SUPPLEMENT TO "A NEYMAN-ORTHOGONALIZATION APPROACH TO THE INCIDENTAL PARAMETER PROBLEM IN LIKELIHOOD MODELS"

STÉPHANE BONHOMME
Department of Economics, University of Chicago

KOEN JOCHMANS
Toulouse School of Economics, Université Toulouse Capitole

MARTIN WEIDNER
Department of Economics and Nuffield College, University of Oxford

This supplement collects material omitted from the main text. Appendix A proves Lemma 2 of the main text. Appendix B describes our diagnostic for the order q of orthogonality. Appendix C gives implementation details for the nonlinear regression model. Appendix D reports a Monte Carlo experiment, Appendix E discusses restrictions that are independent of the individual effects, and Appendix F plots the estimated production function.

APPENDIX A: PROOF OF LEMMA 2 OF THE MAIN TEXT

For convenience we first restate the lemma.

LEMMA 2: *Let $q \in \{1,2,3,\dots\}$, and let x be some realization of the covariates. Remember that ∇_η^q and $w_q(z;\theta,\eta)$ are vectors of dimension $k_q = \sum_{p=1}^q d_p$, and that $\Sigma_{w_q w_q}(x;\theta,\eta)$ is a $k_q \times k_q$ matrix. We assume that $\Sigma_{w_q w_q}(x;\theta,\eta)$ is invertible, and we define*

$$\widetilde{w}_q(z;\theta,\eta) = \Sigma_{w_q w_q}(x;\theta,\eta)^{-1} w_q(z;\theta,\eta).$$

Then,

$$\mathbb{E}_{\theta,\eta}\left[(\nabla_\eta^q)^\top \, \widetilde{w}_q(Z;\theta,\eta) \middle| X = x \right] = \begin{pmatrix} -\mathbb{I}_{d_1} & 0 & 0 & \cdots & 0 \\ 0 & +\mathbb{I}_{d_2} & 0 & \cdots & 0 \\ 0 & 0 & -\mathbb{I}_{d_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (-1)^q \, \mathbb{I}_{d_q} \end{pmatrix}, \tag{A.1}$$

where the diagonal $k_q \times k_q$ matrix on the right hand side is obtained by stacking $(-1)^p \, \mathbb{I}_{d_p}$ on the diagonal for $p = 1,\dots,q$, and $\mathbb{I}_{d_p}$ is the identity matrix of dimensions d_p.

PROOF: By taking derivatives of $\int \ell(y|x;\theta,\eta)dy = 1$ with respect to η, we obtain

$$\int [\nabla_\eta^q \ell(y|x;\theta,\eta)] dy = 0,$$

which can also be written as $\mathbb{E}_{\theta,\eta}[w_q(Z;\theta,\eta)|X = x] = 0$. Since $\Sigma_{w_q w_q}(X;\theta,\eta)$ does not depend on Y we also have $\mathbb{E}_{\theta,\eta}[\widetilde{w}_q(Z;\theta,\eta)|X = x] = 0$. Using this and the definition of $\widetilde{w}_q$ we obtain

$$\begin{aligned}
\mathbb{I}_{k_q} &= \Sigma_{w_q w_q}(x;\theta,\eta)^{-1}\Sigma_{w_q w_q}(x;\theta,\eta) \\
&= \Sigma_{w_q w_q}(x;\theta,\eta)^{-1}\mathbb{E}_{\theta,\eta}[w_q(Z;\theta,\eta)w_q(Z;\theta,\eta)^\top|X=x] \\
&= \mathbb{E}_{\theta,\eta}[\widetilde{w}_q(Z;\theta,\eta)w_q(Z;\theta,\eta)^\top|X=x] \\
&= \int \widetilde{w}_q(z;\theta,\eta)[\nabla_\eta^q \ell(y|x;\theta,\eta)]^\top dy. \qquad\qquad (A.2)
\end{aligned}$$

According to its definition in Section 3.1 of the main text, the elements of the k_q-vector operator ∇_η^q are given by

$$D_\eta^m = \frac{\partial^p}{\partial\eta_{m_1}\cdots\partial\eta_{m_p}},$$

and are uniquely labeled by vectors of integers $m=(m_1,\ldots,m_p)$ of length $p\in\{1,\ldots,q\}$ in the following set

$$\mathcal{C}_q = \bigcup_{p\in\{1,\ldots,q\}} \{m=(m_1,\ldots,m_p): 1\le m_1\le\cdots\le m_p\le d_\eta\}.$$

Analogously, we now introduce the notation $\widetilde{w}_q^m(z;\theta,\eta)$, $m\in\mathcal{C}_q$, to uniquely denote the elements of the vector $\widetilde{w}_q(z;\theta,\eta)$, which is also a vector of length $k_q=|\mathcal{C}_q|$. With that notation, the result in display (A.2) can equivalently be written as

$$\forall r,m\in\mathcal{C}_q: \qquad \int \widetilde{w}_q^r(z;\theta,\eta)\,[D_\eta^m \ell(y|x;\theta,\eta)]\,dy = \mathbb{1}\{r=m\}. \qquad (A.3)$$

Since $\mathbb{E}_{\theta,\eta}[\widetilde{w}_q(Z;\theta,\eta)|X=x]=0$, we also have

$$\int \widetilde{w}_q^r(z;\theta,\eta)\,\ell(y|x;\theta,\eta)\,dy = 0.$$

For the empty vector $()$ of length zero we have $D_\eta^{()}\ell(y|x;\theta,\eta)=\ell(y|x;\theta,\eta)$. Using this notation we can combine the result in the last two displays to find that for all $r\in\mathcal{C}_q$ and all $v\in\mathcal{C}_q\cup\{()\}$ we have

$$\int \widetilde{w}_q^r(z;\theta,\eta)\,[D_\eta^v \ell(y|x;\theta,\eta)]\,dy = \mathbb{1}\{r=v\}.$$

For $k\in\{1,2,3,\ldots\}$, vector $t=(t_1,\ldots,t_k)\in\mathcal{C}_q$, and a subset $S\subseteq\{1,\ldots,k\}$, let t_S denote the vector formed by keeping only the indices in S, and let $t_{-S}=t_{\{1,\ldots,k\}\setminus S}$ be the vector of the remaining elements. Then, by applying D_η^t to the last display and using the product rule for differentiation we obtain

$$\sum_{S\subseteq\{1,\ldots,k\}} \int [D_\eta^{t_S}\widetilde{w}_q^r(z;\theta,\eta)]\,[D_\eta^{t_{-S}}D_\eta^v \ell(y|x;\theta,\eta)]\,dy = 0. \qquad (A.4)$$

