Issues in estimating production functions and capital stock

in a study of the relationship between migration and productivity

Jennifer C Smith

University of Warwick, Department of Economics

ONS UK Data Service UK Business Data User Conference

21 September 2021

Ultimate aim of the research

- To estimate the relationship between migration and productivity / productivity growth using micro data
- Original research for Migration Advisory Committee report "EEA migration in the UK: Final report", September 2018 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/741926/Final_EEA_report.PDF
- Find a large positive impact of migration on productivity
  - 1 pp increase in migrant share results in a 1.6% increase in TFP
  - Second report for MAC (2018) by Julian Costas-Fernandez: marginal migrant around 2.5 times as productive as UK-born workers (LFS 1998-2014)
  - p.58

Steps in the research

1. Choose data sets
2. Estimate production functions

   Involves

   - estimation of capital stock
   - choice of estimation method
3. Estimate impact of migration on TFP
   - Involves procedures intended to deal with endogeneity

Data

- Business data: ARDx 1998-2015
  - uses Annual Business Survey (ABS) and Business Register Employment Survey (BRES)
  - Includes production function data: value added, investment, materials purchases and employment
  - Data at reporting unit level. Holed, unbalanced panel: average 3.4 observations per ‘firm’
- Migration data: LFS
  - Migrant and native population shares, at Government Office Region (of work, where applicable) level
  - Merged by location of reporting unit

Data

- Deflators

Deflator for investment and capital stock

- “Detailed GFCF deflators” disaggregated by industry Section and asset type: ONS (2017), “Volume Index of Capital Services estimates to 2015”

Deflator for materials purchases and turnover

- 2-digit industry Division deflators for producer prices: ONS FOI release “Industry Level Deflators (Experimental), UK 1997 to 2015”

Deflator for value added

- “Gross value added price indices” at 1-digit industry Section level, variable VA_P, from “United Kingdom Basic Tables” from EUKLEMS database (September 2017 release)

Data

- Capital stock variable not included in ARDx
  - Perpetual inventory method used

    Martin (2002), Gilhooly (2009), Harris (2005), UK Data Service ARDx documentation (2018), Dey-Chowdhury (2008)
  - Use all asset types: land and buildings, vehicles, ‘plant and machinery’ = all other types of investment (ICT equipment, computer software and databases, R&D, mineral exploration and extraction, cultivated biological resources, artistic originals, other machinery and equipment)
  - Initial capital stock values at 2-digit level from VICS (Volume Index of Capital Services; based on IDBR universe of firms) are scaled to the sample of firms by sampled/total turnover, and apportioned to individual firms on the basis of ‘purchases’ (of energy, goods, materials and services; rarely missing, never negative)
  - Imputation: Capital expenditure for each asset type is imputed if missing, using employment-weighted average capital expenditure for that asset type by that firm, with employment interpolated for this purpose if missing, to ensure that all observations are attributed some capital expenditure and capital stock. After imputation, capital expenditure for an asset will be zero (only) if the firm has no capital expenditure data for that asset in any period.
  - Depreciation: Assumed constant depreciation rates are equal to the average depreciation rates that have been used in PIM calculations for sectoral aggregates by the ONS, obtained from (Martin, 2002)
  - Negative capital stock values at firm level can arise due to asset disposals
    - Negative values are reduced to a quite small number by aggregating across the 3 asset types
    - No further correction (“backfilling”) was used – I experimented with increasing all or just the first observation of capital stock by most negative observation but feared this was inducing rather than reducing measurement error.

Production function

log TFP *shock to value added*

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \xi_{it}$$

unobserved

- Log value added for firm i in year t is y_{it}.
- k_{it} is log capital input, l_{it} is log labour input.
- The random error affecting value added can be split into two components:
  - ω_{it} is log total factor productivity TFP (also known as multi-factor productivity or technical efficiency)
  - ξ_{it} is an idiosyncratic shock to value added, distributed as white noise

Production function estimation

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \xi_{it}$$

- I focus on whether migration impacted TFP, so need an estimate of ω_{it}.
- Firms' choices (e.g. materials purchases, investment) will be influenced by TFP. ω_{it} is an unobserved state variable driving firm i's decisions at time t.
- $\rightarrow$ endogeneity problem:
  - +/- productivity shocks will ↑/↓ firms' output and also ↑/↓ input demand.
- Olley and Pakes (1996), Levinsohn-Petrin (2003), and Ackerberg, Caves and Frazer (2006)
- Get round the endogeneity problem by using observed materials purchases as a (correlated) proxy for unobserved TFP ω_{it}.
- Additional assumptions – l_{it} freely chosen; k_{it} results from k_{it-1} and last period's investment decision i_{it-1}; materials purchases monotonically increasing with TFP – mean that capital is also a state variable, so materials purchases are influenced by capital and TFP: $m_{it} = f(k_{it}, \omega_{it})$. This function can be inverted to enable estimation of TFP since it is a function of observed materials purchases and capital stock: $\omega_{it} = f^{-1}(m_{it}, k_{it}) = \phi_{it}(m_{it}, k_{it})$, which can be approximated by a polynomial in $m_{it}$and k_{it}.

