

Model robust hybrid likelihood

Ingrid Dæhlen and Nils Lid Hjort

Department of Mathematics, University of Oslo

Supplementary Material

The supplementary material contains proofs of Theorem 1, 2, 3 and 4 in the paper.

We will give formal proofs for the theorems in the article. The arguments utilize techniques for proofs concerning estimators fitted by two-stage procedures and are similar to those given by e.g. Imbens, Spady, and Johnson (1998) and Schennach (2007). Before we begin, however, we will state and prove a result concerning consistency of Z-estimators. The result is a slight modification of Theorem 5.42 in Van der Vaart (2000).

Lemma 1. *Let $Y_1, \dots, Y_n \in \mathbb{R}^d$ be i.i.d. from some distribution F and let $\psi: \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^q$ be a function with $\Theta \subseteq \mathbb{R}^p$. Assume there is a unique solution θ_n to the equation $0 = \Psi_n(\theta) = n^{-1} \sum_{i=1}^n \psi(Y_i, \theta)$ for all n and a unique solution θ_0 to the limit equation $0 = \Psi(\theta) = E[\psi(Y, \theta)]$ when $Y \sim F$. Then θ_n converges in probability to θ_0 , provided*

(C1) the function Ψ is continuously differentiable in a neighborhood of θ_0 and $\Psi'(\theta_0)$ is invertible;

(C2) there exists a neighborhood $\mathcal{N}$ of θ_0 and an F -integrable function p_0 such that $\|\psi(y, \theta)\| \leq p_0(y)$ for all $\theta \in \mathcal{N}$ and F -almost all y .

Proof. Let Ψ and Ψ_n be defined as in the lemma. Close inspection of the proof of Theorem 5.42 in Van der Vaart (2000) shows that the conditions of the theorem are slightly too strict

and that the only thing needed for the arguments to go through is that Ψ is continuously differentiable with non-singular derivative in a neighborhood of θ_0 and that there exists a closed ball around θ_0 inside which Ψ_n converges uniformly to Ψ . The first assessment is ensured by (C1) and continuity of the determinant function. The second condition follows from (C2) and the uniform law of large numbers. Hence, there exists a sequence of estimators converging in probability to θ_0 . By assumption, $\Psi_n(\theta) = 0$ has a unique solution θ_n . Because of this, θ_n must be this consistent sequence, proving that θ_n converges in probability to θ_0 . $\square$

S1 Proving Theorem 1

We start by proving Theorem 1 in the article. First we will demonstrate an initial and related result.

Lemma 2. *Let $Y_1, \dots, Y_n \in \mathbb{R}^d$ be i.i.d. from some distribution F and $m: \mathbb{R}^d \times \mathcal{M} \rightarrow \mathbb{R}^q$, where $\mathcal{M} \subseteq \mathbb{R}^s$ is an open set, be a function continuously differentiable in the last s arguments. Let $\hat{\mu}$ be the unique maximizer of the empirical likelihood function in $\mathcal{M}$ constructed using m and the data $Y_1, \dots, Y_n$. Furthermore, assume $\hat{\lambda}$ is the unique solution to*

$$0 = n^{-1} \sum_{i=1}^n \frac{m(Y_i, \hat{\mu})}{1 + \lambda^T m(Y_i, \hat{\mu})},$$

and $\hat{\alpha} = (\hat{\lambda}, \hat{\mu})$. Set

$$\psi_{\text{EL}}(y, \alpha) = [1 + \lambda^T m(y, \mu)]^{-1} (m(y, \mu)^T, \lambda^T \partial m(y, \mu))^T$$

with ∂m as short hand for $\partial m / \partial \mu$. Lastly assume $\Gamma^{\text{EL}}(\lambda, \mu) = 0$ has a unique solution $\alpha_0 = (\lambda_0, \mu_0)$, where $Y \sim F$. Then $\hat{\alpha}$ is consistent for α_0 and

$$\hat{\alpha} = \alpha_0 - n^{-1} J \Gamma^{\text{EL}}(\alpha_0)^{-1} \sum_{i=1}^n \psi_{\text{EL}}(Y_i, \alpha_0) + o_{\text{pr}}(1/\sqrt{n}) \quad (\text{S1.1})$$

provided the following conditions hold true:

