}E(D|Z=z)=0$ implies $\frac{\partial}{\partial z}E(Y|Z=z)=0$ with $\alpha(z)=c'$ for all $z\in S_Z$,

$$\frac{\partial}{\partial w}E(Y|W=w)=\frac{\partial}{\partial w}\left[E(Y|Z=z)F_{Z|W}(z|w)\Big|_{\underline{z}}^{\bar{z}}-\int_{S_Z}\left[\frac{\partial}{\partial z}E(Y|Z=z)\right]F_{Z|W}(z|w)dz\right]$$
$$=-\int_{S_Z}\alpha(z)\left[\frac{\partial}{\partial z}E(D|Z=z)\right]\frac{\partial}{\partial w}F_{Z|W}(z|w)dz,$$

where the interchange of integral and derivative follows from e.g., Corbae, Stinchcombe, and Zeman (2009, Thm. 7.5.17) or Bartle (1966, Cor. 5.9). Similarly,

$$\frac{\partial}{\partial w}E(D|W=w)=-\int_{S_Z}\left[\frac{\partial}{\partial z}E(D|Z=z)\right]\left[\frac{\partial}{\partial w}F_{Z|W}(z|w)\right]dz.$$

Since $\frac{\partial}{\partial w}E(D|W=w)\neq 0$, we have $\beta(w)=\int_{S_Z}\alpha(z)\mu^{w}(z)dz$ with $\int_{S_Z}\mu^{w}(z)dz=1$.

(c) Since $E(D|Z=z)$ is (strictly) monotonic over S_Z, $0\leq\mu_k^{w,w'}$ for $k=1,...,K$ or $0\leq\mu^{w,w'}(z)$ for a.e. $z\in S_Z$ if (and only if) either $F_{Z|W}(z|w)\geq F_{Z|W}(z|w')$ for all $z\in S_Z$ or $F_{Z|W}(z|w)\leq F_{Z|W}(z|w')$ for all $z\in S_Z$. Similarly, $0\leq\mu^{w}(z)$ for a.e. $z\in S_Z$ if (and only if) either $\frac{\partial}{\partial w}F_{Z|W}(z|w)\leq 0$ for a.e. $z\in S_Z$ or $\frac{\partial}{\partial w}F_{Z|W}(z|w)\geq 0$ for a.e. $z\in S_Z$. ■

Proof of Theorem 3.2. (a) The result obtains from Lemma B.1 and Theorem 3.1 since, for $k=1,...,K$, $\alpha_{k-1,k}=\alpha^*_{k-1,k}$ if $E(D|Z=z_k)\neq E(D|Z=z_{k-1})$ and we let $\alpha^*_{k-1,k}=0$ otherwise. (b) We demonstrate the result for $(b.ii)$. A similar argument applies in $(b.i)$. Since $F_{V_j}(v(\cdot))$ and $G_j(v(\cdot))$, $j=1,...,J$, are absolutely continuous on S_Z, so are $E(D|Z=\cdot)$ and $E(Y|Z=\cdot)$. From Lemma B.1,

$$\frac{\partial}{\partial w}E(D|W=w)=-\int_{S_Z}\left[\frac{\partial}{\partial z}E(D|Z=z)\right]\left[\frac{\partial}{\partial w}F_{Z|W}(z|w)\right]dz.$$

From Theorem 3.1 and since $v(\cdot)$ is differentiable at a.e. $z\in S_Z$, the chain rule gives that for a.e. $z\in S_Z$ (see e.g., Leoni 2009, Cor. 3.49),

$$\frac{\partial}{\partial z}E(D|Z=z)=\sum_{j=1}^{J}\frac{\partial}{\partial z}\int_{-\infty}^{v(z)}f_{V_j}(v)dv=\sum_{j=1}^{J}f_{V_j}(v(z))\frac{\partial}{\partial z}v(z).$$