$$y_{it} = \beta_0 + \beta_l l_{it} + \phi_{it}(m_{it}, k_{it}) + \xi_{it}$$

- The estimate of β_l is consistent. A consistent estimate of β_k can be obtained under relatively mild assumptions.
  - The consistent estimate of β_k can be obtained by estimating $y_{it} - \widehat{\beta}_l l_{it} = \beta_0 + \beta_k k_{it} + g(\omega_{it-1}) + \xi^*_{it}$, where $g(.)$ can be estimated using a random walk assumption, or as a higher order polynomial, or nonparametrically; and error term ξ^*_{it} includes shocks to TFP: $\xi^*_{it} = \xi_{it} + \varepsilon_{it}$, where it's assumed TFP follows a first-order Markov process $\varepsilon_{it} = \omega_{it} - E(\omega_{it} | \omega_{it-1})$.
- So now can calculate $\quad tfp_{it} = y_{it} - \widehat{\beta}_l l_{it} - \widehat{\beta_k} k_{it}$

Production function estimation

Further issues

Measurement error in capital stock

- The perpetual inventory method of calculating capital stock has a good pedigree, but nevertheless it's likely it gives error ridden measurements $k_{it} = k_{it}^* + \varepsilon_{it}^k$ rather than true k_{it}^*.
- Anyone who has estimated production functions using micro data will be familiar with the surprisingly low resulting estimates of the capital coefficient β_k, suggesting downward bias. The labour coefficient can also be affected, and will tend to be upward biased since capital and labour choices are positively correlated so β_l may pick up impacts
- Collard-Wexler and De Loecker (2020) demonstrate that lagged investment i_{it-1} can be used as an instrumental variable for the potentially mis-measured capital stock (even if the measurement error is serially correlated), while maintaining the robustness to endogenous input choices of the control function approach.
    - requires $E(k_{it} i_{it-1}) \neq 0$ to ensure the instrument is orthogonal to the capital stock measurement error, and $E(\omega_{it} i_{it-1}) = 0$ to ensure the instrument is orthogonal to the productivity term

Production function estimation

Further issues

Attrition

- selection due to firm exit. The issue is that larger capital stocks might enable firms to survive, for given productivity, so selection on survival might negatively bias the capital coefficient β_k leading to mismeasurement of TFP.

1. Unbalanced panel helps reduce the problem
2. Could try to calculate exit from the data. However, the ABS survey samples smaller firms only intermittently, resulting in numerous holes in the panel and making inference of true exit from the presence or absence of firms difficult. The best available option is to define exit as the last period of a single spell (ending before the last sample period).
3. Also can use the ABS survey's categorisation of firms' status (variable `resptype`). I recorded exit when firms reported a 'part-year return due to death in year', 'ceased trading' or became 'dormant'.

- Olley and Pakes' (1996) econometric method to control for selection is fairly standard: in a first stage, a probit model of exit as a function of a polynomial in capital and investment is estimated, and then the predicted probabilities (interacted) are included in the final-stage nonlinear model to retrieve the consistent capital coefficient.

Estimating the migration-productivity relationship

Issues

Migration data

- Area level.
  - The underlying assumption is that all firms in an area experience the same increase in labour supply due to migration: migration shocks are changes in the local availability of migrant workers so they are common to all firms in an area.
  - Firm data on migrant employment captures migrant labour use, but this isn't the same as migration. So need an area-level measure.
  - Ideally TTWA level? Best available data source = LFS. For overall migration, TTWA may be just about acceptable in terms of sample size. For any more disaggregated measure of migration, sample size requires a larger regional dimension. I choose GOR, preferring GOR for work where this is available and GOR of residence where not.

Estimating the migration-productivity relationship

Endogeneity

- Migrants may (do) favour high-productivity regions → upward bias in estimates of impact of migration.
  - Migration into an area will depend on productivity shocks since these are positively related to labour demand.
  - The positive correlation of migrant flows with unobserved productivity shocks means that OLS estimates will give an estimate of the true impact of migration on productivity that is upward-biased.
- Use a 'shift-share' type instrument.
  - Exploit the observation that immigrants tend to settle in regions with larger immigrant populations (Bartel, 1989; Lalonde and Topel, 1991; Altonji and Card, 1991; Peri, 2016, and Dustmann, Schönberg and Stuhler, 2016, discuss this approach)
- To improve predictive power, use a variant of the Card (2001) instrument which divides migrants by country of origin.
- To reduce endogeneity, measure the pre-sample initial migrant distribution across regions (here: average settlement pattern during 1993-1997). The instrument is predicted migrant share during the sample, which is constructed using this pre-sample distribution and migrant growth rates…
  - In case this is insufficiently distant from the 1998-2015 sample period, a second variant of the instrument is constructed that uses the settlement pattern of *only those migrants who arrived before 1990* (and who were observed during 1993-1997, prior to the sample). Shocks that might have influenced this *pre-1990 settlement pattern* are even more remote from within-sample shocks, and serial correlation should have diminished very substantially.
- … Also to reduce endogeneity, 'leave-one-out': instead of using national migrant growth rates to inflate the pre-sample migrant distribution, the leave-one-out instrument uses growth rates across all regions except the one whose migrant stock is being predicted.

Migration and productivity

Higher migrant shares appear associated with higher firm productivity.

A 1 percentage point increase in the migrant share results in a 1.6 per cent increase in TFP.

Dependent	**Migrant share coefficient**		
variable			
Total factor		1.57***	
productivity (TFP)		(0.06)	

Table reports the migrant share coefficient from a regression of firm-level TFP on regional migrant share and firm-level control variables *firm* age and *foreign owned*, over 1998-2015 (121,278 observations; 47,426 firms). Migrant share is instrumented with a Bartik-type leave-one-out instrument at region level , distinguishing 3 country of birth groups (EU13+, NMS, NonEU). TFP is estimated using the Collard-Wexler and De Loecker method that controls for both endogeneity and measurement error in capital stock. Standard errors in parentheses are clustered at firm level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.