Online Supplement to

Proxy Variable Estimation of Multi-Product Production Functions

Emir Malikov[1] Gudbrand Lien[2]

[1]University of Nevada, Las Vegas

[2]Inland Norway University

A Derivation of Eq. (4.8)

From (4.7), the first-order condition with respect to M_{it} is

$$\sum_{j=1}^{S} P_{j,t} \frac{\partial F(\mathbf{X}_{it}, M_{it})}{\partial M_{it}} \frac{\exp\{\omega_{it}\} \times \theta}{\widetilde{H}^j(Z_{1,it}, \ldots, Z_{S,it})} = P_{M,t}.$$

Taking logs of both sides and recognizing that, due to the linear homogeneity of $H(\cdot)$,

$$\sum_{j=1}^{S} P_{j,t} \left[\widetilde{H}^j(Z^j_{1,it}, \ldots, Z^j_{S,it})\right]^{-1} = [H(Y_{1,it}, \ldots, Y_{S,it})]^{-1} \sum_{j=1}^{S} P_{j,t} Y_{j,it}, \tag{A.1}$$

from (4.7) we get

$$\ln\left[\frac{\partial F(\mathbf{X}_{it}, M_{it})}{\partial M_{it}}\right] - h(Y_{1,it}, \ldots, Y_{S,it}) + \omega_{it} + \ln\theta = \ln\left[\frac{P_{M,t}}{\sum_{j=1}^{S} P_{j,t} Y_{j,it}}\right]. \tag{A.2}$$

Adding the log of M_{it} to both sides of (A.2) and then subtracting the latter from the normalized production function in logs (4.3) yields

$$\ln\left[\frac{\partial F(\mathbf{X}_{it}, M_{it})}{\partial M_{it}} \frac{M_{it}}{F(\mathbf{X}_{it}, M_{it})}\right] + \ln\theta - \eta_{it} = \ln\left[\frac{P_{M,t} M_{it}}{\sum_{j=1}^{S} P_{j,t} Y_{j,it}}\right]. \tag{A.3}$$

Define $G_M(\mathbf{X}_{it}, M_{it}) \stackrel{\text{def}}{=} \frac{\partial F(\mathbf{X}_{it}, M_{it})}{\partial M_{it}} \frac{M_{it}}{F(\mathbf{X}_{it}, M_{it})}$ which equals the elasticity of $F(\cdot)$ with respect to M_{it}. Next, define $V_{M,it} \stackrel{\text{def}}{=} P_{M,t} M_{it} / \sum_{j=1}^{S} P_{j,t} Y_{j,it}$, which equals the nominal share of intermediate inputs in total revenue. Therefore, from (A.3), we obtain equation (4.8):

$$\ln V_{M,it} = \ln\left[G_M(\mathbf{X}_{it}, M_{it})\theta\right] - \eta_{it}.$$

B Estimation Procedure

This section describes how to operationalize the identification strategy outlined in Section 4. We estimate the unknown production function $f(\cdot)$, the output index function $\widetilde{h}(\cdot)$ and the (unobservable) firm-level persistent productivity ω_{it} via a two-stage procedure. The estimation is nonparametric, with all unknown functions in (4.8) and (4.12) globally approximated using cubic B-spline sieves.

We opt for the method of sieves over kernel methods to avoid the computational burden associated with kernel estimation which is especially relevant here due to the need for observation-specific numerical integration of the material elasticity function in (4.11). In contrast, sieves approximate unknown nonparametric (i.e., infinite-parameter) functions using a sequence of less complex parameter spaces that are characterized by a finite number of "parameters" which effectively reduces the estimation problem to a parametric estimation when implemented in practice, with the caveat being that the complexity of such an approximation increases with the sample size to ensure consistency. Besides the ease of implementation, sieves are also advantageous over kernels when the nonparametric estimands enter the estimation problem nonlinearly as is the case in (4.12). For more on sieves, see an excellent review by Chen (2007).

We opt for cubic B-splines that are a popular type of linear (in parameters) sieves and known for their superior finite-sample performance over other linear sieves, including widely used polynomial series. Essentially, B-splines provide means for a *smooth* piece-wise polynomial approximation to the unknown function, with the number and size of subintervals determined via data-driven procedure.

First Stage. We first estimate the material revenue share equation (4.8). We approximate the unknown function $\ln\left[G_M(\mathbf{X}_{it}, M_{it})\theta\right]$ using a linear series of cubic B-spline basis functions $\{\mathscr{B}_{r,n}(\mathbf{x}_{it}, m_{it})\}$ constructed using logged regressors:

$$\ln\left[G_M(\mathbf{X}_{it}, M_{it})\theta\right] \approx \sum_{r=1}^{R_n} \alpha_r \mathscr{B}_{r,n}(\mathbf{x}_{it}, m_{it}), \tag{B.1}$$

where $R_n \to \infty$ slowly with n. More concretely, for each of the four logged explanatory variables $v_j \in \{\mathbf{x}, m\}$ distributed on a compact interval $[\min(v_{j,it}), \max(v_{j,it})]$, we partition the regressor's domain into $d_n + 1$ subintervals using $d_n \geq 0$ equiquantile interior knots to construct a basis system of $d_n + 4$ univariate B-splines for v_j.[18] A multivariate basis system $\{\mathscr{B}_{r,n}(\mathbf{x}, m)\}$ is then constructed as a tensor product of these univariate basis systems. For tractability, we restrict d_n to be same for all four regressors, implying that $R_n = (d_n + 4)^{\dim(\mathbf{x},m)} = (d_n + 4)^4$.

Given the orthogonality conditions in (4.9) whereby all regressors in the material revenue share equation are exogenous, the approximated function $\ln\left[G(\mathbf{X}_{it}, M_{it})\theta\right]$ is estimated via least squares for a given R_n (or equivalently, d_n). Thus, letting $\boldsymbol{\alpha}(R_n) = [\alpha_1, \ldots, \alpha_{R_n}]'$ and $\mathbf{B}_{R_n}(\mathbf{x}_{it}, m_{it}) = [\mathscr{B}_{1,n}(\mathbf{x}_{it}, m_{it}), \ldots, \mathscr{B}_{R_n,n}(\mathbf{x}_{it}, m_{it})]'$, we have that

$$\widehat{\boldsymbol{\alpha}}(R_n) = \left[\sum_{i,t} \mathbf{B}_{R_n}(\mathbf{x}_{it}, m_{it})\mathbf{B}_{R_n}(\mathbf{x}_{it}, m_{it})'\right]^{-1} \sum_{i,t} \mathbf{B}_{R_n}(\mathbf{x}_{it}, m_{it}) \ln V_{M,it}. \tag{B.2}$$

We select the complexity of approximation R_n indirectly by selecting the optimal number of interior knots d_n via generalized cross-validation of Craven & Wahba (1979):

$$d_n^* = \arg\min_{d_n} \frac{\frac{1}{nT}||(\mathbf{I}_{nT} - \mathbf{P}_{R_n})\mathbf{v}||^2}{\left[1 - \frac{1}{nT}\text{tr}\{\mathbf{P}_{R_n}\}\right]^2} \quad \text{with } R_n = (d_n + 4)^4, \tag{B.3}$$

where $\mathbf{P}_{R_n} = \mathbb{B}_{R_n}(\mathbb{B}'_{R_n}\mathbb{B}_{R_n})^{-1}\mathbb{B}'_{R_n}$ is a projection matrix defined using the matrix of basis functions

[18] Generally, a B-spline of degree a with b interior knots yields $(b + a + 1)$ basis functions with $(a + 1)$ known as the "order" of this B-spline. We use cubic splines, hence $a + 1 = 4$.