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Similarly, Lemma B.1, Theorem 3.1, and the chain rule give

$$\begin{aligned}\frac{\partial}{\partial w}E(Y|W=w) &= -\int_{S_Z}\left[\frac{\partial}{\partial z}E(Y|Z=z)\right]\left[\frac{\partial}{\partial w}F_{Z|W}(z|w)\right]dz\\ &= -\int_{S_Z}\left[\sum_{j=1}^{J}\frac{\partial}{\partial z}\int_{-\infty}^{\nu(z)}E\left[Y(j)-Y(j-1)|V_j=v\right]f_{V_j}(v)dv\right]\\ &\quad\times\left[\frac{\partial}{\partial w}F_{Z|W}(z|w)\right]dz\\ &= -\int_{S_Z}\left[\sum_{j=1}^{J}E\left[Y(j)-Y(j-1)|V_j=\nu(z)\right]f_{V_j}(\nu(z))\frac{\partial}{\partial z}\nu(z)\right]\\ &\quad\times\left[\frac{\partial}{\partial w}F_{Z|W}(z|w)\right]dz\\ &= -\int_{S_Z}\alpha^*(z)\left[\frac{\partial}{\partial z}E(D|Z=z)\right]\left[\frac{\partial}{\partial w}F_{Z|W}(z|w)\right]dz,\end{aligned}$$

where we use $\alpha^*(z)=0$ if $\sum_{j=1}^{J}f_{V_j}(\nu(z))\frac{\partial}{\partial z}\nu(z)=0$, $z\in S_Z$, in the last equality. Since $\frac{\partial}{\partial w}E(D|W=w)\neq 0$, $\beta(w)=\int_{S_Z}\alpha^*(z)\mu^w(z)dz$. ■

Proof of Proposition 3.3. Let $f:S_Z\to S_W$ be a nondecreasing monotonic function then

$$F_{Z|W}(z|w)\equiv\Pr(Z\leq z|W=w)=\begin{cases}0 & \text{if } f(z)<w\\ c\in[0,1] & \text{if } f(z)=w\\ 1 & \text{if } f(z)>w\end{cases}.$$

Then, for all $w,w'\in S_W$ with $w<w'$ and all $z\in S_Z$, we have

$$0\leq F_{Z|W}(z|w)-F_{Z|W}(z|w')=\begin{cases}0 & \text{if } f(z)<w<w'\\ c\in[0,1] & \text{if } w\leq f(z)<w'\\ 1-c',\ c'\in[0,1] & \text{if } w<w'\leq f(z)\end{cases}.$$

Also, we have that for all $w\in S_W\subseteq\mathbb{R}$, $\frac{\partial}{\partial w}F_{Z|W}(z|w)\leq 0$ for a.e. $z\in S_Z$, provided $\frac{\partial}{\partial w}F_{Z|W}(\cdot|\cdot)$ exists. A symmetric argument applies for f a nonincreasing monotonic function.

To show that monotonicity of f is not necessary for intensity preservation, suppose that $S_Z=\{0,1,2,3\}$ and that $P(Z=z)=\frac{1}{4}$ for $z=0,1,2,3$. Let $W=\mathbf{1}\{Z=1 \text{ or } Z=3\}$. We have $F_{Z|W}(z|w)=\sum_{k=0}^{z}\Pr(Z=k|W=w)$, and by Bayes' theorem

$$\Pr(Z=k|W=w)=\frac{\Pr(W=w|Z=k)\Pr(Z=k)}{\sum_{m=0}^{3}\Pr(W=w|Z=m)\Pr(Z=m)}.$$

Thus, $F_{Z|W}(0|0)=\frac{1}{2}>0=F_{Z|W}(0|1)$, $F_{Z|W}(1|0)=\frac{1}{2}=F_{Z|W}(1|1)$, $F_{Z|W}(2|0)=1>\frac{1}{2}=F_{Z|W}(2|1)$, and $F_{Z|W}(3|0)=1=F_{Z|W}(3|1)$ and W is positively intensity preserving for Z. ■

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Proof of Proposition 3.4. The result follows from Lehmann (1966, Example 12). ■

Proof of Corollary 3.5. The result follows from Lemma B.1(c) since A.1–A.2 and (strict) A.4 ensure that $E(D|Z=\cdot)$ is (strictly) monotonic over $S_Z \subseteq \mathbb{R}$. ■