- (1) the function Γ^{EL} is continuously differentiable in a neighborhood of α_0 and its Jacobian matrix at α_0 is non-singular;
- (2) there exists $L > 0$ and a neighborhood $\mathcal{N}$ of α_0 on which $\Pr(1 + \lambda^T m(Y, \mu) > L) = 1$ for all $(\lambda, \mu) \in \mathcal{N}$;
- (3) there exists a neighborhood $\mathcal{N}$ of α_0 and functions p_j for $j = 0, 1, 2$ such that for F -almost all y and all $\alpha \in \mathcal{N}$, the norm of m and all first and second order partial derivatives of m are bounded by p_j for $j = 0, 1, 2$ respectively, with $E[p_0(Y)^4], E[p_1(Y)^4] < \infty$ and $E[p_2(y)^2] < \infty$.

In particular, $\sqrt{n}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N[0, J \Gamma^{\text{EL}}(\alpha_0)^{-1} K_{\text{EL}} J \Gamma^{\text{EL}}(\alpha_0)^{-1}]$, with $K_{\text{EL}} = \text{Var}[\psi_{\text{EL}}(Y, \alpha_0)]$, $Y \sim F$.

Proof. Since m is differentiable and $\mathcal{M}$ is open, the MEL estimator satisfies the following equation

$$0 = \frac{\partial}{\partial \mu} \log \text{EL}_n(\mu) = - \sum_{i=1}^n \frac{\lambda'_n(\mu)^T m(Y_i, \mu) + \partial m(Y_i, \mu)^T \lambda_n(\mu)}{1 + \lambda_n(\mu)^T m(Y_i, \mu)} = - \sum_{i=1}^n \frac{\partial m(Y_i, \mu)^T \lambda_n(\mu)}{1 + \lambda_n(\mu)^T m(Y_i, \mu)},$$

where the last equality follows from definition of $\lambda_n(\mu)$. Hence, $(\lambda_n(\hat{\mu}), \hat{\mu})$ is the solution to $0 = n^{-1} \sum_{i=1}^n \psi_{\text{EL}}(Y_i, \alpha)$, where ψ_{EL} is as defined in the lemma. This shows that $\hat{\alpha}$ is a Z-estimator, and hence, standard results concerning such estimators can be used to study limiting properties of $\hat{\alpha}$.

We start by showing consistency towards α_0 . This is achieved by proving condition (C2) of Lemma 1. (C1) holds by condition (1) of this lemma. Fix a neighborhood $\mathcal{N}$ of α_0 such that all conditions of the lemma hold true on $\mathcal{N}$. By assumption (2) there exists $L > 0$ such that $1 + \lambda^T M(\mu) > L$ with probability 1 for all $\alpha \in \mathcal{N}$. Combining this with condition (3) ensures $\|\psi_{\text{EL}}(y, \alpha)\| \leq C_1 p_0(y) + C_2 p_1(y)$ for some constants $C_1, C_2 \in \mathbb{R}$. The right hand side

of this inequality is a sum of F -integrable functions. This shows condition (C2) of Lemma 1, and guarantees that $\hat{\alpha}$ converges in probability to α_0 .

We will use Theorem 5.21 in Van der Vaart (2000) to show (S1.1). We start with the Lipschitz condition. Direct computations show $\|\frac{\partial \psi}{\partial \alpha}(y, \alpha)\| \leq C_3 p_0(y)^2 + C_4 p_0(y) p_1(y) + C_5 p_1(y) + C_6 p_1(y)^2 + C_7 p_2(y)$, for some constants C_j for $j = 3, \dots, 7$. This is a consequence of condition (3) of the lemma. Since all functions on the right hand side of this equation are square integrable with respect to the measure F , the sum is square integrable as well. This in conjunction with the mean value theorem for multivariate functions, ensures the Lipschitz property of Theorem 5.21 in Van der Vaart (2000). Furthermore, $E[\|\psi(Y, \alpha_0)\|^2] < \infty$ by a similar argument. The rest of the conditions of Theorem 5.21 in Van der Vaart (2000) hold by assumption. (S1.1) follows.

