Uniform Convergence of Weighted Sums of Non- and Semi-parametric Residuals for Estimation and Testing
– Supplemental Materials –

Juan Carlos Escanciano* Indiana University
David T. Jacho-Chávez† Emory University
Arthur Lewbel‡ Boston College

First draft: May 2010. This draft: April 2013

Abstract

This supplement provides additional material that is not included in the main text due to space constraints. This supplement contains the summaries of how our results can be generically applied to derive asymptotic properties of semiparametric estimators such as Ichimura (1993), Klein and Spady (1993) and Rothe (2009), allowing for data-driven bandwidths, random trimming and estimated weights.

1 Some Generic Applications

In this section, we illustrate the general applicability of Theorems 3.1 and 3.2 in the main text to a variety of settings in semiparametric estimation. In particular, we summarize how the asymptotic distribution of semiparametric estimators such as Ichimura (1993), Klein and Spady (1993) and Rothe (2009) may be derived using our results, which allow for data-driven bandwidths, random trimming and estimated weights. Throughout this section technicalities are omitted for the sake of clarity, since our goal here is only to sketch how our results could be used for classes of applications beyond the specific ones provided in the main text.

1.1 Ichimura's (1993) Estimator

Consider the class of functions $\mathcal{W} = \{x \to W(\theta) := v(\theta, x) : \theta \in \Theta \subset \mathbb{R}^{d_X}\}$, where $v(\cdot,\cdot)$ is a known function, i.e. $v(\theta, x) = x^\top\theta$. The class $\mathcal{W}$ trivially satisfies Assumption 2 in the main text.

*Department of Economics, Indiana University, 105 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. E-mail: jescanci@indiana.edu. Web Page: http://mypage.iu.edu/~jescanci/. Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

†Department of Economics, Emory University, Rich Building 306, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA. E-mail: djachocha@emory.edu. Web Page: http://userwww.service.emory.edu/~djachoc/.

‡Corresponding Author: Department of Economics, Boston College, 140 Commonwealth Avenue, Chesnut Hill, MA 02467, USA. E-mail: lewbel@bc.edu. Web Page: http://www2.bc.edu/~lewbel/.

Denote $W_0 := W(\theta_0)$, $W_{i0} := v(\theta_0, X_i)$ and $W_i(\theta) := v(\theta, X_i)$ for $i = 1, \ldots, n$, where $\left\{Y_i, X_i^\top\right\}_{i=1}^n$ is a random sample from the joint distribution of $\left(Y, X^\top\right)^\top$ that fulfills the index restriction $E[Y|X] = E[Y|W_0(X)] =: m(W_0(X))$ for $\theta_0 \in \Theta$. Consider the following semiparametric least squares function

$$\mathcal{S}_n(\theta) := \frac{1}{n}\sum_{i=1}^n \widehat{t}_{ni}\{Y_i - \widehat{m}_{i\theta}\}^2, \tag{1.1}$$

where $\widehat{m}_{i\theta} := \widehat{m}(W_i(\theta)|W(\theta))$, $\widehat{t}_{ni} := \mathbb{I}(\widehat{f}(W_i(\widetilde{\theta})|W(\widetilde{\theta})) \geq \tau_n)$, $\tau_n \to 0$ as $n \to \infty$ at a rate that satisfies Assumption 11 in the main text, and $\widetilde{\theta}$ is a preliminary consistent estimator for θ_0, i.e. $\widetilde{\theta}$ could be an estimator that minimizes (1.1) but with $\widehat{t}_{ni} = \mathbb{I}(X_i \in A)$ for a compact set $A \subset \mathcal{X}_X$, and both $\widehat{m}$ and $\widehat{f}$ defined as in Section 3 in the main text. The proposed estimator $\widehat{\theta}$ of θ_0 is the minimizer of this objective function:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \mathcal{S}_n\left(\theta\right). \tag{1.2}$$

