Estimating production functions with expectations data

Agnes Norris Keiller[1] Aureo de Paula[2] John Van Reenen[1,3]

[1]Programme on Innovation and Diffusion, LSE

[2]UCL

[3]MIT

May 16 2023

Motivation

- Production functions integral to many strands of research

$$y_{it} = f(k_{it}, l_{it}; \theta) + e_{it}$$

Motivation

- Production functions integral to many strands of research

$$y_{it} = f(k_{it}, l_{it}; \theta) + e_{it}$$

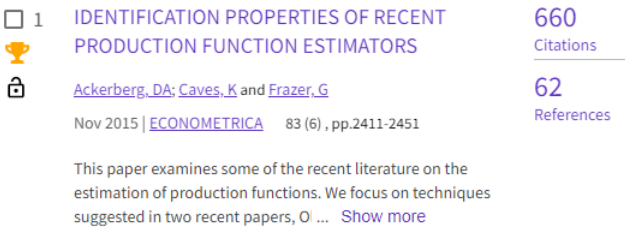
- Extensive literature has presented various estimation methods
  - First order conditions: Solow 1957; Hall 1988
  - Dynamic panel IV: Chamberlain 1982; Blundell and Bond 2000
  - Control functions: Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015

Motivation

- Production functions integral to many strands of research

$$y_{it} = f(k_{it}, l_{it}; \theta) + e_{it}$$

- Extensive literature has presented various estimation methods


1 IDENTIFICATION PROPERTIES OF RECENT PRODUCTION FUNCTION ESTIMATORS

Ackerberg, DA; Caves, K and Frazer, G

Nov 2015 | ECONOMETRICA 83 (6) , pp.2411-2451

This paper examines some of the recent literature on the estimation of production functions. We focus on techniques suggested in two recent papers, Ol ... Show more

660 Citations

62 References

Motivation

- Production functions integral to many strands of research

$$y_{it} = f(k_{it}, l_{it}; \theta) + e_{it}$$

- Extensive literature has presented various estimation methods

1 IDENTIFICATION PROPERTIES OF RECENT PRODUCTION FUNCTION ESTIMATORS

Ackerberg, DA; Caves, K and Frazer, G

As of November/December 2022, this highly cited paper received enough citations to place it in the top 1% of the academic field of **Economics & Business** based on a highly cited threshold for the field and publication year.

660 Citations

62 References

Motivation

- Production functions integral to many strands of research

$$y_{it} = f(k_{it}, l_{it}; \theta) + e_{it}$$

- Extensive literature has presented various estimation methods
    - First order conditions: Solow 1957; Hall 1988
    - Dynamic panel IV: Chamberlain 1982; Blundell and Bond 2000
    - Control functions: Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015
- Surveys increasingly elicit firms' expectations about future inputs and outputs

Motivation

- Production functions integral to many strands of research

$$y_{it} = f(k_{it}, l_{it}; \theta) + e_{it}$$

- Extensive literature has presented various estimation methods
  - First order conditions: Solow 1957; Hall 1988
  - Dynamic panel IV: Chamberlain 1982; Blundell and Bond 2000
  - Control functions: Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015
- Surveys increasingly elicit firms' expectations about future inputs and outputs
- **Can we improve on existing production function estimators using data on firms' expectations?**

Preview

Can we improve on existing production function estimators using data on firms' expectations?

Preview

Can we improve on existing production function estimators using data on firms' expectations?

- Theory
  - Expectations data allow one to relax assumptions of optimal firm choices required by control function estimators

Preview

Can we improve on existing production function estimators using data on firms' expectations?

- Theory
  - Expectations data allow one to relax assumptions of optimal firm choices required by control function estimators
- MC simulations
  - Our proposed estimator is robust to optimisation error in inputs, while other methods are not

Preview

Can we improve on existing production function estimators using data on firms' expectations?

- Theory
  - Expectations data allow one to relax assumptions of optimal firm choices required by control function estimators
- MC simulations
  - Our proposed estimator is robust to optimisation error in inputs, while other methods are not
- UK data over 2017-2020
  - Expectations estimator implies more dispersed productivity distribution than alternatives

Today

Methodology

Performance

- Monte Carlo
- Empirical application

Next steps

Extra results

Today

Methodology

Performance

- Monte Carlo
- Empirical application

Next steps

Extra results

The object of interest

- Consider a general production function of the following form
$$y_{it} = f(k_{it}, l_{it}; \theta) + \omega_{it} + \epsilon_{it} + \upsilon_{it} \quad (1)$$

- ω_{it} = idiosyncratic productivity known by the firm when deciding period t input and investment
- ϵ_{it} and υ_{it} = unanticipated mean-zero disturbances
  - ϵ_{it} = productivity shocks unknown by the firm when making period t decisions
  - υ_{it} = measurement error

Dynamics

- Capital evolves according to

$$K_{it} = (1 - \delta)K_{it-1} + i_{it-1} \tag{2}$$

  - δ = the depreciation rate, i_{it-1} = investment

Dynamics

- Capital evolves according to

$$K_{it} = (1-\delta)K_{it-1} + i_{it-1} \tag{2}$$

  - δ = the depreciation rate, i_{it-1} = investment

- ω_{it} follows a Markov process

$$\omega_{it} = \mathbb{E}[\omega_{it}|\omega_{it-1}] + \xi_{it} = g(\omega_{it-1}) + \xi_{it} \tag{3}$$

  - $\mathbb{E}[\xi_{it}|I_{it-1}] = 0$

Expectations

- Firms form expectations about $t+1$ production and inputs at the end of t conditional on $I_{it} = \{k_{it}, l_{it}, i_{it}, \omega_{it}, \epsilon_{it}, k_{it+1}\}$