$\mathbb{B}_{R_n}$ constructed by stacking up $\mathbf{B}_{R_n}(\mathbf{x}_{it}, m_{it})$ in the ascending order of index i first then index t. The column vector $\mathbf{v}$ is stacked up similarly but using $\{\ln V_{M,it}\}$.

Recall that the residuals from fitting (4.8) are the (consistent) estimates of the random productivity shock η_{it}. Using these estimates $\widehat{\eta}_{it} = \sum_{r=1}^{R_n} \widehat{\alpha}_r \mathscr{B}_{r,n}(\mathbf{x}_{it}, m_{it}) - \ln V_{M,it}$, we consistently estimate θ via

$$\widehat{\theta} = (nT)^{-1} \sum_{i,t} \exp\{\widehat{\eta}_{it}\}. \tag{B.4}$$

This allows us to recover the estimator of the $G_M(\cdot)$ function free of constant θ, i.e.,

$$\widehat{G}_M(\mathbf{X}_{it}, M_{it}) = \exp\left\{ \sum_{r=1}^{R_n} \widehat{\alpha}_r \mathscr{B}_{r,n}(\mathbf{x}_{it}, m_{it}) - \ln \widehat{\theta} \right\}, \tag{B.5}$$

which we then use to numerically evaluate the integral in (4.11) and construct the estimates of $y_{1,it}^*$:

$$\widehat{y}_{it}^* = y_{1,it} - \int_{\min(M)}^{M_{it}} \frac{\widehat{G}_M(\mathbf{X}_{it}, m)}{m} \mathrm{d}m - \widehat{\eta}_{it}. \tag{B.6}$$

Second Stage. With the estimates $\widehat{y}_{it}^*$ in hand, we proceed to estimate (4.12), where just like in the first stage above, we approximate the unknown functions $\widetilde{h}(\cdot)$, $C(\cdot)$ and $\psi[\cdot]$ in logs via cubic B-spline sieves:

$$\begin{aligned} \widetilde{h}(\mathbf{Z}_{it}) &\approx \sum_{r'=1}^{R_n'} \beta_{r'} \mathscr{B}_{r'}(\mathbf{z}_{it}), \\ C(\mathbf{X}_{it}) &\approx \sum_{r''=1}^{R_n''} \gamma_{r''} \mathscr{B}_{r''}(\mathbf{x}_{it}), \\ \psi[\cdot] &\approx \sum_{r'''=1}^{R_n'''} \delta_{r'''} \mathscr{B}_{r'''}(\cdot), \end{aligned} \tag{B.7}$$

where R_n', R_n'' and R_n''' all increase with the sample size. Obviously, to avoid multicollinearity, we remove redundant zero-degree "intercept" basis functions, with the remaining intercept parameter attributed to the mean productivity function $\psi[\cdot]$.

We continue restricting the number of interior knots p_n to be same across all regressors for tractability sake, whereby $R_n' = (p_n + 4)^{(S-1)}$, $R_n'' = (p_n + 4)^3$ and $R_n''' = (p_n + 4)$. Just like earlier, we select the smoothing parameter p_n via generalized cross-validation. Since the estimating equation is nonlinear (in parameters) in this case, we use the following generalized cross-validation function (also see Gao, 2007):

$$p_n^* = \arg\min_{p_n} \frac{\frac{1}{nT} \left|\left| \widehat{\boldsymbol{\zeta}}(R_n', R_n'', R_n''') \right|\right|^2}{\left[1 - \frac{1}{nT} \kappa(R_n', R_n'', R_n''')\right]^2} \tag{B.8}$$

where $\widehat{\boldsymbol{\zeta}}(\cdot)$ is the vector of residuals from (4.12), and $\kappa(\cdot)$ is the total number of sieve parameters in the B-spline approximators which effectively substitutes for the trace of a projection matrix.[19]

Per the moment conditions in (4.13), all regressors in the productivity process (4.12) are weakly exogenous. We therefore estimate the second-stage equation via nonlinear least squares. Namely, letting $\boldsymbol{\vartheta}(R_n', R_n'', R_n''') = [\beta_1, \ldots, \beta_{R_n'}, \gamma_1, \ldots, \gamma_{R_n''}, \delta_1, \ldots, \delta_{R_n'''}]'$ for given R_n', R_n'' and R_n''' (or equivalently, p_n), we have the following:

$$\begin{aligned}\widehat{\boldsymbol{\vartheta}}(R_n', R_n'', R_n''') = \arg\min_{\boldsymbol{\vartheta}} \Bigg[& \widehat{y}_{it}^* - \sum_{r'=1}^{R_n'} \beta_{r'} \mathscr{B}_{r'}(\mathbf{z}_{it}) + \sum_{r''=1}^{R_n''} \gamma_{r''} \mathscr{B}_{r''}(\mathbf{x}_{it}) \\ & - \sum_{r'''=1}^{R_n'''} \delta_{r'''} \mathscr{B}_{r'''} \left(\widehat{y}_{it-1}^* - \sum_{r'=1}^{R_n'} \beta_{r'} \mathscr{B}_{r'}(\mathbf{z}_{it-1}) + \sum_{r''=1}^{R_n''} \gamma_{r''} \mathscr{B}_{r''}(\mathbf{x}_{it-1}) \right) \Bigg]^2 .\end{aligned} \tag{B.9}$$

The second-stage estimation readily provides estimates of the output index function: $\widehat{\widehat{h}}(\mathbf{z}_{it}) = \sum_{r'=1}^{R_n'} \widehat{\beta}_{r'} \mathscr{B}_{r'}(\mathbf{z}_{it})$. The production function $f(\cdot)$ is estimated from (4.11) using the second-stage estimates of $C(\cdot)$:

$$\widehat{f}(\mathbf{X}_{it}, M_{it}) = \int_{\min(M)}^{M_{it}} \frac{\widehat{G}_M(\mathbf{X}_{it}, m)}{m} \mathrm{d}m - \sum_{r''=1}^{R_n''} \widehat{\gamma}_{r''} \mathscr{B}_{r''}(\mathbf{x}_{it}) \ , \tag{B.10}$$

and the unobservable firm-level productivity ω_{it} is estimated using (4.3) as

$$\widehat{w}_{it} = y_{1,it} - \widehat{f}(\mathbf{X}_{it}, M_{it}) - \widehat{\widehat{h}}(\mathbf{Z}_{it}) - \widehat{\eta}_{it} \ . \tag{B.11}$$

Asymptotic Inference. To facilitate an easy-to-implement inference, the sieve least squares estimators in the first [eq. (B.2)] and second [eq. (B.9)] stages can be recast as a sequential two-step nonparametric sieve M-estimator. Essentially, we can rewrite our sequential estimator as a simultaneous multiple-equation just-identified system of nonparametric moment restrictions where the instrument sets vary across equations. A similar method-of-moments interpretation of sequential multi-step, albeit parametric, estimators has also been proposed by Newey (1984).

