Appendix A The Pairwise-Matching Strategy of Chen (1999)

To illustrate Chen's (1999) identification strategy, let us temporarily assume a linear location function $\mu_d = \alpha + \beta d$ and assume a conventional error term e_d in the outcome equation. The target parameter β is the ATE on the mean. Using the notation in (3.1), we rewrite the outcome equation as a partially linear model, $y = \alpha + \beta d + \lambda_d(m(z)) + w_d$ for $d = 0$ or 1, where the non-linear part is selectivity bias $\lambda_d(m(z))$, defined by $E[e_d|m(z)] = E[e_1 d + e_0(1-d)|m(z)]$, given d. By construction, the new error term $w_d \equiv e_d - \lambda_d(m(z))$ has zero mean for z, given d. The key to identification of α and β is to cancel out the nuisance term within symmetric pairs (i,j), whose selection indexes are symmetric around zero, $m(z_i) = -m(z_j) = m$. Without loss of generality, we suppose that $m > 0$:

$$\begin{aligned}
\lambda_{d_i}(m) - \lambda_{d_j}(-m) &= E[e_{1i}I\{\eta_i \leq m\} + e_{0i}I\{\eta_i > m\}|m] \\
&- E[e_{1j}I\{\eta_j \leq -m\} + e_{0j}I\{\eta_j > -m\}|m] \\
&= E[(e_{1i} - e_{0i})I\{-m < \eta_i \leq m\}|m] = 0.
\end{aligned}$$

Because $I\{-m < \eta_i \leq m\} + I\{\eta_i \leq -m\} = I\{\eta_i \leq m\}$, the second equality is valid if (e_{di}, η_i) is independent and identically distributed across individuals. The final equality holds if e_{di} and η_i are symmetrically distributed. The bounded form for the latent index in the final equality is similar to the identification-at-infinity argument (Chamberlain, 1986) and the marginal-treatment-effect concept (Heckman and Vytlacil, 1999).

Appendix B Proof of the Equivalence Result

To proof the equivalence result in equation (4.13) , consider symmetric pairs (i,j):

$$\begin{aligned}
p(x_i, z_i) &= F_\eta(m(x_i, z_i)|x_i, z_i) = 1 - p(x_j, z_j) = 1 - F_\eta(m(x_j, z_j)|x_j, z_j) \\
&= F_\eta(-m(x_j, z_j)|x_j, z_j).
\end{aligned}$$

The first and third equalities are from the latent-index selection model, the second equality is from $p(x_i, z_i) + p(x_j, z_j) = 1$, and the last equality is because of symmetry. Because z and η are independent by Assumption 5(1), $m(x_i, z_i) = -m(x_j, z_j)$ by holding $x_i = x_j = x$. The proof for the converse is similar.

Appendix C Large Sample Behaviour

Here we explore the asymptotic properties of an averaged scale ratio and an averaged variance ratio, both of which are obtained by averaging across covariates of the estimators proposed in Section 4. We focus primarily on the IV estimator and the symmetric quantile estimator ($\widehat{r}_{IV}$ and $\widehat{r}_q$), simply noting that the properties of the other proposed estimator ($\widehat{r}_v$) follow from similar arguments and thus are omitted here.

Let $\widehat{r}_{IV}$ be the averaged variance ratio of the IV estimator based on model (3.1) and (3.6) and Assumptions 1 and 2, our first theorem is as follows:

Theorem 1 *Our variance-based IV estimator $\hat{r}_{IV}$ of the average variance ratio r_{IV}, defined by $E_X[\sigma_1^2(x)/\sigma_0^2(x)]$, has the following linear representation:*

$$\hat{r}_{IV} - r_{IV} = \frac{1}{n}\sum_{i=1}^{n}(\psi_{ai} + \psi_{bi}) + o_p(n^{-1/2})$$

where $\psi_{ai} \equiv V_1(x_i)/V_0(x_i) - E_X[\sigma_1^2(x)/\sigma_0^2(x)]$ and

$$\psi_{bi} \equiv V_{0i}^{-1}(\psi_{12i} - 2\lambda_{1i}\psi_{11i}) - V_{1i}V_{0i}^{-2}(\psi_{02i} - 2\lambda_{0i}\psi_{01i}) + o_p(n^{-1/2}), \quad \text{(Appendix C.15)}$$

where $\lambda_{1i} \equiv E(y_{1i}|d_{1i} > d_{0i}, x_i)$ and $V_{1i} \equiv V(y_{1i}|d_{1i} > d_{0i}, x_i)$; λ_{0i} and V_{0i} are defined analogously. For $k = 1, 2$, ψ_{1ki} is of the form

$$\begin{aligned}&\frac{[y_i^k d_i z_i - E(y_i^k d_i|z_i = 1, x_i)] - [y_i^k d_i(1 - z_i) - E(y_i^k d_i|z_i = 0, x_i)]}{E(d_i|z_i = 1, x_i) - E(d_i|z_i = 0, x_i)}\\ &-[E(y_i^k d_i|z_i = 1, x_i) - E(y_i^k d_i|z_i = 0, x_i)]\\ &\times\frac{[d_i z_i - E(d_i|z_i = 1, x_i)] - [d_i(1 - z_i) - E(d_i|z_i = 0, x_i)]}{[E(d_i|z_i = 1, x_i) - E(d_i|z_i = 0, x_i)]^2}\end{aligned}$$

and ψ_{0ki} can be defined analogously by replacing d_i with $(1 - d_i)$. The root-n consistency and asymptotic normality of the estimator follow from this linear representation.