Of course, we have $D_\eta^t D_\eta^v = D_\eta^{(t,v)}$, and instead of distinguishing between t and v we can also just write m for (t,v) combined. The last display equation then implies that for any nonempty subset $T \subseteq \{1, 2, \ldots, |m|\}$ we have (just set $t = m_T$ and $v = m_{-T}$ in the last display result—in doing so, it was important that above we allowed for v to be the empty vector):

$$\sum_{S \subseteq T} \int [D_\eta^{m_S} \widetilde{w}_q^r(z; \theta, \eta)] \, [D_\eta^{m_{-S}} \ell(y|x; \theta, \eta)] \, dy = 0.$$

Now, consider the following linear combination of the result in the last display,

$$\sum_{\substack{T \subseteq \{1,2,\ldots,|m|\} \\ T \neq \emptyset}} (-1)^{|T|} \sum_{S \subseteq T} \int [D_\eta^{m_S} \widetilde{w}_q^r(z; \theta, \eta)] \, [D_\eta^{m_{-S}} \ell(y|x; \theta, \eta)] \, dy = 0. \tag{A.5}$$

For a fixed $S \subseteq \{1, 2, \ldots, |m|\}$, the total coefficient of the term $\int [D_\eta^{m_S} \widetilde{w}_q^r] \, [D_\eta^{m_{-S}} \ell] \, dy$ in this linear combination is given by

$$\sum_{\substack{T \,:\, S \subseteq T \subseteq \{1,\ldots,|m|\} \\ T \neq \emptyset}} (-1)^{|T|} = \begin{cases} -1 & \text{if } S = \emptyset, \\ (-1)^{|m|} & \text{if } S = \{1, \ldots, |m|\}, \\ 0 & \text{otherwise.} \end{cases} \tag{A.6}$$

To see that the result in the last display holds, notice first that for $S = \emptyset$ the sum is simply

$$\sum_{T \,:\, \emptyset \neq T \subseteq \{1,\ldots,|m|\}} (-1)^{|T|} = -1 + \underbrace{\sum_{T \subseteq \{1,\ldots,|m|\}} (-1)^{|T|}}_{=0} = -1,$$

where $\sum_{T \subseteq \{1,\ldots,|m|\}} (-1)^{|T|} = \sum_{r=0}^{|m|} \binom{|m|}{r} (-1)^r = (1-1)^{|m|} = 0$ is a classic alternating sum result, which holds for all $|m| \geq 1$. Next, for $S = \{1, \ldots, |m|\}$, the left hand side of (A.6) only sums over one element, $T = \{1, \ldots, |m|\}$, and we thus get $(-1)^{|T|} = (-1)^{|m|}$ for the sum. Finally, if $S \neq \emptyset$ and $S \neq \{1, \ldots, |m|\}$, then the left hand side of (A.6) can be written as

$$\sum_{T \,:\, S \subseteq T \subseteq \{1,\ldots,|m|\}} (-1)^{|T|} = \sum_{R \subseteq -S} (-1)^{|S|+|R|} = (-1)^{|S|} \underbrace{\sum_{R \subseteq -S} (-1)^{|R|}}_{=0} = 0$$

where we replaced the sum over T by a sum over R such that $T = S \cup R$, with $-S = \{1, \ldots, |m|\} \setminus S$, and in the final step we used the alternating sum result again.

Using (A.6), our linear combination in (A.5) equals

$$-\int [D_\eta^m \widetilde{w}_q^r(z; \theta, \eta)] \, \ell(y|x; \theta, \eta) \, dy + (-1)^{|m|} \int \widetilde{w}_q^r(z; \theta, \eta) [D_\eta^m \ell(y|x; \theta, \eta)] \, dy = 0.$$

Together with (A.3) we thus find

$$\int [D_\eta^m \widetilde{w}_q^r(z; \theta, \eta)] \, \ell(y|x; \theta, \eta) \, dy = (-1)^{|m|} \, \mathbb{1}\{r = m\},$$

which in vector notation can be written as (A.1). *Q.E.D.*

APPENDIX B: ORDER q OF ORTHOGONALITY

B.1. *A diagnostic for q*

Consider a moment function $u(Z_i;\theta,\eta_i,\mu)$, and denote its q-orthogonal counterparts based on Theorem 1 of the main text as $u_q^*(Z_i;\theta,\eta_i,\mu)$, for $q=1,2,\ldots$. Given some order q we propose the following diagnostic:

$$S_q=\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\left[u_q^*(Z_i;\widehat{\theta},\widehat{\eta}_i,\widehat{\mu})-u_{q+1}^*(Z_i;\widehat{\theta},\widehat{\eta}_i,\widehat{\mu})\right]\right\|_{\widehat{V}^{-1}}^2,$$

where $\widehat{\theta}$ and $\widehat{\mu}$ are based on $(q+1)$-orthogonal functions, and $\widehat{V}$ is a consistent estimator of the variance

$$V=\lim_{n\to\infty}\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{N}(u_q^*(Z_i;\widehat{\theta},\eta_i,\widehat{\mu})-u_{q+1}^*(Z_i;\widehat{\theta},\eta_i,\widehat{\mu}))\right].$$

As a special case, suppose that one wishes to estimate θ, and $u(Z_i;\theta,\eta_i)$ does not depend on μ. Then, since $\widehat{\theta}$ is based on u_{q+1}^* that is uncorrelated with $u_q^*-u_{q+1}^*$, a consistent estimator of V is

$$\widehat{V}=\frac{1}{n}\sum_{i=1}^{N}(u_q^*(Z_i;\widehat{\theta},\widehat{\eta}_i)-u_{q+1}^*(Z_i;\widehat{\theta},\widehat{\eta}_i))(u_q^*(Z_i;\widehat{\theta},\widehat{\eta}_i)-u_{q+1}^*(Z_i;\widehat{\theta},\widehat{\eta}_i))^\top.$$

More generally, a consistent estimator of V is available using standard Taylor expansions and Lemma 1 of the main text. We provide an example in the case of the Neyman-Scott model in the next subsection.