The asymptotic distribution of the estimator will be established here by a combination of standard methods and our Theorem 3.1. Consider the first order conditions

$$0 = \sqrt{n}\partial_\theta \mathcal{S}_n(\widehat{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - \widehat{m}_{i\widehat{\theta}}\}\partial_\theta \widehat{m}_{i\widehat{\theta}}\widehat{t}_{ni}, \tag{1.3}$$

where $\partial_\theta \widehat{m}_{i\widehat{\theta}} := \partial \widehat{m}(W_i(\theta)|W(\theta))/\partial\theta|_{\theta=\widehat{\theta}}$. By a Taylor series expansion,

$$\sqrt{n}(\widehat{\theta} - \theta_0) = G_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - \widehat{m}_{i\theta_0}\}\partial_\theta \widehat{m}_{i\theta_0}\widehat{t}_{ni} + o_P\left(1\right),$$

where $G_n = n^{-1}\sum_{i=1}^n \widehat{t}_{ni}\partial_\theta \widehat{m}_{i\widehat{\theta}}\partial_\theta^\top \widehat{m}_{i\bar{\theta}}$ and $\bar{\theta}$ is such that $|\bar{\theta} - \theta_0| \leq |\widehat{\theta} - \theta_0|$ a.s. By the uniform consistency results in Appendix B, and the continuous mapping theorem, it follows that

$$G_n \to_P \Lambda_0 =: E[\partial_\theta m(W_{i0})\partial_\theta^\top m(W_{i0})], \tag{1.4}$$

where $\partial_\theta m(W_{i0}) := \partial m(W_i(\theta)|W(\theta))/\partial\theta|_{\theta=\theta_0}$. By another application of the results in the main text with the uniform consistency of $\partial_\theta \widehat{m}_{i\theta_0}$ shown in Appendix B, we have

$$\Lambda_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - \widehat{m}_{i\theta_0}\}\partial_\theta \widehat{m}_{i\theta_0}\widehat{t}_{ni} = \Lambda_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - m(W_{0i})\}\partial_\theta m(W_{i0}) + o_P\left(1\right),$$

where the equality above follows from the fact that $E[\partial_\theta m(W_{i0})|W_{i0}] = 0$, see Ichimura (1993, Lemma 5.6, p. 95). An application of Linderberg-Lévy CLT then yields

$$\sqrt{n}(\widehat{\theta} - \theta_0) \to_d N(0, \Lambda_0^{-1}\Xi\Lambda_0^{-1}),$$

where $\Xi = E[\text{var}(Y|X = X_i)\partial_\theta m(W_{i0})\partial_\theta^\top m(W_{i0})]$.

Remark 1.1 Delecroix, Hristache, and Patilea (2006) have also derived the asymptotic properties of Ichimura's (1993) estimator with random (non-vanishing) trimming and uniformly in the bandwidth, while Härdle, Hall, and Ichimura (1993) have shown the first order asymptotic properties of the semiparametric least squares estimator are not affected when plugging in a data-dependent bandwidth chosen jointly with $\widehat{\theta}$ in (1.2) using fixed trimming. Our result above essentially combines these features, while extending them to vanishing trimming.

1.2 Klein and Spady's (1993) Estimator

Klein and Spady (1993) consider the binary choice model of the form

$$Y = \mathbb{I}(X^\top \theta_0 - u > 0), \tag{1.5}$$

where u and $X \in \mathcal{X}_X \subseteq \mathbb{R}^{d_X}$ are independent. Consider the class of functions $\mathcal{W} = \{x \to W(\theta) := v(\theta, x) : \theta \in \Theta \subset \mathbb{R}^{d_X}\}$, where $v(\cdot,\cdot)$ is a known function, i.e. $v(\theta, x) = x^\top \theta$. As before, denote $W_0 := W(\theta_0)$, $W_{i0} := v(\theta_0, X_i)$ and $W_i(\theta) := v(\theta, X_i)$ for $i = 1, \ldots, n$, where $\left\{Y_i, X_i^\top\right\}_{i=1}^n$ is a random sample from the joint distribution of $\left(Y, X^\top\right)^\top$ generated from model (1.5). In this case, the regression of Y given X satisfies the index restriction $E[Y|X] = E[Y|W_0(X)] =: m(W_0(X))$ for $\theta_0 \in \Theta$. Our results can be used here to establish the asymptotic properties of a variant of Klein and Spady's (1993) estimator. First, define the semiparametric likelihood function as