Remark 1 *Thus we can see that the estimator is root-n consistent and asymptotically normal:*

$$\sqrt{n}(\hat{r}_{IV} - r_{IV}) \Rightarrow N(0, E[(\psi_{ai} + \psi_{bi})^2]).$$

The asymptotic variance $E[(\psi_{ai} + \psi_{bi})^2]$ can be estimated for inference purposes, though an alternative approach, used in our application, is to bootstrap. We recommend bootstrapped confidence intervals, rather than estimating the asymptotic variances, because the calculations require additional nonparametric steps that make the estimation cumbersome. Subbotin (2009) has proven that the confidence intervals and the standard errors of an estimator like ours can be consistently estimated by bootstrapping. Thus, we recommend bootstrapped confidence intervals particularly for practitioners.

Turning our attention to asymptotic theory for matching procedures, as mentioned, we will focus on the quantile estimator, as identical arguments can be used for the variance-based matching estimator. Furthermore, to focus on the asymptotic arguments pertaining to our kernel-weighted matching, we will focus on a conditional (not averaged) scale ratio. In order to circumvent the identification and dimensionality conditions as mentioned previously, our proofs assume that the scale parameters depend on treatment but not on covariates.

While the regularity conditions for the quantile-based estimator are standard when compared to existing work (Ahn and Powell 1993; Chen and Khan 2003), they are still quite detailed, particularly as multiple semiparametric steps are involved. To ease the notational burdens, letting $w = (x, z)$, we impose the parametric restriction $m(w) = w'\delta \equiv v$. It is noteworthy that a non-parametric $m(\cdot)$ would still allow our matching estimators of r to be root-n consistent and asymptotically normal, analogous to Ahn and Powell's (1993) results. To simplify the notation, we define

$$\begin{aligned} q_\tau^{(d)}(w) &\equiv q_\tau(y_i|d_i = d, w_i = w), \\ \Delta q_\tau^{(d)}(w) &\equiv IQ_\tau(y_i|d_i = d, w_i = w), \end{aligned}$$

for $d = 0, 1$. The asymptotic distribution of the quantile-based estimator of the scale ratio, defined here as $r_q = \sigma_1/\sigma_0$, is as follows:

Theorem 2 *Under regularity conditions (I), (KH1), (S0), (RD2), (S2), and (H1) in the Appendix, if the coefficient estimators in the selection equation has the following linear representation:*[4]

$$\hat{\delta} - \delta = \frac{1}{n}\sum_{i=1}^{n} \psi_i^+ + o_p(n^{-1/2}),$$

then we have

$$\sqrt{n}(\hat{r}_q - r_q) \Rightarrow N(0, \Sigma_0^{-2} E[(\psi_i^- + \mathcal{M}\psi_i^+)^2]),$$

where $\Sigma_0 \equiv E[p(v_i)^2 f(p(v_i))]$, $p(v_i)$ *is the propensity score, and* $f(p(v_i))$ *is its density. Furthermore,*

$$\psi_i^- \equiv (1 - p(v_i))^2 f_V(v_i) f_W(w_i) \left[d_i \Delta q_\tau^{(0)}(w_i)^{-1} \phi_{1i} - (1 - d_i) \frac{\Delta q_\tau^{(1)}(w_i)}{\Delta q_\tau^{(0)}(w_i)^2} \phi_{0i} \right].$$

For $d = 0$ *and* 1 *we define*

$$\phi_{di} \equiv f_{U_{1d}|W}(0|w_i)^{-1}\{I[y_i \le q_{1-\tau}^{(d)}(w_i)] - (1 - \tau)\} - f_{U_{0d}|W}(0|w_i)^{-1}\{I[y_i \le q_\tau^{(d)}(w_i)] - \tau\},$$

where $\tau \in (0, 1/2)$, $f_{U_{1d}|W}$ *and* $f_{U_{0d}|W}(0|w_i)$ *are conditional density functions of residuals associated with the conditional quantile function for* $d = 1$ *and* 0*, given the upper and lower*

[4] We are taking the linear representation for the selection equation estimator as given. This is because any semi-parametric binary choice estimator for δ that is root-n consistent can be used, and most of these have already established linear representations. The purpose of our theorem here is to establish how the influence function for an estimator of δ affects the influence function of our quantile estimator of r_q.

quantiles $(1-\tau$ and $\tau)$ respectively. Finally, define $\mathcal{M}$ as

$$E\{[1-p(v_i)] \times \left[\mathcal{G}_2(v_i,v_i)p(-v_i)f_V(-v_i)+\mathcal{G}(v_i,v_i)p'(-v_i)f_V(-v_i)+\mathcal{G}(v_i,v_i)p(-v_i)f'_V(-v_i)\right]\}, \tag{Appendix C.16}$$

where $\mathcal{G}(v_i,v_j) \equiv \int\int \left[\Delta q_\tau^{(1)}(w_i) - \Delta q_\tau^{(0)}(w_j)\right](w_i+w_j)'dF(w_i|v_i)dF(w_j|v_j)$, and $\mathcal{G}_2(\cdot,\cdot)$ is the partial derivative of $\mathcal{G}(\cdot)$ with respect to its second argument.

Remark 2 *While the form of the influence function in the above linear representation is very complicated, we can see how its form relates to similar estimation procedures in the literature. For example, term Σ_0 corresponds to the propensity score matching weights and is very similar to the form attained in Ahn and Powell (1993). The component ψ_i^- corresponds to the variance introduced by replacing true quantile functions with their nonparametric estimators; consequently, its form is similar to other estimators which use first-stage nonparametric quantile regression estimators, such as those developed by Chaudhuri et al. (1991) and Khan (2001). The component ψ_i^+ corresponds to the noise induced by the selection equation estimator, and $\mathcal{M}$ corresponds to the weighting term induced from the linear expansion of the kernel weighted matching function. As a result, this second piece in its entirety $(\mathcal{M}\psi_i^+)$ is of a form similar to that found in Powell (2001).*

Appendix D Regularity Conditions and Proofs

Here we state the regularity conditions and proofs for the theorems stated in Appendix B. We note that under identification assumptions, the support issues for the IV estimator are not as severe as those for our matching-based estimators, so part of the proof derives local linear representations for variance ratios as a function of covariates (i.e., $\widehat{r}_{IV}(x)$).