Under the conditions of Theorem 2 of the main text, which in particular guarantee that the estimates $\widehat{\eta}_i$ converge at a fast enough rate, S_q converges in distribution to a χ_r^2 random variable, where $r=\dim u$. In practice, we thus suggest comparing S_q to the quantiles of the χ_r^2 distribution.

B.2. *Illustration in the Neyman-Scott model*

Consider the estimation of $\mu={}^1\!/\!{}_N\sum_{i=1}^N h(\eta_i)$ in the Neyman-Scott model introduced in Section 2.1 of the main text. Here we assume that T – the number of periods used in estimation – is fixed as N tends to infinity, and consequently we use $\sqrt{N}$ in order to scale all cross-sectional average quantities. Suppose first that σ is known. By equation (5.1) of the main text we have

$$\begin{aligned}&u_q^*(Y_i;\sigma^2,\eta_i,\mu)-u_{q+1}^*(Y_i;\sigma^2,\eta_i,\mu)\\&=-\sigma^{q+1}T^{-\frac{q+1}{2}}\frac{1}{(q+1)!}\nabla_\eta^{(q+1)}h(\eta_i)H_{q+1}\left(\frac{\sum_{j=1}^{T}(Y_{ij}-\eta_i)}{\sqrt{T}\sigma}\right),\end{aligned}$$

and

$$V = \lim_{N\to\infty} \sigma^{2q+2} T^{-q-1} \frac{1}{(q+1)!} \frac{1}{N} \sum_{i=1}^{N} [\nabla_\eta^{(q+1)} h(\eta_i)]^2.$$

Hence, again treating σ as known, the diagnostic is

$$S_q = \frac{\left[\sum_{i=1}^{N} \nabla_\eta^{(q+1)} h(\widehat{\eta}_i) H_{q+1} \left(\frac{\sum_{j=1}^{T} (Y_{ij} - \widehat{\eta}_i)}{\sqrt{T}\sigma} \right) \right]^2}{(q+1)! \sum_{i=1}^{N} [\nabla_\eta^{(q+1)} h(\widehat{\eta}_i)]^2}. \tag{B.1}$$

With σ unknown, we have, using $\widehat{\sigma}$ in equation (2.11) of the main text,

$$\begin{aligned}
&u_q^*(Y_i; \widehat{\sigma}^2, \eta_i, \mu) - u_{q+1}^*(Y_i; \widehat{\sigma}^2, \eta_i, \mu) \\
&= -\widehat{\sigma}^{q+1} T^{-\frac{q+1}{2}} \frac{1}{(q+1)!} \nabla_\eta^{(q+1)} h(\eta_i) H_{q+1} \left(\frac{\sum_{j=1}^{T} (Y_{ij} - \eta_i)}{\sqrt{T}\widehat{\sigma}} \right) \\
&= -\sigma^{q+1} T^{-\frac{q+1}{2}} \frac{1}{(q+1)!} \nabla_\eta^{(q+1)} h(\eta_i) H_{q+1} \left(\frac{\sum_{j=1}^{T} (Y_{ij} - \eta_i)}{\sqrt{T}\sigma} \right) \\
&\quad + \frac{1}{2T} \widetilde{\sigma}^{q-1} T^{-\frac{q-1}{2}} \frac{1}{(q-1)!} \nabla_\eta^{(q+1)} h(\eta_i) H_{q-1} \left(\frac{\sum_{j=1}^{T} (Y_{ij} - \eta_i)}{\sqrt{T}\widetilde{\sigma}} \right) (\widehat{\sigma}^2 - \sigma^2),
\end{aligned}$$

for $\widetilde{\sigma}$ between σ and $\widehat{\sigma}$, where we have used the expression for the derivatives of Hermite polynomials.

Hence, for $q \geq 2$ we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(u_q^*(Y_i; \widehat{\sigma}^2, \eta_i, \mu) - u_{q+1}^*(Y_i; \widehat{\sigma}^2, \eta_i, \mu) \right)$$

$$= -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma^{q+1}T^{-\frac{q+1}{2}}\frac{1}{(q+1)!}\nabla_\eta^{(q+1)}h(\eta_i)H_{q+1}\left(\frac{\sum_{j=1}^{T}(Y_{ij}-\eta_i)}{\sqrt{T}\sigma}\right)+o_P(1),$$

and we rely on the diagnostic S_q given by (B.1) even though σ is estimated. Now, for $q=1$ we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(u_1^*(Y_i;\widehat{\sigma}^2,\eta_i,\mu)-u_2^*(Y_i;\widehat{\sigma}^2,\eta_i,\mu)\right)\\
&= -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma^2T^{-1}\frac{1}{2}\nabla_\eta^{(2)}h(\eta_i)\left(\left(\frac{\sum_{j=1}^{T}(Y_{ij}-\eta_i)}{\sqrt{T}\sigma}\right)^2-1\right)\\
&\quad+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{2T}\nabla_\eta^{(2)}h(\eta_i)\left(\frac{1}{N(T-1)}\sum_{i'=1}^{N}\sum_{j=1}^{T}(Y_{i'j}-\overline{Y}_{i'})^2-\sigma^2\right)+o_P(1),
\end{aligned}$$

where we have used the expressions of H_2 and $\widehat{\sigma}^2$. That is,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(u_1^*(Y_i;\widehat{\sigma}^2,\eta_i,\mu)-u_2^*(Y_i;\widehat{\sigma}^2,\eta_i,\mu)\right)\\
&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{2T}\nabla_\eta^{(2)}h(\eta_i)\left[\frac{1}{N(T-1)}\sum_{i'=1}^{N}\sum_{j=1}^{T}(Y_{i'j}-\overline{Y}_{i'})^2-\left(\frac{\sum_{j=1}^{T}(Y_{ij}-\eta_i)}{\sqrt{T}}\right)^2\right]+o_P(1).
\end{aligned}$$