Proof of Proposition 3.6. For all $z \in S_Z$ and $w \in S_W$, $Z \perp (V, \epsilon)$ and $V \perp \epsilon$ give

$$\begin{aligned} F_{W|Z}(w|z) &= \Pr(W \le w|Z=z) = \Pr((1-\epsilon)Z + \epsilon V \le w|Z=z) = \Pr((1-\epsilon)z + \epsilon V \le w) \\ &= \Pr(z \le w)\Pr(\epsilon = 0) + \Pr(V \le w)\Pr(\epsilon = 1) \\ &= \begin{cases} \Pr(V \le w)\Pr(\epsilon = 1) & \text{if } w < z \\ \Pr(\epsilon = 0) + \Pr(V \le w)\Pr(\epsilon = 1) & \text{if } z \le w \end{cases}. \end{aligned}$$

It follows that

$$f_{W|Z}(w|z) = f_V(w)\Pr(\epsilon = 1) \text{ for all } z \in S_Z \text{ and } w \in S_W \backslash \{z\}.$$

Similarly, for all $w \in S_W$, $Z \perp (V, \epsilon)$ and $V \perp \epsilon$ give $V \perp (\epsilon, Z)$, and we have

$$\begin{aligned} F_W(w) &= \Pr(W \le w) = \Pr((1-\epsilon)Z + \epsilon V \le w) \\ &= \int_{S_Z} \Pr(Z \le w|\epsilon = 0, Z = z)\Pr(\epsilon = 0|Z = z) f_Z(z) dz \\ &\quad + \int_{S_Z} \Pr(V \le w|\epsilon = 1, Z = z)\Pr(\epsilon = 1|Z = z) f_Z(z) dz \\ &= \Pr(\epsilon = 0) \int_{S_Z} \Pr(z \le w) f_Z(z) dz + \Pr(V \le w)\Pr(\epsilon = 1) \int_{S_Z} f_Z(z) dz \\ &= \Pr(\epsilon = 0)\Pr(Z \le w) + \Pr(V \le w)\Pr(\epsilon = 1) = F_Z(w), \end{aligned}$$

since Z and V are identically distributed. Thus, for all $z \in S_W$ and $s \in S_Z \backslash \{z\}$, we have

$$f_{Z|W}(s|z) = \frac{f_{W|Z}(z|s) f_Z(s)}{f_W(z)} = \frac{f_V(z)\Pr(\epsilon = 1) f_Z(s)}{f_Z(z)} = \Pr(\epsilon = 1) f_Z(s).$$

It follows that, for all $z \in S_W$ and $s \in S_Z$,

$$F_{Z|W}(s|z) = \Pr(\epsilon = 1) F_Z(s) + \Pr(\epsilon = 0) H(s - z) = \begin{cases} \Pr(\epsilon = 1) F_Z(s) & \text{if } s < z \\ \Pr(\epsilon = 0) + \Pr(\epsilon = 1) F_Z(s) & \text{if } z \le s \end{cases},$$

where $H(v) = \begin{cases} 0 \text{ if } v < 0 \\ 1 \text{ if } 0 \le v \end{cases}$ is the Heaviside step function. Thus, $\frac{\partial}{\partial w} F_{Z|W}(s|z) = -\Pr(\epsilon = 0)\delta(s - z)$ for all $z \in S_W$ and $s \in S_Z$ (see e.g., Bracewell 1986, p. 72), with $\Pr(\epsilon = 0) > 0$. ■