Expectations

- Firms form expectations about $t+1$ production and inputs at the end of t conditional on $I_{it} = \{k_{it}, l_{it}, i_{it}, \omega_{it}, \epsilon_{it}, k_{it+1}\}$
- If firms' expectations align with the true production technology

$$\mathbb{E}_{it}[y_{it+1}|I_{it}] = \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1}) + \mathbb{E}_{it}[\omega_{it+1}|I_{it}] + \mathbb{E}_{it}[\epsilon_{it+1}|I_{it}] + \mathbb{E}_{it}[\upsilon_{it+1}|I_{it}] \quad (4)$$

  - $F_{it}(l_{it+1})$ = firm i's subjective probability distribution over their next-period labour input

Expectations

- Firms form expectations about $t+1$ production and inputs at the end of t conditional on $I_{it} = \{k_{it}, l_{it}, i_{it}, \omega_{it}, \epsilon_{it}, k_{it+1}\}$
- If firms' expectations align with the true production technology

$$\begin{aligned} \mathbb{E}_{it}[y_{it+1}|I_{it}] &= \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1}) \\ &+ \mathbb{E}_{it}[\omega_{it+1}|I_{it}] + \mathbb{E}_{it}[\epsilon_{it+1}|I_{it}] + \mathbb{E}_{it}[\upsilon_{it+1}|I_{it}] \quad (4) \\ &= \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1}) + g(\omega_{it}) \end{aligned}$$

  - $F_{it}(l_{it+1})$ = firm i's subjective probability distribution over their next-period labour input

Recovering ω_{it}

- Rearranging equation 4 for $g(\omega_{it})$ obtains

$$g(\omega_{it}) = \mathbb{E}_{it}[y_{it+1}|I_{it}] - \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1}) \quad (5)$$

- Assuming the RHS of equation 5 is strictly increasing in ω_{it}

$$\omega_{it} = g^{-1}\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1})\right) \quad (6)$$

- Ψ = a non-parametric representation of g^{-1}

- $\rightarrow$ a moment condition we can use to recover θ

$$\mathbb{E}[\epsilon_{it} + \upsilon_{it}]$$

$$= \mathrm{E}[y_{it} - f(k_{it}, l_{it}; \theta) - \omega_{it}]$$
$$= \mathrm{E}\left[y_{it} - f(k_{it}, l_{it}; \theta) - \Psi\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1})\right)\right]$$
$$= 0$$

(6)

- Ψ = a non-parametric representation of g^{-1}
- Combining

$$\omega_{it} = g^{-1}\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1})\right) \qquad (6)$$

with

$$y_{it} = f(k_{it}, l_{it}; \theta) + \omega_{it} + \epsilon_{it} + \upsilon_{it} \qquad (1)$$

- $\rightarrow$ a moment condition we can use to recover θ

$\mathbb{E}[\epsilon_{it} + \upsilon_{it}]$

$$= \mathrm{E}[\mathrm{y}_{it} - f(k_{it}, l_{it}; \theta) - \omega_{it}]$$
$$= \mathrm{E}\left[y_{it} - f(k_{it}, l_{it}; \theta) - \Psi\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \int f(k_{it+1}, l_{it+1}; \theta) dF_{it}(l_{it+1})\right)\right]$$
$$= 0$$

(6)

Estimation

Cobb-Douglas production implies

$$\begin{aligned} y_{it} &= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \omega_{it} + \epsilon_{it} + \upsilon_{it} \\ &= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \Psi\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_0 - \beta_k k_{it+1} - \beta_l \mathbb{E}_{it}[l_{it+1}|I_{it}]\right) + \epsilon_{it} + \upsilon_{it} \end{aligned}$$

Estimation

Cobb-Douglas production implies

$$
\begin{aligned}
y_{it} &= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \omega_{it} + \epsilon_{it} + \upsilon_{it} \\
&= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \Psi\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_0 - \beta_k k_{it+1} - \beta_l \mathbb{E}_{it}[l_{it+1}|I_{it}]\right) + \epsilon_{it} + \upsilon_{it}
\end{aligned}
$$

- Assuming Ψ is a smooth function, this is an example of a *generalized additive model*
  - See Hastie and Tibshirani (1986) and Robinson (1988)

Estimation

Cobb-Douglas production implies

$$
\begin{aligned}
y_{it} &= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \omega_{it} + \epsilon_{it} + \upsilon_{it} \\
&= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \Psi\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_0 - \beta_k k_{it+1} - \beta_l \mathbb{E}_{it}[l_{it+1}|I_{it}]\right) + \epsilon_{it} + \upsilon_{it}
\end{aligned}
$$

- Assuming Ψ is a smooth function, this is an example of a *generalized additive model*
  - See Hastie and Tibshirani (1986) and Robinson (1988)
- Problem 1: we require Ψ to be monotonic
  - Impose constraints on the 1st and 2nd derivatives of the smooth functions that comprise Ψ (Pya and Wood 2015)

Estimation

Cobb-Douglas production implies

$$\begin{aligned} y_{it} &= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \omega_{it} + \epsilon_{it} + \upsilon_{it} \\ &= \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \Psi\left(\mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_0 - \beta_k k_{it+1} - \beta_l \mathbb{E}_{it}[l_{it+1}|I_{it}]\right) + \epsilon_{it} + \upsilon_{it} \end{aligned}$$

- Assuming Ψ is a smooth function, this is an example of a *generalized additive model*
    - See Hastie and Tibshirani (1986) and Robinson (1988)
- Problem 1: we require Ψ to be monotonic
    - Impose constraints on the 1st and 2nd derivatives of the smooth functions that comprise Ψ (Pya and Wood 2015)
- Problem 2: Ψ's argument is a function of the linear parameters
    - Use an iterative 'backfitting' algorithm (Friedman and Stuetzle 1981)