In this case,

$$V=\lim_{N\to\infty}\sigma^4T^{-2}\frac{1}{2}\left[\frac{1}{N}\sum_{i=1}^{N}[\nabla_\eta^{(2)}h(\eta_i)]^2+\frac{1}{T-1}\left(\frac{1}{N}\sum_{i=1}^{N}\nabla_\eta^{(2)}h(\eta_i)\right)^2\right],$$

and we rely on

$$S_1 = \frac{\left\{ \sum_{i=1}^{N} \nabla_{\eta}^{(2)} h(\widehat{\eta}_i) \left[\left(\frac{\sum_{j=1}^{T} (Y_{ij} - \widehat{\eta}_i)}{\sqrt{T}\widehat{\sigma}} \right)^2 - \frac{1}{N(T-1)} \sum_{i'=1}^{N} \sum_{j=1}^{T} \left(\frac{Y_{i'j} - \overline{Y}_{i'}}{\widehat{\sigma}} \right)^2 \right] \right\}^2}{2 \left[\sum_{i=1}^{N} [\nabla_{\eta}^{(2)} h(\widehat{\eta}_i)]^2 + \frac{1}{N(T-1)} \left(\sum_{i=1}^{N} \nabla_{\eta}^{(2)} h(\widehat{\eta}_i) \right)^2 \right]}.$$

REMARK B.1—(Sample splitting): When implementing the diagnostic in this example, we use a small number T of time periods in estimation. Keeping T fixed – while at the same time having $\widehat{\eta}_i \xrightarrow{p} \eta_i$ – is important for the theory in order to guarantee that the impact of the estimation error in $\widehat{\eta}_i$ does not affect the asymptotic distribution of S_q. In practice, one can cross-fit by taking multiple subsets of T time periods (along with different $\widehat{\eta}_i$'s) and average over them to reduce variability.

APPENDIX C: IMPLEMENTATION IN NONLINEAR REGRESSION

C.1. *Expression for M*

Following Constantine and Savits (1996), consider a multivariate function

$$h(x_1, \ldots, x_d) = f\left[g^{(1)}(x_1, \ldots, x_d), \ldots, g^{(m)}(x_1, \ldots, x_d)\right],$$

$h_\nu = D^\nu_{\mathbf{x}} h(\mathbf{x}^0), f_{\boldsymbol{\lambda}} = D^{\boldsymbol{\lambda}}_{\mathbf{y}} f(\mathbf{y}^0), g^{(i)}_{\boldsymbol{\mu}} = D^\mu_{\mathbf{x}} g^{(i)}(\mathbf{x}^0), \mathbf{g}_{\boldsymbol{\mu}} = \left(g^{(1)}_{\boldsymbol{\mu}}, \ldots, g^{(m)}_{\boldsymbol{\mu}}\right)$. The Faà di Bruno formula is (Theorem 2.1 in Constantine and Savits, 1996):

$$h_{\boldsymbol{\nu}} = \sum_{1 \le |\boldsymbol{\lambda}| \le n} f_{\boldsymbol{\lambda}} \underbrace{\sum_{s=1}^{n} \sum_{p_s(\boldsymbol{\nu}, \boldsymbol{\lambda})} (\boldsymbol{\nu}!) \prod_{j=1}^{s} \frac{\left[\mathbf{g}_{\boldsymbol{\ell}_j}\right]^{\mathbf{k}_j}}{(\mathbf{k}_j!) \left[\boldsymbol{\ell}_j!\right]^{|\mathbf{k}_j|}}}_{\text{elements of } M},$$

where $n = |\boldsymbol{\nu}|$, and

$$p_s(\boldsymbol{\nu}, \boldsymbol{\lambda}) = \Big\{ (\mathbf{k}_1, \ldots, \mathbf{k}_s; \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_s) : |\mathbf{k}_i| > 0,$$

$$\mathbf{0} \prec \boldsymbol{\ell}_1 \prec \cdots \prec \boldsymbol{\ell}_s, \sum_{i=1}^{s} \mathbf{k}_i = \boldsymbol{\lambda} \text{ and } \sum_{i=1}^{s} |\mathbf{k}_i| \, \boldsymbol{\ell}_i = \boldsymbol{\nu} \Big\}.$$

C.2. *Useful properties of the normal distribution*

We first consider the univariate normal case.

LEMMA C.1: *For $m \in \mathbb{R}$ and $\sigma \in (0, \infty)$, let $Y \sim \mathcal{N}(m, \sigma^2)$, with corresponding likelihood function*

$$\ell(y \,|\, m, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\,\sigma^2} \right).$$

Let $j, k \in \{0, 1, 2, \ldots\}$, and define

$$\kappa_{jk} := \mathbb{E}_{m,\sigma}\left[\frac{1}{\ell(Y\,|\,m,\sigma)}\frac{\partial^j \ell(Y\,|\,m,\sigma)}{(\partial m)^j}\frac{1}{\ell(Y\,|\,m,\sigma)}\frac{\partial^k \ell(Y\,|\,m,\sigma)}{(\partial m)^k}\right],$$

$$\rho_j := \mathbb{E}_{m,\sigma}\left[\frac{1}{\ell(Y\,|\,m,\sigma)}\frac{\partial^j \ell(Y\,|\,m,\sigma)}{(\partial m)^j}\frac{\partial \log \ell(Y\,|\,m,\sigma)}{\partial \sigma}\right].$$

Then,

$$\kappa_{jk} = \mathbb{1}\{j = k\}\,\frac{j!}{\sigma^{2j}}\,, \qquad \rho_j = \mathbb{1}\{j = 2\}\,\frac{2}{\sigma^3}\,.$$

Let $\phi(y) = \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)$ and $\phi^{(j)}(y) = \frac{d^j \phi(y)}{dy^j}$. Hermite polynomials are defined by $h_j(y) = (-1)^j [\phi(y)]^{-1}\,\phi^{(j)}(y)$. The proof of Lemma C.1 is given in Section C.4. It crucially relies on the following orthogonality property of Hermite polynomials:

$$\int_{-\infty}^{\infty} h_j(y)\,h_k(y)\,\phi(y)\,dy = \mathbb{1}\{j = k\}\,j!\,. \tag{C.1}$$

The result in Lemma C.1 is sufficient for our purposes, but more general results can be derived.[1]

Next, we consider a vector of independent normal variables with heteroscedastic means and variances.