Strictly speaking, we have only showed that the sequence of solutions to $n^{-1} \sum_{i=1}^n \psi(Y_i, \alpha)$ satisfies the properties of the lemma. Since, $\log \text{EL}_n(\mu)$ is equal to $-\sum_{i=1}^n \log[1 + \lambda_n(\mu)^T m(Y_i, \mu)]$ only when $1 + \lambda_n(\mu)^T m(Y_i, \mu) > 0$ for all $i = 1, \dots, n$, $\partial \log \text{EL}_n(\hat{\mu}) / \partial \mu = 0$ is not immediately equivalent to $n^{-1} \sum_{i=1}^n \psi(Y_i, \hat{\alpha}) = 0$. However, by construction the empirical likelihood function is always non-negative and equal to zero when $1 + \lambda_n(\mu)^T m(Y_i, \mu) \leq 0$ for some $i = 1, \dots, n$. Hence, the maximizer of $-n^{-1} \sum_{i=1}^n \log[1 + \lambda_n(\mu)^T m(Y_i, \mu)]$ is equal to the maximizer of EL_n as long as it satisfies $1 + \lambda_n(\mu)^T m(Y_i, \mu) > 0$ for all $i = 1, \dots, n$. Furthermore, since $1 + \lambda^T m(Y_i, \mu) > 0$ for all $i = 1, \dots, n$ holds on $\mathcal{N}$ by assumption (2), $\hat{\alpha}$ will satisfy $1 + \lambda_n(\mu)^T m(Y_i, \mu) > 0$ for all $i = 1, \dots, n$ and be the maximizer of $\log \text{EL}_n(\mu)$ with probability tending to one. $\square$

S1.1 Proof of Theorem 1

By Lemma 2, $\hat{\mu}$ converges in probability to the last s components of α_0 , the parameter solving $0 = E[\psi_{\text{EL}}(Y, \alpha)]$. By definition $\lambda(\mu)$ is the root of $E\{m(Y, \mu)/[1 + \lambda^T m(Y, \mu)]\}$, the first q component functions in $E[\psi_{\text{EL}}(Y, \alpha)]$. Hence, $\alpha_0 = (\lambda(\mu_0), \mu_0)$ where μ_0 is the root of $E\{\partial m(Y, \mu)^T \lambda(\mu)/[1 + \lambda(\mu)^T m(Y, \mu)]\}$. Furthermore, the maximizer of $-E\{\log[1 + \lambda(\mu)^T m(Y, \mu)]\}$ satisfies $0 = E\{\lambda'(\mu)^T m(Y, \mu) + \partial m(Y, \mu)^T \lambda(\mu)/[1 + \lambda(\mu)^T m(Y, \mu)]\}$, since $\mathcal{M}$ is open and m is differentiable. Since $\lambda(\mu)$ is the root of $E\{m(Y, \mu)/[1 + \lambda^T m(Y, \mu)]\}$, the right hand side of this equation is equal to $E\{\partial m(Y, \mu)^T \lambda(\mu)/[1 + \lambda(\mu)^T m(Y, \mu)]\}$ which, by definition of μ_0 , has μ_0 as a root. In particular, μ_0 maximizes $-E\{\log[1 + \lambda(\mu)^T m(Y, \mu)]\}$ in the interior of the parameter space. Since $\mathcal{M}$ is open and the maximizer of $-E\{\log[1 + \lambda(\mu)^T m(Y, \mu)]\}$ is unique, μ_0 is the unique maximizer in $\mathcal{M}$.

For (2.7) in the main part of the article, let $(0, I)$ be the $(q + s) \times (q + s)$ block matrix where I is the $s \times s$ -identity matrix. Then, by (S1.1),

$$\hat{\mu} = (0, I)\hat{\alpha} = \mu_0 - (0, I)\Gamma^{\text{EL}}(\alpha_0)^{-1}n^{-1}\sum_{i=1}^n \psi(Y_i, \alpha_0) + o_{\text{pr}}(1/\sqrt{n}).$$