$$\mathcal{L}_n(\theta) := \frac{1}{n}\sum_{i=1}^n \{Y_i \log[\widehat{m}_{i\theta}] + (1 - Y_i)\log[1 - \widehat{m}_{i\theta}]\} \widetilde{t}_{in}, \tag{1.6}$$

where $\widehat{m}_{i\theta}$, and $\widetilde{t}_{in}$ are like those in Section 1.1 above but with $\widehat{g}$ replaced by one, and $\widetilde{\theta}$ is a preliminary consistent estimator for θ_0 (see Klein and Spady, 1993, footnote 4, p. 399 for examples). Similarly, both $\widehat{m}$ and $\widehat{f}$ are defined as in Section 3. The proposed estimator $\widehat{\theta}$ of θ_0 is the maximizer of this objective function:

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}_n(\theta). \tag{1.7}$$

The asymptotic distribution of the estimator can be established here by a combination of standard methods and Theorem 3.1 in the main text. Now consider the first order conditions

$$0 = \sqrt{n}\partial_\theta \mathcal{L}_n(\widehat{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - \widehat{m}_{i\widehat{\theta}}\}\widehat{\psi}_i \widetilde{t}_{in} \tag{1.8}$$

where $\widehat{\psi}_{i\widehat{\theta}} := \partial_\theta \widehat{m}_{i\widehat{\theta}}[\widehat{m}_{i\widehat{\theta}}(1 - \widehat{m}_{i\widehat{\theta}})]^{-1}$ and $\partial_\theta \widehat{m}_{i\widehat{\theta}}$ is as in Section 1.1 above. Now by a Taylor series expansion,

$$\sqrt{n}(\widehat{\theta} - \theta_0) = H_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - \widehat{m}_{i\theta_0}\}\widehat{\psi}_{i\widehat{\theta}}\widetilde{t}_{in} + o_P(1),$$

where

$$H_n = \frac{1}{n}\sum_{i=1}^n \frac{\partial_\theta \widehat{m}_{i\widehat{\theta}}\partial_\theta^\top \widehat{m}_{i\bar{\theta}}}{\widehat{m}_{i\widehat{\theta}}(1 - \widehat{m}_{i\widehat{\theta}})}\widetilde{t}_{in}, \text{ and } |\bar{\theta} - \theta_0| \le |\widehat{\theta} - \theta_0| \text{ a.s.}$$

From results in the main text, the uniform consistency results in Appendix B, and the continuous mapping theorem, it follows that

$$H_n \to_P \Delta_0 \equiv E[\partial_\theta m(W_{i0})\partial_\theta^\top m(W_{i0})[m(W_{i0})(1 - m(W_{i0}))]^{-1}], \tag{1.9}$$

where $\partial_\theta m(W_{i0})$ is as in Section 1.1. By another application of the results in the paper along with the uniform consistency of $\widehat{\psi}_{i\theta}$ we have, using that $E[\partial_\theta m(W_{i0})|W_{i0}] = 0$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - \widehat{m}_{i\theta_0}\}\widehat{\psi}_{i\theta_0}\widetilde{t}_{in} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{Y_i - m(W_{0i})\}\psi_{i\theta_0} + o_P(1),$$

where $\psi_{i\theta_0} = \partial_\theta m(W_{i0})/[m(W_{i0})(1-m(W_{i0}))]^{-1}$. It then follows from (1.9) that

$$\sqrt{n}(\widehat{\theta}-\theta_0) = \Delta_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{Y_i - m(W_{i0})}{m(W_{i0})(1-m(W_{i0}))}\right]\partial_\theta m(W_{i0}) + o_P(1),$$

and an application of Linderberg-Lévy CLT yields

$$\sqrt{n}(\widehat{\theta}-\theta_0) \to_d N(0, \Delta_0^{-1}).$$

Remark 1.2 The trimming function $\widetilde{t}_{in}$ here is slightly different from Klein and Spady's (1993) in that the latter behaved like a smoothed indicator. As in Klein and Spady (1993), the asymptotic variance Δ_0^{-1} equals the semiparametric efficiency bound for the binary choice model as originally derived by Chamberlain (1986) and Cosslett (1987). The main novelty here is that this asymptotic distribution is shown to hold for potentially data-driven bandwidths.