Proof of Corollary 3.7. (a) Since $F_{Z|W}(z_m|w) = F_{Z|W}(z_m|w')$ for all $z_m \ne z_{k-1}$ and $F_{Z|W}(z_{k-1}|w) \ne F_{Z|W}(z_{k-1}|w')$, we have in Theorem 3.2 that $\mu_m^{w,w'} = 0$ for all $m \ne k$ and $\mu_k^{w,w'} = 1$, and thus $\beta_{w,w'} = \sum_{m=1}^K \mu_m^{w,w'} \alpha_{m-1,m}^* = \alpha_{k-1,k}^*$. (b) The result obtains from Theorem 3.2(b) by considering the limit as $\tau \to 0$ of the integrals involving $\frac{\partial}{\partial w} F_{Z|W}^{\tau}(s|w)$. Under the conditions of Theorem 3.1(b) and with $v(\cdot)$ continuously differentiable at z, $\frac{\partial}{\partial z} E(D|Z=\cdot)$ is continuous at z. Thus, by the "sifting property" of the

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

Dirac delta function (e.g., Bracewell 1986, p. 75; Li and Racine, 2007, p. 679)

$$\frac{\partial}{\partial w}E(D|W=w)=\lim_{\tau\to 0}-\int_{S_Z}\left[\frac{\partial}{\partial z}E(D|Z=s)\right]\left[\frac{\partial}{\partial w}F^{\tau}_{Z|W}(s|w)\right]ds$$
$$=-\int_{S_Z}\left[\frac{\partial}{\partial z}E(D|Z=s)\right]\gamma\,\delta(s-z)ds=-\gamma\,\frac{\partial}{\partial z}E(D|Z=z)\neq 0.$$

Similarly, $\frac{\partial}{\partial z}E(Y|Z=\cdot)$ is continuous at z and by the sifting property

$$\frac{\partial}{\partial w}E(Y|W=w)=-\int_{S_Z}\left[\frac{\partial}{\partial z}E(Y|Z=s)\right]\gamma\,\delta(s-z)ds=-\gamma\,\frac{\partial}{\partial z}E(Y|Z=z).$$

Since $\frac{\partial}{\partial z}E(D|Z=z)\neq 0$, Theorem 3.1 then gives $\beta(w)=\alpha(z)=\alpha^*(z)$. ■

Proof of Theorem 4.1. By A.3 $Cov(Y,g(W))=Cov(Y,\tilde{g}(Z))$ and $Cov(D,g(W))=Cov(D,\tilde{g}(Z))$. (a) Theorem 2 of IA gives that

$$Cov(Y,\tilde{g}(Z))=\sum_{k=1}^{K}[E(Y|Z=z_k)-E(Y|Z=z_{k-1})]\sum_{h=k}^{K}f_Z(z_h)\ddot{g}(z_h)$$
$$=\sum_{k=1}^{K}\alpha^*_{k-1,k}[E(D|Z=z_k)-E(D|Z=z_{k-1})]E(\ddot{g}(Z)|Z\geq z_k)\Pr(Z\geq z_k),$$

since, for $k=1,...,K$, Theorem 3.1(a) gives $\alpha_{k-1,k}=\alpha^*_{k-1,k}$ if $E(D|Z=z_k)\neq E(D|Z=z_{k-1})$ and we let $\alpha^*_{k-1,k}=0$ otherwise. Similarly,

$$Cov(D,\tilde{g}(Z))=\sum_{m=1}^{K}[E(D|Z=z_m)-E(D|Z=z_{m-1})]E(\ddot{g}(Z)|Z\geq z_m)\Pr(Z\geq z_m).$$

Since $Cov(D,g(W))\neq 0$, we have $\beta^g_W=\sum_{k=1}^{K}\phi^g_k\alpha^*_{k-1,k}$. Clearly, $\sum_{k=1}^{K}\phi^g_k=1$.

(b) From the proof of Theorem 3.2, $E(X|Z=\cdot)$, $X=Y$, D is absolutely continuous on S_Z. Further, $E(\ddot{g}(Z)|Z\geq z)\Pr(Z\geq z)=-\int_{\underline{z}}^{z}\ddot{g}(s)f_Z(s)ds$ is absolutely continuous on S_Z. Integration by parts and arguments similar to the proof of Theorem 3.2 give (see also Theorem 4 in AGI or Appendix C in HUV)