Let H_{xy} be the matrix obtained by first differentiating $E\{\log[1 + \lambda^T M(\mu)]\}$ with respect to y and afterwards with respect to x evaluated at (λ_0, μ_0) . Then $J\Gamma^{\text{EL}}(\alpha_0)$ can be written as the following block matrix

$$J\Gamma^{\text{EL}}(\alpha_0) = \begin{pmatrix} H_{\lambda\lambda} & H_{\mu\lambda} \\ H_{\lambda\mu} & H_{\mu\mu} \end{pmatrix}.$$

Furthermore, $\lambda'(\mu_0) = -H_{\lambda\lambda}^{-1}H_{\mu\lambda}$ by the implicit function theorem. Utilizing this and the formula for inverses of block matrices, we find $(0, I)\Gamma^{\text{EL}}(\alpha_0) = (A^{-1}\lambda'(\mu_0)^T, A^{-1}) = A^{-1}(\lambda'(\mu_0)^T, I)$, where $A = H_{\mu\mu} - H_{\lambda\mu}H_{\lambda\lambda}^{-1}H_{\mu\lambda}$. Direct computation shows that $A = J_{\text{EL}}$. Putting all of this together shows $-(0, I)\Gamma^{\text{EL}}(\alpha_0)^{-1}\psi(y, \alpha_0) = J_{\text{EL}}^{-1}(\lambda'(\mu_0)^T, I)\psi(y, \lambda_0, \mu_0) =$

$J_{\text{EL}}^{-1}\psi_\mu(y, \mu_0)$, proving (2.7) in the main part of the article.

S2 Proving Theorem 2

We proceed similarly as for the proof of Theorem 1, by first showing a lemma.

Lemma 3. *Let $Y_1, \dots, Y_n$ be i.i.d. from some distribution F and $m: \mathbb{R}^{d+r} \times \mathcal{M} \rightarrow \mathbb{R}^q$, where $\mathcal{M} \subseteq \mathbb{R}^s$ is open, be a function continuously differentiable in the last s arguments. Assume furthermore, that $\hat{P}$ is some estimator satisfying $n^{-1} \sum_{i=1}^n h(Y_i, \hat{P}) = o_p(1/\sqrt{n})$ and that $\hat{\mu}$ is the unique maximizer of the empirical likelihood function in $\mathcal{M}$ constructed using the estimating function $(y, \mu) \mapsto m(y, \hat{P}, \mu)$. Lastly, let $\hat{\lambda}$ be the unique solution to*

$$0 = n^{-1} \sum_{i=1}^n \frac{m(Y_i, \hat{P}, \hat{\mu})}{1 + \lambda m(Y_i, \hat{P}, \hat{\mu})}$$

and set

$$\psi_{\text{EL}}^P(y, \lambda, \mu, P) = [1 + \lambda^T m(y, P, \mu)]^{-1} (m(y, P, \mu)^T, \lambda^T \partial_\mu m(y, P, \mu))^T$$

with $\partial_\mu m$ being shorthand for $\partial m / \partial \mu$. Then, if $\alpha_{P,0} = (\mu_0, \lambda_0, P_0)$ is the unique solution to $0 = \Gamma_P(\lambda, \mu, P) = (\mathbb{E}[\psi_{\text{EL}}^P(Y, \alpha_P)]^T, \mathbb{E}[h(Y, P)]^T)^T$, the vector $\hat{\alpha}_P = (\hat{\mu}^T, \hat{\lambda}^T, \hat{P}^T)^T$ converges in probability to $\alpha_{P,0}$ and

$$\hat{\alpha}_P = \alpha_{P,0} - n^{-1} J \Gamma_P(\alpha_{P,0})^{-1} \sum_{i=1}^n \begin{pmatrix} \psi_{\text{EL}}^P(Y_i, \alpha_{P,0}) \\ h(Y_i, P_0) \end{pmatrix} + o_p(1/\sqrt{n}), \quad (\text{S2.2})$$

provided the following conditions hold true:

- (1) the function Γ_P is continuously differentiable in a neighborhood of $\alpha_{P,0}$ and its Jacobian at $\alpha_{P,0}$ is non-singular;
- (2) there exists $L > 0$ and a neighborhood $\mathcal{N}$ of $\alpha_{P,0}$ on which $\Pr(1 + \lambda^T m(Y, P, \mu) > L) = 1$ for all $(\lambda, \mu, P) \in \mathcal{N}$;

