Migration, Productivity and Firm Performance
A Report for the Migration Advisory Committee

Jennifer Smith, University of Warwick[1]

11 September 2018

Abstract

The links between migration and firm performance are analysed using data from UK firms between 1998 and 2015 combined with regional measures of migration. Firm productivity is used as a primary measure of firm performance, because of its economic and policy importance. A key focus is total factor productivity (TFP) – productivity after accounting for the quantities of labour and capital employed – since TFP is thought to capture underlying growth. The analysis carefully assesses the best measure of TFP, providing robustness checks using a wide variety of estimates.

Contents

[1] I would like to thank Timo Leidecker for excellent research assistance.

1. Aim

This report aims to present evidence on the empirical links between migration and firm productivity. Firm level data from UK administrative surveys are combined with regional migration measures calculated from the Labour Force Survey. The empirical approach is fairly simple, adopting what is known as a “reduced form” approach: the empirical links are examined via econometric regressions of measures of firm performance on the key migration variable, with other variables added to the regression to control for observed and unobserved characteristics of firms, their industry, and the region in which they operate. Care is taken to provide estimates that are as robust as possible. Instrumental variable techniques are used to control for the endogeneity of migration – the fact that migrants might be attracted to areas where firm performance is good. Extensive alternative estimates are provided in relation to total factor productivity, which is considered a key indicator of firm performance but is unobserved and must be estimated.

Robustness is essential for studies involving migration effects. The appendix to this document presents a very wide range of production function estimates, and a broad selection of TFP estimates calculated from these production functions are used to investigate the relationship between migration and productivity, attempting to further ensure robustness by using a variety of specifications for that relationship. As discussed in the appendix, the finding of a positive migration-productivity relationship is found to be robust to different estimation techniques. It is also robust to productivity estimates derived from aggregate or industry production functions.

2. Key results: Relationship between migration and productivity

To investigate links between firm performance and migration, firm-level estimates of productivity, as measured by the logarithm of total factor productivity TFP, are regressed on migrant share (measured at region-year level). These estimates are based on data covering firms in Great Britain during 1998-2015. Migrant share is the share of migrants in the population at Government Office Region level, calculated using LFS data in the Secure Lab of the UK Data Service. Migrants are defined as those not born in UK or Ireland.

Firm productivity is regressed on migrant share in the firm’s region. The idea is that migration into the firm’s local area will alter the labour supply available to the firm. The alteration in labour supply could be simply a matter of greater quantity of labour, if migrants and natives are perfect substitutes and work with the same efficiency (quality). It is also possible that the arrival of migrants will alter the quality of labour available to the firm – either by enhancing native labour input, if migrants and natives are complements, or because migrant productivity is higher than natives’.

The simple message of the results is that an increased migrant share in the region in which firms operate is correlated with higher productivity. This general result is robust to the inclusion of dummy variables for region, dummy variables for industry, region time trends and industry time trends, and firm fixed effects (dummy variables for each firm). It is robust to adjusting standard errors for heteroscedasticity by clustering at firm level or region level. Because TFP has to be estimated, for robustness many different TFP measures have been used as alternate dependent variables, varying according to the production function estimates, and the positive TFP-migration relationship is consistent across these different measures.

Log TFP is calculated as log real value added less log real capital stock and log employment multiplied by their respective estimated elasticities. These elasticities are the parameter estimates for *k* and *l* from estimated production functions. Production function estimates can be found in Appendix 4.

Table 2.1: Relationship between migrant share and productivity

Dependent variable	Migrant share coefficient			Observations
	(1)	(2)	(3)	(firms)
Total factor productivity (TFP)	1.57^{***} (0.06)	4.21^{***} (0.17)	1.24^{***} (0.28)	121,278 (47,726)
Region dummies		Y	Y	
Industry dummies		Y	Y	
Firm fixed effects			Y	

Notes to Table 2.1
Table 2.1 reports the migrant share coefficient from a regression of firm-level TFP on regional migrant share, over 1998-2015. Migrant share is instrumented with a Bartik-type leave-one-out instrument at region level. TFP is estimated using the Collard-Wexler and DeLoecker method that controls for both endogeneity and measurement error in capital stock. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The numbers in Table 2.1 relate to estimates using the preferred (CWDL) estimate of TFP. They indicate that a 1 percentage point increase in the share of migrants is associated with TFP that is 1.57 per cent higher (column (1)). Estimates including region and industry dummies show that a firm's productivity is about 4 percent higher compared to the average over time for all firms in the same region and industry, as migrant share increases in that firm's region (column (2)). Estimates additionally including region and industry time trends suggest a very similar relationship: as migrant share increases faster than average in a firm's region, that firm's TFP tends to be about 4 per cent higher compared to the average for all firms in the same region and industry (see column (3) in Appendix 3 Table A3.1). Estimates including firm fixed effects indicate that this positive migration-TFP relationship comes at least in part from higher TFP growth for a given firm when migration within its region increases (column (3) of Table 2.1 and also column (5) of Appendix 3 Table A3.1).

In principle, the use of instrumental variables should enable the estimates to be given a causal interpretation, but the literature tends to be cautious about doing so, primarily because it is difficult to find a fully convincing instrument. As is well known, the reason it is desirable to instrument the migrant share variable is that migration into a region might be affected by productivity shocks that also influence TFP. This endogeneity problem can bias the estimated migrant share effect. Several instruments were investigated. All are Bartik-type instruments that extrapolate the pre-sample (1993-97) regional migrant distribution, estimated from the Labour Force Survey, using national (Great Britain) growth rates for migrant groups defined by country of birth. Country of birth groups used are those relevant to the MAC commission on EEA migration, namely "EU13+" (the 13 pre-2004 EU members, excluding UK and Ireland, plus the members of the EEA – Norway, Iceland and Liechtenstein – plus Switzerland, which is not technically a member of the EEA), "NMS" (newer EU member states that joined in or after 2004), and "NonEU" (non-EEA countries). In all cases here, the instruments pass normal statistical tests for instrument validity. A "leave-one-out" instrument is preferred: it was found to be adequately robust to serial correlation in region-level shocks. A further step to enhance exogeneity by restricting the initial migrant share to those that arrived many years before the sample start gave very similar results.

Two observable control variables, controlling for firm characteristics that might influence productivity, are included in all regressions: firm age and a foreign ownership dummy.

Firm age	Firm age is year of observation minus birth year. The birth year variable is taken directly from ARDx.
Foreign owned	The foreign ownership dummy takes value 1 if the firm has an ultimate owning company that is not based in Britain. The foreign ownership dummy takes value 0 otherwise, which includes firms whose ultimate owning company is British or which are independent entities. The foreign ownership dummy is derived from ARDx variable `ultfoc`.

Note: “firm” is a shorthand for “ruref” (which represents a Reporting Unit in UK business data).

Robustness of results to different methods of estimating TFP estimates was checked. TFP estimates were used from production functions estimated by OLS, IV instrumenting capital to control for measurement error, the control function approaches of Olley and Pakes, Levinsohn and Petrin and Ackerberg, Caves and Frazer, Wooldridge's one-step GMM version of Levinsohn-Petrin, the IV-plus-ACF control function approach of Collard-Wexler and De Loecker using either lagged investment or energy and water purchases as instruments for capital, and most of these re-estimated imposing constant returns to scale by constraining capital and labour coefficients to sum to unity. Reassuringly, results were very similar in pattern no matter what method of estimating production functions and TFP was used. In addition to TFP estimates from a production function in which capital and labour elasticities are assumed to be the same for all firms, all specifications are also estimated allowing production function coefficients to vary by industry.

Box 1: Bias in TFP estimates: its impact and how to avoid it

The need for careful choice of estimator for production functions is emphasized clearly when attempting to capture heterogeneity across firms in characteristics that relate to capital or labour. This arises because TFP can be seriously mismeasured in ways that correlate with capital or labour, if methods are used to estimate production functions that result in biased estimates of capital or labour elasticities.

These predicted impacts of TFP bias are borne out empirically in estimates using the varied production function estimates reported in Appendix 4 to estimate TFP and investigate heterogeneity in the relationship between TFP and migration. Methods that overestimate TFP for capital-intensive firms could attribute to those firms excessive benefit from migration, which might affect estimates of heterogeneous migration affects across categories for capital intensity, capital-labour ratio, and TFP. Methods that underestimate TFP for labour-intensive firms will do likewise, and might also affect estimates ordered by firm size when this is measured by employment. In contrast, for characteristics unrelated to capital or labour, such as region or industry, and perhaps international trade, R&D or size as measured by turnover, migration and TFP relationships should be unaffected.