First, analogous to $\boldsymbol{\alpha}$, we define $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{R_n'}]'$, $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_{R_n''}]'$, $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_{R_n'''}]'$. Then, with the residual function from the second-stage estimation given by $\zeta_{it}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \theta) =$

[19]Essentially, $\kappa(\cdot)$ is the sum of R_n', R_n'' and R_n''' less the removed redundant "intercept" basis functions.

$y_{it}^*(\boldsymbol{\alpha},\theta)-\boldsymbol{\beta}'\mathbf{B}_{R_n'}(\mathbf{z}_{it})+\boldsymbol{\gamma}'\mathbf{B}_{R_n''}(\mathbf{x}_{it})-\boldsymbol{\delta}'\mathbf{B}_{R_n'''}\left(y_{it-1}^*(\boldsymbol{\alpha},\theta)-\boldsymbol{\beta}'\mathbf{B}_{R_n'}(\mathbf{z}_{it-1})+\boldsymbol{\gamma}'\mathbf{B}_{R_n''}(\mathbf{x}_{it-1})\right)$, we have

$$\mathbb{E}\begin{bmatrix} \Big(\ln V_{M,it}-\boldsymbol{\alpha}'\mathbf{B}_{R_n}(\mathbf{x}_{it},m_{it})\Big)\mathbf{B}_{R_n}(\mathbf{x}_{it},m_{it}) \\ \exp\left\{\boldsymbol{\alpha}'\mathbf{B}_{R_n}(\mathbf{x}_{it},m_{it})-\ln V_{it}\right\}-\theta \\ \zeta_{it}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\theta)\begin{bmatrix}\partial\zeta_{it}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\theta)/\partial(\boldsymbol{\beta}',\boldsymbol{\gamma}')' \\ -\mathbf{B}_{R_n'''}\left(y_{it-1}^*(\boldsymbol{\alpha},\theta)-\boldsymbol{\beta}'\mathbf{B}_{R_n'}(\mathbf{z}_{it-1})+\boldsymbol{\gamma}'\mathbf{B}_{R_n''}(\mathbf{x}_{it-1})\right)\end{bmatrix}\end{bmatrix}=\mathbf{0}_{\dim(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\theta)}, \tag{B.12}$$

where the moment vector consists of three blocks: the first block corresponds to the first-stage (linear) least squares, the second block is the moment identifying the constant θ, and the third block corresponds to the second-stage nonlinear least squares.

Let the moment vector in (B.12) be concisely written as $\mathbb{E}[\boldsymbol{\rho}(\Theta)]=\mathbf{0}$, where $\Theta=(\boldsymbol{\alpha}',\boldsymbol{\beta}',\boldsymbol{\gamma}',\boldsymbol{\delta}',\theta)'$ is a collection of all sieve parameters. Our two-stage nonparametric estimator $\widehat{\Theta}$ is the solution to a system in (B.12). Under mild regularity conditions, this two-step sieve M-estimator is consistent and asymptotically normal (see Hahn et al., 2018). Furthermore, given the just-identification of the model and so long as we use linear sieves (such as B-splines in our case), we can make use of the numerical equivalence between the consistent estimator of the asymptotic variance of *nonparametric* "parameters" $\widehat{\Theta}$ and a consistent estimator of the asymptotic variance derived for these parameter estimators as if the estimated model were of a *parametric* form specified in (B.2) and (B.9). Thus, in practice, one can use the variance formula for a parametric two-step estimator to consistently estimate the variance of a nonparametric sieve estimator.[20] The asymptotic variance for such a parametric two-step estimator can in turn be derived following Newey's (1984) suggestion by making use of the optimal GMM covariance formula: $\mathbb{V}ar\big[\widehat{\Theta}\big]=\big[\mathbb{E}\frac{\partial\boldsymbol{\rho}(\Theta)}{\partial\Theta'}\big]^{-1}\mathbb{E}[\boldsymbol{\rho}(\Theta)\boldsymbol{\rho}(\Theta)']\big[\mathbb{E}\frac{\partial\boldsymbol{\rho}(\Theta)}{\partial\Theta}\big]^{-1}$. This streamlines asymptotic inference. Alternatively, one can use bootstrap methods.

C Productivity Estimates

Table C.1 reports estimates of the annual growth in the average aggregate productivity for all farms in our sample (see the column labeled "$\Delta\mathrm{p}_t$"). The aggregate productivity is computed as the total-revenue-weighted average of farm-level productivities for each year, i.e., $\mathrm{p}_t=\sum_i\varrho_{it}\omega_{it}$, where $\varrho_{it}\equiv R_{it}/\sum_j R_{jt}\ \forall\ t$, with the corresponding annual productivity growth rates, termed "total productivity growth", computed as the log difference: $\Delta\mathrm{p}_t=\mathrm{p}_t-\mathrm{p}_{t-1}$.

Based on the multi-product model, the average annual growth rate in total productivity in the Norwegian farm production from 2003 to 2018 was around 0.73% *p.a.* with some rather sharp, both up- and downward, fluctuations over the course of years, yielding a modest cumulative fifteen-year increase of 11.8%. In comparison to the multi-product productivity growth estimates reported in the literature for dairy sectors in other European countries like Germany, Iceland, Finland and the

[20] Note that this equivalence applies to finite samples only because, asymptotically, the number of sieve "parameters" will diverge to infinity with the sample size whereas the number of parameters in a parametric specification will stay a finite constant. For more, see Hahn et al. (2018).

Netherlands (e.g., Brümmer et al., 2002; Atsbeha et al., 2012; Sipiläinen et al., 2014) which share institutional environments similar to that of Norway, these are quite conservative estimates.

The finding of a relatively slow productivity growth among milk-producing farms in Norway may be (at least) partly attributed to the regulatory environment in the sector dominated by the Norwegian Agricultural Marketing Board (Brunstad et al., 2005). The regulatory market distortions may well discourage individual farms from improving their efficiency/productivity. Farms in Norway may have even less incentive to improve productivity in the face of the government-supported "multi-functionality" policies, where the latter refer to policies aimed to support non-agricultural aspects of farming such as environmental conservation, cultural heritage tourism and the maintenance of countryside scenery. Furthermore, higher productivity growth rates reported in the above-cited papers likely suffer from simultaneity biases because these studies ignore the endogeneity of inputs/outputs.

The growth in the aggregate productivity can be attributed to two primary sources: (*i*) a secular increase in average productivity across all farms and (*ii*) the reallocation of factors towards more productive farms which would enable the latter to produce more output (Olley & Pakes, 1996). To differentiate between these two sources, we decompose the growth in productivity (i.e., $\Delta \mathrm{p}_t$) into two components. First, we decompose the aggregate productivity p_t itself:

$$\mathrm{p}_t = \sum_i \varrho_{it}\omega_{it} = \overline{\omega}_t + \sum_i \left(\varrho_{it} - \overline{\varrho}_t\right)\left(\omega_{it} - \overline{\omega}_t\right) \equiv \overline{\omega}_t + \mathrm{cov}_t \quad \forall\ t, \tag{C.1}$$

where $\overline{\omega}_t = 1/n_t \sum_i \omega_{it}$ and $\overline{\varrho}_t = 1/n_t$ are the unweighted average productivity and the unweighted output share (a uniform weight), respectively.