(3) *there exists a neighborhood $\mathcal{N}$ of α_{P_0} and functions p_j for $j = 0, 1, 2$ such that for F -almost all y and all $\alpha \in \mathcal{N}$, the norm of m and all first and second order partial derivatives of m are bounded by p_j for $j = 0, 1, 2$ respectively, with $E[p_0(Y)^4], E[p_1(Y)^4] < \infty$ and $E[p_2(y)^2] < \infty$.*

(4) *The norm of $h(y, P)$ and all first order partial derivatives of h with respect to P is bounded by functions q_1 and q_2 not depending on P and with $E[q_1(Y)], E[q_2(Y)^2] < \infty$.*

Proof. By assumption, the plug-in estimator $\hat{P}$ satisfies $n^{-1} \sum_{i=1}^n h(Y_i, \hat{P}) = o_p(1/\sqrt{n})$.

Hence, the vector $\hat{\alpha}_P$ is the solution to

$$0 = n^{-1} \sum_{i=1}^n \begin{pmatrix} \psi_{EL}^P(y, \lambda, \mu, P) \\ h(Y_i, P) \end{pmatrix} + \delta_n$$

where δ_n is a term of order $o_p(1/\sqrt{n})$ which does not depend on α_P . The lemma now follows from arguments completely analogue to those given in the proof of lemma 2. $\square$

As with Lemma 2, the above result give asymptotic properties of the full vector $\hat{\alpha}_P$. In practice, however, $\hat{\mu}$ alone is of interest, and because of this, a marginal result for $\hat{\mu}$ alone is given in Theorem 2 the paper. We give the proof in the following subsection

S2.1 Proof of Theorem 2

Let $\alpha_{P,0} = (\lambda_0^T, \mu_0^T, P_0^T)^T$ be the root of $E[\Gamma_P^{EL}(\alpha_P)] = (E[\psi_{EL}^P(Y, \alpha_P)]^T, E[h(Y, P)]^T)^T$. Then, by definition, P_0 is the solution to $E[h(Y, P)] = 0$ and λ_0 the solution to

$$E\left(\frac{m(Y, P_0, \mu_0)}{1 + \lambda^T m(Y, P_0, \mu_0)}\right) = 0.$$

Now let μ_1 be the maximizer of $-\log[1 + \lambda(\mu, P_0)^T m(Y, P_0, \mu)]$ in $\mathcal{M}$ where $\lambda(\mu, P)$ is the solution to.

$$E\left(\frac{m(Y, P, \mu)}{1 + \lambda^T m(Y, P, \mu)}\right) = 0.$$

Then μ_1 satisfies

$$0 = E \left(\frac{\partial_{\mu_1} \lambda(\mu, P_0)^T m(Y, P_0, \mu_1) + \lambda(\mu, P_0)^T \partial_{\mu_1} m(Y, P_0, \mu)}{1 + \lambda(\mu, P_0)^T m(Y, P_0, \mu_1)} \right). \quad (\text{S2.3})$$

By definition of $\lambda(\mu_1, P)$ the second term equals zero, implying

$$0 = E \left(\frac{\lambda(\mu, P_0)^T \partial_{\mu_1} m(Y, P_0, \mu)}{1 + \lambda(\mu, P_0)^T m(Y, P_0, \mu_1)} \right), \quad (\text{S2.4})$$

but by assumption this equation has a unique root at μ_0 . Hence, $\mu_1 = \mu_0$ and μ_0 is the maximizer of $-E\{\log[1 + \lambda(\mu, P_0)^T m(Y, P_0, \mu)]\}$. This proves that $\alpha_{P,0} = (\lambda(\mu_0, P_0)^T, \mu_0^T, P_0^T)^T$.