$$Cov(Y,\tilde{g}(Z))=Cov(E(Y|Z),\tilde{g}(Z))=E[E(Y|Z)\ddot{g}(Z)]=\int_{S_Z}E(Y|Z=z)\ddot{g}(z)f_Z(z)dz$$
$$=E(Y|Z=z)\int_{\underline{z}}^{z}\ddot{g}(s)f_Z(s)ds\Big|_{\underline{z}}^{\bar{z}}-\int_{S_Z}\left[\frac{\partial}{\partial z}E(Y|Z=z)\right]$$
$$\times\left[\int_{\underline{z}}^{z}\ddot{g}(s)f_Z(s)ds\right]dz$$
$$=-\int_{S_Z}\left[\sum_{j=1}^{J}E\left[Y(j)-Y(j-1)|V_j=v(z)\right]f_{V_j}(v(z))\frac{\partial}{\partial z}v(z)\right]$$

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

$$\times \left[\int_{\underline{z}}^{z} \ddot{g}(s) f_Z(s) ds \right] dz$$

$$= \int_{S_Z} \alpha^*(z) \left[\frac{\partial}{\partial z} E(D|Z=z) \right] E(\ddot{g}(Z)|Z \geq z) \Pr(Z \geq z) dz,$$

with $\underline{z}$ and $\bar{z}$ the (possibly infinite) infimum and supremum over S_Z and since we let $\alpha^*(z) = 0$ if $\sum_{j=1}^{J} f_{V_j}(\nu(z)) \frac{\partial}{\partial z} \nu(z) = 0$ for $z \in S_Z$. Similarly,

$$Cov(D, \tilde{g}(Z)) = \int_{S_Z} \left[\frac{\partial}{\partial z} E(D|Z=m) \right] E(\ddot{g}(Z)|Z \geq m) \Pr(Z \geq m) dm.$$

Thus, $\beta_W^g = \int_{S_Z} \alpha^*(z) \phi^g(z) dz$ with $\int_{S_Z} \phi^g(z) dz = 1$. ■

Proof of Proposition 4.2. Under positive intensity preservation, for all $z \in S_Z$ and $w \in S_W$, $\Pr(Z \leq z|W=w)$ is nonincreasing in w. Lemmas 4 and 1(v) in Lehmann (1966) give that, for all $z \in S_Z$ and $w \in S_W$, $\Pr(Z \leq z|W \leq w) \geq \Pr(Z \leq z)$ or equivalently $\Pr(W \geq w|Z \geq z) \geq \Pr(W \geq w)$. Let g be nondecreasing then Theorem 1.A.3(a) and property 1.A.7 in Shaked and Shanthikumar (1994) give that $\Pr(g(W) \geq m|Z \geq z) \geq \Pr(g(W) \geq m)$ for all $z \in S_Z$ and $m \in S_g$ and thus $E(g(W)|Z \geq z) \geq E(g(W))$ for all $z \in S_Z$. Similar arguments apply for g nonincreasing or negative intensity preservation. ■

Proof of Corollary 4.3. Since $U_W \perp_m Z$, we have $E(W|Z \geq z) - E(W) = E(f(Z)|Z \geq z) - E(f(Z))$. Let f be nondecreasing, Proposition 3.3 gives that $f(Z)$ is positively intensity preserving for Z. By Proposition 4.2, it follows that $f(Z)$, and thus W, are positively mean quadrant dependent on Z. A similar argument applies for nonincreasing f. The counterexample in Proposition 3.3 gives a nonmonotonic f such that $f(Z)$ is positively intensity preserving for Z and thus, by Proposition 4.2, $f(Z)$, and therefore W, are positively mean quadrant dependent on Z. ■

Proof of Corollary 4.4. Under A.1, A.2, and (strict) A.4, $E(D|Z=z)$ is (strictly) monotonic over S_Z. Then, $0 \leq \phi_k^g$ for $k = 1, ..., K$ or $0 \leq \phi^g(z)$ for a.e. $z \in S_Z$ if (and only if) $E(\ddot{g}(Z)|Z \geq z)$ does not change sign over S_Z, or, by the law of iterated expectations and $\ddot{g}(Z) \equiv E[g(W)|Z] - E[g(W)]$, if (and only if) $g(W)$ is mean quadrant dependent on Z. ■