1.3 Rothe's (2009) Estimator

Rothe (2009) considers the estimation of θ_0 in the 'endogenous' binary choice model of the form

$$Y = \mathbb{I}(\widetilde{X}^\top\theta_0 - u > 0), \tag{1.10}$$

where u is independent of $\widetilde{X} := (\widetilde{X}^e, \widetilde{X}^{-e})^\top \in \mathcal{X}_{\widetilde{X}^e}\times\mathcal{X}_{\widetilde{X}^{-e}} \subseteq \mathbb{R}^{d_{\tilde{X}}}$ only conditionally on V where $\widetilde{X}^e = g_0(\widetilde{X}^{-e}, Z) + V$, $E[V|\widetilde{X}^{-e}, Z] = 0$, g_0 is a vector of conditional mean functions of each of the $d_{\widetilde{X}^e}$-'endogenous' components of $\widetilde{X}$, i.e. $\widetilde{X}^e$, given the $d_{\widetilde{X}^{-e}}$-'exogenous' components of $\widetilde{X}$, i.e. $\widetilde{X}^{-e}$, and some d_Z–vector of exogenous instruments Z. Notice that $d_{\widetilde{X}} = d_{\widetilde{X}^e} + d_{\widetilde{X}^{-e}}$. Let $X := (\widetilde{X}^\top, Z^\top)^\top$, and consider the class of functions $\mathcal{W} = \{x \to W(\theta, g) := v(\theta, g, x) : \theta\in\Theta\subset\mathbb{R}^{d_{\tilde{X}}},\ g\in\mathcal{G}\}$, where $v(\cdot,\cdot,\cdot)$ is a $(1+d_{\widetilde{X}^e})$-dimensional known function, i.e. $v(\theta, g, x) = (\widetilde{x}^\top\theta, \widetilde{x}^e - g(\widetilde{x}^{-e}, z))^\top$. As before, denote $W_0 := W(\theta_0, g_0)$, $W_{i0} := v(\theta_0, g_0, X_i)$ and $W_i(\theta, g) := v(\theta, g, X_i)$ for $i = 1,\ldots,n$, for an iid sample $\{Y_i, X_i^\top\}_{i=1}^n$ from the joint distribution of $(Y, X^\top)$. In this case, the regression of Y given X satisfies the index restriction $E[Y|X] = E[Y|W_0(X)] =: m(W_{i0})$ for $\theta_0\in\Theta$, and $g_0\in\mathcal{G}$. Our results can be used here to establish the asymptotic properties of an efficient version of Rothe's (2009) estimator.

Rothe (2009) proposes estimating θ_0 as in (1.7) where $\widehat{m}_{i\theta} := \widehat{m}(W_i(\theta,\widehat{g})|W(\theta,\widehat{g}))$, $\widehat{t}_{ni} = \mathbb{I}(\widetilde{X}_i\in A)$ for a compact set $A\subset\mathcal{X}_{\widetilde{X}}$, and both $\widehat{m}$ and $\widehat{f}$ defined as in Section 3 in the main text. This type of fixed trimming affects the asymptotic distribution of his estimator. Consider instead $\widehat{t}_{ni} := \mathbb{I}(\widehat{f}(W_i(\widetilde{\theta},\widehat{g})|W(\widetilde{\theta},\widehat{g})) \geq \tau_n)$, $\tau_n\to 0$ as $n\to\infty$ at a rate that satisfies Assumption 7 in the main text, and $\widetilde{\theta}$ is a preliminary consistent estimator for θ_0, i.e. Rothe's (2009) original estimator, which uses fixed trimming $\widehat{t}_{ni} = \mathbb{I}(\widetilde{X}_i\in A)$ for a compact set $A\subset\mathcal{X}_{\widetilde{X}}$. The first order conditions are like (1.8). Similar arguments as in Section 1.2 and repeated application of Theorem 3.1 and Theorem 3.2 in the main text yields