Secondly, consider the matrix $J \Gamma_P(\alpha)$. By definition of the Jacobian matrix, we can write $J \Gamma_P(\alpha)$ as the following block matrix:

$$J \Gamma_P(\alpha) = \begin{pmatrix} J \Gamma_P^{\text{EL}}(\lambda, \mu) & \partial_P \Gamma_P(\lambda, \mu, P) \\ 0 & \partial_P E h(Y, P) \end{pmatrix},$$

where $J \Gamma_P^{\text{EL}}(\lambda, \mu)$ is equal to $J \Gamma^{\text{EL}}(\lambda, \mu)$ defined in lemma 3 where the estimating function used is $(y, \mu) \mapsto m(y, P, \mu)$ for each fixed P and $\partial_P f$ is short-hand for $\partial f / \partial P$ for a general function f . By inversion of block matrices and condition (1) in theorem Theorem 2 in the paper, this matrix is continuously differentiable in a neighborhood of $(\lambda_0^T, \mu_0^T, P_0^T)^T$ with

$$[J \Gamma_p(\alpha_{P,0})]^{-1} = \begin{pmatrix} [J \Gamma_P^{\text{EL}}(\lambda, \mu)]^{-1} & \alpha'(P_0)[\partial_P E h(Y, P)]^{-1} \\ 0 & [\partial_P E h(Y, P)]^{-1} \end{pmatrix}$$

non-singular. In the above $\alpha(P)$ is the root of $(\lambda, \mu) \mapsto \Gamma_P(\lambda, \mu, P)$ for each fixed P . Hence, by arguments similar to those given in the proof of Theorem 1,

$$- (0 \ I_s \ 0) (J \Gamma_p(\alpha_{P,0}))^{-1} \psi_P(y, \alpha_{P,0}) =$$

$$J_{\text{EL}}^{-1} \psi_{\mu}^{P_0}(y, \mu_0) - \mu'(P_0)[\partial_P E h(Y, P)]^{-1} h(y, P_0),$$

where $\psi_{\mu}^{P_0}$ is equal to ψ_{μ} defined in Theorem 1 in the article when replacing m by $(y, \mu) \mapsto m(y, \mu, P_0)$. Lastly, by condition (4) in the theorem and theorem 5.21 in Van der Vaart (2000),

$\hat{P} - P_0 = -[\partial_P E h(Y, P)]^{-1} \sum_{i=1}^n h(Y_i, P_0) + o_p(1/\sqrt{n})$. Hence, $-\mu'(P_0)n^{-1}[\partial_P E h(Y, P)]^{-1} n^{-1} \sum_{i=1}^n h(Y_i, P_0) = \mu'(P_0)(\hat{P} - P_0) + o_p(1/\sqrt{n})$. This concludes the proof.

S3 Proving Theorem 3

Again we first show a lemma.

Lemma 4. *Let $Y_1, \dots, Y_n \in \mathbb{R}^d$ be i.i.d. from some distribution F and $m: \mathbb{R}^{d+s} \rightarrow \mathbb{R}^q$ a function continuously differentiable in the last s arguments. Assume that m is used in construction of the empirical likelihood function EL_n . Let f_θ for $\theta \in \Theta \subseteq \mathbb{R}^p$, where Θ is an open set, be some parametric family with score function u and $\mu: \mathbb{R}^p \rightarrow \mathbb{R}^s$ such that $E\{m[Y, \mu(\theta)]\} = 0$ when $Y \sim f_\theta$. Fix $a \in [0, 1)$ and assume that $\hat{\theta}$ is the unique maximizer of h_n . Furthermore, let $\hat{\lambda}(\hat{\theta})$ be the unique solution to*

$$0 = n^{-1} \sum_{i=1}^n \frac{m[Y_i, \mu(\hat{\theta})]}{1 + \lambda^T m[Y_i, \mu(\hat{\theta})]} = 0$$

Lastly, let $\beta = (\theta, \lambda)$, $\hat{\beta} = (\hat{\theta}, \hat{\lambda})$, and

$$\psi_{\text{HL}}(y, \beta) = (1 - a) \begin{pmatrix} 0 \\ u(y, \theta) \end{pmatrix} - a \begin{pmatrix} I & 0 \\ 0 & \mu'(\theta)^T \end{pmatrix} \psi_{\text{EL}}[y, \lambda, \mu(\theta)].$$