Define $\widehat{V}_d(x)$ as the kernel estimators of $V_d(x) \equiv V[y_d|d_1 > d_0, x]$, for $d = 0$ or 1, based on equations in (4.12). Our proof for the IV estimator is to consider the case of an averaged scale ratio. Our variance-based IV estimator can be expressed as

$$\hat{r}_{IV} \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{V}_1(x_i)/\widehat{V}_0(x_i).$$

In Theorem 1, we linearise $\widehat{r}_{IV}$ around the true value $r_{IV} \equiv E_x[\sigma_1^2(x)/\sigma_0^2(x)]$. To do so, we decompose $(\widehat{r}_{IV} - r_{IV})$ into two components: $(\widehat{r}_{IV} - r_{nIV})$ and $(r_{nIV} - r_{IV})$, where

$$r_{nIV} \equiv \frac{1}{n}\sum_{i=1}^{n} V_1(x_i)/V_0(x_i).$$

We notice that r_{nIV} converges to r_{IV} in probability by the condition of invariant normalization (Assumption 2) and the Law of Large Numbers. In what follows, we establish the linearisation of $(\widehat{r}_{IV} - r_{nIV})$ in Lemma 1 with the following regularity conditions. From Lemma 1, the proof of Theorem 1 follows immediately.

Regularity Conditions for Theorem 1:

Assumption RS (Random sampling) The vector $(y_i, z_i, d_i, x_i')'$ is i.i.d.

Assumption RD (Regressor distribution) The regressor vector x_i has support which is a compact subset of $\mathbf{R}^k$. x_i may have discrete and continuous components, and we let k_c denote the number of continuous components. We assume the conditional density function of the continuous components given the discrete components is continuously differentiable of order p, where $p > 5k_c/2$.

Assumption K (Kernel function and bandwidth) The kernel function is of order p and the bandwidth satisfies $\sqrt{n}h_n^p \to 0$ and $nh_n^{k_c} \to \infty$.

Assumption MF (Moment functions) The moment functions $E[y_i^l|x_i, d_i, z_i = 1]$ are p times continuously differentiable for $l = 1, 2$.

Lemma Appendix D.1 *Under Assumptions (RS), (RD), (K), and (MF), our variance-based IV estimator $\hat{r}_{IV}$ has the following linear representation:*

$$\hat{r}_{IV} - r_{nIV} = \frac{1}{n}\sum_{i=1}^{n}\psi_{bi} + o_p(n^{-1/2}), \qquad \text{(Appendix D.17)}$$

where ψ_{bi} is defined in equation (Appendix C.15).

Proof: By the definitions of $\hat{r}_{IV}$ and r_{nIV}, the linearisation of

$$\hat{r}_{IV} - r_{nIV} = \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{V}_1(x_i)/\widehat{V}_0(x_i) - V_1(x_i)/V_0(x_i)\right], \qquad \text{(Appendix D.18)}$$

can be obtained by expanding the ratios inside the summation. As in Newey and McFadden (1994, p.2204), the first-order expansion of ratio $\widehat{a}/\widehat{b}$ around a/b is $b^{-1}[\widehat{a} - a - (a/b)(\widehat{b} - b)]$, so the linearisation of $\widehat{V}_1/\widehat{V}_0$ around V_1/V_0 is $V_0^{-1}(\widehat{V}_1 - V_1) - V_1 V_0^{-2}(\widehat{V}_0 - V_0)$. Therefore, from equation (Appendix D.18), $\hat{r}_{IV} - r_{nIV}$ can also be expressed as

$$\hat{r}_{IV} - r_{nIV} = \frac{1}{n}\sum_{i=1}^{n} V_{0i}^{-1}(\widehat{V}_{1i} - V_{1i}) - \frac{1}{n}\sum_{i=1}^{n} V_{1i}V_{0i}^{-2}(\widehat{V}_{0i} - V_{0i}). \qquad \text{(Appendix D.19)}$$

Both terms on the right hand side can be further linearised separately using the following properties: (i) Variance equals the second moment minus the square of the first moment. (ii) The first and second moments of outcomes of compliers involve the ratio of differences; that is, $E[h(y_d)|d_1 > d_0, x] = (a_1 - a_0)/(b_1 - b_0)$ for $d = 0$ or 1, as in equations (3.3) and (3.4), where $h(y) = y$ or y^2, and a_1, a_0, b_1, and b_0 denote $E[h(y)d|z = 1, x]$, $E[h(y)d|z = 0, x]$, $E[d|z = 1, x]$, and $E[d|z = 0, x]$ respectively. (iii) The first-order expansion of the kernel estimator $(\widehat{a}_1 - \widehat{a}_0)/(\widehat{b}_1 - \widehat{b}_0)$ around $(a_1 - a_0)/(b_1 - b_0)$ is

$$(b_1 - b_0)^{-1}[(\hat{a}_1 - a_1) - (\hat{a_0} - a_0)] - (a_1 - a_0)(b_1 - b_0)^{-2}[(\hat{b}_1 - b_1) - (\hat{b}_0 - b_0)].$$

As suggested in Chen and Khan (2003), the kernel estimator has the fourth root consistency followed by assumptions (RS), (RD), and (MF). By the the consistency of those kernel estimators, the linearisation of the kernel estimators for the ratios implies that we can derive a linear representation for $\frac{1}{n}\sum_{i=1}^{n}[\widehat{E}(h(y_{1i})|d_{1i} > d_{0i}, x_i) - E(h(y_{0i})|d_{1i} > d_{0i}, x_i)]$ as follows:

$$\frac{1}{n}\sum\nolimits_{i=1}^{n}(b_{1i}-b_{0i})^{-1}[(\hat{a}_{1i}-a_{1i})-(\hat{a_{0i}}-a_{0i})]-\frac{1}{n}\sum\nolimits_{i=1}^{n}(a_{1i}-a_{0i})(b_{1i}-b_{0i})^{-2}[(\hat{b}_{1i}-b_{1i})-(\hat{b}_{0i}-b_{0i})] \tag{Appendix D.20}$$

since the remainder term is $o_p(n^{-1/2})$. We focus on the first summation for the case of the second moment (e.g., $h(y) = y^2$), as similar arguments can be used for the other components. Applying the results from Newey and McFadden (1994), we can represent the first summation in the linearisation of the first term in (Appendix D.19) as