TFP will tend to be overestimated for capital-intensive firms by methods that understate the capital coefficient. This capital coefficient understatment can be due to failure to take account of measurement error in capital stock, which affects OLS and the Levinsohn-Petrin and Ackerberg-Caves-Frazer control function approaches. The Olley-Pakes control function approach seems somewhat less affected; as suggested by Collard-Wexler and DeLoecker (2016), this is probably due to that method's use of lagged investment (rather than materials inputs) as a proxy for unobserved TFP. IV methods that instrument capital stock, including simple IV and Collard-Wexler-DeLoecker's hybrid IV-control function approach, should not be subject to this overestimation of TFP for capital intensive firms.

OLS, and IV methods that just instrument capital stock, will overstate the labour coefficient due to their failure to treat TFP as endogenous. TFP is positively correlated with the dependent variable (output) and with inputs chosen after TFP is realized (labour), so OLS will understate TFP for labour-intensive firms (as well as understating TFP for capital-intensive firms as discussed in the preceding paragraph). The control function approaches of Olley-Pakes, Levinsohn-Petrin, Ackerberg-Caves-Frazer and Collard-Wexler-DeLoecker should not understate TFP for labour-intensive firms.

In summary, having used a wide range of methods to estimate production functions, the evidence seems to support the use of the Collard-Wexler-DeLoecker method which controls for endogeneity of TFP and measurement error in capital stock. This method is relied upon for the key results relating to TFP reported in the main body of this paper.

3. Variation across regions and industries in the relationship between migration and TFP

All regions exhibit a positive relationship between migration and the productivity of firms in that region – see Table 3.1. The largest positive relationship between migration and productivity appears to be in the North East, with Scotland also exhibiting a relatively large positive relationship, and the smallest in London.

The relationship between firm productivity and migration is also positive in all industries. The largest positive relationship appears in Agriculture.

Table 3.1: How does the migrant share effect vary by region?

Dependent variable: Log TFP

Region	Migrant share effect (1)	Coefficient (s.e.) (2)	
North East (reference region)	9.51	9.51^{***}	(1.14)
North West	4.82	-4.69^{***}	(1.20)
Yorkshire and Humberside	6.81	-2.70^{*}	(1.24)
East Midlands	4.39	-5.12^{***}	(1.19)
West Midlands	5.91	-3.60^{**}	(1.21)
East England	4.61	-4.90^{***}	(1.21)
Greater London	3.36	-6.15^{***}	(1.16)
Rest of South East	5.40	-4.11^{***}	(1.19)
South West	6.08	-3.43^{**}	(1.27)
Wales	---	-1.77	(1.61)
Scotland	7.02	-2.49^{**}	(1.26)

Notes to Table 3.1

Table 3.1 reports the migrant share coefficient and interaction terms from a 2SLS regression of firm-level TFP on regional migrant share interacted with a 11-category region dummy, using 121,384 observations on 47,773 firms over 1998-2015. TFP is estimated using the Collard-Wexler and DeLoecker method that controls for both endogeneity and measurement error in capital stock. Migrant share is instrumented with a Bartik-type leave-one-out instrument at region level, distinguishing 3 country of birth groups (EU13+, NMS, NonEU). Standard errors (s.e.) in parentheses are clustered at firm level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The equation includes region and industry dummies. No separate migration effect is reported for Wales; the estimate indicates the effect for Wales is insignificantly different from the reference region (North East).

Table 3.2: How does the migrant share effect vary by industry?

Dependent variable: Log TFP

Industry Section	Migrant share effect (1)	Coefficient (s.e.) (2)		Observations per industry (3)
AB (reference industry)	7.27	7.27^{***}	(0.85)	1,473
C	4.00	-3.27^{***}	(0.84)	45,537
DE	2.61	-4.66^{***}	(0.99)	1,585
F	4.66	-2.61^{**}	(0.86)	6,326
G	3.88	-3.39^{***}	(0.84)	18,727
H	4.72	-2.55^{**}	(0.86)	6,003
I	3.93	-3.34^{***}	(0.84)	1,968
J	4.31	-2.96^{***}	(0.85)	5,244
K	4.63	-2.64^{*}	(1.03)	246
L	5.38	-1.89^{*}	(0.88)	1,976
M	5.00	-2.27^{**}	(0.84)	9,349
N	4.62	-2.65^{**}	(0.85)	7,931
P	3.64	-3.63^{***}	(0.89)	5,877
Q	3.62	-3.65^{***}	(0.87)	4,281
RST	4.26	-3.01^{***}	(0.86)	5,401

Notes to Table 3.2

Table 3.2 reports the migrant share coefficient and interaction terms from a 2SLS regression of firm-level TFP on regional migrant share interacted with a 15-category industry dummy, using 121,384 observations on 47,773 firms over 1998-2015. Observations per industry are reported for the estimation sample. TFP is estimated using the Collard-Wexler and DeLoecker method that controls for endogeneity and for measurement error in capital stock. Migrant share is instrumented with a Bartik-type leave-one-out instrument at region level, distinguishing 3 country of birth groups (EU13+, NMS, NonEU). Standard errors (s.e.) in parentheses are clustered at firm level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The equation includes region and industry dummies.

Industry Section descriptions

Industry	Description
AB	Agriculture; forestry and fishing; Mining and quarrying
C	Manufacturing
DE	Electricity, gas and water supply
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food services
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support activities
P	Public administration and defence; compulsory social security
Q	Education
RST	Human health and social work; Other service activities; Activities of households as employers

Appendix 1: Data sources

Firm level data

The focus of this study is the relationship between migration and the performance of 'firms' (which will be defined precisely below). Firm data is drawn from the ARDx dataset, which is compiled from two sources, each the result of separate administrative surveys of UK businesses. The Annual Business Survey (ABS) provides financial and accounting variables, and the Business Register Employment Survey (BRES) covers employment.

The data span 18 years, from 1998 to 2015. Firms are included in the sample only if they make a "full return" to the administrative surveys. Production function data on value added, capital stock, materials purchases and employment is available for just over 336,000 firm-years. The average number of observations on a firm is fairly low at 3.4, reflecting the fact that the dataset includes a large number of very small firms for whom the sampling scheme entails they are observed only

The Inter-Departmental Business Register, which includes 100% of firms operating in Great Britain, forms the sampling frame for the datasets underlying ARDx. Large firms are surveyed each year, but only a subset of smaller firms is included in the sample. One quarter of firms employing less than 10 people, and half of firms employing 10-25 people, are sampled each year. The sampling probability of a firm employing 25-100 people ranges from about 50% to 100%. If a small firm is sampled they are not sampled again for at least 3 years. So there is a full balanced panel of information on large firms, but for smaller firms the panel is holed and unbalanced.

Firms are matched to migration data using the region of the reporting unit. While it is acknowleged that there are some multi-establishment firms where production takes place in local units located separately from the reporting unit, the financial data from ABS relate to the reporting unit, and there is no information about the performance of local units beyond their employment. In order to use geographical information on local units, it would be necessary to make assumptions on how financial variables should be apportioned among local units. Richard Harris has done so using employment.

Labour force survey data

The Labour Force Survey is used to estimate migrant and native population shares, using a sample including all survey respondents aged between 16 and 70. The sample is weighted to obtain population totals using the most recent vintage of population weights. Region is region of work where applicable, and region of residence otherwise, using Government Office Region definitions.

Other data

Deflators

Deflator for investment and capital stock

"Detailed GFCF deflators" disaggregated by industry Section and asset type are taken from the following source: ONS (2017), "Volume Index of Capital Services estimates to 2015" available from https://www.ons.gov.uk/file?uri=/economy/economicoutputandproductivity/output/datasets/capitalservicesestimates/estimatesto2015/rftunlinked.xls. An aggregate deflator is constructed weighting asset type deflators by the relevant productive (nominal) capital stocks, also obtained from the VICS dataset.

Deflator for materials purchases and turnover

2-digit industry Division deflators for producer prices were taken from the ONS FOI release "Industry Level Deflators (Experimental), UK 1997 to 2015". Simple averaging was used to combine 3-digit Group-level data into 2-digit Division-level data where necessary. These industry (division) level deflators are a mixture of product and implied industry (division) level deflators, produced by allocating and aggregating industry product deflators across industries, based on their use of these products in line with the supply-use framework. The product level deflators are based on the GDP(O) sources catalogue. https://www.ons.gov.uk/file?uri=/economy/inflationandpriceindices/adhocs/006718industryleveldeflatorsexperimentaluk1997to2015/experimentalindustryleveldeflators.xls. PPI data are missing for industry groups 42 Civil engineering and 43 Specialised construction activities, so the PPI for these groups is taken to be the same as 41 Construction of buildings; and the simple average PPI over 62 Computer programming and consultancy, 63 Information service activities, 68 Real estate activities and 69 Legal and accounting activities is used to replace missing PPI data for 64 Financial ex insurance and pension, 65 Insurance, reinsurance and pension, and 66 Auxiliary to financial and insurance.