Estimation

Adapting the Friedman and Stuetzle (1981) algorithm to our setting

Estimation

Adapting the Friedman and Stuetzle (1981) algorithm to our setting

1. Initialise the parameter vector at $\hat{\theta}_0 = (\hat{\beta}_{00}, \hat{\beta}_{k0}, \hat{\beta}_{l0})$

Estimation

Adapting the Friedman and Stuetzle (1981) algorithm to our setting

1. Initialise the parameter vector at $\hat{\theta}_0 = (\hat{\beta}_{00}, \hat{\beta}_{k0}, \hat{\beta}_{l0})$
   - *"A good starting value might set... [$\hat{\theta}_0$] equal to the values predicted by a linear in x least squares regression of Y on a constant and all the regressors."* (Ichimura and Todd 2007)

In maths

Estimation

Adapting the Friedman and Stuetzle (1981) algorithm to our setting

1. Initialise the parameter vector at $\hat{\theta}_0 = (\hat{\beta}_{00}, \hat{\beta}_{k0}, \hat{\beta}_{l0})$
   - *"A good starting value might set... [$\hat{\theta}_0$] equal to the values predicted by a linear in x least squares regression of Y on a constant and all the regressors."* (Ichimura and Todd 2007)

     In maths
2. For iteration j, calculate
   $Z_{ij} = \mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_{0j-1} - \beta_{kj-1}k_{it+1} - \beta_{lj-1}\mathbb{E}_{it}[l_{it+1}|I_{it}]$

Estimation

Adapting the Friedman and Stuetzle (1981) algorithm to our setting

1. Initialise the parameter vector at $\hat{\theta}_0 = (\hat{\beta}_{00}, \hat{\beta}_{k0}, \hat{\beta}_{l0})$
   - *"A good starting value might set... [$\hat{\theta}_0$] equal to the values predicted by a linear in x least squares regression of Y on a constant and all the regressors."* (Ichimura and Todd 2007)

     In maths
2. For iteration j, calculate
   $Z_{ij} = \mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_{0j-1} - \beta_{kj-1}k_{it+1} - \beta_{lj-1}\mathbb{E}_{it}[l_{it+1}|I_{it}]$
3. Fit the model $y_{it} = \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \Psi\left(Z_{ij}\right) + \epsilon_{it} + \upsilon_{it}$ using the shape constrained estimator of Pya and Wood (2015) to obtain $\hat{\theta}_j$

Estimation

Adapting the Friedman and Stuetzle (1981) algorithm to our setting

1. Initialise the parameter vector at $\hat{\theta}_0 = (\hat{\beta}_{00}, \hat{\beta}_{k0}, \hat{\beta}_{l0})$
   - *"A good starting value might set... [$\hat{\theta}_0$] equal to the values predicted by a linear in x least squares regression of Y on a constant and all the regressors."* (Ichimura and Todd 2007)

     In maths
2. For iteration j, calculate
   $Z_{ij} = \mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_{0j-1} - \beta_{kj-1}k_{it+1} - \beta_{lj-1}\mathbb{E}_{it}[l_{it+1}|I_{it}]$
3. Fit the model $y_{it} = \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \Psi(Z_{ij}) + \epsilon_{it} + \upsilon_{it}$ using the shape constrained estimator of Pya and Wood (2015) to obtain $\hat{\theta}_j$
4. Calculate the Euclidean distance between $\hat{\theta}_j$ and $\hat{\theta}_{j-1}$. If the distance is below some tolerance level, stop and treat $\hat{\theta}_j$ as the model's parameter estimates. If not then set $j \leftarrow j+1$ and repeat from step 2

For the remainder of these slides, this algorithm is referred to as **'NPR'**

Estimation

Adapting the Friedman and Stuetzle (1981) algorithm to our setting

1. Initialise the parameter vector at $\hat{\theta}_0 = (\hat{\beta}_{00}, \hat{\beta}_{k0}, \hat{\beta}_{l0})$
   - *"A good starting value might set... [$\hat{\theta}_0$] equal to the values predicted by a linear in x least squares regression of Y on a constant and all the regressors."* (Ichimura and Todd 2007)

     In maths
2. For iteration j, calculate
   $Z_{ij} = \mathbb{E}_{it}[y_{it+1}|I_{it}] - \beta_{0j-1} - \beta_{kj-1}k_{it+1} - \beta_{lj-1}\mathbb{E}_{it}[l_{it+1}|I_{it}]$
3. Fit the model $y_{it} = \beta_0 - \beta_k k_{it} - \beta_l l_{it} + \Psi(Z_{ij}) + \epsilon_{it} + \upsilon_{it}$ using the shape constrained estimator of Pya and Wood (2015) to obtain $\hat{\theta}_j$
4. Calculate the Euclidean distance between $\hat{\theta}_j$ and $\hat{\theta}_{j-1}$. If the distance is below some tolerance level, stop and treat $\hat{\theta}_j$ as the model's parameter estimates. If not then set $j \leftarrow j + 1$ and repeat from step 2

For the remainder of these slides, this algorithm is referred to as npr

Why bother?

Why bother?

- NPR requires
  1. Firms expectations align with the true production technology
  2. ω follows a first-order Markov process
  3. The LOM for ω is monotonic

Why bother?

- NPR requires
  1. Firms expectations align with the true production technology
  2. ω follows a first-order Markov process
  3. The LOM for ω is monotonic
- 1-2 are also required by OP\LP\ACF
- The equivalent of point 3 for OP\LP\ACF is that firm decisions (conditional on observables) are monotonic in ω

Why bother?