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} V_{0i}^{-1}[E(d_i|z_i &= 1, x_i) - E(d_i|z_i = 0, x_i)]^{-1} \times \\ \{[y_i^2 d_i z_i - E(y_i^2 d_i|z_i &= 1, x_i)] - [y_i^2 d_i(1 - z_i) - E(y_i^2 d_i|z_i = 0, x_i)]\} + o_p(n^{-1/2}). \end{aligned} \tag{Appendix D.21}$$

Similarly, the second summation is of the form:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} V_{0i}^{-1}[E(y_i^2 d_i|z_i &= 1, x_i) - E(y_i^2 d_i|z_i = 0, x_i)] \times [E(d_i|z_i = 1, x_i) - E(d_i|z_i = 0, x_i)]^{-2} \\ \{[d_i z_i - E(d_i|z_i &= 1, x_i)] - [d_i(1 - z_i) - E(d_i|z_i = 0, x_i)]\} + o_p(n^{-1/2}). \end{aligned} \tag{Appendix D.22}$$

Subtracting (Appendix D.22) from (Appendix D.21) establishes the linearisation of the component for the second moment of the outcome variable, $\frac{1}{n}\sum_{i=1}^{n} V_{0i}^{-1}[\widehat{E}(y_{1i}^2|d_{1i} > d_{0i}, x_i) - E(y_{1i}^2|d_{1i} > d_{0i}, x_i)]$ in equation (Appendix D.19). Denote the term in the resulting summation (excluding V_{0i}^{-1}) by ψ_{12i}. Turning our attention to the square of the first moment, the linear representation of the component for the first moment would be the same as above simply replacing y_i^2 with y_i. Let ψ_{11i} denote this term. Letting λ_1 denote $E[y_1|d_1 > d_0, x]$, the linear representation for the square of the first moment can be denoted by

$$\frac{1}{n}\sum_{i=1}^{n} 2V_{0i}^{-1}\lambda_{1i}\psi_{11i} + o_p(n^{-1/2}), \tag{Appendix D.23}$$

which is a straight forward application of the delta method.

Collecting all of our results, we can conclude that we have the following linear representation for the variance of the treated group:

$$\frac{1}{n}\sum_{i=1}^{n} V_{0i}^{-1}(\widehat{V}_{1i} - V_{1i}) = \frac{1}{n}\sum_{i=1}^{n} V_{0i}^{-1}(\psi_{12i} - 2\lambda_{1i}\psi_{11i}) \quad + o_p(n^{-1/2}). \qquad \text{(Appendix D.24)}$$

Note we can derive an analogous linear representation for $(\widehat{V}_{0i} - V_{0i})$ with analogous terms in the summation, which we will replace $(\psi_{12}, \lambda_1, \psi_{11})$ with $(\psi_{02}, \lambda_0, \psi_{01})$.

Next, we conclude that the linearisation of $\hat{r}_{IV} - r_{nIV}$ in equation (Appendix D.18) or (Appendix D.19) is

$$\frac{1}{n}\sum_{i=1}^{n}\{V_{0i}^{-1}(\psi_{12i} - 2\lambda_{1i}\psi_{11i}) - V_{1i}V_{0i}^{-2}(\psi_{02i} - 2\lambda_{0i}\psi_{01i})\} \ + o_p(n^{-1/2}) \qquad \text{(Appendix D.25)}$$

where λ_0 above is $E[y_0|d_1 > d_0, x]$. This completes the proof of Lemma 1 and thus concludes the proof of Theorem 1.

Regularity Conditions for Theorem 2: Let h_{0n} and h_{1n} be the bandwidths for the selection equation estimation and the pairwise matching kernel-weighting scheme in the first stage, and let h_{2n} denote the bandwidth for the local polynomial quantile regressions in the second stage. The regularity conditions for Theorem 2 are summarized below.

Assumption I (Identification) $\Sigma_0 > 0$.

We next impose conditions on the kernel function used to match propensity score values and its bandwidth sequence:

Assumption KH1 The kernel function $K_{1n}(\cdot)$ is assumed to have the following properties: (i) $K_{1n}(\cdot)$ is twice continuously differentiable with a bounded second derivative and has a compact support; (ii) symmetric about zero; and (iii) a fourth-order kernel with $\int u^l K_{1n}(u)du = 0$ for $l = 1, 2, 3$ and $\int u^4 K_{1n}(u)du \neq 0$. The bandwidth sequence h_{1n} is of the form: $h_{1n} = c_1 n^{-\gamma_1}$, where c_1 is a constant and $\gamma_1 \in (\frac{1}{8}, \frac{1}{6})$.

The following assumption characterizes the smoothness of the density and the conditional expectation functions of the selection index:

Assumption S0 The function $f_V(\cdot)$ has an order of differentiability of four, with the fourth-order derivative bounded.

We next impose three conditions associated with the estimation of interquartile spreads. This involves smoothness assumptions on the conditional quantile functions and on the distributions of $w_i = (x_i, z_i)$ and the residuals associated with the quantile functions. For notational convenience, we describe the conditions in terms of w, whose support is denoted by $\mathcal{W}$.