$$\begin{aligned}\sqrt{n}(\widehat{\theta}-\theta_0) &= H_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{Y_i - \widehat{m}_{i\theta_0}\}\widehat{\psi}_{i\theta_0}\widehat{t}_{ni} + o_P(1),\\ &= \Delta_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{Y_i - m(W_{0i})\}\psi_{i\theta_0} - \Delta_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E[\psi_{i\theta_0}\partial_{\overline{g}}m(W_{0i})|\widetilde{X}^{-e}, Z]V_i + o_P(1),\end{aligned}$$

where $\partial_{\overline{g}} m(W_{0i}) := \partial m(W_i(\theta_0, \overline{g})|W_0)/\partial\overline{g}|_{\overline{g}=g_0}$, and the last equality holds uniformly in the bandwidth from Theorem 3.2 in the main text with $\psi_{i\theta_0} = \partial_\theta m(W_{0i})/[m(W_{0i})(1-m(W_{0i}))]^{-1}$. Finally, an application of Linderberg-Lévy CLT yields

$$\sqrt{n}(\widehat{\theta} - \theta_0) \to_d N(0, \Delta_0^{-1} + \Delta_0^{-1}\Psi_0\Delta_0^{-1}),$$

where $\Psi_0 = E[\xi(W_{0i})V_iV_i^\top \xi^\top(W_{0i})]$ and

$$\xi(W_{0i}) = E\left[\frac{\partial_\theta m(W_{0i})\partial_{\overline{g}} m(W_{0i})}{m(W_{0i})(1-m(W_{0i}))}\middle| \widetilde{X}^{-e}, Z\right].$$

Remark 1.3 The asymptotic variance of the estimator here is different from Rothe's (2009) because Rothe (2009) uses a fixed trimming function, i.e. $\widehat{t}_{ni} = \mathbb{I}(\widetilde{X}_i \in A)$ for a compact set $A \subset \mathcal{X}_{\widetilde{X}}$, that appears everywhere in the limiting distribution. However, if we neglect his trimming effect, taking $\widehat{t}_{ni} = 1$ for all $i = 1, \ldots, n$, then both expressions for the asymptotic variance will coincide, (see Rothe, 2009, Theorem 3, p. 55). Unlike Rothe's (2009) original calculations that use results in Chen, Linton, and van Keilegom (2003), the results in the main text can be used to allow for plug-in data driven bandwidths and random trimming, while avoiding the need to explicitly calculate pathwise derivatives.

References

Chamberlain, G. (1986). "Asymptotic Efficiency in Semiparametrid Models with Censoring," *Journal of Econometrics*, 32(2), 189–218.

Chen, X., O. B. Linton, and I. van Keilegom (2003): "Estimation of Semiparametric Models when the Criterion Function Is Not Smooth," *Econometrica*, 71(5), 1591–1608.

Cosslett, S. R. (1987). "Efficiency Bounds for Distribution-free Estimators of the Binary Choice and Censored Regression Models," *Econometrica*, 55(3), 559–586.

Delecroix, M., M. Hristache, and V. Patilea (2006): "On Semiparametric M-Estimation in Single-Index Regression," *Journal of Statistical Planning and Inference*, 136(3), 730–769.

Härdle, W., P. Hall, and H. Ichimura (1993): "Optimal Smoothing in Single-Index Models," *The Annals of Statistics*, 21(1), 157–178.

Ichimura, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58(1-2), 71–120.

Klein, R. and R. Spady (1993) "An efficient Semiparametric Estimator for Discrete Choice Models", *Econometrica*, 61(2), 387-421.

Rothe, C. (2009): "Semiparametric Estimation of Binary Response Models with Endogenous Regressors," *Journal of Econometrics*, 153(1), 51–64.