- NPR requires
  1. Firms expectations align with the true production technology
  2. ω follows a first-order Markov process
  3. The LOM for ω is monotonic
- 1-2 are also required by OP\LP\ACF
- The equivalent of point 3 for OP\LP\ACF is that firm decisions (conditional on observables) are monotonic in ω
  - OP: firms' investment policy $\rightarrow \omega = \Phi^{OP}(i_{it}, k_{it})$
  - LP: firms' material input policy $\rightarrow \omega = \Phi^{LP}(k_{it})$
  - ACF: firms' material input policy $\rightarrow \omega = \Phi^{ACF}(l_{it}, k_{it})$
  - Typically justified by a model of optimal firm decisions

Why bother?

- NPR requires
  1. Firms expectations align with the true production technology
  2. ω follows a first-order Markov process
  3. The LOM for ω is monotonic
- 1-2 are also required by OP\LP\ACF
- The equivalent of point 3 for OP\LP\ACF is that firm decisions (conditional on observables) are monotonic in ω
  - OP: firms' investment policy $\rightarrow \omega = \Phi^{OP}(i_{it}, k_{it})$
  - LP: firms' material input policy $\rightarrow \omega = \Phi^{LP}(k_{it})$
  - ACF: firms' material input policy $\rightarrow \omega = \Phi^{ACF}(l_{it}, k_{it})$
  - Typically justified by a model of optimal firm decisions
- NPR assumes nothing about the optimality of firms' decisions

Today

Methodology

Performance

- Monte Carlo
- Empirical application

Next steps

Extra results

Today

Methodology

Performance

- Monte Carlo
- Empirical application

Next steps

Extra results

Monte Carlo setup

- Following ACF
  - y a Leontief composite of m and a 'value added' function of l and k
  - ω follows an AR(1) process
  - Investment subject to a firm-specific convex adjustment cost
  - Allow for l to be chosen at an intermediate period without full knowledge of ω
  - Allow for firm-specific wage shocks

In maths

Monte Carlo setup

- Following ACF
  - y a Leontief composite of m and a 'value added' function of l and k
  - ω follows an AR(1) process
  - Investment subject to a firm-specific convex adjustment cost
  - Allow for l to be chosen at an intermediate period without full knowledge of ω
  - Allow for firm-specific wage shocks

In maths

- Firms' optimal decisions have an analytical solution
- Simulate 1000 firms over 100 periods, use data from last 10
- Compare OLS\LP\ACF\NPR across various DGPs

Measurement error in materials

Optimisation error in labour

M Meas. Error	β_l	β_k	β_l	β_k
	OLS		LP	
0.0				
0.1				
0.25				
0.5				
	ACF		NPR	
0.0				
0.1				
0.25				
0.5				

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications.

Serially correlated wages and intermediate l | Serially correlated wages and optimisation error in l

Measurement error in materials

Optimisation error in labour

M Meas. Error	β_l	β_k	β_l	β_k
	OLS		LP	
0.0	0.381 (0.007)	0.919 (0.002)	0.600 (0.003)	0.399 (0.014)
0.1	0.381 (0.007)	0.919 (0.002)	0.611 (0.003)	0.391 (0.013)
0.25	0.381 (0.007)	0.919 (0.002)	0.655 (0.003)	0.355 (0.012)
0.5	0.381 (0.007)	0.919 (0.002)	0.746 (0.004)	0.276 (0.010)
	ACF		NPR	
0.0	0.600 (0.009)	0.400 (0.016)	0.649 (0.099)	0.336 (0.524)
0.1	0.601 (0.009)	0.401 (0.016)	0.649 (0.099)	0.336 (0.524)
0.25	0.605 (0.010)	0.407 (0.016)	0.649 (0.099)	0.336 (0.524)
0.5	0.617 (0.012)	0.411 (0.017)	0.649 (0.099)	0.336 (0.524)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications.

Serially correlated wages and intermediate I | Serially correlated wages and optimisation error in l

Measurement error in materials

Optimisation error in labour

M Meas. Error	β_l	β_k	β_l	β_k
	OLS		LP	
0.0	0.381 (0.007)	0.919 (0.002)	0.600 (0.003)	0.399 (0.014)
0.1	0.381 (0.007)	0.919 (0.002)	0.611 (0.003)	0.391 (0.013)
0.25	0.381 (0.007)	0.919 (0.002)	0.655 (0.003)	0.355 (0.012)
0.5	0.381 (0.007)	0.919 (0.002)	0.746 (0.004)	0.276 (0.010)
	ACF		NPR	
0.0	0.600 (0.009)	0.400 (0.016)	0.600 (0.003)	0.400 (0.005)
0.1	0.601 (0.009)	0.401 (0.016)	0.600 (0.003)	0.400 (0.005)
0.25	0.605 (0.010)	0.407 (0.016)	0.600 (0.003)	0.400 (0.005)
0.5	0.617 (0.012)	0.411 (0.017)	0.600 (0.003)	0.400 (0.005)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications. NPR results from initialisation at $\beta_{l0} = 0.45$ and $\beta_{k0} = 0.55$.

Serially correlated wages and intermediate I Serially correlated wages and optimisation error in l

Optimisation error

Optimisation error in labour and other inputs

Optim. Error	β_l	β_k	β_l	β_k
	OLS		LP	
m				
i				
(i, m)				
	ACF		NPR	
m				
i				
(i, m)				

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications. All DGPs feature optimisation error in labour.