Assume that $\beta_0 = (\lambda_0, \mu_0)$ is the unique root of $\Gamma^{\text{HL}}(\lambda, \mu) = E[\psi(Y, \beta)]$, where $Y \sim F$. Then $\hat{\beta}$ is consistent for β_0 and

$$\hat{\beta} = \beta_0 - n^{-1} J \Gamma^{\text{HL}}(\beta_0)^{-1} \sum_{i=1}^n \psi_{\text{HL}}(Y_i, \beta_0) + o_{\text{pr}}(1/\sqrt{n}) \quad (\text{S3.5})$$

provided condition (1-3) of Lemma 2 hold true with $\alpha_0 = (\lambda_0^T, \mu(\theta_0)^T)^T$ and the following additional assumptions

(4) The function $\mu(\theta)$ is twice continuously differentiable in a neighborhood of θ_0 .

(5) *There exists a neighborhood of θ_0 on which all components of u and $\partial u/\partial \theta$ are bounded by some F -square integrable functions q_1 and q_2 respectively not depending on θ .*

In particular, $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, J \Gamma^{\text{HL}}(\beta_0)^{-1} K_{\text{HL}} J \Gamma^{\text{HL}}(\beta_0)^{-1})$, with $K_{\text{HL}} = \text{Var}[\psi_{\text{HL}}(Y, \beta_0)]$, $Y \sim F$.

Proof. Notice that

$$\begin{aligned} \frac{\partial h_n(\theta)}{\partial \theta} &= \sum_{i=1}^n \left\{ (1-a)u(Y_i, \theta) - a\mu'(\theta)^T \frac{\partial m[Y_i, \mu(\theta)]^T \lambda_n(\theta) + \lambda_n'(\theta)^T m[Y_i, \mu(\theta)]}{1 + \lambda_n(\theta)^T m[Y_i, \mu(\theta)]} \right\} \\ &= \sum_{i=1}^n \left\{ (1-a)u(Y_i, \theta) - a\mu'(\theta)^T \frac{\partial m[Y_i, \mu(\theta)]^T \lambda_n(\theta)}{1 + \lambda_n(\theta)^T m[Y_i, \mu(\theta)]} \right\}, \end{aligned}$$

where the last equality follows from the definition of $\lambda_n(\theta)$ and ∂m is short hand for $\partial m/\partial \mu$. Hence, since Θ is open and the maximizer of h_n in Θ unique, $\hat{\beta}$ is the solution to $0 = n^{-1} \sum_{i=1}^n \psi_{\text{HL}}(Y_i, \beta)$ where ψ_{HL} is as defined in the lemma. Because of this $\hat{\beta}$ is a Z-estimator and, just as for the MEL estimator, standard results can be used to study its limiting behavior.

Let $m_2: \mathbb{R}^{d+p} \rightarrow \mathbb{R}^q$ be defined as $m_2(y, \theta) = m[y, \mu(\theta)]$. Since μ is twice continuously differentiable as a function of θ , direct computations shows that conditions (1-3) hold for m_2 .

By the proof of lemma 2 this ensures that there exists a neighborhood on which the norm of

$$\begin{pmatrix} I & 0 \\ 0 & \mu'(\theta)^T \end{pmatrix} \psi_{\text{EL}}[y, \lambda, \mu(\theta)] = \begin{pmatrix} -m_2(y, \theta) \\ -\partial m_2(y, \theta)^T \lambda \end{pmatrix} [1 + \lambda^T m_2(y, \theta)]^{-1} \quad (\text{S3.6})$$

is bounded some square F -integrable function depending on neither λ nor μ . Combined with assumption (5), this proves that the norm of $\psi(y, \beta)$ is bounded by some square integrable function not depending on β . In particular, (C2) of Lemma 1 holds true. As (C1) holds by assumption, consistency of $\hat{\beta}$ towards β_0 follows.

To show (S4.7), we will again use Theorem 5.21 in Van der Vaart (2000). The only condition which remains to show is the Lipschitz condition. By the chain rule and proof of

Lemma 2, the norm of all partial derivatives of (S3.6) are bounded by some square integrable function. Combining this with assumption (5) of this lemma and the mean value theorem, ensures that there exists a square integrable function r such that $\|\psi_{\text{HL}}(y, \beta_1) - \psi_{\text{HL}}(y, \beta_2)\| \leq r(y)\|\beta_1 - \beta_2\|$. This proves the Lipschitz condition of Theorem 5.21 in Van der Vaart (2000) and concludes the proof. $\square$

S3.1 Proof of Theorem 3

The proof of Theorem 3 now follows directly from Lemma 4 by performing computations similar to those in Section S1.1.