LEMMA C.2: *Let $d \in \{1, 2, 3, \ldots\}$. For $m \in \mathbb{R}^d$ and $\sigma \in (0,\infty)^d$, let $\Sigma(\sigma)$ be the $d \times d$ diagonal matrix with diagonal entries σ_i^2, and let $Y \sim \mathcal{N}(m, \Sigma(\sigma))$. The corresponding likelihood function reads*

$$\ell(y\,|\,m,\sigma) = \prod_{i=1}^d \ell(y_i\,|\,m_i,\sigma_i), \qquad \ell(y_i\,|\,m_i,\sigma_i) = \left[\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{(y_i - m_i)^2}{2\,\sigma_i^2}\right)\right].$$

Let $j, k \in \{0, 1, 2, \ldots\}^d$, $j^* = \sum_{i=1}^d j_i$, $k^* = \sum_{i=1}^d k_i$, and define

$$\kappa(j,k) := \mathbb{E}_{m,\sigma}\left[\frac{1}{\ell(Y\,|\,m,\sigma)}\frac{\partial^{j^*} \ell(Y\,|\,m,\sigma)}{\prod_{i=1}^d (\partial m_i)^{j_i}}\frac{1}{\ell(Y\,|\,m,\sigma)}\frac{\partial^{k^*} \ell(Y\,|\,m,\sigma)}{\prod_{i=1}^d (\partial m_i)^{k_i}}\right],$$

[1] More generally, for $k_1, k_2 \in \{0, 1, 2, \ldots\}$ and $j_1, j_2 \in \{0, 1\}$, let

$$\begin{aligned}\kappa_{k_1,k_2,j_1,j_2} &:= \mathbb{E}_{m,\sigma}\left[\frac{1}{\ell(Y\,|\,m,\sigma)}\frac{\partial^{k_1+j_1}\ell(Y\,|\,m,\sigma)}{(\partial m)^{k_1}(\partial\sigma)^{j_1}}\frac{1}{\ell(Y\,|\,m,\sigma)}\frac{\partial^{k_2+j_2}\ell(Y\,|\,m,\sigma)}{(\partial m)^{k_2}(\partial\sigma)^{j_2}}\right] \\ &= \int_{-\infty}^{\infty}\frac{1}{\ell(y\,|\,m,\sigma)}\frac{\partial^{k_1+j_1}\ell(y\,|\,m,\sigma)}{(\partial m)^{k_1}(\partial\sigma)^{j_1}}\frac{\partial^{k_2+j_2}\ell(y\,|\,m,\sigma)}{(\partial m)^{k_2}(\partial\sigma)^{j_2}}dy.\end{aligned}$$

One then finds

$$\kappa_{k_1,k_2,j_1,j_2} = \mathbb{1}\{k_1 + 2j_1 = k_2 + 2j_2\}\,(k_1 + 2j_1)!\,\sigma^{-[2(k_1+2j_1)-j_1-j_2]}.$$

$$\rho(j,i') := \mathbb{E}_{m,\sigma}\left[\frac{1}{\ell(Y\,|\,m,\sigma)}\,\frac{\partial^{j^*}\ell(Y\,|\,m,\sigma)}{\prod_{i=1}^{d}(\partial m_i)^{j_i}}\,\frac{\partial \log \ell(Y\,|\,m,\sigma)}{\partial \sigma_{i'}}\right],$$

where $i' \in \{1,\dots,d\}$ in the last line. Then,

$$\kappa(j,k) = \mathbb{1}\{j=k\}\prod_{i=1}^{d}\frac{j_i!}{\sigma_i^{2j_i}}, \qquad \rho(j,i) = \begin{cases} \dfrac{2}{\sigma_i^3} & \text{if } j_i = 2\text{, and all other entries of } j \text{ are zero,} \\ 0 & \text{otherwise.} \end{cases}$$

Lemma C.2 is an immediate corollary of Lemma C.1. Using the independence of the components of Y we find

$$\frac{\partial^{j^*}\ell(y\,|\,m,\sigma)}{\prod_{i=1}^{d}(\partial m_i)^{j_i}} = \prod_{i=1}^{d}\frac{\partial^{j_i}\ell(y_i\,|\,m_i,\sigma_i)}{(\partial m_i)^{j_i}},$$

and

$$\kappa(j,k) = \prod_{i=1}^{d}\kappa_{j_i,k_i}.$$

Plugging in the result for κ_{jk} in Lemma C.1 then gives the result for $\kappa(j,k)$ in Lemma C.2. Analogously for $\rho(j,i')$.

C.3. *Nonlinear regression with normal errors*

Consider the model

$$Y_i = m(X_i;\theta,\eta) + \sigma(X_i;\theta)\,U_i, \qquad U_i\,|\,X \sim iid\mathcal{N}(0,1), \qquad i=1,\dots,d,$$

where $m(\cdot;\cdot,\cdot)$ and $\sigma(\cdot;\cdot)$ are known functions, and θ and η are unknown parameters. Ignore θ for the moment and write

$$Y_i = m_i(\eta) + \sigma_i\,U_i.$$

Let $m=(m_1,\dots,m_d)$ and $\sigma=(\sigma_1,\dots,\sigma_d)$. The likelihood for $y=(y_1,\dots,y_d)$ is then given by

$$\ell(y\,|\,\eta) = \ell(y\,|\,m(\eta),\sigma),$$

where $\ell(y\,|\,m,\sigma)$ is given in Lemma C.2. Let ∇_η^p be the vector operator that collects all unique derivatives with respect to η up to order p. Let ∇_m^p be the vector operator that collects all unique derivatives with respect to m up to order p. Then, there exists a matrix valued function $M(\eta)$, which only depends on η and on the function $m(\eta)$, such that

$$\nabla_\eta^p\,\ell(y\,|\,\eta) = M(\eta)\,\nabla_m^p\,\ell(y\,|\,m(\eta),\sigma). \tag{C.2}$$

We want to calculate

$$\mathbb{E}_\eta\left[\frac{\nabla^p_\eta \ell(Y\,|\,\eta)}{\ell(Y\,|\,\eta)}\,\frac{\nabla^p_\eta \ell(Y\,|\,\eta)^\top}{\ell(Y\,|\,\eta)}\right].$$

Lemma C.2 gives us explicit expressions for all the components of

$$\mathbb{E}_{m,\sigma}\left[\frac{\nabla^p_m \ell(Y\,|\,m,\sigma)}{\ell(Y\,|\,m,\sigma)}\,\frac{\nabla^p_m \ell(Y\,|\,m,\sigma)^\top}{\ell(Y\,|\,m,\sigma)}\right].$$

Using (C.2) we have

$$\begin{aligned}
&\mathbb{E}_\eta\left[\frac{\nabla^p_\eta \ell(Y\,|\,\eta)}{\ell(Y\,|\,\eta)}\,\frac{\nabla^p_\eta \ell(Y\,|\,\eta)^\top}{\ell(Y\,|\,\eta)}\right]\\
&= M(\eta)\,\mathbb{E}_{m(\eta),\sigma}\left[\frac{\nabla^p_m \ell(Y\,|\,m(\eta),\sigma)}{\ell(Y\,|\,m(\eta),\sigma)}\,\frac{\nabla^p_m \ell(Y\,|\,m(\eta),\sigma)^\top}{\ell(Y\,|\,m(\eta),\sigma)}\right] M(\eta)^\top.
\end{aligned}$$

Thus, by combining Lemma C.2 with the multidimensional Faà di Bruno's formula we get explicit expressions for all the matrices we need.