According to this decomposition, the aggregate productivity p_t is a sum of the (unweighted) average of farm-level productivity $\overline{\omega}_t$ and a sample covariance between the farm-level total output and productivity $\mathrm{cov}_t \equiv \sum_i \left(\varrho_{it} - \overline{\varrho}_t\right)\left(\omega_{it} - \overline{\omega}_t\right)$. From (C.1), it immediately follows that $\Delta \mathrm{p}_t = \Delta\overline{\omega}_t + \Delta \mathrm{cov}_t\ \forall\ t$, where the average productivity growth $\Delta\overline{\omega}_t$ represents a secular change in productivity capturing temporal shifts in the productivity distribution via the change in its first moment, and $\Delta \mathrm{cov}_t$ measures the reallocation of market share from less productive to more productive farms thus capturing "reshuffling" within the joint distribution of the productivity and market share. The larger the covariance term, the larger the output share of more productive farms in the sector. The decomposition results are reported in Table C.1.

The results suggest that, between the two components, the change in average productivity is by far the dominant contributor to the total productivity growth. This is particularly evident from Figure C.1 which summarizes the productivity decomposition in the form of a bar chart. Over the course of our sample period, 11.4 out of 11.8 percentage points of the cumulative growth in the aggregate productivity came from the improvement in the average farm productivity, with the reallocation effect amounting to negligible 0.4 percentage points.

Table C.1. Productivity Growth Decomposition, %

Year	$\Delta \mathrm{p}_t$	$\Delta \overline{\omega}_t$	$\Delta \mathrm{cov}_t$
2004	–5.08	–3.69	–1.39
2005	–1.42	–0.84	–0.58
2006	5.73	5.27	0.46
2007	–2.09	–1.73	–0.37
2008	8.69	7.25	1.44
2009	6.27	7.36	–1.09
2010	–4.39	–5.31	0.92
2011	2.58	2.29	0.29
2012	–0.40	–0.58	0.18
2013	2.94	3.35	–0.41
2014	–0.35	0.00	–0.35
2015	1.20	1.72	–0.52
2016	–0.99	–2.08	1.09
2017	–3.21	–3.17	–0.05
2018	2.28	1.50	0.78
2003–2008†	0.73		

The results are based on the multi-product model. p_t is the revenue-share-weighted average of firm-level persistent productivity (in logs), i.e., a sector-level aggregate productivity; $\Delta \mathrm{p}_t$ is the total productivity growth; $\Delta \overline{\omega}_t$ is the growth in an unweighted average productivity; $\Delta \mathrm{cov}_t$ is the reallocation effect in the productivity growth; $\Delta \mathrm{p}_t = \Delta \overline{\omega}_t + \Delta \mathrm{cov}_t \; \forall \; t$.
†Annualized average rate of growth.

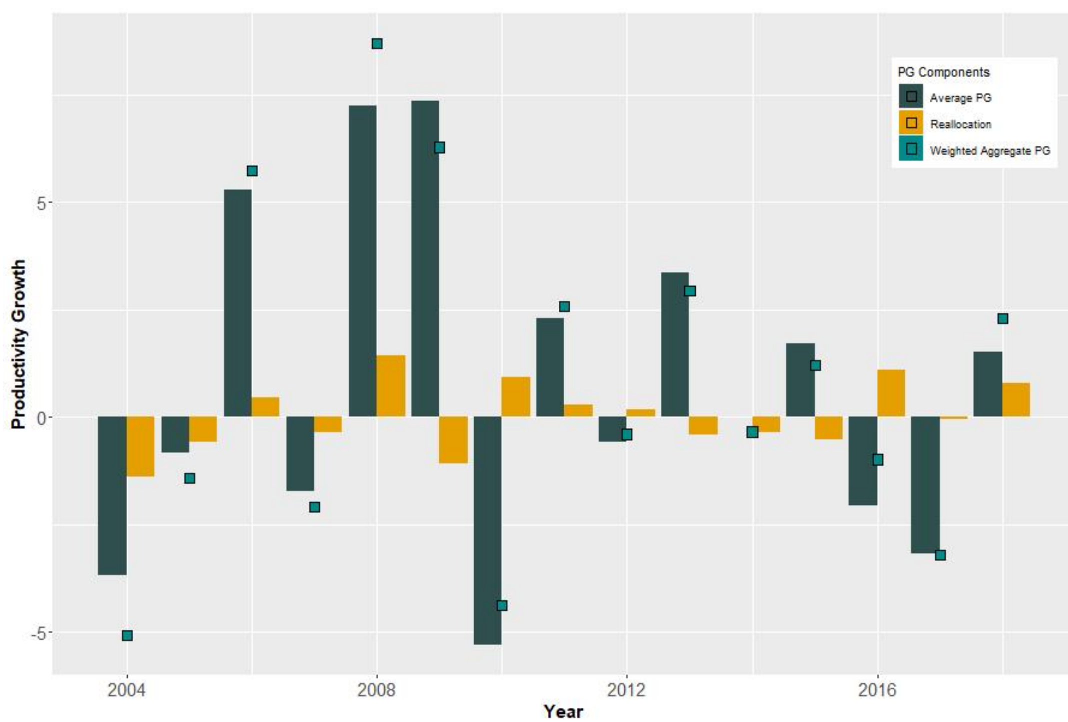

Figure C.1. Productivity Growth Decomposition

D Relation to Other Proxy-Variable Estimators of Multi-Product Production Functions

To facilitate a meaningful comparability of Grieco & McDevitt's (2017) and Dhyne et al.'s (2016) estimators [hereafter "GM" and "DPSW," respectively] with ours, we do the following. First, we discuss these two alternative estimators while treating production function nonparametrically. Second, we adapt them to match the production process considered in this paper by including land as an input and treating labor as being quasi-fixed. Third, since they are not pivotal to the below discussion, we assume any non-production-function factors [e.g., referred to as "x" variables by Grieco & McDevitt (2017)] influencing input/output decisions away. Lastly, to avoid notational clutter, throughout this Appendix, we adhere by our notation wherever possible and define new variables only when necessary.

D.1 The GM Estimator

In this section, we show that the GM estimator cannot be easily extended—due to under-identification issues—to the popular practice whereby the freely-varying intermediate inputs (materials) are used to "invert out" the latent productivity. To show this result, we first summarize GM's identification arguments under their modeling assumption according to which *no* static intermediate inputs are present inside the production function.