Alternative initialisation

Optimisation error

Optimisation error in labour and other inputs

Optim. Error	β_l	β_k	β_l	β_k
	OLS		LP	
m	0.380 (0.006)	0.919 (0.002)	0.815 (0.004)	0.462 (0.028)
i	-0.044 (0.016)	0.806 (0.006)	0.000 (0.004)	0.403 (0.016)
(i, m)	-0.044 (0.016)	0.806 (0.006)	0.636 (0.009)	0.396 (0.006)
	ACF		NPR	
m	0.688 (0.014)	0.350 (0.017)	0.650 (0.100)	0.339 (0.527)
i	0.367 (0.785)	0.404 (0.012)	0.616 (0.063)	0.400 (0.002)
(i, m)	-286.016 (3815.097)	-19.295 (508.348)	0.616 (0.061)	0.400 (0.002)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications. All DGPs feature optimisation error in labour.

Alternative initialisation

Optimisation error

Optimisation error in labour and other inputs

Optim.	β_l	β_k	N runs	β_l	β_k	N runs
Error	OLS			LP		
m	0.380 (0.006)	0.919 (0.002)	500	0.815 (0.004)	0.462 (0.028)	500
i	0.144 (0.186)	0.860 (0.053)	2	0.003 (0.002)	0.402 (0.002)	260
(i, m)	0.144 (0.186)	0.860 (0.053)	2	0.636 (0.009)	0.396 (0.006)	500
Error	ACF			NPR		
m	0.688 (0.014)	0.350 (0.017)	500	0.633 (0.096)	0.407 (0.076)	393
i	0.580 (0.087)	0.400 (0.002)	455	0.616 (0.063)	0.400 (0.002)	500
(i, m)	0.351 (0.054)	0.563 (0.224)	2	0.616 (0.061)	0.400 (0.002)	500

Note: table restricted to replications with both β_l and β_k in the range 0 to 1. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the number of replications stated in the table. All DGPs feature optimisation error in labour.

Monte Carlo summary

1. NPR robust to measurement error in materials
   - Outperforms OLS\LP\ACF as measurement error $\sigma \uparrow$
2. NPR robust to optimisation errors
   - Outperforms OLS\LP\ACF as optimisation error $\sigma \uparrow$
3. NPR algorithm sensitive to initialisation
   - Far more imprecise than OLS\LP\ACF using the Ichimura and Todd (2007) initialisation
   - Outperforms OLS\LP\ACF across all DGPs when initialisation adequately close to true values

Today

Methodology

Performance

- Monte Carlo
- Empirical application

Next steps

Extra results

Data

1. Management and Expectations Survey* (MES)
    - Voluntary survey of a representative sample of firms in 2017 and 2020
    - Output, labour, materials and one-period-ahead expectations
2. ABI/ABS**
    - Detailed questions on capital expenditure
    - Match with MES to obtain investment and impute capital

* Office for National Statistics (2022)

** University of West of England et al. (2022); Office for National Statistics (2023).

Expectations in the MES

- The MES elicits expectations by asking
  - *Looking ahead to the 2018 calendar year, what is the approximate pound sterling value of turnover you would anticipate for this business in the following scenarios [Lowest, Low, Medium, High, Highest], and what likelihood do you assign to each scenario?*

Expectations in the MES

- The MES elicits expectations by asking
    - *Looking ahead to the 2018 calendar year, what is the approximate pound sterling value of turnover you would anticipate for this business in the following scenarios [Lowest, Low, Medium, High, Highest], and what likelihood do you assign to each scenario?*
- MES 2017: turnover, employment, capital expenditure and expenditure on energy, goods and services
- MES 2020: turnover, employment

Expectations in the MES

- The MES elicits expectations by asking
  - *Looking ahead to the 2018 calendar year, what is the approximate pound sterling value of turnover you would anticipate for this business in the following scenarios [Lowest, Low, Medium, High, Highest], and what likelihood do you assign to each scenario?*
- MES 2017: turnover, employment, capital expenditure and expenditure on energy, goods and services
- MES 2020: turnover, employment
- To get what we need
  - Convert scenario responses into 5 points on a CDF
  - Estimate the parameters of a lognormal distribution to fit these points via minimisation

GDP Turnover Employment

Sample selection

	2017			2020		
		Share			Share	
	N	N	TO	N	N	TO
All	8970	1	1	10014	1	1
(1) $\{y_{it}, l_{it}\}$ obs.	8541	0.95	0.99	9899	0.99	1.00
(2) $\mathbb{E}[y_{it+1}, l_{it+1}]$ obs.	6301	0.70	0.71	7007	0.70	0.71
(3) k_{it} obs.	7802	0.87	0.95	6373	0.64	0.91
(4) k_{it+1} obs.	5939	0.66	0.89	4258	0.43	0.84
Estimation sample (1-4)	4388	0.49	0.66	3127	0.31	0.58

Note: table shows the number and share of firms that comply with various sample selection criteria and the share of turnover these firms account for. Characteristics

Production function estimates

Full sample

	OLS	OP	LP	ACF	NPR
β_l	0.72	0.69	0.47	0.71	0.74
	(0.01)	(0.01)	(0.01)	(0.00)	(0.05)
β_k	0.28	0.26	0.28	0.28	0.22
	(0.01)	(0.08)	(0.03)	(0.00)	(0.04)
N Obs.	13763	13533	13763	13763	6941
N firms	6249	6249	6249	6249	6249

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications.

Translog Manufacturing Non-manufacturing

Production function estimates

2017 sample

	OLS	OP	LP	ACF	NPR
β_l	0.69	0.65	0.36	0.69	0.43
	(0.02)	(0.02)	(0.01)	(0.00)	(0.11)
β_k	0.30	0.29	0.30	0.29	0.10
	(0.02)	(0.12)	(0.02)	(0.00)	(0.07)
N Obs.	8039	7897	8039	8039	4075
N firms	4075	4075	4075	4075	4075

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications.