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. Where it is necessary to combine Sections (A and B, R-S and T), the components are weighted by their respective "Gross value added, volume (2010 prices)" (variable VA_Q) from the same source. http://www.euklems.net/TCB/2017/UK_output_17i.xlsx

Capital stock

The ABS survey does not ask firms about their capital stock, so a measure of capital stock has to be derived from other available data. Estimation of the capital stock is not straightforward and involves several decisions where judgement and investigation are required to ensure the estimates make best use of available data and generate reasonable results. In constructing the estimates of capital stock used in production function estimation, I draw upon major previous work by Martin (2002) and Gilhooly (2009) and am informed by others such as Harris (2005) and some STATA code available in UK Data Service

(2018) ARDx documentation. Differences and similarities between my method and others are described and explained below.

The capital stock is derived using the perpetual inventory method (for details on how this method is used by the UK Office for National Statistics, see Dey-Chowdhury, 2008). The perpetual inventory method is based on the 'law of motion' for capital K

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

which can be integrated forward from the initial conditions for capital stock K_{i0} to give an expression for capital stock in period t in terms of initial capital stock and subsequent capital expenditure flows I_{it} given the sequence of depreciation rates $\{\delta_{it}\}$. If the depreciation rate is constant (as assumed here) and equal to δ, the perpetual inventory calculation of the capital stock is as follows:

$$K_{it} = \sum_{s=0}^{t-1} (1-\delta)^{s} I_{it-s} + (1-\delta)^{t} K_{i0} \tag{2}$$

The ABS survey includes capital expenditure by three types of asset: land and buildings, vehicles, and a category known as 'plant and machinery' that includes 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). ABS records gross investment (acquisitions less disposals, not accounting for depreciation, which is actually termed "net" capital expenditure in the ABS survey).

At firm level, investment is frequently negative, since it is common in a single year for a firm's disposals of a particular asset type to exceed their acquisitions. Using the perpetual inventory method to calculate capital stock can therefore lead to the problem that the estimated capital stock appears negative. I minimise the impact of negative capital expenditure by aggregating across asset types for each firm, after imputation (discussed below) but before the perpetual inventory calculation.

An estimate of the initial capital stock K_{i0} – the capital stock relating to the first period a firm is observed – has to be derived from external data. I use Office for National Statistics estimates of "Volume index of UK capital services" (VICS). These estimates are still called 'experimental' although they have been constructed for some years now and are normally the preferred variable on which to base firm-level estimates (e.g. Martin, 2007).

The estimation of initial capital stock for micro data gives rise to the same issues as for aggregate data: the initial capital stock estimates can dominate investment flows in determining current capital stock, and the extent to which they do depends negatively on sample length and assumed asset lifespan (see, for example, Burda and Severgnini, 2008). Fortunately, the UK has a long time series of investment data which the Office for National Statistics use to estimate capital stock: UK data extend back to 1948, which means the impact of assumptions about initial capital stock on figures for capital stock at industry and asset level during our sample period (1998-2015) will be low. The challenge then is to use data aggregated at industry-asset level most efficiently to represent initial firm-level capital stock.

ONS VICS capital services data relate to the universe of firms. These data have to be rescaled to match the proportion of firms that appear in the ARDx sample. I do this on the basis of the proportion of turnover summed over the universe of firms that is accounted for by the sample, for each 2-digit industry. This measure of turnover is available for the universe of firms: it is sourced from the IDBR at the time firms are selected for the ABS survey. (Two alternative scaling variables were available. It would be possible to scale on the basis of IDBR employment (used by Attanasio, Pacelli and dos Reis, 2003); but the proportion of employment in the sample relative to the population is lower than for turnover, which would exacerbate the problem of negative estimated capital stock for given subsequent negative investment flows. A further possibility, suggested in ARDx documentation, would be to use observed Section-level capital expenditure (in the ARDx sample) relative to ONS Section-level capital expenditure from the VICS survey. This faces the serious problem that both VICS and observed Section-level capital expenditure can be negative for some asset types, so scaling induces negative values for initial capital stock.

The allocation of scaled 2-digit industry capital stock to a firm is done on the basis of that firm's share of industry 'purchases' (total purchases of energy, goods, materials and services). 'Purchases' is used because it is rarely missing and never negative (contrasting with capital expenditure, for example, which can be negative if there are net disposals).

During the sample, each firm's capital stock (of a particular asset) is increased by capital expenditure (for that particular asset), less assumed depreciation (for that particular asset). Capital expenditure is imputed if missing. The imputation consists of replacing missing capital expenditure for an asset type with employment-weighted average capital expenditure for that asset type by that firm. For the purposes of full imputation of capital expenditure, to obtain continuous capital stock estimates for the duration of that firm's time in the sample, total employment is interpolated and extrapolated, with missing end-points replaced by a small positive value (0.1) to ensure that all observations are attributed some capital expenditure and capital stock (in accordance with the procedure in UK Data Service, 2018, ARDx documentation). Hence, 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. Assumed depreciation rates are equal to the average depreciation rates used in PIM calculations for sectoral aggregates by the ONS, obtained from Martin (2002): buildings 0.02; vehicles 0.20; other 0.06; aggregate 0.11.

I experimented with corrections for negative capital stock values (which can arise if a firm's capital expenditure is negative due to disposals exceeding acquisitions, to an extent that exceeds initial capital stock). However, I do not use these manipulated capital stock values, since I found that in practice the procedures did not add many observations to the sample (cases with negative capital stock are excluded once logarithms are taken), and the corrected variables led to a worsening of results which I suspected might be due to induced measurement error. The relevant procedures are known as 'backfilling' (Gilhooly, 2009; UK Data Service, 2018, ARDx documentation). Where a firm's capital stock is estimated to be negative, two different methods are suggested to 'correct' this error: *missing capex* or *rebase initial stock*. *missing capex* raises all capital stock observations by an amount equal to the most-negative capital stock. *rebase initial stock* alters only the initial capital stock observation by this amount. The idea behind *missing capex* is that negative capital stock is caused by one or more 'lumps' of missing investment. *rebase initial stock* is based on the alternative supposition that negative capital stock is caused by the initial value for capital stock being too low.

References and Citations: Data

Jäger, Kirsten (2017), "EU KLEMS Growth and Productivity Accounts 2017 release - Description of Methodology and General Notes", The Conference Board.

Office for National Statistics (2017), *Business Register and Employment Survey, 2009-2016: Secure Access* [data collection], 6th Edition, UK Data Service: SN 7463, http://doi.org/10.5255/UKDA-SN-7463-6

Office for National Statistics (2018), *Annual Business Survey, 2008-2016: Secure Access* [data collection], 9th Edition, UK Data Service: SN 7451, http://doi.org/10.5255/UKDA-SN-7451-9

Office for National Statistics, Virtual Microdata Laboratory (VML), and University of the West of England, Bristol (2017), *Annual Respondents Database X, 1998-2014: Secure Access* [data collection], 4th Edition, Office for National Statistics [original data producer(s)], UK Data Service: SN 7989, http://doi.org/10.5255/UKDA-SN-7989-4

Office for National Statistics Social Survey Division, Northern Ireland Statistics and Research Agency and Central Survey Unit (2018), *Quarterly Labour Force Survey, 1992-2017: Secure Access* [data collection], 12th Edition, UK Data Service: SN 6727, http://doi.org/10.5255/UKDA-SN-6727-13

Office for National Statistics Social Survey Division, Northern Ireland Statistics and Research Agency and Central Survey Unit (2010-2018), *Quarterly Labour Force Survey, October-December 1992 to October-December 2017* [data collections], UK Data Service various SNs: 5889-8326, http://doi.org/10.5255/UKDA-SN-5889-1}-\url{http://doi.org/10.5255/UKDA-SN-8326-1

References: Capital Stock

Attanasio, Orazio P, Pia Pacelli and Isabel Reduto dos Reis (2003), "Investment patterns in UK manufacturing establishments", UK Data Archive Study Number 6698 – Capital Stock Dataset, 1979 – 2005: Secure Data Service Access.

Burda, Michael and Battista Severgnini (2008), "Solow Residuals without capital stocks", SFB Discussion Paper 2008-040.

Chen, Xi and Tatiana Plotnikova (2014), "Measuring the initial capital stock: a generalized framework", *Procedia Economics and Finance* 14 147-153.

Dey-Chowdhury, Sumit (2008), “Perpetual inventory method”, *Economic and Labour Market Review* 2 (9) 48-52.

Gilhooly, Bob (2009), “Firm-level estimates of capital stock and productivity”, *Economic and Labour Market Review* 3 (5) 36-41.