Let the multi-product production function be a function of dynamic inputs only:

$$H\left(\widetilde{Y}_{1,it},\ldots,\widetilde{Y}_{S,it}\right) = F(K_{it},N_{it},L_{it})\exp\{\omega_{it}^{q}\}, \tag{D.1}$$

where $\widetilde{Y}_{s,it}$ is the "target" level of output which differs from the realized output $Y_{s,it}$ by the value of an $i.i.d.$ measurement error $\epsilon_{s,it}$ such that $\mathbb{E}[\epsilon_{s,it}|\Omega_{it}]=0$, i.e., $Y_{s,it}=\widetilde{Y}_{s,it}\exp\{\epsilon_{s,it}\}$ for all s.

To remind the reader of GM's assumptions, the target output quantities are said to be decided based on quasi-fixed inputs and ω_{it}^{q}: $\widetilde{Y}_{s,it}=\widetilde{\mathbb{Y}}_{s,t}(K_{it},N_{it},L_{it},\omega_{it}^{q})\ \forall\ s$. The persistent productivity then gets updated to $\omega_{it}^{h}=\omega_{it}^{q}+\epsilon_{it}^{\omega}$, where $\mathbb{E}[\epsilon_{it}^{\omega}|\Omega_{it-1}]=0$, before becoming a part of the information set for the hiring ($\mathcal{H}_{it}$) and investment (I_{it}) decisions. Hence, $\mathcal{H}_{it}=\mathbb{H}_t(K_{it},N_{it},L_{it},\omega_{it}^{h})$ and $I_{it}=\mathbb{I}_t(K_{it},N_{it},L_{it},\omega_{it}^{h})$. The productivity ω_{it}^{h} then evolves to the next period according to an exogenous first-order Markov process: $\omega_{it}^{q}=\mathbb{E}[\omega_{it}^{q}|\ \omega_{it-1}^{h}]+\zeta_{it}$, where $\mathbb{E}[\zeta_{it}|\ \omega_{it-1}^{h}]=0$.

Assuming that the output index function is multiplicatively separable in one of the outputs, say $\widetilde{Y}_{1,it}$, as GM do, one has that $H\left(\widetilde{Y}_{1,it},\ldots,\widetilde{Y}_{S,it}\right)=\widetilde{Y}_{1,it}H_{-1}\left(\widetilde{Y}_{2,it},\ldots,\widetilde{Y}_{S,it}\right)$, from where we obtain in logs:

$$\begin{aligned}\widetilde{y}_{1,it} &= \widetilde{h}\left(\widetilde{Y}_{2,it},\ldots,\widetilde{Y}_{S,it}\right)+f(K_{it},N_{it},L_{it})+\omega_{it}^{q}\\ &= \widetilde{h}\left(\widetilde{Y}_{2,it},\ldots,\widetilde{Y}_{S,it}\right)+f(K_{it},N_{it},L_{it})+\omega_{it}^{h}-\epsilon_{it}^{\omega},\end{aligned} \tag{D.2}$$

where, in the second line, we make use of $\omega_{it}^h = \omega_{it}^q + \epsilon_{it}^\omega$, and $\widetilde{h}(\cdot) \overset{\text{def}}{=} -h_{-1}(\cdot)$.

Following GM, we can proxy for ω_{it}^h via the contemporaneous conditional hiring function, i.e., $\omega_{it}^h = \mathbb{H}_t^{-1}(K_{it}, N_{it}, L_{it}, I_{it}, \mathcal{H}_{it})$,[21] which is functionally dependent with the production function $f(\cdot)$ that is also a function of (K_{it}, N_{it}, L_{it}). Hence, both are not separable:

$$\widetilde{y}_{1,it} = \widetilde{h}\left(\widetilde{Y}_{2,it}, \ldots, \widetilde{Y}_{S,it}\right) + \Phi(K_{it}, N_{it}, L_{it}, I_{it}, \mathcal{H}_{it}) - \epsilon_{it}^\omega, \tag{D.3}$$

where $\Phi(K_{it}, N_{it}, L_{it}, I_{it}, \mathcal{H}_{it}) \overset{\text{def}}{=} f(K_{it}, N_{it}, L_{it}) + \mathbb{M}_t^{-1}(K_{it}, N_{it}, L_{it}, I_{it}, \mathcal{H}_{it})$.

The remaining question is if the output index $\widetilde{h}(\cdot)$ is also perfectly functionally dependent with $\Phi(\cdot)$. It is *not* and therefore separably identified under the GM assumptions that differentiate between ω_{it}^q and ω_{it}^h. While the outputs inside $\widetilde{h}(\cdot)$ in (D.3) are also functions of the same arguments as is $\Phi(\cdot)$, they still exhibit independent variation. Specifically, $\widetilde{Y}_{s,it} = \widetilde{\mathbb{Y}}_{s,t}(K_{it}, N_{it}, L_{it}, \omega_{it}^q) = \widetilde{\mathbb{Y}}_{s,t}(K_{it}, N_{it}, L_{it}, \omega_{it}^h - \epsilon_{it}^\omega) = \widetilde{\mathbb{Y}}_{s,t}(K_{it}, N_{it}, L_{it}, \mathbb{H}_t^{-1}(K_{it}, N_{it}, L_{it}, I_{it}, \mathcal{H}_{it}) - \epsilon_{it}^\omega)$, from where it is evident that ϵ_{it}^ω constitutes a source of (unobserved) independent variation necessary to rule out Ackerberg et al.'s (2015) critique. This breaks the perfect dependence between the two unknown functions and thus identifies $\widetilde{h}(\cdot)$.

To establish our result, we now proceed to the identification strategy that relies on the use of statically optimized, freely varying inputs such as materials to proxy for the latent productivity. The latter is the predominant choice for a proxy variable approach to the estimation of production functions in the applied literature owing to impracticality of the investment or net hiring proxies because they tend to take on too many zero values rending large chunks of data unusable. For instance, GM themselves have to discard some data before being able to operationalize their methodology.

Generalizing GM's production function formulation to include freely varying materials, we have

$$H\left(\widetilde{Y}_{1,it}, \widetilde{Y}_{2,it}\right) = F(K_{it}, N_{it}, L_{it}, M_{it}) \exp\{\omega_{it}^q\}. \tag{D.4}$$

Now, one needs to specify the timing of the firm's decisions about freely varying inputs in the GM framework. Provided that the firm chooses the target output quantities in advance (given the already predetermined quantities of quasi-fixed inputs), it is only reasonable to assume that all freely varying inputs, such as materials, must also be determined at this stage. From (D.4), it is clear that the target output quantities depend on the already predetermined (K_{it}, N_{it}, L_{it}) and productivity ω_{it}^q, which are all observable by the firm when it chooses $(\widetilde{Y}_{1,it}, \widetilde{Y}_{2,it})$. To ensure that the firm's choice of $(\widetilde{Y}_{1,it}, \widetilde{Y}_{2,it})$ is meaningful, one is bound to accept that materials are chosen at this stage too. That is, $M_{it} = \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega_{it}^q)$.

Assuming that the output index function is multiplicatively separable in $\widetilde{Y}_{1,it}$ and proxying for

[21] To guarantee the invertability of the $\mathbb{H}_t$ function, it is conditioned on the level of investment I_{it}.