C.4. *Proof of Lemma C.1*

We already introduced $\phi(y)=\frac{1}{\sqrt{2\pi}}\exp\left(-y^2/2\right)$ and $\phi^{(j)}(y)=\frac{d^j\phi(y)}{dy^j}$ above. Let $j,k\in\{0,1,2,3,\dots\}$. The well-known orthogonality property of Hermite polynomials in (C.1) can be rewritten as

$$\int_{-\infty}^{\infty}\frac{\phi^{(j)}(y)\,\phi^{(k)}(y)}{\phi(y)}\,dy=\mathbb{1}\{j=k\}\,j!\,. \tag{C.3}$$

Another well-known property of Hermite polynomials is the recurrence relation $h_{j+1}(y)=yh_j(y)-\frac{d}{dy}h_j(y)$. Using this, it is easy to show that for $j>k$ we have

$$\int_{-\infty}^{\infty}\frac{y\,\phi^{(j)}(y)\,\phi^{(k)}(y)}{\phi(y)}\,dy=-\,\mathbb{1}\{j=k+1\}\,j!\,. \tag{C.4}$$

Next, we have

$$\ell(y\,|\,m,\sigma)=\frac{1}{\sigma}\,\phi\left(\frac{y-m}{\sigma}\right),\qquad\qquad \frac{\partial^j\ell(y\,|\,m,\sigma)}{(\partial m)^j}=\frac{(-1)^j}{\sigma^{j+1}}\,\phi^{(j)}\left(\frac{y-m}{\sigma}\right).$$

Using this we obtain

$$\begin{aligned}
\kappa_{jk}:=&\,\mathbb{E}_{m,\sigma}\left[\frac{1}{\ell(Y\,|\,m,\sigma)}\,\frac{\partial^j\ell(Y\,|\,m,\sigma)}{(\partial m)^j}\,\frac{1}{\ell(Y\,|\,m,\sigma)}\,\frac{\partial^k\ell(Y\,|\,m,\sigma)}{(\partial m)^k}\right]\\
=&\int_{-\infty}^{\infty}\frac{1}{\ell(y\,|\,m,\sigma)}\,\frac{\partial^j\ell(y\,|\,m,\sigma)}{(\partial m)^j}\,\frac{\partial^k\ell(y\,|\,m,\sigma)}{(\partial m)^k}\,dy
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(-1)^{j+k}}{\sigma^{j+k+1}} \int_{-\infty}^{\infty} \frac{1}{\phi\left(\frac{y-m}{\sigma}\right)} \phi^{(j)}\left(\frac{y-m}{\sigma}\right) \phi^{(k)}\left(\frac{y-m}{\sigma}\right) dy \\
&= \frac{(-1)^{j+k}}{\sigma^{j+k}} \int_{-\infty}^{\infty} \frac{\phi^{(j)}(y)\, \phi^{(k)}(y)}{\phi(y)}\, dy \\
&= \mathbb{1}\{j=k\}\, \frac{j!}{\sigma^{2j}}\,,
\end{aligned}
$$

where the second to last step employs a change of variables in the integral ($\frac{y-m}{\sigma} \mapsto y$), and the last step uses (C.3). Similarly, for

$$
\frac{\partial \ell(y\,|\,m,\sigma)}{\partial \sigma} = -\frac{1}{\sigma^2}\,\phi\left(\frac{y-m}{\sigma}\right) - \left(\frac{y-m}{\sigma^3}\right) \phi^{(1)}\left(\frac{y-m}{\sigma}\right),
$$

one finds

$$
\begin{aligned}
\rho_j &:= \mathbb{E}_{m,\sigma}\left[\frac{1}{\ell(Y\,|\,m,\sigma)} \frac{\partial^j \ell(Y\,|\,m,\sigma)}{(\partial m)^j} \frac{\partial \log \ell(Y\,|\,m,\sigma)}{\partial \sigma}\right] \\
&= \int_{-\infty}^{\infty} \frac{1}{\ell(y\,|\,m,\sigma)} \frac{\partial^j \ell(y\,|\,m,\sigma)}{(\partial m)^j} \frac{\partial \ell(y\,|\,m,\sigma)}{\partial \sigma}\, dy \\
&= \frac{(-1)^{1+j}}{\sigma^{2+j}} \int_{-\infty}^{\infty} \frac{1}{\phi\left(\frac{y-m}{\sigma}\right)} \phi^{(j)}\left(\frac{y-m}{\sigma}\right) \left[\phi\left(\frac{y-m}{\sigma}\right) + \left(\frac{y-m}{\sigma}\right) \phi^{(1)}\left(\frac{y-m}{\sigma}\right)\right] dy \\
&= \frac{(-1)^{1+j}}{\sigma^{1+j}} \int_{-\infty}^{\infty} \frac{1}{\phi(y)}\, \phi^{(j)}(y)\, \left[\phi(y) + y\, \phi^{(1)}(y)\right] dy \\
&= \mathbb{1}\{j=2\} \frac{(-1)}{\sigma^3} \int_{-\infty}^{\infty} \frac{y\, \phi^{(1)}(y)\, \phi^{(2)}(y)}{\phi(y)}\, dy \\
&= \mathbb{1}\{j=2\} \frac{2}{\sigma^3},
\end{aligned}
$$

where we again employed the same change of variables in the integration and also used (C.3) and (C.4).

APPENDIX D: MONTE CARLO SIMULATION

In this section of the appendix we report on the results of a Monte Carlo experiment. We specify a CES model of team production with log-normal errors, where we take the network structure (i.e., the set $\mathcal{K}$ in equation (5.3) of the main text) as given from the empirical data. We fix the true value of the substitution parameter to $\gamma_0 = 1$, the team size parameter to $\beta_0 = 1$, the log-error variance in teams of size 2 to $\sigma_0^2(2) = 1$, and the variance in teams of size 1 to $\sigma_0^2(1) = 1$. This data generating process is designed to approximate what we found on the empirical data.