Including intermediates Value added

Production function estimates

2020 sample

	OLS	OP	LP	ACF	NPR
β_l	0.76	0.74	0.55	0.77	0.80
	(0.02)	(0.02)	(0.02)	(0.00)	(0.07)
β_k	0.26	-0.04	0.25	0.27	0.22
	(0.02)	(0.11)	(0.04)	(0.04)	(0.05)
N Obs.	5724	5636	5724	5724	2866
N firms	2866	2866	2866	2866	2866

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications.

Productivity dispersion

Productivity dispersion implied by full sample estimates

	OLS	OP	LP	ACF	NPR
All					
75/25 ratio	2.51	2.56	2.99	2.50	3.96
90/10 ratio	7.13	7.36	9.05	7.08	13.22
90/50 ratio	2.84	2.84	2.98	2.85	4.07
50/10 ratio	2.51	2.59	3.04	2.48	3.25
N	13840				

By sector By year

Today

Where we are, where we're going

- Theoretical framework showing how to use expectations data to recover production function parameters
- MC simulations showing relative strength of NPR estimator
  - Develop initialisation protocol to improve NPR precision
  - Experiment with other DGPs e.g. non-linear ω dynamics
- Empirical application on UK data
  - Understand implausibly low RTS observed in 2017
  - Examine industry heterogeneity at finer resolution
  - Analyse implied productivity
- Any suggestions welcome!

Today

Ichimura and Todd (2007) initialisation

- $y_{it} = \beta_1 + \beta_2 l_{it} + \beta_3 k_{it} + \beta_4 \mathbb{E}[y_{it+1}] + \beta_5 \mathbb{E}[l_{it+1}] + \beta_6 k_{t+1}$
- $\hat{\theta}_0 = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$

Back

Monte Carlo setup

- Following ACF
  - $Y = \min\{\beta_0 K_{it}^{\beta_k} L_{it}^{\beta_l} e^{\omega_{it}}, \beta_m M_{it}\} e^{\epsilon_{it}}$
  - $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$
  - $c_i(I_{it}) = \frac{\phi_i}{2} I_{it}^2$
  - $\omega_{it-b} = \rho^{1-b}\omega_{it-1} + \xi_{it}^A$, $\omega_{it} = \rho^b\omega_{it-b} + \xi_{it}^B$
  - $\ln W_{it} = 0.3 \ln W_{it-1} + \xi_{it}^W$

Back

Monte Carlo setup

- Following ACF
  - $Y = \min\{\beta_0 K_{it}^{\beta_k} L_{it}^{\beta_l} e^{\omega_{it}}, \beta_m M_{it}\} e^{\epsilon_{it}}$
  - $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$
  - $c_i(I_{it}) = \frac{\phi_i}{2} I_{it}^2$
  - $\omega_{it-b} = \rho^{1-b}\omega_{it-1} + \xi_{it}^A$, $\omega_{it} = \rho^b \omega_{it-b} + \xi_{it}^B$
  - $\ln W_{it} = 0.3 \ln W_{it-1} + \xi_{it}^W$

Back

- Firms' optimal decisions have an analytical solution
- Set $K_{i0} = 0$, simulate the model for 100 periods, use data from the last 10
- Compare OLS, LP, ACF, NPR performance across various DGPs

Measurement error in materials

Serially correlated wages and labour set at $t-b$

M Meas. Error	β_l	β_k	β_l	β_k
	OLS		LP	
0.0	0.402 (0.013)	0.966 (0.004)	-0.000 (0.003)	1.432 (0.028)
0.1	0.402 (0.013)	0.966 (0.004)	0.056 (0.004)	1.353 (0.027)
0.25	0.402 (0.013)	0.966 (0.004)	0.270 (0.007)	1.054 (0.023)
0.5	0.402 (0.013)	0.966 (0.004)	0.587 (0.009)	0.601 (0.018)
	ACF		NPR	
0.0	0.599 (0.009)	0.401 (0.021)	0.611 (0.048)	0.394 (0.363)
0.1	0.602 (0.009)	0.410 (0.020)	0.611 (0.048)	0.394 (0.363)
0.25	0.616 (0.009)	0.432 (0.018)	0.611 (0.048)	0.394 (0.363)
0.5	0.652 (0.008)	0.428 (0.015)	0.611 (0.048)	0.394 (0.363)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications.

Back

Measurement error in materials

Serially correlated wages and labour set at $t - b$

M Meas. Error	β_l	β_k	β_l	β_k
	OLS		LP	
0.0	0.402 (0.013)	0.966 (0.004)	-0.000 (0.003)	1.432 (0.028)
0.1	0.402 (0.013)	0.966 (0.004)	0.056 (0.004)	1.353 (0.027)
0.25	0.402 (0.013)	0.966 (0.004)	0.270 (0.007)	1.054 (0.023)
0.5	0.402 (0.013)	0.966 (0.004)	0.587 (0.009)	0.601 (0.018)
	ACF		NPR	
0.0	0.599 (0.009)	0.401 (0.021)	0.600 (0.002)	0.403 (0.006)
0.1	0.602 (0.009)	0.410 (0.020)	0.600 (0.002)	0.403 (0.006)
0.25	0.616 (0.009)	0.432 (0.018)	0.600 (0.002)	0.403 (0.006)
0.5	0.652 (0.008)	0.428 (0.015)	0.600 (0.002)	0.403 (0.006)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications. NPR results from initialisation at $\beta_l 0 = 0.55$ and $\beta_k 0 = 0.45$.