Harris, Richard (2005), ``Deriving measures of plant-level capital stock in UK manufacturing, 1973-2001'', Final Report to DTI.

Martin, Ralf (2002), ``Building the capital stock'', CeRiBA mimeo.

ONS (2017) “Volume Index of Capital Services estimates to 2015” https://www.ons.gov.uk/file?uri=/economy/economicoutputandproductivity/output/datasets/capitalservicesestimates/estimatesto2015/rftunlinked.xls

UK Data Service (2018), “ARDx capital stock: data manager guide” and “ARDx capital stock user guide”, UK Data Archive Study Number 7989 – Annual Respondents Database X: Secure Access.

Appendix 2: Methods

Estimating TFP

TFP, or total factor productivity, captures influences on a firm's productivity that are not due to the amount of observable inputs, such as capital or labour, employed by that firm. TFP embodies the contribution to value added of factors unobservable in economic data. TFP is represented by variable ω_{it} in the following production function:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \xi_{it} \tag{3}$$

Equation (3) shows firm i 's production function at time t. y_{it} is log value added, 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_{it} (also known as multi-factor productivity or technical efficiency). ξ_{it} is an idiosyncratic shock to value added, distributed as white noise.

Control function approach

The 'control function' approach, also termed the 'structural proxy estimator', is a popular method of controlling for endogeneity in capital and labour by constructing a proxy for unobservable TFP from observable variables. Key papers in developing the approach include Olley and Pakes (1996), Levinsohn-Petrin (2003), and Ackerberg, Caves and Frazer (2006).

Olley and Pakes (1996) pointed out that both output and input choices might be affected by factors that are unobserved by the econometrician but observable by the firm. In particular, firms' choices will be influenced by TFP. Thus ω_{it} is an unobserved state variable driving firm i's decisions at time t. Olley and Pakes (1996) focused on firms' investment decisions, whereas Levinsohn and Petrin (2003) focused on firms' purchases of intermediate inputs. Both of these will be influenced by TFP, assuming the firm can observe this. The problem for researchers is that we cannot observe the TFP variable driving firms' decisions; the problem, therefore, is one of omitted variables in the production function to be estimated.

The existence of variable ω_{it} which can be observed by firms but not in data available to researchers gives rise to an endogeneity problem. Positive productivity shocks will raise firms' output and lead to greater input demand. Negative shocks lead to a decline in output and demand for input. The positive correlation between the observable input levels and the unobservable productivity shocks is a source of bias when estimating production functions using techniques such as OLS.

Olley and Pakes (1996) pointed out that under fairly minimal assumptions, the problematic unobserved variable can be modelled using observable variables: the observable variables can act as a 'proxy' for the unobserved component.

To proceed, the models need to make assumptions about the statistical processes of the labour and capital variables. The assumptions made are generally regarded as intuitively appealing. Olley and Pakes (1996) and Levinsohn and Petrin (2003) assume that labour is a “free” variable – that is, it is freely chosen by firms each period after the productivity shock is observed, and thus labour does not follow a dynamic process.

The capital stock follows a dynamic process. Changes in the capital stock are assumed to depend on investment decisions taken last period, and last period's capital stock:

$$k_{it} = \kappa\left(k_{it-1}, i_{it-1}\right) \tag{4}$$

This structure implies that the capital stock is uncorrelated with this period’s innovation in the productivity process (ξ_{it} in equation (3)). Capital is a fixed factor of production, affected only by past values of unobserved (log) total factor productivity ω. The assumption that ensures this is that log TFP evolves according to a Markov process:

$$\omega_{it} = E\left(\omega_{it} \middle| \Omega_{it-1}\right) + \varepsilon_{it} = E\left(\omega_{it} \middle| \omega_{it-1}\right) + \varepsilon_{it} = g\left(\omega_{it-1}\right) + \varepsilon_{it} \tag{5}$$

where Ω_{it-1} is the information set at $t-1$ and ε_{it} is the shock to unobserved (log) productivity (log TFP). ω acts as an unobserved state variable affecting the firm’s decisions, including those concerning investment and purchases of intermediate inputs.

With that structure and under those assumptions, the firm’s decisions about the amount of intermediate inputs to purchase will be a function of the two state variables, capital and TFP:

$$m_{it} = f\left(k_{it}, \omega_{it}\right)$$

where m_{it} is the natural log of purchases of intermediate inputs (materials). The same structure would apply to decisions about investment, i_{it} : this is also a function of those two state variables. Here I focus on materials, following Levinsohn and Petrin (2003), using a technique paralleling that of Olley and Pakes (1996) which, instead, involved investment. The Olley-Pakes formulation has some disadvantages. It requires that investment be strictly positive, in order that the function f be invertible. As they acknowledge, due to asset disposals this is quite often not the case. A further problem with investment is that its demand might be ‘lumpy’, rather than smooth, in the face of adjustment costs. Levinsohn and Petrin (2003) pointed out that materials purchases are positive for the vast majority of firms, and are less likely to be lumpy.

A proxy for the unobserved productivity shock ω_{it} can be developed by assuming that m_{it} (Levinsohn-Petrin) is monotonically increasing in ω_{it} . Because of the timing assumptions embodied , materials purchases are orthogonal to the observed and unobserved state variables at t :

$E\left[m_{it} \left| (k_{it}, \omega_{it}) \right.\right] = 0$. Purchases depend only on past values of capital and TFP. So the purchases function can be inverted to give the relationship between the unobserved state variable, the observed state variable and the firm's observed purchases.

$$\omega_{it} = f^{-1}(m_{it}, k_{it}) \tag{6}$$

The production function can then be written as

$$y_{it} = \beta_0 + \beta_l l_{it} + \phi_{it}(m_{it}, k_{it}) + \xi_{it} \tag{7}$$

where $\phi_{it}(m_{it}, k_{it}) = \beta_k k_{it} + f^{-1}(m_{it}, k_{it}) = \beta_k k_{it} + \omega_{it}$.

Equation (7) can be estimated by OLS using a higher-order polynomial to approximate the unknown functional form of inverted purchases function ϕ. Second-, third- or fourth-order polynomials in m and k have variously been used in the literature, sometimes also including interactions between these variables. This is now more common than the Levinsohn-Petrin suggestion of using locally weighted (quadratic least squares) regression. Hellerstein and Neumark (2007) show that – although in principle this choice could affect the consistency of production function estimates, and in particular the ability to identify the separate effects of inputs and firm-specific unobservables – empirically the choice of order of polynomial appears to matter very little, with results being almost unchanged with different specifications.

Estimation of equation (7) is able to consistently identify effects of labour (and any other 'free' variables) on productivity, but not capital since it appears in ϕ not only on its own account but also as a part of the proxy for unobserved TFP.

So a second stage is needed. Under the assumptions set out above, that ω_{it} follows a first-order Markov process and capital (investment) does not respond to unforeseen innovations in TFP, $\varepsilon_{it} = \omega_{it} - E\left(\omega_{it} \left| \omega_{it-1} \right.\right) = \omega_{it} - \omega_{it}$, the capital coefficient can be consistently estimated using a consistent estimate of ω_{it} in the following:

$$\begin{aligned} y_{it}^{*} &= \beta_0 + \beta_k k_{it} + \omega_{it} + \xi_{it} \\ &= \beta_0 + \beta_k k_{it} + E\left(\omega_{it} \left| \omega_{it-1} \right.\right) + \xi_{it}^{*} \\ &= \beta_0 + \beta_k k_{it} + g\left(\omega_{it-1}\right) + \xi_{it}^{*} \end{aligned} \tag{8}$$

where y_{it}^{*} is value added net of labour's contribution, $y_{it}^{*} = y_{it} - \beta_l l_{it}$, and the error term ξ_{it}^{*} includes shocks to TFP: $\xi_{it}^{*} = \xi_{it} + \varepsilon_{it}$. Using the estimates of ϕ_{it}, the polynomial expression, from the first stage, a consistent estimate of ω_{it} can be defined, since $\omega_{it} = \phi_{it} - \beta_k k_{it}$. The following expression uses the additional assumption that $g(\cdot)$ follows a random walk (it is alternatively possible to leave $g(\cdot)$ unspecified and approximate using a higher-order polynomial as in the first stage, or estimate nonparametrically).

$$\begin{aligned} y_{it}^{*} &= \beta_0 + \beta_k k_{it} + g\left(\phi_{it-1} - \beta_k k_{it-1}\right) + \xi_{it}^{*} \\ &= \beta_0 + \beta_k\left(k_{it} - k_{it-1}\right) + \phi_{it-1} + \xi_{it}^{*} \end{aligned}$$

Log TFP can be calculated

$$tfp_{it} = y_{it} - \beta_l l_{it} - \beta_k k_{it} .$$

Ackerberg, Caves and Frazer (2015) (ACF) proposed a correction to the Levinsohn-Petrin (2003) method to deal with the possibility that the effects of labour and materials in the production function cannot be separately identified. In the model, employment and purchases decisions are modelled as if taken simultaneously by the firm, after the productivity shock is realised. ACF proposed to change timing assumptions so the labour input was selected after capital is determined but before the realisation of the productivity shock. In practice, this means that the labour coefficient of the production function can no longer be consistently identified in the first stage regression. The first stage regression is now used simply to remove the white noise error from the production function, and the labour coefficient is then identified along with the capital coefficient in the second stage. The second stage of the ACF procedure involves GMM estimation based on a moment condition that stems from the assumption (common also to Olley-Pakes and Levinsohn-Petrin) that unobserved productivity ω follows a Markov process:

$$\omega_{it} = \omega_{it-1} + \xi_{it}$$

Combined with the first step of the ACF procedure this implies the moment condition $E\left(\xi_{it} k_{it}\right) = 0$. The ACF version of the control function approach has been widely used. However, Mollisi and Rovigatti (2017) noted potential instability due to variation of results with starting values, which suggests caution in some applications.