Assumption RD2 (Distribution of regressors and instruments) The vector w can be decomposed as $w = (w^{(c)\prime}, w^{(ds)\prime})'$ where the k_c-dimensional vector $w^{(c)}$ is continuously distributed, and the k_{ds}-dimensional vector $w^{(ds)}$ is discretely distributed. Letting $f_{W^{(c)}|W^{(ds)}}(\cdot|w^{(ds)})$ denote the conditional density function of $w_i^{(c)}$, we assume it is bounded away from zero and is Lipschitz continuous on $\mathcal{W}$. Letting $f_{W^{(ds)}}(\cdot)$ denote the mass function of $w^{(ds)}$, we assume that there is a finite number of mass points on $\mathcal{W}$. Finally, we let $f_W(\cdot)$ denote $f_{W^{(c)}|W^{(ds)}}(\cdot|\cdot) f_{W^{(ds)}}(\cdot)$.

Assumption S2 (Smoothness of conditional quantile functions)

S2.1 The polynomial used for the second-stage quantile function estimators is of order m.

S2.2 For all values of $w^{(ds)}$, the quantile functions $q_{\tau_1}^{(d)}(\cdot)$ and $q_{\tau_2}^{(d)}(\cdot)$ $d = 0, 1$ are bounded and m times continuously differentiable with bounded m^{th} derivatives with respect to $w^{(c)}$ on $\mathcal{W}$.

Assumption H1 (Second-stage bandwidth sequence for interquartile spread estimation). The bandwidth sequence used to estimate the conditional interquantile spread is of the form: $h_{2n} = c_2 n^{-\gamma_2}$, where c_2 is a constant, and $\gamma_2 \in ((\gamma_1 + 0.5)/m, (1 - 4\gamma_1)/3k_c)$, where γ_1, m and k_c are given in Assumptions KH1, S2 and RD2 respectively.

Proof of Theorem 2: The arguments used to derive the limiting distribution theory are very similar to those used in Chen and Khan (2003), who impose similar regularity conditions, hereafter referred to as CK. We thus only provide a sketch of the main arguments, referring readers interested in technical details to CK. It is there were the technical conditions, such as those imposed on the bandwidths, are used.

We note we can write $\hat{r}_q = \hat{\Sigma}_1/\hat{\Sigma}_0$, where

$$\begin{aligned}\hat{\Sigma}_1 &= \frac{1}{n(n-1)}\sum_{i\neq j} d_j(1-d_i)\hat{\omega}_{ij}\Delta\hat{q}_\tau^{(1)}(w_i)/\Delta\hat{q}_\tau^{(0)}(w_i),\\ \hat{\Sigma}_0 &= \frac{1}{n(n-1)}\sum_{i\neq j} d_j(1-d_i)\hat{\omega}_{ij}.\end{aligned}$$

Recall that $r_q = \sigma_1/\sigma_0$. We will establish a linear representation for $\hat{r}_q - r_q$. Our proof strategy is to establish the probability limit of the denominator and establish a linear representation for the numerator. The probability limit of the denominator follows from similar arguments used in proving Theorem 3.1(ii) in Ahn and Powell (1993) and Lemma A.6 in CK:

$$\hat{\Sigma}_0 \xrightarrow{p} \Sigma_0 \equiv E[p(v_i)^2 f(p(v_i))].$$

Turning attention to $(\hat{\Sigma}_1 - r\hat{\Sigma}_0)$, we consider an expansion of $\hat{\omega}_{ij}$ around ω_{ij} defined by $h_{2n}^{-1}K_{1n}\left(\left(w_i'\delta_0 + w_j'\delta_0\right)/h_{2n}\right)$. After using this expansion, $(\hat{\Sigma}_1 - r\hat{\Sigma}_0)$ equals:

$$\frac{1}{n(n-1)}\sum_{i\neq j} d_j(1-d_i)\omega_{ij}\left[\Delta\hat{q}_\tau^{(1)}(w_i)/\hat{q}_\tau^{(0)}(w_i) - r_q\right]. \qquad \text{(Appendix D.26)}$$

We note that if we replace $\Delta\hat{q}_\tau^{(1)}(w_i)$ and $\Delta\hat{q}_\tau^{(0)}(w_i)$ with $\Delta q_\tau^{(1)}(w_i)$ and $\Delta q_\tau^{(0)}(w_i)$ in the above expression, the term is $o_p(n^{-1/2})$ by arguments similar to those the proof of Lemma

A.4 in CK. Linearization of $\Delta\hat{q}_\tau^{(1)}(w_i)/\hat{q}_\tau^{(0)}(w_i)$ around the true value gives:

$$\frac{1}{n(n-1)}\sum_{i\neq j} d_j(1-d_i)\omega_{ij}\Delta q_\tau^{(0)}(w_i)^{-1}[\Delta\hat{q}_\tau^{(1)}(w_i)-\Delta q_\tau^{(1)}(w_i)] \quad \text{(Appendix D.27)}$$
$$-\frac{1}{n(n-1)}\sum_{i\neq j} d_j(1-d_i)\omega_{ij}\Delta q_\tau^{(1)}(w_i)\Delta q_\tau^{(0)}(w_i)^{-2}[\Delta\hat{q}_\tau^{(0)}(w_i)-\Delta q_\tau^{(0)}(w_i)].$$

We further establish a linear representation for the term involving $[\Delta\hat{q}_\tau^{(1)}(w_i)-\Delta q_\tau^{(1)}(w_i)]$. Following the arguments used in Lemma A.4 in CK, we can rephrase the first line of (Appendix D.27) as the following

$$\frac{1}{n}\sum_{i=1}^{n}(1 - p(v_i))^2 f_V(v_i)d_i f_W(w_i)\Delta q_\tau^{(0)}(w_i)^{-1}\{f_{U_{11}|W}(0|w_i)^{-1}(I[y_i\leq q_{1-\tau}^{(1)}(w_i)]-(1-\tau))$$
$$- f_{U_{01}|W}(0|w_i)^{-1}(I[y_i\leq q_\tau^{(1)}(w_i)]-\tau)\}+o_p(n^{-1/2}).$$

where $\tau\in(0,1/2)$, $f_{U_{11}|W}$ and $f_{U_{01}|W}(0|w_i)$ are conditional density functions of residuals associated with the conditional quantile function, for upper quantile $(1-\tau)$ and lower quantile (τ), respectively.