ω^q_{it} (not ω^h_{it}) using the material demand, i.e., $\omega^q_{it} = \mathbb{M}_t^{-1}(K_{it}, N_{it}, L_{it}, M_{it})$, from (D.4) we have that

$$\widetilde{y}_{1,it} = \widetilde{h}\left(\widetilde{Y}_{2,it}\right) + \Phi(K_{it}, N_{it}, L_{it}, M_{it}), \tag{D.5}$$

where we redefine $\Phi(K_{it}, N_{it}, L_{it}, M_{it}) \stackrel{\text{def}}{=} f(K_{it}, N_{it}, L_{it}, M_{it}) + \mathbb{M}_t^{-1}(K_{it}, N_{it}, L_{it}, M_{it})$.

We next examine if the output index $\widetilde{h}_t(\cdot)$ remains identified. Under GM's assumptions, $\widetilde{Y}_{2,it} = \widetilde{\mathbb{Y}}_{2,t}(K_{it}, N_{it}, L_{it}, \omega^q_{it}) = \widetilde{\mathbb{Y}}_{2,t}(K_{it}, N_{it}, L_{it}, \mathbb{M}_t^{-1}(K_{it}, N_{it}, L_{it}, M_{it}))$, from where one can clearly see that $\widetilde{Y}_{2,it}$ is a function of same inputs $(K_{it}, N_{it}, L_{it}, M_{it})$ which already enter $\Phi(\cdot)$ implying that the two unknown function are perfectly functionally dependent and thus are *not* separably identified. One therefore *cannot* proceed with the GM methodology in this instance.

What if one assumes that the materials decision is made *after* the target output quantities had been determined by the firm, say, if it is made along with the decisions about hiring and investment, which are based on the updated productivity ω^h_{it}? That is, consider the case when $M_{it} = \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega^h_{it})$ instead of $M_{it} = \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega^q_{it})$. Under this scenario, while the output index function *is* identified, the production function however is *not*. Furthermore, the entire method is operational under a rather dubious assumption that M_{it} is determined *after* the firm chooses the target *levels* of outputs, which is hardly reasonable given that the quantity of materials directly determines the quantity of outputs. While it might seem at first that this assumption is identical to the one we make in this paper, the latter is certainly *not* the case. In our methodology, while we also assume that the firm determines the quantity of materials based on the updated productivity, this decision is however made after the choice of the output *ratios* thus allowing for the newly chosen quantity of materials to affect the *levels* of outputs.

Despite a dubious nature of the $M_{it} = \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega^h_{it})$ assumption, we nevertheless show identification results based on it. Specifically, making use of $\omega^h_{it} = \omega^q_{it} + \epsilon^\omega_{it}$, from (D.4) we have that

$$\begin{aligned}
\widetilde{y}_{1,it} &= \widetilde{h}\left(\widetilde{Y}_{2,it}\right) + f(K_{it}, N_{it}, L_{it}, M_{it}) + \omega^h_{it} - \epsilon^\omega_{it} \\
&= \widetilde{h}\left(\widetilde{Y}_{2,it}\right) + f(K_{it}, N_{it}, L_{it}, M_{it}) + \mathbb{M}_t^{-1}(K_{it}, N_{it}, L_{it}, M_{it}) - \epsilon^\omega_{it} \\
&\stackrel{\text{def}}{=} \widetilde{h}\left(\widetilde{Y}_{2,it}\right) + \Phi(K_{it}, N_{it}, L_{it}, M_{it}) - \epsilon^\omega_{it}.
\end{aligned} \tag{D.6}$$

Analogous to the case when hiring (or investment) is used to proxy for latent productivity, here we have that $\widetilde{Y}_{s,it} = \widetilde{\mathbb{Y}}_{s,t}(K_{it}, N_{it}, L_{it}, M_{it}, \omega^q_{it}) = \widetilde{\mathbb{Y}}_{s,t}(K_{it}, N_{it}, L_{it}, M_{it}, \mathbb{M}_t^{-1}(K_{it}, N_{it}, L_{it}, M_{it}) - \epsilon^\omega_{it})$, from where it is clear that ϵ^ω_{it} provides independent variation to $\widetilde{Y}_{s,it}$ necessary to break the perfect functional dependence between the two unknown functions and thus separate $\widetilde{h}_t(\cdot)$ from $\Phi_t(\cdot)$. However, the identification is not yet achieved in this case because M_{it} is now correlated with the unobserved error ϵ^ω_{it}. We tackle this problem below.

Before attempting to estimate (D.6), one needs to address the fact that the target output quantities are not observable, but the realized quantities $Y_{s,it}$ $\forall$ s are. From (D.6), we then obtain

$$y_{1,it} = \widetilde{h}\left(Y_{2,it}\exp\{-\epsilon_{2,it}\}\right) + \Phi(K_{it}, N_{it}, L_{it}, M_{it}) - \epsilon^{\omega}_{it} + \epsilon_{1,it}. \tag{D.7}$$

To be able to estimate the above equation, GM implicitly (by virtue of adopting the Cobb-Douglas specification) assume that the output index is such that $H_{-1}\left(Y_{2,it}\exp\{-\epsilon_{2,it}\}\right) = H_{-1}\left(Y_{2,it}\right) \times \exp\{-\epsilon_{2,it}\}$, which yields

$$y_{1,it} = \widetilde{h}\left(Y_{2,it}\right) + \Phi(K_{it}, N_{it}, L_{it}, M_{it}) - \epsilon^{\omega}_{it} + \epsilon_{1,it} - \epsilon_{2,it}, \tag{D.8}$$

where all right-hand-side covariate besides $Y_{2,it}$ and M_{it} are mean-independent of the composite error term $(\epsilon_{1,it} - \epsilon^{\omega}_{it} - \epsilon_{2,it})$. While it might seem that M_{it} can be instrumented via numerous exogenous lags of inputs and outputs, only $(K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})$ are the relevant instruments. To see this, consider the conditional material demand function:

$$\begin{aligned} M_{it} &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega^h_{it}) \\ &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega^q_{it} + \epsilon^{\omega}_{it}) \\ &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \mathbb{E}[\omega^q_{it}|\omega^h_{it-1}] + \zeta_{it} + \epsilon^{\omega}_{it}) \\ &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \psi[\mathbb{M}^{-1}_{t-1}(K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})] + \zeta_{it} + \epsilon^{\omega}_{it}), \end{aligned} \tag{D.9}$$

where, in the third line, we have made use of the Markovian process for ω^q_{it}, and $\psi[\cdot] \stackrel{\text{def}}{=} \mathbb{E}[\omega^q_{it}|\omega^h_{it-1}]$.

To instrument for $Y_{2,it}$, which suffers from the measurement-error-induced endogeneity, one needs a different noisy measure of $\widetilde{Y}_{2,it}$, as also suggested by GM. Denote this instrument as $Y^{\ddagger}_{2,it}$. Then, both unknown functions $\widetilde{h}(\cdot)$ and $\Phi(\cdot)$ in (D.8) are identified (up to a constant) on the basis of the following moment conditions:

$$\mathbb{E}\left[\epsilon_{1,it} - \epsilon^{\omega}_{it} - \epsilon_{2,it} \;\middle|\; K_{it}, N_{it}, L_{it}, Y^{\ddagger}_{2,it}, K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}\right] = 0. \tag{D.10}$$

While one is able to successfully identify $\Phi(\cdot)$ in (D.8), the GM approach generally however cannot meaningfully disentangle this function into the production function $f(\cdot)$ and productivity ω^q_{it} that it consists of. The latter happens because the second stage of the GM) methodology now suffers from the under-identification problem à la Gandhi et al. (2020) due to the presence of endogenous M_{it} in the production function.