Back

Measurement error in materials

Optimisation error in labour and serially correlated wages

M Meas. Error	β_l	β_k	β_l	β_k
	OLS		LP	
0.0	0.294 (0.015)	0.909 (0.004)	0.357 (0.004)	0.881 (0.020)
0.1	0.294 (0.015)	0.909 (0.004)	0.373 (0.004)	0.863 (0.020)
0.25	0.294 (0.015)	0.909 (0.004)	0.444 (0.005)	0.780 (0.018)
0.5	0.294 (0.015)	0.909 (0.004)	0.593 (0.006)	0.593 (0.016)
	ACF		NPR	
0.0	0.608 (0.005)	0.374 (0.021)	0.605 (0.028)	0.410 (0.334)
0.1	0.610 (0.005)	0.383 (0.021)	0.605 (0.028)	0.410 (0.334)
0.25	0.616 (0.006)	0.416 (0.018)	0.605 (0.028)	0.410 (0.334)
0.5	0.634 (0.006)	0.447 (0.015)	0.605 (0.028)	0.410 (0.334)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications.

Back

Measurement error in materials

Optimisation error in labour and serially correlated wages

M Meas. Error	β_l	β_k	β_l	β_k
	OLS		LP	
0.0	0.294 (0.015)	0.909 (0.004)	0.357 (0.004)	0.881 (0.020)
0.1	0.294 (0.015)	0.909 (0.004)	0.373 (0.004)	0.863 (0.020)
0.25	0.294 (0.015)	0.909 (0.004)	0.444 (0.005)	0.780 (0.018)
0.5	0.294 (0.015)	0.909 (0.004)	0.593 (0.006)	0.593 (0.016)
	ACF		NPR	
0.0	0.608 (0.005)	0.374 (0.021)	0.600 (0.002)	0.402 (0.006)
0.1	0.610 (0.005)	0.383 (0.021)	0.600 (0.002)	0.402 (0.006)
0.25	0.616 (0.006)	0.416 (0.018)	0.600 (0.002)	0.402 (0.006)
0.5	0.634 (0.006)	0.447 (0.015)	0.600 (0.002)	0.402 (0.006)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications. NPR results from initialisation at $\beta_l 0 = 0.55$ and $\beta_k 0 = 0.45$.

Back

Optimisation error

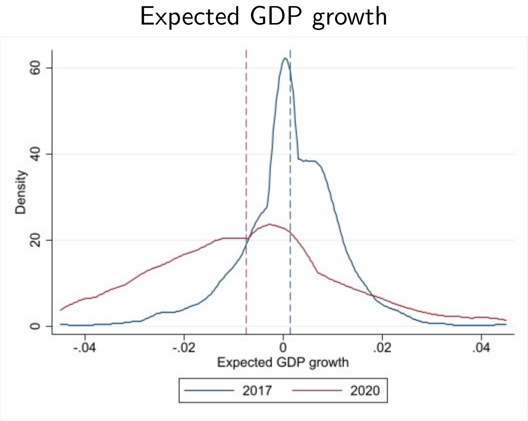
Optimisation error in labour and other inputs

Optim. Error	β_l	β_k	β_l	β_k
	OLS		LP	
i	-0.044 (0.016)	0.806 (0.006)	0.000 (0.004)	0.403 (0.016)
m	0.380 (0.006)	0.919 (0.002)	0.815 (0.004)	0.462 (0.028)
(i, m)	-0.044 (0.016)	0.806 (0.006)	0.636 (0.009)	0.396 (0.006)
	ACF		NPR	
i	0.367 (0.785)	0.404 (0.012)	0.601 (0.017)	0.400 (0.001)
m	0.688 (0.014)	0.350 (0.017)	0.600 (0.003)	0.400 (0.005)
(i, m)	-286.016 (3815.097)	-19.295 (508.348)	0.601 (0.017)	0.400 (0.001)

Note: 500 replications. True values of β_l and β_k are 0.6 and 0.4 respectively. Parentheses contain standard deviations which are calculated for the parameter estimates over the 500 replications. NPR results from initialisation at $\beta_l 0 = 0.55$ and $\beta_k 0 = 0.45$.

Back

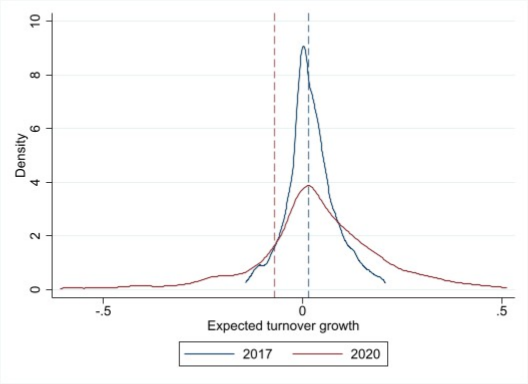
Aggregate expectations far more dispersed in 2020



Note: dashed lines denote means of 0.1% in 2017 and -0.7% in 2020. Sample sizes are 4268 in 2017 and 3127 in 2020. Back

Turnover expectations far more pessimistic in 2020

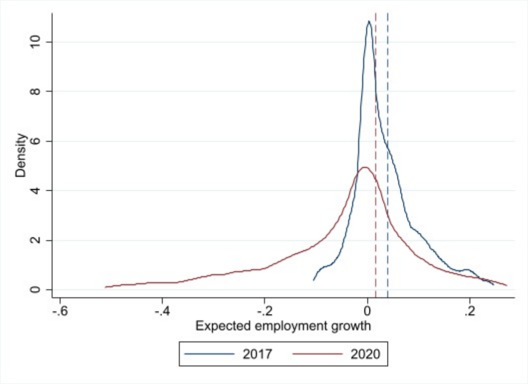
Expected turnover growth



Note: distribution is trimmed at the top and bottom 5%. Dashed lines denote means calculated across the entire distribution of 1.5% in 2017 and -7.1% in 2020. Sample sizes are 4388 in 2017 and 3127 in 2020. Back

Employment expectations slightly more pessimistic in 2020

Expected employment growth



Note: distribution is trimmed at the top and bottom 5%. Dashed lines denote means calculated across the entire distribution of 4.0% in 2017 and 1.7% in 2020. Sample sizes are 4388 in 2017 and 3127 in 2020. Back

Sample characteristics

2017	MES: est. samp.	MES: all
Age	25.22	23.45
Turnover (£th)	63.03	52.28
Employment	325	318
Capital (£th)	20.08	19.75
N firms	4388	7532
2020	**MES: est. samp.**	**MES: all**
Age	27.98	26.80
Turnover (£th)	44.86	34.17
Employment	297	206
Capital (£th)	25.19	16.13
N firms	3127	6373

Note: table shows mean values calculated over the firms indicated by ‘N firms’.

