Office for National Statistics

Economic & Labour Market Review

Volume 5, Number 4

Edition No.: **April 2011**

Editor: **Graeme Chamberlin**

Office for National Statistics

April 2011

ISSN 1751–8334

A National Statistics publication

National Statistics are produced to high professional standards set out in the Code of Practice for Official Statistics. They are produced free from political influence.

About us

The Office for National Statistics

The Office for National Statistics (ONS) is the executive office of the UK Statistics Authority, a non-ministerial department which reports directly to Parliament. ONS is the UK government's single largest statistical producer. It compiles information about the UK's society and economy, and provides the evidence-base for policy and decision-making, the allocation of resources, and public accountability. The Director-General of ONS reports directly to the National Statistician who is the Authority's Chief Executive and the Head of the Government Statistical Service.

The Government Statistical Service

The Government Statistical Service (GSS) is a network of professional statisticians and their staff operating both within the Office for National Statistics and across more than 30 other government departments and agencies.

Contacts

This publication

For information about the content of this publication, contact ELMR team
Email: elmr@ons.gsi.gov.uk

Other customer enquiries

ONS Customer Contact Centre
Tel: 0845 601 3034
International: +44 (0)845 601 3034
Minicom: 01633 815044
Email: info@statistics.gsi.gov.uk
Fax: 01633 652747
Post: Room 1.101, Government Buildings,
Cardiff Road, Newport, South Wales NP10 8XG
www.ons.gov.uk

Media enquiries

Tel: 0845 604 1858
Email: press.office@ons.gsi.gov.uk

Copyright and reproduction

Economic & Labour Market Review

Volume 5, Number 4
April 2011

Contents

Regulars

Articles

Data and support

In brief

Special announcement regarding the last edition of Economic & Labour Market Review and the launch of the new ONS website

It was announced in last month's edition of Economic & Labour Market Review (ELMR) that this would be the final edition of the journal. The discontinuation of ELMR was timed to coincide with the lunch of ONS's new website on 30 April 2011, which would offer a new and improved way for ONS to deliver statistics and articles online. However, it was announced on 7 April 2011 that this launch will be postponed until 28 August 2011 as more time is needed for transition to a new web design and new business processes for publishing.

As a result, there will be one more edition of ELMR, published on Tuesday 10 May 2011. This is entirely due to a backlog of submissions which it was not possible to clear in the April edition. On line Labour Market Tables will be published on Monday 18 April 2011, and the Economic Tables provisionally on Friday 6 May 2011. These will be confirmed and declared on the Publication Hub shortly.

The final edition of ELMR will be purely article–based with most of the regular content removed. No new submissions are being accepted for this. A full list of all ELMR articles published since the start of the journal in January 2007 will also be provided.

Further information
ONS's announcement regarding the postponement of the new website's launch and other details about the ONS's Web Development Project can be found at:
www.ons.gov.uk/about/what-we-do/programmes-projects/web-development/index.html

Contact
web.development.programme@ons.gov.uk
elmr@ons.gov.uk

Census 2011: understanding how it works

The latest *Population Trends* from the Office for National Statistics marks Census Day on 27 March 2011 with a special Census edition. This includes an overview from the Census Director covering the questionnaire, delivery and return methods, field staff deployment and community engagement. Other technical articles included in the edition provide a methodological background to the Census covering cover quality assurance and field design. Finally, the edition includes a review of a book 'Census and Sensitivity' by Ian White of ONS tracing the history of the Census and its predecessors.

Further information
Population Trends 143 is available at www.statistics.gov.uk/populationtrends/

Contact
vsob@ons.gsi.gov.uk

New Social Trends chapters to be published in April 2011

On 14 April 2011 *Social Trends* will publish the next 2 articles of the 41st edition of Social Trends. These are part of a rolling quarterly release of articles. The articles to be released in April are:

- Population – containing information on population and the components of population change, international migration and British citizenship, population movement within the UK and changes in age structure.
- Households and Families – detailing household composition, people living in households, families, family formation, civic partnerships, divorces and dissolution, births and conceptions and abortions.