Wooldridge (2009) showed that the Olley-Pakes and Levinsohn-Petrin proposals can be implemented in a single-step generalised method of moments (GMM) estimation method which overcomes identification problems and improves standard errors. Mollisi and Rovigatti (2017) suggested an enhanced set of moment conditions in a Wooldridge-type GMM procedure, to improve robustness and efficiency. However, Mollisi and Rovigatti (2017) provided Monte Carlo evidence that these non-linear GMM models can also give biased estimates in empirical applications, with results varying substantially depending on optimization starting points. From a practical point of view, unfortunately the Mollisi-Rovigatti estimator itself cannot be used here as it only works with Stata versions higher than the 14.0 available in the UK Data Service Secure Lab.

Attrition and firm exit

Olley and Pakes (1996) addressed a second issue in the estimation of production functions using firm-level data, namely selection (or attrition) 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. This problem is often considered adequately solved by the use of an unbalanced panel of firms, as here; Olley and Pakes themselves and subsequent authors have confirmed this. To investigate any remaining impact of attrition I investigate exit defined in two ways. One method calculates exit from observed 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). A second definition of exit uses the ABS survey's categorisation of firms' status (variable `resptype`): firms are considered to exit if they report a 'part-year return due to death in year', 'cease trading' or become 'dormant'. The Olley-Pakes 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 capital coefficient.

The size of the capital coefficient in firm-level production functions

Recently, attention has been paid and solutions proposed to the issue that the capital coefficient in firm-level production functions appears to be seriously underestimated using many popular techniques. Underestimating the capital coefficient would lead to capital intensive firms appearing more productive than they really are, and potentially biased estimates of TFP.

Collard-Wexler and De Loecker (2016) (henceforth CWDL) suggest that standard estimates of the capital coefficient are downward biased because capital is measured with error in firm-level data. CWDL also provide supportive Monte Carlo evidence.

If capital is measured with error, then we observe k_{it}^{*} rather than the true capital stock k_{it}:

$$k_{it}^{*} = k_{it} + \varepsilon_{it}^{k}$$

where a star indicates error-ridden data available to the researcher. Non-starred true values are assumed to be observable by the firm, but not the researcher. It is assumed that ε_{it}^{k} is classical measurement error in the capital stock, uncorrelated with the true capital stock k_{it}, so $E\left(\varepsilon_{it}^{k}\right) = 0$, but ε_{it}^{k} can be serially correlated within firms (which makes sense given that capital is constructed using historical information).

CWDL suggest that investment can be used as an instrument to deal with error in the observed capital stock. Statistically, the CWDL method requires $E\left(k_{it} i_{it}\right) \neq 0$ and $E\left(\varepsilon_{it}^{k} i_{it}\right) = 0$. Intuitively, investment satisfies the first of these requirements for being a good instrument, since it will be correlated with the capital stock. To be a valid instrument, investment must be uncorrelated with the capital measurement error (the second statistical condition). Intuitively, this is possible if the source of measurement error in capital is accumulated errors in depreciation (rather than errors in the investment data used to calculate capital stock via the perpetual inventory method). Depreciation is undoubtedly a substantial source of measurement error in capital stock. It is normal, as here, to use depreciation rates that do not vary across firms, over time, over capital stock vintages, and over types of asset. These non-firm-specific and non-time-dependent depreciation rates are bound to be a source of error in estimated capital stock. In contrast, current investment data, used in perpetual inventory method calculation of capital stock, is likely to be relatively well measured by firms' reports of capital expenditure on various asset types.

References: Estimating TFP

Ackerberg, Daniel A, Kevin Caves and Garth Frazer (2015), "Identification properties of recent production function estimators", *Econometrica* 83 (6) 2411-2451.

Collard-Wexler, Allan, and De Loecker, Jan (2016), "Production function estimation with measurement error in inputs", mimeo.

Levinsohn, James and Amil Petrin (2003), "Estimating production functions using inputs to control for unobservables", *Review of Economic Studies* 70 (2) 317-341.

Marshak, Jacob and W H Andrews (1944), "Random simultaneous equations and the theory of production", *Econometrica* 12 143-205.

Mollisi, Vincenzo and Gabriele Rovigatti (2017), "Theory and Practice of TFP Estimation: the Control Function Approach Using Stata", CEIS Tor Vergata Research Paper Series 15 (2) No 399.

Olley, Steven and Ariel Pakes (1996), "The dynamics of productivity in the telecommunications equipment industry", *Econometrica* 64 (6) 1263-1297.

Van Beveren, Ilke (2012), "Total factor productivity estimation: a practical review", *Journal of Economic Surveys* 26 (1) 98-128.

Wooldridge, Jeffrey M (2009), "On estimating firm-level production functions using proxy variables to control for unobservables", *Economics Letters* 104 (3) 112-114.

Instrumental variables

A key aim of this paper is to analyse correlations between firm productivity (and other aspects of firm performance) and changes in labour supply due to migration. These correlations will be able to be viewed as causal, and as impacts of an increase in migrant labour supply, if those changes in migrant labour supply can be treated as exogenous.

Labour supply changes due to migration are measured by migration into the firm's geographical area. 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 (as in Mitiratonna, Orefice and Peri, 2014). Firms can respond differently to such changes in labour supply, depending on their particular characteristics.

Migration into an area will depend on productivity shocks since these are positively related to labour demand. Productivity shocks are not observed and form part of the error term in the productivity equation. 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.

So it is necessary to instrument migration flows to identify the component that represents labour supply changes and to remove confounding effects from that part of migration flows that is attributable to productivity-shock-driven labour demand changes.

Like many migration studies, the instruments used in this paper are ‘shift-share’ (or ‘enclave-based’) instruments which 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). Again, this effectively defines the relevant labour market as region-based. (Alternatively, it would be possible to use differences in the concentration of immigrants across skill (education-experience groups) (Borjas, 2003). However, this ‘skill-cell’ approach involves judgement about appropriate skill groups, and has been criticised for identifying only relative effects (Dustmann, Schönberg, and Stuhler, 2016).)

To improve predictive power, the particular instruments used here are variants of the Card (2001) instrument which divides migrants by country of origin, rather than using the geographic distribution of all migrants as in the earlier work of Altonji and Card (1991). This improvement exploits the fact that migrants tend to locate near previous migrants from the same country of origin. Reasons for new migrants settling in areas in which their origin-country migrant concentration is higher include cheaper information flows (about facilities such as housing, for example), cheaper job search (via word of mouth and contacts), and higher wellbeing due to settlement alongside similar others.

The variable to be instrumented is region-year-varying migrant share $Mshare_{rt} = M_{rt} / \left(M_{rt} + N_{rt} \right)$.

Calculation of the instrument, $Mshare_{rt}$, involves two steps. The first step is to make a prediction of current migrant stocks at region-year level, M_{rt}, on the basis of the pre-sample regional migrant distribution and subsequent national migrant population growth rates. The instrument is easiest to explain by starting with the early version used by Altonji and Card (1991) that does not differentiate migrants by country of origin. Regional migrant stocks M_{rt} can be predicted as:

$$M_{rt} = M_{r0} \frac{M_t}{M_0} = \frac{M_{r0}}{M_0} M_t = \lambda_{r0} M_t \tag{9}$$

where M_{r0} is the number of migrants in region r at time 0, M_t is the number of migrants in the country as a whole at time t and M_0 is the number of migrants in the country as a whole at time 0. M_t/M_0 measures national-level growth in the stock of migrants between periods 0 and t, emphasising that the instrument extrapolates the pre-sample regional migrant distribution using national migrant population growth rates. It is also useful to rearrange the expression to emphasise a slightly different interpretation: $\lambda_{r0} = M_{r0}/M_0$ is the fraction of migrants at time 0 (pre-sample) who were (working or residing) in area r, which the instrument uses to allocate contemporaneous migrant stocks M_t across regions.