We report results based on 300 simulations. In each simulated sample, we estimate the parameters using plug-in method-of-moments and the Neyman-orthogonalized method-of-moments estimates of degree $q = 1$ to $q = 6$. As we did in our empirical study, we compute

TABLE I

MONTE CARLO SIMULATION

	Median	Mean	2.5%	97.5%
	Substitution γ (true value = 1)			
Plug-in	0.5054	0.5089	0.4150	0.5966
$q = 1$	0.9723	0.9864	0.7523	1.3259
$q = 2$	1.0310	1.0625	0.7787	1.5571
$q = 3$	1.0100	1.0408	0.7455	1.5795
$q = 4$	0.9964	1.0227	0.7364	1.5556
$q = 5$	0.9898	1.0161	0.7363	1.4904
$q = 6$	0.9897	1.0154	0.7362	1.4904
	Team size β (true value = 1)			
Plug-in	1.0578	1.0595	1.0277	1.1003
$q = 1$	1.0459	1.0466	0.9971	1.1038
$q = 2$	1.0029	1.0012	0.9299	1.0605
$q = 3$	1.0007	0.9979	0.9249	1.0577
$q = 4$	1.0013	0.9980	0.9226	1.0590
$q = 5$	1.0007	0.9983	0.9203	1.0566
$q = 6$	1.0007	0.9983	0.9214	1.0565

Note: 300 simulations.

sample-split preliminary estimates of the author fixed-effects based on all their sole-authored publications except for one, selected at random. However, in the simulation exercise we do not cross-fit the estimators, and simply choose a random selection of sole-authored publications for each author in each Monte Carlo run.

In Tables I and II we show the median, mean, 2.5% quantile, and 97.5% quantile of each estimate across simulations. Starting with the substitution parameter γ, we see that the plug-in estimator is severely biased, with median and mean biases of -50% (expressed in proportion of the true value). For this parameter, all Neyman-orthogonalized estimators are substantially less biased, with a median bias below 3%, and below 1% for $q \geq 4$. However, in some replications the orthogonalized estimates tend to have large values, which is reflected in a somewhat larger mean bias, ranging from 6% at $q = 2$ down to 1.5% at $q = 5$ and $q = 6$, and quantile bands that are not symmetric around the true value.

Turning next to the team size parameter β, we see that both the plug-in and first-order Neyman-orthogonalized estimators are biased, with a median and mean bias of 4%–6%. All orthogonalized estimates of order $q \geq 2$ are virtually unbiased, both for the mean and the median. Moreover, in this case the quantile bands are symmetric around the true parameter value.

Shifting attention to the variance in teams of size 2, $\sigma^2(2)$, we see that both the plug-in and first-order Neyman-orthogonalized estimators are severely biased, with a median and mean bias of 16%–18%. All orthogonalized estimates of order $q \geq 2$ are virtually unbiased, both for the mean and the median, and the quantile bands are centered around the true parameter value.

Lastly, turning to the variance in teams of size 1, $\sigma^2(1)$, the plug-in estimator exhibits a large bias of 33%. First-order orthogonalization only decreases the bias slightly, to 27%. In contrast, the Neyman-orthogonalized estimators continue to show good performance. In particular, when $q \geq 4$ the estimates are virtually unbiased, and the quantile bands are symmetric around the true value.

TABLE II

MONTE CARLO SIMULATION (CONTINUED)

	Median	Mean	2.5%	97.5%
Variance $\sigma^2(2)$ (true value = 1)				
Plug-in	1.1623	1.1618	1.1284	1.1944
$q = 1$	1.1824	1.1825	1.1456	1.2228
$q = 2$	0.9964	0.9969	0.9615	1.0334
$q = 3$	1.0001	1.0009	0.9665	1.0366
$q = 4$	0.9995	1.0006	0.9637	1.0373
$q = 5$	0.9993	1.0000	0.9639	1.0365
$q = 6$	0.9982	0.9996	0.9637	1.0374
Variance $\sigma^2(1)$ (true value = 1)				
Plug-in	1.3361	1.3368	1.2840	1.3856
$q = 1$	1.2732	1.2723	1.2140	1.3299
$q = 2$	1.0158	1.0178	0.9478	1.0925
$q = 3$	1.0087	1.0119	0.9435	1.0842
$q = 4$	0.9996	1.0033	0.9335	1.0761
$q = 5$	0.9980	1.0002	0.9321	1.0731
$q = 6$	0.9976	0.9992	0.9272	1.0718

Note: 300 simulations.

APPENDIX E: RESTRICTIONS INDEPENDENT OF INDIVIDUAL EFFECTS

Model (5.3) of the main text implies restrictions on parameters $\gamma_0, \beta_0, \sigma_0^2(1), \sigma_0^2(2)$ that do not depend on the author-specific effects η_{i0}.[2] As an example, the model implies the following alternative expression for the team size parameter β_0:

$$\beta_0 = \left(\frac{\mathbb{E}[Y_j^{\gamma_0} \mid s_j = 2]}{\mathbb{E}[Y_j^{\gamma_0} \mid s_j = 1]} \right)^{\frac{1}{\gamma_0}} \exp\left(\frac{1}{2} \gamma_0 [\sigma_0^2(1) - \sigma_0^2(2)] \right), \tag{E.1}$$

where the expectation given $s_j = 1$ is taken over the co-authors' own sole-authored articles, and (E.1) does not involve the fixed-effects η_{i0}. Note that, if $\gamma_0 = 0$ and log-output is additive in worker inputs, then $\log \beta_0$ is simply the difference between average log-outputs in teams of size 2 and 1, respectively. As a check, in Figure 1 we report estimates of the left-hand side of (E.1), against the estimates of β_0 shown in Table I of the main text, for various orders of orthogonalization. We see that the estimates of the two sides of (E.1) tend to agree with each other well irrespective of the orthogonalization order.