An analogous linear representation can be derived for the term involving $[\Delta\hat{q}_\tau^{(0)}(w_i)-\Delta q_\tau^{(0)}(w_i)]$, where we would replace d_i with $(1-d_i)$ in the above expression, and superscripts (1) with superscripts (0). Collecting both these terms, this can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\psi_i^- + o_p(n^{-1/2}).$$

We next consider the linear term of $\hat{\omega}_{ij}$ around ω_{ij}. This is of the form

$$\frac{1}{n(n-1)}\sum_{i\neq j} d_j(1-d_i)\omega'_{ij}(w_i+w_j)'(\hat{\delta}-\delta_0)\left[\Delta\hat{q}_\tau^{(1)}(w_i)/\Delta\hat{q}_\tau^{(0)}(w_i)-r\right],$$

where $\omega'_{ij}=h_{2n}^{-2}K'_{1n}\left(\left(w_i'\delta_0+w_j'\delta_0\right)/h_{2n}\right)$.

Note we can replace $\Delta\hat{q}_\tau^{(1)}(w_i)/\Delta\hat{q}_\tau^{(0)}(w_i)$ with $\Delta q_\tau^{(1)}(w_i)/\Delta q_\tau^{(0)}(w_i)$ in the above expression. The resulting remainder term is $o_p(n^{-1/2})$ by the root-n consistency of $\hat{\delta}$ and the uniform consistency of the quantile estimators. In what follows, we derive an expression for the probability limit of

$$\frac{1}{n(n-1)}\sum_{i\neq j} d_j(1-d_i)\omega'_{ij}\left[\Delta q_\tau^{(1)}(w_i)/\Delta q_\tau^{(0)}(w_i)-r\right](w_i+w_j)'.$$

Using standard U-statistic projection theorems and the change of variables, the above term converges in probability to $\mathcal{M}$, defined in (Appendix C.16). Thus, the linear term in the expansion has the linear representation:

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{M}\psi_{\delta i}+o_p(n^{-1/2}).$$

Finally we note higher-order terms in the expansion of $\hat{\omega}_{ij}$ around ω_{ij} are asymptotically negligible by the uniform rates of convergence of the quantile estimators and the root-n consistency of $\hat{\delta}$. This completes the linear representation for $\hat{r}_q - r_q$.

Appendix E Monte Carlo Study

The Design and Specifications

In the previous sections we explored the conditions under which the proposed estimators can be identified. In this section we assess their small-sample performance through Monte Carlo simulations. We independently draw outcome errors $(\epsilon_{1i}, \epsilon_{0i})$ from the same distribution to satisfy the condition of invariant normalization (Assumptions 2 and 3). The dependence between outcome and selection errors (ϵ_{di}, η_i) is constructed by setting their correlation coefficient to 0.5 for $d = 0$ and 1. To satisfy the symmetry condition (Assumption 5(2)), we generate a sample (ϵ_{di}, η_i) of size n from three distributions: bivariate normal, bivariate Student-t with 10 degrees of freedom, and bivariate Cauchy distributions. We then increase the sample size from $n = 190$ to $n = 760$, to $n = 3,040$ (the last sample size being about the same as that of one of the empirical studies in Section 5). Finally, we iterate the estimation process $3,041$ times. We consider three cases:

Case 1: Constant Scales First, we consider a case of constant scale ratios, $(\sigma_0, \sigma_1) = (1, e^{0.2})$, where the scale parameters are constant, given the treatment status. The outcome and selection equations are given by

$$\begin{aligned} y_i &= x_{1i} + d_i[2 + \epsilon_{1i}] + (1 - d_i)[e^{0.2}\epsilon_{0i}], \\ d_i &= I\{x_{1i} + z_i \geq \eta_i\}, \end{aligned}$$

where covariate x_{1i} is uniformly distributed between -1 and 1, and instrument z_i is a binary variable with a 50-50 chance of equalling 0 or 1.

Case 2: Covariate-Dependent Scales In the second case, we allow the scale parameters to vary with both treatment d_i and covariate x_{2i} such that $(\sigma_0(x_{2i}), \sigma_1(x_{2i})) = (1, e^{0.2})$ if $x_{2i} = 0$, and $(1.05, 1.05e^{0.3})$ if $x_{2i} = 1$. We assume that x_{2i} has a 50-50 chance of equalling 0 or 1. The outcome and selection equations are given by

$$\begin{aligned} y_i &= x_{1i} + 0.3x_{2i} + d_i[2 + \sigma_1(x_{2i})\epsilon_{1i}] + (1 - d_i)[\sigma_0(x_{2i})\epsilon_{0i}], \\ d_i &= I\{x_{1i} - 0.2x_{2i} + z_i \geq \eta_i\}. \end{aligned}$$

Both cases satisfy Assumption 1. Letting $x_i \equiv (x_{1i}, x_{2i})$, we use Abadie's (2003) Lemma 2.1 to calculate the fraction of compliers, given $x_i = x$:

$$\begin{aligned} \Pr\{d_{1i} > d_{0i}|x\} &= E[d_i|z_i = 1, x] - E[d_i|z_i = 0, x] \quad \text{(Appendix E.28)} \\ &= F_\eta(m(x_i, 1)) - F_\eta(m(x_i, 0)), \end{aligned}$$

where F_η is the cumulative distribution function of η_i. The fraction of compliers is strictly positive in both cases, since the fraction of compliers, given that x, equals $F(x_1+1)-F(x_1)$ in case 1 and $F(x_1 - 0.2x_2 + 1) - F(x_1 - 0.2x_2)$ in case 2. In addition, the symmetric first-stage condition (Assumption 5(3)) is also satisfied in both cases because the conditional probability of symmetric pairs, given that $x_{2i} = x_2$, is strictly positive:

$$\Pr\{(i,j) : m(x_{1i}, x_2, z_i) = -m(x_{1j}, x_2, z_j)\} = \Pr\{(i,j) : x_{1i} + x_{1j} = 0.4x_2 + z_i - z_j\}. \tag{Appendix E.29}$$

For example, if (x_2, z_i, z_j) are all equal to zero, then $\Pr\{(i,j) : x_{1i} = -x_{1j}\} = (\sqrt{2}/2) > 0$.