Taking account of country of birth gives the following predicted migrant stocks in region r at time t:

$$M_{rt}^{c} = \sum_{c} \frac{M_{r0}^{c}}{M_{0}^{c}} M_{t}^{c} = \sum_{c} \lambda_{r0}^{c} M_{t}^{c} \tag{10}$$

where M_{r0}^{c} is the number of migrants from country c in region r at time 0, and $\lambda_{r0}^{c} = M_{r0}^{c}/M_{0}^{c}$ is the fraction of all migrants from country c who were (working or residing) in area r at time 0.

In the second step of the calculation of the instrument for migrant share $Mshare_{rt}$, the predicted migrant stocks are used to calculate predicted migrant share. Differentiating migrants by country of birth group, this migrant share instrument is calculated as follows:

$$Mshare_{rt}^{c} = \frac{M_{rt}^{c}}{M_{rt}^{c} + N_{r0}} \tag{11}$$

The use of predicted migrant stocks takes account of the fact that the distribution of migrants across regions is affected by region-specific demand shocks. These shocks will also influence the distribution of natives across regions, so the pre-sample regional distribution of natives N_{r0} is used in calculation of the predicted migrant share.

The shift-share instrument used here is a ‘Bartik’-type instrument, differing from Bartik’s (1991) original in using country-of-birth rather than industry. Bartik (1991) suggested a within-region by-industry shift-share instrument for regional employment. The instrumental variable he used was predicted employment, taking an initial pre-sample distribution of employment across industry-region cells and extrapolating it forward on the basis of national employment growth for the relevant industry. Like Bartik’s original, the instrument used in this paper avoids contamination from within-sample region-specific shocks by using national growth rates, and the combination of the initial region-country-of-birth distribution and subsequent country-of-birth growth should mean the instrument correlates well enough with actual region-country-of-birth values. As Goldsmith-Pinkham, Sorkin and Swift (2018) put it, the initial industry shares identify the supply shock, and the industry growth rates boost instrument relevance.

To further reduce potential instrument endogeneity, a 'leave-one-out' instrument is constructed, as recommended by Goldsmith-Pinkham, Sorkin and Swift (2018) and also used by Wozniak and Murray (2012) and Hunt (2017): the area's own inflows are removed from the national inflow rate to reduce endogeneity to local conditions. Instead of using national 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. Summing migrant growth over all but one region means that growth rate will be free of influence from the left-out region's specific productivity shocks, so exogeneity should be improved. The leave-one-out version of the instrument for region-year varying migrant stock using variation over country of birth group can be written:

$$M^c_{r(-r)t} = \sum_c M^c_{r0} \frac{M^c_{(-r)t}}{M^c_{(-r)0}} = \sum_c \frac{M^c_{r0}}{M^c_{(-r)0}} M^c_{(-r)t} = \sum_c \lambda^c_{r(-r)0} M^c_{(-r)t} \tag{12}$$

where $M^c_{(-r)t}$ is the stock of migrants from country c at time t, measured over the whole economy excluding region r. Using this, the leave-one-out version of the instrument for region-year varying migrant share, using variation over country of birth group, is:

$$Mshare^c_{r(-r)t} = \frac{M^c_{r(-r)t}}{M^c_{r(-r)t} + N_{r0}} \tag{13}$$

Amior and Manning (2017) point out that it is likely that productivity or other labour demand shocks are serially correlated, leading to a potential correlation between current migrant flows and the initial distribution of immigrants. Similarly, Goldsmith-Pinkham, Sorkin, and Swift (2017) argue that, because a shift-share instrument is a function of national inflow rates and regional migrant shares, it might not be exogenous in the sense of satisfying the exclusion restriction required for a valid instrument, if the shares are correlated with persistent unobserved regional conditions, even if the national inflow rates are unrelated to those conditions. Borjas (1999) also noted that the exclusion restriction may be violated if local demand shocks are serially correlated, leading to correlation between the immigrant shares used in the construction of the instrument and subsequent demand shocks.

If the pre-sample distribution of migrants across regions is not independent of (orthogonal to) subsequent demand changes and productivity shocks in those regions, this would render the instrument ineffective since it would not distinguish only supply changes.

This is addressed here in two ways. The first method aims to ensure that the instrument is constructed using settlement patterns that are sufficiently lagged. The main instrument used here is based on the average settlement pattern during 1993-1997, prior to the start of the sample. 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. The benefits of using earlier arrivals depends on their not moving as freely as new arrivals in response to later shocks, which seems a reasonable assumption. In the full 1998-2015 sample, pre-1990 arrivals account for nearly 94% of all migrants observed during 1993-1997, suggesting that the extra care to exclude later arrivals from the pre-sample distribution might have limited empirical impact.

References: Instrumental variables

Altonji, Joseph G and David Card (1991), ”The effects of immigration on the labor market outcomes of less-skilled natives” 201--234 in eds John M Abowd and Richard B Freeman *Immigration, trade, and the labor market*, University of Chicago Press.

Amior, Michael and Alan Manning (forthcoming), “The persistence of local joblessness”, *American Economic Review*.

Bartel, Ann (1989), “Where do the new US immigrants live?”, *Journal of Labor Economics* 7 (4) 371-391.

Bartik, Timothy J (1991), *Who Benefits from State and Local Economic Development Policies?*, Kalamzoo, MI: W E Upjohn Institute for Employment Research.

Borjas, George J (1999), “The economic analysis of immigration”, 1697-1760 in eds Orley Ashenfelter and David Card, *Handbook of Labor Economics*, Vol 3A, Elsevier.

Dustmann, Christian, Tommaso Frattini and Ian P Preston (2013), “The effect of immigration along the distribution of wages”, *Review of Economic Studies* 80 (1) 145-173.

Dustmann, Christian, Uta Shönberg and Jan Stuhler (2016), “The impact of immigration: why do studies reach such different results?” *Journal of Economic Perspectives* 30 (4) 31-56.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2018), “Bartik instruments: what, when, why and how?”, NBER Working Paper 24408.

Hunt, Jennifer (2017), “The impact of immigration on the educational attainment of natives”, *Journal of Human Resources* 52 (4) 1060-1118.

Lalonde, Robert J and Robert H Topel (1991), “Immigrants in the American labor market: quality, assimilation and distributional effects”, *American Economic Review* 81 (2) 297-302.

Peri, Giovanni (2016), “Immigrants, productivity and labor markets”, *Journal of Economic Perspectives* 30 (4) 3-30.

Wozniak, Abigail, and Thomas J Murray (2012), “Timing is everything: short-run population impacts of immigration in US cities”, *Journal of Urban Economics* 72 (1) 60-78.

Appendix 3: Robustness of the migration-TFP relationship to alternative TFP estimates

The tables in Appendix 3 report estimates of the relationship between migration and TFP using a wide range of alternative TFP estimates.

- Table A3.1 repeats the estimates of the migration-TFP relationship for the Collard-Wexler-DeLoecker TFP estimate, as reported in Table 2.1 in Section 2, to compare this with the migration relationship to TFP when TFP is calculated from alternative production function estimates. Each line of Table A3.1 refers to the relevant production function estimates , which can be found in Appendix 4.
- Table A3.2 presents estimates using only the measure of TFP derived from OLS estimates (a measure corresponding to the production function estimates in Table A4.2, column 1), to show in more detail how the migration-TFP relationship varies with different econometric modelling specifications.