More specifically, consider the second-stage estimator based on the Markovian nature of the productivity. First, define $y^*_{1,it} \stackrel{\text{def}}{=} y_{1,it} - \widetilde{h}\left(Y_{2,it}\right)$. From (D.4), we have that $\omega^q_{it} = y^*_{1,it} - f(K_{it}, N_{it}, L_{it}, M_{it})$. By the definition of $\Phi(\cdot)$, we also have that $\omega^h_{it} = \Phi(K_{it}, N_{it}, L_{it}, M_{it}) - f(K_{it}, N_{it}, L_{it}, M_{it})$. Then, using the Markov process for productivity $\omega^q_{it} = \mathbb{E}[\omega^q_{it}|\;\omega^h_{it-1}] + \zeta_{it}$ yields

$$\begin{aligned} & y^*_{it} - f(K_{it}, N_{it}, L_{it}, M_{it}) = \\ & \quad \psi\left[\Phi(K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}) - f(K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})\right] + \zeta_{it}. \end{aligned} \tag{D.11}$$

Given the already identified output index (and hence y^*_{it}) and composite function $\Phi(\cdot)$, one may hope to also identify the unknown production function $f(\cdot)$ from (D.11), where all variables except M_{it} (which is correlated with ζ_{it}) instrument for themselves and additional lagged inputs can supposedly be used to instrument for M_{it}. Unfortunately, that is *not* the case. From (D.9), one can see that, conditional on the already included self-instrumenting $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})$, the only sources of exogenous variation in M_{it} are unobservable innovations ζ_{it} and ϵ^{ω}_{it}, which by GM's) assumptions are not correlated with any lags. Hence, the "reduced-form" nonparametric regression implied by (D.11), i.e.,

$$y^*_{it} = f(K_{it}, N_{it}, L_{it}, M_{it}) + \widetilde{\psi}(K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}) + \zeta_{it}, \tag{D.12}$$

lacks an excluded relevant instrument for M_{it} and thus is *un*identified.

D.2 The DPSW Estimator

In this section, we show the following two main results about the DPSW methodology: *(i)* one can meaningfully estimate the generalized production function only for *one* output without otherwise violating a scalar unobservability condition necessary for invertability of the proxy function; and *(ii)* it likely suffers from the under-identification problem à la Gandhi et al. (2020) regardless whether one employs investment or materials to proxy for latent productivity.

To remind the reader, the authors propose considering S separate product-specific "generalized" production functions which measures the maximal quantity of a given output producible from the firm's total input quantities while holding quantities of other outputs fixed:

$$Y_{j,it} = F_j(K_{it}, N_{it}, L_{it}, M_{it}, Y_{1,it}, \ldots, Y_{s\neq j,it}, \ldots, Y_{S,it}) \exp\{\omega_{j,it} + \eta_{j,it}\} \qquad \forall\ j = 1, \ldots, S, \tag{D.13}$$

where $F_j(\cdot)$ is positively/negatively monotone in inputs/outputs and concave/quasi-concave in freely varying/quasi-fixed inputs; $\omega_{j,it}$ is the persistent productivity following the first-order Markovian process $\omega_{j,it} = \mathbb{E}[\omega_{j,it}|\omega_{j,it-1}]+\zeta_{j,it}$ with $\mathbb{E}[\zeta_{j,it}|\omega_{j,it-1}] = 0$; and $\eta_{j,it}$ is a random *i.i.d.* transitory productivity shock such that $\mathbb{E}[\eta_{j,it}|\Omega_{it}] = 0$.

The obvious advantage of such a formulation of the multi-output production process is that it allows for product-specific production functions and productivity as well as does not require knowing the input allocations across product lines. Following the literature, DPSW assume that quasi-fixed inputs are chosen in period $t-1$ and thus predetermined, whereas investment and materials decisions are made in the time period t based on quasi-fixed inputs and the realized updated value of persistent productivity.

To arrive at the first result, one needs to carefully think about the firm's investment and materials decisions because they provide the foundation for a control function approach to proxying for unobserved productivity. To use either investment or materials as a proxy for $\omega_{j,it}$ when estimating the model in (D.13), one needs to appeal [as DPSW do in their paper] to the investment or material demand functions of the following forms: $I_{it} = \mathbb{I}_t(K_{it}, N_{it}, L_{it}, \omega_{j,it})$ and $M_{it} = \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega_{j,it})$, respectively. That is, the total *firm*-level investment and the demand for materials are to be postulated as functions of the *j-product-specific*, not the firm-level, latent productivity. Assume one is comfortable with such an assumption. Then, when estimating the generalized production function for some different $s(\neq j)$th output, in order to make use of the proxy approach, one is to similarly assume that the firm-level investment and materials are functions of the s-product-specific latent productivity: $I_{it} = \mathbb{I}_t(K_{it}, N_{it}, L_{it}, \omega_{s,it})$ and $M_{it} = \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega_{s,it})$ for $s \neq j$. In doing so, one arrives at the *contradiction* which can be reconciled by either *(i)* assuming that persistent productivity is the same across all products, i.e., $\omega_{j,it} = \omega_{s,it}$ for all $s(\neq j) = 1, \ldots, S$, which however undermines one of the advertised advantages of the DPSW model whereby it identifies differential firm-product efficiency measures, or *(ii)* reasonably postulating that the firm-level demands for investment and materials are functions of *all* product-specific productivities, i.e., $I_{it} = \mathbb{I}_t(K_{it}, N_{it}, L_{it}, \omega_{1,it}, \ldots, \omega_{S,it})$ and $M_{it} = \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega_{1,it}, \ldots, \omega_{S,it})$. While the latter assumption implies quite a realistic production framework, it however violates a scalar unobservability condition necessary to guarantee invertability of the investment/materials demand function in a single $\omega_{j,it}$. Hence, the proxy method would be invalid in this instance.

Given the above, perhaps the easiest way out which provides a consistent story about the firm's production decision making process is to assume that both the investment and materials are determined based on *one* product-specific productivity for which the researcher wishes to estimate the generalized production function. However, the latter implies that the estimation of production functions for the remaining outputs via the same proxy approach will generally be infeasible. In what follows, we abide by this assumption and focus on the estimation of production process for the *first* output only. For notational simplicity, we also set $S = 2$.