Case 3: Covariate-Dependent Scales (No Compliers) Finally, consider the case where the symmetric first-stage condition is satisfied but there is no first stage for some subgroup. We maintain the same outcome equation as in (Appendix E.28) but change the selection equation to

$$d_i = I\{x_{1i} - 0.2x_{2i} + x_{2i}z_i \geq \eta_i\}. \tag{Appendix E.30}$$

When $x_{2i} = 0$, there is no first stage, so there are no compliers for the subgroup of individuals with $x_{2i} = 0$. The symmetric first-stage condition is still satisfied because the fraction of symmetric pairs is $\Pr\{(i,j) : x_{1i} = -x_{1j}\} = \sqrt{2}/2 > 0$ for the same subgroup.

Simulation Results

To examine the properties of the proposed estimators, we use four assessment measures: mean bias, median bias, root mean-squared errors, and mean absolute deviations. We examine the rate of convergence of mean absolute deviations for the quantile-based estimator and the rate of convergence of root mean-squared errors for variance-based estimators, including both the IV estimator and the variance-based matching method. Hence, the following analysis focuses on these two convergences. An estimator is said to be consistent or converging if the corresponding assessment measure undergoes root-n convergence. The results of simulations suggest that the proposed quantile-based matching method performs best in small samples, even when there is no first stage or when the underlying distributions have heavy tails. The findings are summarized in online tables A1 and A2 and are discussed below.[5]

In the presence of exogenous covariates, we note that the first-stage condition (Assumption 1) is required for the IV estimator, but not for the pairwise-matching methods. The comparison of these two approaches is presented in online table A2 for the results of Case 3, where there is no first stage even though symmetry conditions (Assumptions 5(2)(3)) are satisfied. As expected, the IV estimator cannot identify the scale ratios in this case, because there are no compliers, but the pairwise-matching estimators still converge at the rate of root-n.

The simulation results in online tables A1 and A2 suggest that the best performing estimator is the quantile-based matching method: all of its assessment measures converge rapidly, even in the presence of outliers or in the absence of compliers. In cases 1 and 2, the variance-based methods perform well in bivariate normal and bivariate Student t(10) models, but they fail to converge in the bivariate Cauchy model because moments are undefined. In case 3, where there are no compliers, the IV methods do not apply, but the variance-based matching method converges at the parametric rate for both bivariate normal and bivariate Student-t(10) models. The quantile-based matching method converges rapidly at the rate of root n in all cases.

It is noteworthy that the symmetry conditions (Assumption 5) are neither stronger nor weaker than the LATE conditions (Assumption 1). The IV methods and the pairwise-matching methods use different subgroups to identify the treatment effect on dispersion

[5] In an earlier version of this paper, we attempted to establish a benchmark by estimating the models using maximum-likelihood estimators, assuming the joint distribution of the error terms to be bivariate normal, although the true distribution can be Student-t or Cauchy. The results showed that under the correct specification, maximum-likelihood estimators had much smaller assessment measures than the proposed estimators. But if the model is misspecified (as bivariate Student t(10)), then maximum-likelihood estimators failed to converge or occasionally generated larger assessment measures than the proposed methods.

in potential outcomes. The IV estimator uses compliers, while the pairwise-matching methods use symmetric pairs. Since both methods use a subpopulation for identification, the precision of estimation depends heavily on the size of the subpopulation. The pairwise-matching methods are more efficient than the IV methods if the instrument is multi-valued or continuously distributed, in the sense that the number of symmetric pairs exceeds the number of compliers, as discussed earlier in Section 5.1. The empirical study described in the next section illustrates this point.

Online Table A1. Simulation Results of Constant Scale Ratios

Sample	Bivariate Normal			Bivariate Student-t(10)			Bivariate Student-t(3)			Bivariate Cauchy		
		Pairwise-Matching			Pairwise-Matching			Pairwise-Matching			Pairwise-Matching	
	IV	Variance	Quantile	IV	Variance	Quantile	IV	Variance	Quantile	IV	Variance	Quantile
Size	Estimator	-Based	-Based	Estimator	-Based	-Based	Estimator	-Based	-Based	Estimator	-Based	-Based
Case 1. Constant Scale Ratios:												
Mean Bias												
5000	.035	.009	-.019	.001	-.047	-.078	-.365	.032	-.036	1.650	70.349	.044
190	3.225	.176	.278	2.144	.189	.611	.952	.417	.318	-	-	.621
760	.537	.032	.049	.168	.040	.050	-.071	.188	.055	-	-	.104
3040	.148	.008	.013	-.043	.014	.013	-.203	.095	.016	-	-	.032
Median Bias												
190	.914	.128	.200	5.354	.137	.185	.217	.294	.206	1.855	25	.382
760	.271	.029	.039	.066	.036	.039	-.126	.133	.045	1.747	28	.082
3040	.133	.007	.010	-.051	.013	.010	-.222	.071	.013	1.866	30	.031
RMSE												
190	6.622	.356	.670	.523	.433	13.0	3.690	.776	1.460	-	-	1.762
760	1.373	.094	.145	.753	.113	.150	.378	.431	.159	-	-	.236
3040	.187	.043	.067	.094	.054	.070	.261	.243	.075	-	-	.103
MAD												
190	.914	.156	.235	.525	.172	.228	.375	.312	.249	1.855	25	.413
760	.271	.061	.092	.128	.072	.095	.200	.163	.105	1.747	28	.141
3040	.133	.030	.044	.068	.036	.046	.226	.099	.050	1.866	30	.069