S4 Proving Theorem 4

We start with a lemma, yet again.

Lemma 5. *Let $Y_1, \dots, Y_n \in \mathbb{R}^d$ be i.i.d. from some distribution F and $m: \mathbb{R}^{d+r+s} \rightarrow \mathbb{R}^q$ a function continuously differentiable in the last s arguments. Let $\hat{P}$ be some estimator satisfying $n^{-1} \sum_{i=1}^n h(Y_i, \hat{P}) = o_p(1/\sqrt{n})$ and assume that $(y, \mu) \mapsto m(y, \hat{P}, \mu)$ is used in construction of the empirical likelihood function EL_n . Let f_θ for $\theta \in \Theta \subseteq \mathbb{R}^p$, where Θ is an open set, be some parametric family with score function u and $\mu: \mathbb{R}^{p+r} \rightarrow \mathbb{R}^s$ such that $\text{E}\{m[Y, \mu(\theta, P), P]\} = 0$ when $Y \sim f_\theta$. Fix $a \in [0, 1)$ and assume that $\hat{\theta}(\hat{P})$ is the unique maximizer of h_n . Furthermore, let $M_i(\mu, P)$ be short-hand for $m(Y_i, P, \mu)$ and $\hat{\lambda}$ the unique solution to*

$$0 = n^{-1} \sum_{i=1}^n \frac{M_i[\mu(\hat{\theta}, \hat{P})]}{1 + \lambda^T M_i[\mu(\hat{\theta}, \hat{P})]} = 0$$

Lastly, let $\beta_P = (\theta, \lambda, P)$, $\hat{\beta} = (\hat{\theta}(\hat{P}), \hat{\lambda}, \hat{P})$, and

$$\psi_{\text{HL}}^P(y, \beta) = (1 - a) \begin{pmatrix} 0 \\ u(y, \theta) \\ 0 \end{pmatrix} - a \begin{pmatrix} I & 0 & 0 \\ 0 & \mu'(\theta)^T & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \psi_{\text{EL}}[y, \lambda, \mu(\theta), P] \\ h(y, P) \end{pmatrix}.$$

Then if $\beta_{P,0} = (\theta_0, \lambda_0, P_0)$ is the unique root of $\Gamma_P^{\text{HL}}(\lambda, \mu) = \mathbb{E}[\psi_{\text{HL}}^P(Y, \beta)]$ where $Y \sim F$, $\hat{\beta}_P$ is consistent for $\beta_{P,0}$ and

$$\hat{\beta} = \beta_0 - n^{-1} J \Gamma_P^{\text{HL}}(\beta_0)^{-1} \sum_{i=1}^n \psi_{\text{HL}}^P(Y_i, \beta_0) + o_{\text{pr}}(1/\sqrt{n}) \quad (\text{S4.7})$$

provided condition (1-4) of Lemma 3 hold true with $\alpha_{P,0} = (\mu(\theta_0)^T, \lambda_0^T, P_0^T)^T$ and the following additional assumptions

(5) For all P in an neighborhood of P_0 , function $\mu(\theta, P)$ is twice continuously differentiable with respect to θ in a neighborhood of θ_0 .

(6) There exists a neighborhood of θ_0 on which all components of u and $\partial u / \partial \theta$ are bounded by some F -square integrable functions q_1 and q_2 respectively not depending on θ .

Proof. The result follows from arguments similar to those given in the proof of Lemma 3 and Lemma 4 and is left out. $\square$

S4.1 Proof of Theorem 4

Theorem 4 now follows directly from the above lemma by arguing and performing computations similar to those in section S2.1.

Bibliography

Imbens, G. W., R. H. Spady, and P. Johnson (1998). Information theoretic approaches to inference in moment condition models.

Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics* 35, 634–672.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge, England: Cambridge University Press.