Proof of Theorem 4.5. By A.3 $Cov(Y, g(W)) = Cov(Y, \tilde{g}(Z)) = E(Y\ddot{g}(Z))$. We have

$$Cov(Y, g(W)) = E(Y\ddot{g}(Z)) = E\left[\ddot{g}(Z) \sum_{j=0}^{J} Y(j) \left[\mathbf{1}\{V_j \leq \nu(Z)\} - \mathbf{1}\{V_{j+1} \leq \nu(Z)\} \right] \right]$$

$$= \sum_{j=0}^{J} E\left[\ddot{g}(Z) \mathbf{1}\{V_j \leq \nu(Z)\} E\left[Y(j)|Z, V_j \right] \right]$$

$$- E\left[\ddot{g}(Z) \mathbf{1}\{V_{j+1} \leq \nu(Z)\} E\left[Y(j)|Z, V_{j+1} \right] \right]$$

$$= \sum_{j=0}^{J} E\left[E\left[\ddot{g}(Z) \mathbf{1}\{V_j \leq \nu(Z)\} | V_j \right] E\left[Y(j)|V_j \right] \right]$$

$$- E\left[E\left[\ddot{g}(Z) \mathbf{1}\{V_{j+1} \leq \nu(Z)\} | V_{j+1} \right] E\left[Y(j)|V_{j+1} \right] \right],$$

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press

where we apply the law of iterated expectations in the third and fourth equalities and A.2 in the fourth equality. Let

$$\vartheta_j^g(v) \equiv E\left[\ddot{g}(Z)\mathbf{1}\left\{V_j \leq \nu(Z)\right\}|V_j = v\right] = E[\ddot{g}(Z)|v \leq \nu(Z)]\Pr(v \leq \nu(Z)),$$

where we use A.2 in the last equality. Then

$$\begin{aligned} Cov(Y, g(W)) &= \sum_{j=0}^{J} E\left[\vartheta_j^g\left(V_j\right) E\left(Y(j)|V_j\right)\right] - E\left[\vartheta_{j+1}^g\left(V_{j+1}\right) E\left(Y(j)|V_{j+1}\right)\right] \\ &= E\left[\vartheta_0^g(V_0)E(Y(0)|V_0)\right] \\ &\quad + \sum_{j=1}^{J} E\left[\vartheta_j^g(V_j)E\left[Y(j) - Y(j-1)|V_j\right] - E\left[\vartheta_{J+1}^g(V_{J+1})E(Y(J)|V_{J+1})\right]\right. \\ &\qquad \left. = \sum_{j=1}^{J}\int_{S_{V_j}} \vartheta_j^g(v)E(Y(j) - Y(j-1)|V_j = v\right] f_{V_j}(v)dv, \end{aligned}$$

since $\vartheta_0^g(V_0) = \vartheta_{J+1}^g(V_{J+1}) = 0$. Similarly,

$$Cov(D, g(W)) = E(D\ddot{g}(Z)) = \sum_{j=1}^{J} E\left[\vartheta_j^g\left(V_j\right)(j - (j-1))\right] = \sum_{j=1}^{J}\int_{S_{V_j}} \vartheta_j^g(v) f_{V_j}(v)dv.$$

Dividing the two expressions gives the expression for β_W^g. Clearly, $\sum_{j=1}^{J}\int_{S_{V_j}} \psi_j^g(v) f_{V_j}(v)dv = 1$. The weights $\psi_j^g(v) \geq 0$ for all $v \in S_{V_j}$, $j = 1, ..., J$, if and only if $E[\ddot{g}(Z)\mathbf{1}\{v \leq \nu(Z)\}] \leq 0$ for all $v \in S_{V_j}$, $j = 1, ..., J$, or $E[\ddot{g}(Z)\mathbf{1}\{v \leq \nu(Z)\}] \geq 0$ for all $v \in S_{V_j}$, $j = 1, ..., J$. ■

https://doi.org/10.1017/S0266466615000390 Published online by Cambridge University Press