Online Table A2. Simulation Results of Covariate-Dependent Scale Ratios

Sample Size	Bivariate Normal			Bivariate Student-t(10)			Bivariate Student-t(3)			Bivariate Cauchy		
		Pairwise-Matching			Pairwise-Matching			Pairwise-Matching			Pairwise-Matching	
	IV Estimator	Variance -Based	Quantile -Based	IV Estimator	Variance -Based	Quantile -Based	IV Estimator	Variance -Based	Quantile -Based	IV Estimator	Variance -Based	Quantile -Based
Case 2. r(x2=0), satisfying LATE assumptions:												
Mean Bias												
380	3.198	17.065	.298	1.991	3.197	.272	.953	18.605	.316	-	-	.655
760	1.415	.460	.108	.674	.890	.101	.178	1.505	.119	-	-	.229
3040	.255	.314	.026	.011	.335	.027	-.149	.693	.030	-	-	.058
Median Bias												
380	.941	.615	.182	.531	.657	.176	.208	1.126	.185	1.670	25.6	.391
760	.514	.412	.085	.211	.440	.072	.004	.766	.097	1.710	23.6	.170
3040	.188	.309	.020	-.016	.322	.021	-.182	.523	.024	1.802	28.9	.053
RMSE												
380	6.496	619.331	1.751	5.030	60.360	.701	3.369	837.411	1.806	-	-	3.164
760	3.370	.620	.239	2.164	21.482	.252	1.290	16.936	.274	-	-	.606
3040	.457	.349	.098	.169	.381	.100	.355	2.084	.109	-	-	.159
MAD												
380	.941	.615	.229	.532	.657	.233	.377	1.126	.243	1.670	25.590	.421
760	.514	.412	.139	.236	.441	.138	.241	.766	.150	1.710	23.575	.229
3040	.188	.309	.064	.082	.322	.064	.200	.523	.072	1.802	28.927	.102
Case 3. r(x2=0), as there is no complier:												
Mean Bias												
380	-	4.453	.280	-	3.176	.290	-	3.724	.328	-	-	.738
760	-	1.601	.111	-	.520	.109	-	1.593	.132	-	-	.250
3040	-	.317	.027	-	.340	.030	-	.779	.034	-	-	.066
Median Bias												
380	-	.635	.202	-	.709	.203	-	1.277	.234	-	31.4	.454
760	-	.433	.098	-	.465	.092	-	.878	.115	-	28.6	.213
3040	-	.311	.024	-	.328	.026	-	.549	.029	-	38.3	.061
RMSE												
380	-	126.434	1.254	-	89.443	1.163	-	54.784	1.649	-	-	3.376
760	-	62.341	.214	-	.716	.224	-	13.190	.242	-	-	.396
3040	-	.341	.085	-	.374	.087	-	3.438	.094	-	-	.139
MAD												
380	-	.635	.225	-	.709	.221	-	1.277	.251	-	31.4	.458
760	-	.433	.128	-	.465	.128	-	.878	.144	-	28.6	.230
3040	-	.311	.056	-	.328	.057	-	.549	.061	-	38.3	.091

Online Table A3: Mean and Standard Deviation

	All Schooling Levels		High School or Less		Some College	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
	(1)	(2)	(3)	(4)	(5)	(6)
Years of schooling	13.26	2.68	11.11	1.61	15.37	1.64
(a) Outcome						
Log hourly earnings 1976	1.66	.44	1.56	.44	1.75	.43
(b) Instrument						
4-yr college in county	.68	.47	.63	.48	.73	.44
(c) Covariates						
Age in 1976	28.12	3.14	28.21	3.24	28.03	3.04
Work experience in 1976	8.86	4.14	11.09	3.79	6.67	3.20
Proportion black	.23	.42	.32	.47	.15	.35
Residence south in 1976	.40	.49	.46	.50	.35	.48
Live in metropolitan 1976	.71	.45	0.64	0.48	.78	.41
Both parents attended college	.06	.24	.01	.09	.11	.32
Live in metropolitan 1966	.65	.48	.60	.49	.70	.46
(d) Size	3010	3010	1489	1489	1521	1521

Online Table A4. IV Estimates of Return to Schooling

Estimation method	OLS (1)	2SLS (2)	2SLS (3)	2SLS (4)	OLS (5)
(a) Coefficient on years of schooling	.073	.132	.133	.122	.075
	(.004)	(.049)	(.049)	(.053)	(.004)
Quadratic experience	yes	yes	no	no	no
Full set of demographics	yes	yes	yes	no	no
Card (1995): Table-Column	T2-C2	T3-C5a	-	-	-
Kling (2001): Table-Column	-	-	T1-C4	-	-
(b) Coefficient on college attendance	.216	.688	.693	.673	.213
	(.018)	(.279)	(.280)	(.329)	(.017)
Quadratic experience	yes	yes	no	no	no
Full set of demographics	yes	yes	yes	no	no

Note: The number of observations in all regressions=3010. Standard errors in parentheses. In columns (2)-(5), schooling and experience are treated as endogenous variables, with college proximity and age as excluded instruments. All the columns include a racial black indicator, residence in Southern states in 1976, residence in a metropolitan area in 1966 and 1976, and college attendance of both parents. Following Kling's (2001) suggestion, we exclude the quadratic term for work experience, take work experience as an endogenous variable in the model, and use age as an excluded instrument for work experience. Card (1995) and Kling (2001) additionally control for the following variables that are excluded in columns (4)-(5): eight regional dummies, indicators for living with both parents and for living only with mother, and eleven interaction terms of parental education variables. We apply the models in columns (4)-(5) to estimate the degree of dispersion in wages.