Back

Production function estimates

Full sample

	OLS	ACF	NPR
β_l	0.74	1.07	0.74
β_k	0.27	-0.13	0.20
N Obs.	14786	14786	7515
N firms	6786	6786	6786

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications. Back

Production function estimates

Manufacturing sample

	OLS	OP	LP	ACF	NPR
β_l	0.86	0.80	0.56	0.85	0.92
	(0.03)	(0.02)	(0.05)	(0.00)	(0.08)
β_k	0.28	0.28	0.26	0.28	0.22
	(0.02)	(0.05)	(0.00)	(0.00)	(0.07)
N Obs.	3452	3396	3452	3452	1734
N firms	1553	1553	1553	1553	1553

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications. Back Translog

Production function estimates

Manufacturing sample

	OLS	ACF	NPR
β_l	0.88	1.40	0.96
β_k	0.24	-0.33	0.19
N Obs.	3633	3633	1836
N firms	1646	1646	1646

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications. Back

Production function estimates

Non-manufacturing sample

	OLS	OP	LP	ACF	NPR
β_l	0.69	0.67	0.45	0.68	0.68
	(0.02)	(0.01)	(0.01)	(0.00)	(0.06)
β_k	0.27	0.27	0.27	0.26	0.21
	(0.02)	(0.02)	(0.02)	(0.00)	(0.04)
N Obs.	10311	10137	10311	10311	5207
N firms	4696	4696	4696	4696	4696

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications. Back Translog

Production function estimates

Non-manufacturing sample

	OLS	ACF	NPR
β_l	0.71	0.96	0.72
β_k	0.27	-0.11	0.18
N Obs.	11153	11153	5679
N firms	5140	5140	5140

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications. Back

Production function estimates

2017 sample

	OLS	OP	LP	ACF	NPR
β_l	0.39	0.39	0.37	0.38	0.32
	(0.02)	(0.01)	(0.02)	(0.38)	(0.09)
β_m	0.46	0.45	0.43	0.44	0.38
	(0.02)	(0.01)	(0.02)	(0.00)	(0.07)
β_k	0.14	0.01	0.14	0.13	0.06
	(0.01)	(0.07)	(0.02)	(0.00)	(0.06)
N Obs.	8347	7838	8347	8347	4235
N firms	4235	4235	4235	4235	4235

Note: dependent variable is log turnover. Parentheses contain standard errors. NPR standard errors calculated from 100 bootstrap replications.

Back

Production function estimates

2017 sample

	OLS	OP	LP	ACF	NPR
β_l	0.63	0.59	0.67	0.63	0.34
	(0.03)	(0.02)	(0.02)	(0.00)	(0.10)
β_k	0.33	0.32	0.69	0.33	0.46
	(0.02)	(0.04)	(0.13)	(0.00)	(0.09)
N Obs.	7679	7214	7679	7679	3976
N firms	3976	3976	3976	3976	3976

Note: dependent variable is log value added. Parentheses contain standard errors.

NPR standard errors calculated from 100 bootstrap replications.

Back

Productivity dispersion

Productivity dispersion implied by sector-specific estimates

	OLS	OP	LP	ACF	NPR
Manufacturing					
75/25 ratio	2.18	2.21	2.48	2.17	2.48
90/10 ratio	4.48	4.61	5.95	4.48	5.61
90/50 ratio	2.29	2.32	2.45	2.29	2.52
50/10 ratio	1.96	1.99	2.43	1.95	2.23
N			3466		
Non-manufacturing					
75/25 ratio	2.64	2.68	3.13	2.64	4.25
90/10 ratio	8.35	8.37	10.41	8.29	14.95
90/50 ratio	3.07	3.02	3.14	3.06	4.41
50/10 ratio	2.72	2.77	3.32	2.71	3.39
N			10374		

Back

Productivity dispersion

Productivity dispersion implied by year-specific estimates

	OLS	OP	LP	ACF	NPR
2017					
75/25 ratio	2.46	2.47	3.15	2.46	4.34
90/10 ratio	6.98	7.10	9.96	6.98	22.47
90/50 ratio	2.89	2.86	3.15	2.88	6.12
50/10 ratio	2.42	2.48	3.16	2.43	3.67
N			8062		
2020					
75/25 ratio	2.59	5.86	2.94	2.59	3.47
90/10 ratio	7.34	30.80	8.71	7.33	11.04
90/50 ratio	2.79	6.66	2.83	2.80	4.18
50/10 ratio	2.63	4.63	3.08	2.62	2.64
N			5778		

Back

Data citations

Office for National Statistics (2022). *Management and Expectations Survey, 2016-2020: Secure Access.* [data collection]. *3rd Edition*. UK Data Service. SN: 8557. University of West of

England, Bristol, Office for National Statistics, Virtual Microdata Laboratory (VML). (2022). *Annual Respondents Database X, 1998-2014: Secure Access.* [data collection]. *4th Edition*. Office for National Statistics, [original data producer(s)]. Office for National Statistics. SN: 7989. Office for National Statistics

(2023). *Annual Business Survey, 2005-2020: Secure Access.* [data collection]. *16th Edition*. UK Data Service. SN: 7451.