Table A3.1: Relationship between migrant share and TFP: migrant share coefficient, using different estimates of TFP

Dependent variable: Log TFP

	Migrant share coefficient					TFP estimation method	Observations
	(1)	(2)	(3)	(4)	(5)		(firms)
	1.57***	4.21***	3.98	2.70***	8.29***	Collard-Wexler-DeLoecker	121,278
	(0.06)	(0.17)	(2.26)	(0.46)	(1.87)	*Table A4.6 column (1)*	(47,726)
						(repeats Table 2.1 row 1)	
	1.57***	3.51***	5.46*	1.24***	5.09***	OLS	399,107
	(0.02)	(0.12)	(2.56)	(0.28)	(1.95)	*Table A4.2 column (1)*	(99,252)
	1.51***	4.08***	3.45	2.59***	7.84***	IV (investment)	139,222
	(0.05)	(0.16)	(2.16)	(0.44)	(1.70)	*Table A4.3 column (1)*	(55,861)
	1.44***	3.93***	3.01	2.47***	7.75***	IV (investment) imposing CRS	139,222
	(0.05)	(0.16)	(2.18)	(0.44)	(1.70)	*Table A4.3 column (3)*	(55,861)
	1.66***	3.61***	13.90***	1.77***	7.97***	Olley-Pakes	251,110
	(0.04)	(0.12)	(2.82)	(0.28)	(1.98)	*Table A4.4 column (1)*	(80,389)
	1.82***	2.57***	11.56***	1.90***	7.82***	Levinsohn-Petrin	336,413
	(0.04)	(0.12)	(2.82)	(0.27)	(1.75)	*Table A4.5 column (1)*	(98,608)
	2.92***	5.84***	9.40***	4.59***	8.50***	Ackerberg-Caves-Frazer (ACF)	173,818
	(0.08)	(0.23)	(2.78)	(0.48)	(1.62)	*Table A4.6 column (1)*	(71,155)
Region dummies		Y	Y	Y	Y		
Industry dummies		Y	Y	Y	Y		
Region time trends			Y		Y		
Industry time trends			Y		Y		
Firm fixed effects				Y	Y		

Notes: CRS: constant returns to scale. Standard errors in parentheses are clustered at firm level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period 1998-2015. The 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*. Migrant share is instrumented with a Bartik-type leave-one-out instrument at region level, distinguishing 3 country of birth groups (EU13+, NMS, NonEU).

Table A3.2a: OLS estimates of migrant share effect on an OLS estimate of ln(TFP) from Table A4.2, column (1)

Dependent variable: Log TFP

	(1) No controls	(2) Controls	(3) Region cluster	(4) Firm cluster	(5) Controls, region cluster	(6) Controls, firm cluster
Migrant share	1.74***	1.59***	1.74***	1.74***	1.59***	1.59***
	(0.02)	(0.02)	(0.04)	(0.16)	(0.04)	(0.13)
Firm age		0.001***			0.001***	0.001***
		(0.0002)			(0.0003)	(0.0004)
Foreign owned		0.40***			0.40***	0.40***
		(0.005)			(0.01)	(0.03)
Observations	339449	339107	339449	339449	339107	339107
Number of firms	99382	99252	99382	99382	99252	99252
R^2	0.027	0.049	0.027	0.027	0.049	0.049

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period 1998-2015.

The table reports results from regressions of firm-level TFP on regional migrant share and firm-level control variables *firm* age and *foreign owned*.

Table A3.2b: IV and IV-FE estimates of migrant share effect on an OLS estimate of ln(TFP) from Table A4.2, column (1)

Dependent variable: Log TFP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Migrant share	1.57***	1.57***	1.57***	3.51***	3.51***	5.46	5.46*	1.19***	1.24***	5.09**
	(0.02)	(0.11)	(0.04)	(0.58)	(0.12)	(3.73)	(2.56)	(0.28)	(0.28)	(1.95)
Region dummies				Y	Y	Y	Y	Y	Y	Y
Industry dummies				Y	Y	Y	Y		Y	Y
Region time trends						Y	Y			Y
Industry time trends						Y	Y			Y
Firm fixed effects								Y	Y	Y
Standard errors clustered at:										
firm level			Y		Y		Y	Y	Y	Y
region level		Y		Y		Y				

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period 1998-2015.

The table reports the migrant share coefficient from regressions of firm-level TFP on regional migrant share and firm-level control variables *firm* age and *foreign owned*. Migrant share is instrumented with a Bartik-type leave-one-out instrument at region level, distinguishing 3 country of birth groups (EU13+, NMS, NonEU). 399,107 observations on 99,252 firms.

Appendix 4: Production function estimates

Issues in production function estimation

As discussed in the survey article by Van Beveren (2012), production function estimation involves several key issues: the correct functional form for the relationship F between output Y and inputs capital, labour and materials, $Y = F(K, M, L)$; omitted variables and potential endogeneity, and appropriate measures of inputs. The last three issues are addressed at greater length below. Like much of the literature, we assume the production function is Cobb-Douglas: $Y = AK^{b_k}M^{b_m}L^{b_l}$ so, using lower-case letters to denote natural logarithms, $y = \ln A + b^k k + b^m m + b^l l$. In practice, I estimate production functions using value added as the dependent variable (approximately, output less materials inputs), so the inputs of interest are capital and labour.

Description and summary of production function estimates

The tables in Appendix 4 report the results of regressions of log real value added *y* on log real capital stock *k* and and log employment *l*, to show the variation in estimated elasticities of value added with respect to capital and labour as the technique to estimate the production function is altered.

Table A4.1: Description of variables used in production function estimation

log real value added *y*	Firm-level value added variable `gvabasic` is obtained directly from ARDx in the Secure Lab, and is deflated by industry value added deflators from ONS.
log real capital stock *k*	A firm-level capital stock measure is calculated by a perpetual inventory method using ARDx data on capital expenditure by asset type, ONS estimates of depreciation rates, and initial capital stocks and deflators taken from ONS VICS data on value of capital services by industry and asset.
log employment *l*	Firm-level employment (IDBR measure) is taken directly from ARDx.

Note: The word "firm" is a shorthand for `ruref` and represents a Reporting Unit in UK business data.

Input coefficients in estimated production functions are in line with expectations based on previous results using firm-level data. The elasticity of value added with respect to labour is estimated at about 0.6 when the preferred Levinsohn-Petrin (2003) method is used. The elasticity of value added with respect to capital is about 0.06. As expected, OLS elasticities are biased upwards due to endogeneity of inputs: inputs are positively correlated with unobserved productivity shocks that also boost value added. Two firm-level variables are included as controls: whether the firm is foreign-owned and how many years the firm has been in existence. Value added is higher in reporting units that are part of a foreign-owned enterprise. Value added rises with firm age, which is included in simple linear form. As has previously been found (Mollisi and Rovigatti, 2015), adjusting for attrition (firm exit) makes little difference to production function estimates: compare the Olley-Pakes method with and without an attrition correction. The estimated capital elasticity is in line with previous firm-level estimates. For example, estimates range from 0.05 to 0.07 in Hellerstein and

Neumark (2007) using WECD (Worker-Establishment Characteristics Database) and DEED (Decennial Employer-Employee Dataset). The capital elasticity using firm data is substantially lower than is typically found using aggregate or industry data, which might well reflect downwards attenuation due to measurement error in firm-level capital stock estimates (Griliches and Mairesse, 1998). Interestingly and somewhat reassuringly, Mairesse and Jaumandreu (2005) provide strong evidence based on French and Spanish micro data that removing errors due to the use of industry-level rather than firm-level price deflators makes no difference to estimated input elasticities, suggesting that other errors in variables or specification errors might be responsible for the low capital elasticities found in firm-level studies. If the capital elasticity is biased downwards this could lead to bias in estimated TFP, and will tend to overstate the productivity of capital-intensive firms. As a next step it will be interesting to apply the method of Collard-Wexler and De Loecker (2016) which combines instrumenting capital with investment and the control function approach. Their Monte Carlo simulations suggest this is able to successfully address measurement error in capital while still controlling for unobserved productivity shocks.

The Ackerberg, Caves and Frazer (2015) variant of Levinsohn-Petrin, which is designed to correct for possible remaining collinearity between labour and materials, which can prevent the coefficient on labour being identified in the first stage of the procedure and hence impact estimates of both labour and capital elasticities. The differences between the uncorrected Levinsohn-Petrin estimates in Table A4.5 column (1) and the ACF-corrected ones in Table A4.6 column (1) could indicate that this problem is biting. However, the Monte-Carlo study of Mollisi and Rovigatti (2015) suggests caution in the use of the ACF correction, since they find it susceptible to bias if starting values away from the true ones are used in the first-stage's optimisation procedure.

- All estimates are based on an unbalanced panel of firms covering 1998-2015.
- OLS, instrumental variables techniques instrumenting capital stock, and variants of the ‘control function’ approach designed to account for endogenous unobserved TFP, are used to estimate production functions. Some estimates control for selection (firm exit).
- Between 139,000 and 339,000 firm-year observations are used in the regressions, involving between 56,000 and 99,000 firms. Each firm is observed around 3 times on average during the 1998-2015 sample period, some firms being observed only once, and others in all of the 18 years. Certain specifications necessitate dropping observations without strictly positive capital expenditure, or with missing data on lagged variables: certain estimation methods require various combinations of lags of value added, capital stock, employment, capital expenditure or purchases of materials inputs.
- The choice of estimation method for production functions has a very large impact on estimated coefficients. The capital coefficient is particularly variable: in OLS estimates the elasticity of value added with respect to capital is a little under 0.2, and the labour elasticity is close to 0.8. Currently-popular measures to estimate production functions controlling for endogeneity, including Olley-Pakes (OP), Levinsohn-Petrin (LP) and Ackerberg-Caves-Frazer (ACF), show substantially lower capital elasticities of 0.09, 0.06 and 0.13 respectively. I suggest it is sensible to prefer methods that control for measurement error in capital stock – doing so produces capital coefficients above 0.3 (simple IV: 0.31; Collard-Wexler-DeLoecker (CWDL): 0.38).