To derive the second result, we first focus on the case when investment is used to invert out unobserved persistent productivity, a choice favored by DPSW. The first output's generalized production function in logs takes the following form:

$$
\begin{aligned}
y_{1,it} &= f_1(K_{it}, N_{it}, L_{it}, M_{it}, Y_{2,it}) + \omega_{1,it} + \eta_{1,it} \\
&= f_1(K_{it}, N_{it}, L_{it}, M_{it}, Y_{2,it}) + \psi_1[\omega_{1,it-1}] + \zeta_{1,it} + \eta_{1,it} \\
&= f_1(K_{it}, N_{it}, L_{it}, M_{it}, Y_{2,it}) + \psi_1\left[\mathbb{I}_{t-1}^{-1}(K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1})\right] + \zeta_{1,it} + \eta_{1,it} \\
&\overset{\text{def}}{=} f_1(K_{it}, N_{it}, L_{it}, M_{it}, Y_{2,it}) + \varphi_1(K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1}) + \zeta_{1,it} + \eta_{1,it},
\end{aligned}
\tag{D.14}
$$

where we have made use of the first-order Markovian nature of $\omega_{1,it}$ in the second equality by letting $\mathbb{E}[\omega_{1,it}|\ \omega_{1,it-1}]$ be an unknown function $\psi_1[\cdot]$ and have defined $\varphi_1(\cdot) \overset{\text{def}}{=} \psi_1\left[\mathbb{I}_{t-1}^{-1}(\cdot)\right]$ in the last equality.

One might think that the model in (D.14) is identified based on

$$\mathbb{E}\left[\zeta_{1,it}+\eta_{1,it}\;\middle|\;\begin{array}{l} K_{it}, N_{it}, L_{it}, \\ K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}, I_{it-1}, Y_{2,it-1}, \ldots, \\ K_{i1}, N_{i1}, L_{i1}, M_{i1}, I_{i1}, Y_{2,i1} \end{array}\right]=0, \tag{D.15}$$

where $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1})$ instrument for themselves, and one has the abundance of lagged inputs and outputs to instrument for M_{it} and $Y_{2,it}$. This happens to be DPSW's argument.

While the variables in the conditioning set of (D.15) are indeed exogenous, they must also be relevant once conditioned on the already included $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1})$. To examine the latter, consider $Y_{2,it}$ appearing inside the production function $f_1(\cdot)$, which we can write as follows by making use of its own generalized production function:

$$\begin{aligned} Y_{2,it} &= F_2(K_{it}, N_{it}, L_{it}, M_{it}, Y_{1,it})\exp\{\omega_{2,it}+\eta_{2,it}\} \\ &= F_2\Big(K_{it}, N_{it}, L_{it}, M_{it}, F_1(K_{it}, N_{it}, L_{it}, M_{it}, Y_{2,it})\exp\{\omega_{1,it}+\eta_{1,it}\}\Big)\exp\{\omega_{2,it}+\eta_{2,it}\} \\ &= F_2\Big(K_{it}, N_{it}, L_{it}, M_{it}, F_1(K_{it}, N_{it}, L_{it}, M_{it}, Y_{2,it})\exp\{\varphi_1(K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1})+\zeta_{1,it}+\eta_{1,it}\}\Big)\times \\ &\qquad \exp\{\omega_{2,it}+\eta_{2,it}\}. \end{aligned} \tag{D.16}$$

Also, recognize that the materials

$$\begin{aligned} M_{it} &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \omega_{1,it}) \\ &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \varphi_1(K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})+\zeta_{1,it}), \end{aligned} \tag{D.17}$$

when using materials as a productivity proxy (instead of investment), can be written as a function of $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})$. This demonstrates that M_{it-1} is an excluded [from (D.14)] source of exogenous *relevant* variation in M_{it}. Hence, M_{it-1} is a candidate for valid instrument for M_{it}.

Next, by substituting (D.17) into (D.16), it is evident that $Y_{2,it}$ is a function of observable $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}, I_{it-1})$ and unobservable $(\zeta_{1,it}, \eta_{1,it}, \omega_{2,it}, \eta_{2,it})$ with all observable arguments being already included in the model or instrumenting for M_{it}. Hence, in order to identify the production function $f_1(\cdot)$ in (D.14), one needs another *excluded* instrument that correlates with the unobservable determinants $(\zeta_{1,it}, \eta_{1,it}, \omega_{2,it}, \eta_{2,it})$ of $Y_{2,it}$. All additional lags are by definition mean-independent of $(\zeta_{1,it}, \eta_{1,it}, \eta_{2,it})$.[22] What about potential correlates of $\omega_{2,it}$?

This brings us back to the earlier discussion about the treatment of product-specific productivities. If one were to ignore the violation of a scalar unobservability assumption by also inverting $\omega_{2,it}$ using total investment (or materials) [like in DPSW], i.e., $\omega_{2,it}=\psi_2[\omega_{2,it-1}]+\zeta_{2,it}=\psi_2\left[\mathbb{I}_{t-1}^{-1}(K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1})\right]+\zeta_{2,it}$, we find that, conditional on $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1})$,

[22] If they had correlated with $(\zeta_{1,it}, \eta_{1,it}, \eta_{2,it})$, they would have no longer been exogenous.

no additional lag is correlated with $\omega_{2,it}$. Hence, the rank condition is unmet, and the model is unidentified. Alternatively, we could assume that product-specific productivity is the same across all outputs. If so, $\omega_{2,it} = \omega_{1,it} = \varphi_1(K_{it-1}, N_{it-1}, L_{it-1}, I_{it-1}) + \zeta_{2,it}$ follows directly, implying that $Y_{2,it}$ still lacks a valid relevant instrument, and the model remains unidentified.

The under-identification problem worsens further if M_{it} is instead employed as a proxy for $\omega_{1,it}$. In this case, as shown in detail in the main text, M_{it} inside the production function also lacks a valid instrument. Therefore, the generalized production function $F_1(\cdot)$ would be under-identified because both $Y_{2,it}$ and M_{it} are missing relevant instruments.

References

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83, 2411–2451.

Atsbeha, D. M., Kristofersson, D., & Rickertsen, K. (2012). Animal breeding and productivity growth of dairy farms. *American Journal of Agricultural Economics*, 94, 996–1012.

Brümmer, B., Glauben, T., & Thijssen, G. (2002). Decomposition of productivity growth using distance functions: The case of dairy farms in three European countries. *American Journal of Agricultural Economics*, 84, 628–644.

Brunstad, R. J., Gaasland, I., & Vårdal, E. (2005). Efficiency losses in milk marketing boards – the importance of exports. *Nordic Journal of Political Economy*, 31, 77–97.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics*, volume 6B. North Holland.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, 13, 377–403.

Dhyne, E., Petrin, A., Smeets, V., & Warzynski, F. (2016). Multi-product firms, import competition, and the evolution of firm-product technical efficiencies. Working Paper, Aarhus University.

Gandhi, A., Navarro, S., & Rivers, D. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128, 2973–3016.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* Taylor & Francis Group, LLC.

Grieco, P. L. E. & McDevitt, R. C. (2017). Productivity and quality in health care: Evidence from the dialysis industry. *Review of Economic Studies*, 84, 1071–1105.

Hahn, J., Liao, Z., & Ridder, G. (2018). Nonparametric two-step sieve M estimation and inference. *Econometric Theory*, 34, 1281–1324.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, 14 (2), 201–206.

Olley, G. S. & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64, 1263–1297.

Sipiläinen, T., Kumbhakar, S. C., & Lien, G. (2014). Performance of dairy farms in Finland and Norway from 1991 to 2008. *European Review of Agricultural Economics*, 41, 63–86.