Model (5.3) of the main text also implies restrictions on γ_0 alone. To see this, let us write that model, within teams of size 2 only, as

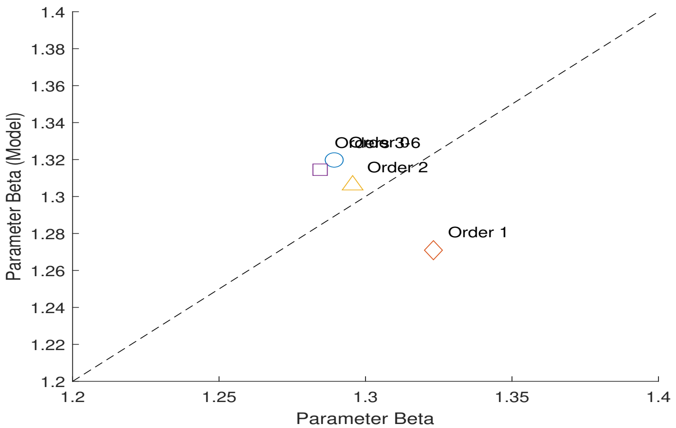
$$Y_j^{\gamma_0} = \frac{1}{2} \beta_0^{\gamma_0} \left(\eta_{k(j,1)0}^{\gamma_0} + \eta_{k(j,2)0}^{\gamma_0} \right) \varepsilon_j^{\gamma_0 \sigma_0(2)},$$

which we write in vector form as

$$Y(\gamma_0) = A\widetilde{\eta}_0 + \widetilde{\varepsilon}, \tag{E.2}$$

[2]The analysis in this section was inspired by discussions with Bo Honoré.

FIGURE 1.—Comparing two estimates of β_0



Notes: Estimate of β_0 on the x-axis, model-based estimate of β_0 based on the right-hand side of (E.1) *on the y-axis. Each point corresponds to an order of orthogonalization.*

where $Y(\gamma_0)$ has elements $Y_j^{\gamma_0}$, A is a matrix of zeros and ones,

$$\widetilde{\eta}_{k0} = \frac{1}{2}\beta_0^{\gamma_0}\eta_{k0}^{\gamma_0}\exp\left(\frac{1}{2}\gamma_0^2\sigma_0^2(2)\right),$$

and

$$\widetilde{\varepsilon}_j = \frac{1}{2}\beta_0^{\gamma_0}\left(\eta_{k(j,1)0}^{\gamma_0} + \eta_{k(j,2)0}^{\gamma_0}\right)\left[\varepsilon_j^{\gamma_0\sigma_0(2)} - \exp\left(\frac{1}{2}\gamma_0^2\sigma_0^2(2)\right)\right].$$

Since $\mathbb{E}[\widetilde{\varepsilon}_j \,|\, A] = 0$, (E.2) implies the conditional moment equalities

$$\mathbb{E}\left[(I - AA^{\dagger})Y(\gamma_0) \,|\, A\right] = 0, \tag{E.3}$$

which only depend on γ_0.

To use (E.3) for estimation, we rely on a set of instruments. For this purpose, we use interacted preliminary estimates $Z_j = \widehat{\eta}_{k(j,1)}\widehat{\eta}_{k(j,2)}$ for $k(j,1), k(j,2)$ the co-authors of j. Since we assume the preliminary estimates are constructed from an independent sample, we have

$$\mathbb{E}\left[Z'(I - AA^{\dagger})Y(\gamma_0)\right] = 0. \tag{E.4}$$

Note these restrictions remain valid when ε_j are not Gaussian or not mutually independent, provided they are independent of A. We use GMM estimation based on (E.4), that is,

$$\widehat{\gamma}^{\mathrm{GMM}} = \underset{\gamma}{\operatorname{argmin}}\,|Z'(I - AA^{\dagger})Y(\gamma)|. \tag{E.5}$$

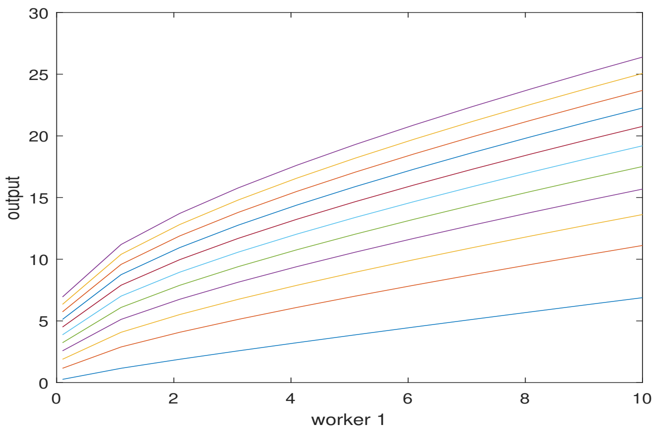
We implement this estimator in the same way we have implemented our Neyman-orthogonalized equations. Specifically, we construct preliminary estimates of author effects using all but one sole-authored paper for each author, where we select the held-out sole-authored paper at random.

Using the same Monte Carlo simulation design as in Appendix D tends to give noisy estimates. For example, when the true value is $\gamma_0 = 1$, and $\sigma_0(1) = 1/5$ and $\sigma_0(2) = 1/5$, we obtain a mean GMM estimate of 1.0218, a median estimate of 0.9725, and a standard deviation of 0.2261 across 300 simulations. Moreover, out of the 300 simulations, in 24 cases we are unable to find another minimum in (E.5) other than $\gamma = 0$ (which is always a minimum by construction). Note these findings correspond to a model with error variances that are *25 times* smaller than the variances we used for our main simulation design in Appendix D. This suggests this estimation approach, at least for this particular choice of instruments, is considerably less precise than our likelihood-based approach.

Lastly, computing the GMM estimator on the empirical data, cross-fitting 100 times, we obtain $\widehat{\gamma}^{\mathrm{GMM}} = 0.5807$. This is of a comparable magnitude to the estimates of γ reported in Table I of the main text, when using a sufficiently high order of orthogonalization. However, it is worth noting that, out of the 100 random splits of the sole-authored productions, in 5 cases we are unable to find another minimum in (E.5) other than $\gamma = 0$, again reflecting the instability of this method in our setting.

APPENDIX F: ESTIMATED PRODUCTION FUNCTION

FIGURE 2.—Production function estimate



Notes: Worker 1's type η_1 on the x-axis, average output Y_j on the y-axis. Each curve corresponds to a different worker 2's type η_2. Figure based on the point estimates for $q = 6$ reported in Table I of the main text.

REFERENCES

CONSTANTINE, GEORGE M., AND THOMAS H. SAVITS (1996): "A Multivariate Faa di Bruno Formula with Applications," *Transactions of the American Mathematical Society*, 348, 503–520.
[7]