- It is well known that OLS coefficients are biased due to the omission of the firm- and time-varying unobserved “total factor productivity”. Higher TFP will raise value added and will also prompt firms to use more labour and materials inputs, and undertake more investment. This simultaneity problem was first noted by Marshak and Andrews (1944). The OLS estimate of the labour coefficient will be biased upwards. The bias in the capital coefficient will depend on the statistical process followed by TFP. Following Olley and Pakes (1996), control function approaches to production function estimation often assume that TFP follows a Markov, or general autoregressive, process, and capital is assumed to be a pre-determined state variable, suggesting that the bias in the capital coefficient stemming from omitted TFP might not be as severe.
- The capital stock coefficient will be biased if it is measured with error, as it undoubtedly is in firm-level studies. Following the suggestion of Collard-Wexler and DeLoecker (2016) and controlling for that measurement error using lagged investment as an instrument restores the capital coefficient to intuitively more reasonable values (above 0.3).
- Selection (or attrition) bias due to firm exit can also lead to understatement of the capital coefficient (Olley and Pakes, 1996). It has been previously found that using an unbalanced panel goes a long way to control for that selection bias. Additional controls via selection (probit) models to control for exit are found to lead to little change in production function estimates. One exit measure used raises the Olley-Pakes capital coefficient from 0.09 to 0.11.
    - Estimates all imply some degree of diminishing returns to scale. For example, OLS capital (0.16) and labour (0.79) coefficients sum to 0.95. Estimates are presented below in which coefficients in some models are constrained to impose constant returns to scale, restricting the capital and labour coefficients to sum to unity. The main purpose is to produce estimates of TFP from such constant returns to scale production functions to test the robustness of later findings of the migration-productivity relationship to the imposition of constant returns. Production function parameters are estimated very precisely due to the large number of observations, and in no case is the restriction accepted at a normal level of confidence, even for specifications where the unrestricted coefficients “appear” to sum to something “close” to unity.
    - Preferred estimates for capital and labour elasticities, and hence TFP, are obtained from the following methods (capital coefficient + labour coefficient):
    - CWDL, a control function and IV hybrid approach using lagged material inputs to proxy for unobserved TFP and instrumenting for measurement error in capital using lagged investment: 0.38 + 0.58 = 0.96;
    - ACF, a popular control function approach using lagged material inputs to proxy for unobserved TFP: 0.13 + 0.81 = 0.94;
    - IV, instrumenting for measurement error in capital using lagged investment: 0.31 + 0.66 = 0.97.

Table A4.2: OLS and FE, without and with the imposition of constant returns to scale

Dependent variable: Log of real value added

	(1)	(2)	(3)	(4)
			Imposing constant returns to scale	
	OLS	FE	OLS	FE
k	0.16^{***}	0.02^{***}	0.15^{***}	0.12^{***}
	(0.002)	(0.002)	(0.002)	(0.001)
l	0.79^{***}	0.57^{***}	0.85^{***}	0.88^{***}
	(0.003)	(0.006)	(0.002)	(0.001)
Observations	339449	339449	339449	339449
Number of firms	99382	99382	99382	99382
Firm fixed effects	No	Yes	No	Yes
Imposes constant returns to scale	No	No	Yes	Yes
R^2	0.747	0.128		

Standard errors in parentheses are clustered at firm level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period: 1998-2015.
k is log real capital stock. l is log employment.
Specifications (3) and (4) impose constant returns to scale: capital and labour coefficients are constrained to sum to 1.

Table A4.3: Controlling for measurement error in capital

Dependent variable: Log of real value added

	(1) IV	(2) IV	(3) IV
k	0.31^{***}	0.35^{***}	0.30^{***}
	(0.004)	(0.005)	(0.004)
l	0.66^{***}	0.54^{***}	0.70^{**}
	(0.005)	(0.007)	(0.004)
Observations	139347	171348	139347
Number of firms	55916	70533	55916
Instrument	Lagged investment	Lagged energy and water purchases	Lagged investment
Imposes constant returns to scale	No	No	Yes
First stage coefficient	0.676^{***}	0.626^{***}	0.676^{***}
(s.e.)	(0.004)	(0.006)	(0.004)
R^2	0.764	0.724	

Standard errors in parentheses are clustered at firm level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period: 1998-2015.
k is log real capital stock. l is log employment.
All columns are estimated using 2SLS. Column (3) imposes constant returns to scale.

Table A4.4: Controlling for endogeneity using lagged investment to proxy for unobserved TFP, without and with controls for selection (firm exit)

Dependent variable: Log of real value added

	(1)	(2)	(3)	(4)
			Control for selection (firm exit)	
Method	Olley-Pakes	Olley-Pakes, no imputation in investment	Olley-Pakes	Olley-Pakes
k	0.09^{***}	0.09^{***}	0.09^{***}	0.11^{***}
	(0.001)	(0.003)	(0.006)	(0.002)
l	0.71^{***}	0.71^{***}	0.71^{***}	0.71^{***}
	(0.003)	(0.003)	(0.002)	(0.004)
Observations	139347	138823	141648	141576
Number of firms	55916	55753	56414	56401
Proxy variable	Lagged investment	Lagged investment excluding 524 imputed investment values	Lagged investment	Lagged investment
Control for firm exit	No	No	Yes	Yes
Exit variable			Exit(ARDx)	Exit(spell)

Bootstrapped standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period: 1998-2015.
k is log real capital stock. l is log employment. "Lagged investment" is the lag of log real capital expenditure.
Controls for selection ("Exit variable"). Exit(ARDx) =1 if ARDx indicates firm becomes dormant, ceases trading or gives a partial return due to death in year. Exit(spell) =1 in the last period of single spells ending before the sample end.
Estimates in columns (1) and (2) are generated by Mollisi and Rovigatti's (2017) *prodest* code. Estimates in columns (3) and (4) are generated by Yasar, Raciborski and Poi's (2012) *opreg* code.

Table A4.5: Controlling for endogeneity using materials inputs to proxy for unobserved TFP

Dependent variable: Log of real value added

	(1)	(2)	(3)
Method	Levinsohn-Petrin	Wooldridge (Levinsohn-Petrin) one-step	Levinsohn-Petrin
k	0.06***	0.06***	0.20***
	(0.001)	(0.001)	(0.003)
l	0.63***	0.61***	0.80***
	(0.001)	(0.001)	(0.003)
Observations	174449	174449	174449
Number of firms	71541	71541	71541
Proxy variable	Lagged materials inputs	Lagged materials inputs	Lagged materials inputs
Imposes constant returns to scale	No	No	Yes

Standard errors in parentheses are bootstrapped (column (1)) and clustered at firm level (columns (2) and (3)). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period: 1998-2015.

k is log real capital stock. l is log employment. "Lagged materials inputs" is the lag of log real materials purchases. Estimates in column (1) are generated by Mollisi and Rovigatti's (2017) *prodest* code. Column (2) uses Collard-Wexler and DeLoecker's (2016) code to estimate a Levinsohn-Petrin-type specification involving a 2-step method instrumenting labour with its lag and including lagged capital and materials. Column (3) uses the same method as column (2) but constrains the coefficients to impose constant returns to scale.

Table A4.6: Controlling for endogeneity using materials inputs to proxy for unobserved TFP and ensuring the labour coefficient is identified, without and with controlling for measurement error in capital stock

Dependent variable: Log of real value added

	(1)	(2)	(3)
Method	Ackerberg-Caves-Frazer	Ackerberg-Caves-Frazer	Collard-Wexler-DeLoecker
k	0.13^{***} (0.001)	0.12	0.38
l	0.81^{***} (0.000)	0.88	0.58
Observations	173973	173973	139020
Number of firms	71220	71220	55707
Proxy variable	Lagged materials inputs	Lagged materials inputs	Lagged materials inputs
Control for measurement error in capital stock	No	No	Yes
Instrument for capital stock			Lagged investment
Impose constant returns to scale	No	Yes	No

Standard errors in parentheses are clustered at firm level. No standard errors are currently available for estimates in columns (2) and (3). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Sample period 1998-2015.

k is log real capital stock. l is log employment. "Lagged materials inputs" is the lag of log real materials purchases. "Lagged investment" is the lag of log real capital expenditure.

Estimates in (1) are generated by Mollisi and Rovigatti's (2017) *prodest* code. Estimates in (2) are generated by a modified version of Collard-Wexler and DeLoecker's (2016) code. Estimates in (3) are generated by Collard-Wexler and DeLoecker's (2016) code