The next group of articles are planned for release in May and June and will cover Education and Training and Labour Market.

Further information

www.statistics.gov.uk/socialtrends/

Contact
social.trends@ons.gsi.gov.uk

Impact of the VAT increase on the Consumer Prices Index

On the 4 January 2011, the standard rate of Value Added Tax (VAT) increased from 17.5 per cent to 20 per cent. ONS estimates that the impact on the January 2011 Consumer Prices Index (CPI) from retailers and service providers passing on the VAT increase (between the December 2010 and January 2011 collection dates) was to increase the 12–month rate by around 0.76 percentage points. This means that if VAT had remained at 17.5 per cent in January 2011, the CPI 12–month rate would have been around 0.76 percentage points lower than the published figure of 4.0 per cent.

The impact of 0.76 percentage points on the January 2011 index compares with an impact of 0.40 percentage points on the January 2010 index when VAT increased from 15 per cent to 17.5 per cent. The most significant differences in impact in January 2011 compared with January 2010 came from:

- transport: where the impact of the increase in VAT on fuel prices and the cost of new cars had more of an upward effect on the CPI in January 2011. A contributing factor to this is due to the weights for these components being higher in January 2011 compared with January 2010, which means that price changes in fuel and new cars have a larger impact on the January 2011 CPI
- alcohol and tobacco: a contributing factor, again, is that the weights for these components are higher in January 2011 compared with January 2010. This means that price changes in alcohol and tobacco have a larger impact on the January 2011 CPI. In addition there was evidence that there was a greater pass–through of the VAT increase in January 2011 on alcohol compared to January 2010

The approach and methods used to measure the impact in January 2011 are consistent with those that were used to estimate the impacts on the CPI of the temporary reduction in the standard rate of VAT in December 2008 and subsequent reversion in January 2010. Further details on this and on the CPI more generally can be found in the article 'Impact of VAT reduction on the consumer price indices' Pike, Lewis and Turner (2009).

Further information

Lewis M, Pike R and Turner D (2009) 'Impact of the VAT reduction on consumer price indices', Economic & Labour Market Review 3(8), pp 17–21 and available at www.statistics.gov.uk/cci/article.asp?ID=2258

Contact

darren.morgan@ons.gsi.gov.uk
rowan.kelsoe@ons.gsi.gov.uk

CPI and the budget

On 23 March 2011, the Chancellor of the Exchequer in his Budget statement announced a number of changes to excise duties. ONS subsequently produced a short article for the National Statistics website which estimated the impact on the consumer prices index (CPI) and retail prices index (RPI) of these measures and those previously announced that come into force in 2011/12. The article also included a comparison with those measures implemented in 2010/11.

It is estimated that the changes in excise duties that will come into force in 2011/12 will add 0.29 percentage points to the CPI, if the duty changes were passed on in full to consumers as soon as they come into effect. The breakdown of this overall impact and the date of the implementation of the changes are as follows:

• tobacco	+0.15 percentage points (pp)	23 March 2011
• road fuel	-0.04 pp	23 March 2011
• alcohol	+0.10 pp	28 March 2011
• road fuel	+0.08 pp	1 January 2012

The measures implemented in 2010/11 were estimated to add 1.77 percentage points to the CPI, if they had been passed on in full to consumers as soon as they came into effect. The estimated impact in 2011/12 is lower than that in 2010/11 mainly due to the increase in the standard rate of Value Added Tax that took place in 2010/11.

It is estimated that the measures due to come into force in 2011/12 will add 0.37 percentage points to the RPI. The impact on the RPI is greater than that for the CPI due to the fact that the:

- tobacco items that are subject to excise duties have a higher weight in the RPI compared to the CPI, which means that the increase in these duties have a larger impact on the RPI
- RPI includes the cost of the road fund licence whereas the CPI does not. The increases in vehicle excise duties therefore impact on the RPI but not the CPI

This compares with an estimated increase of 1.67 percentage points to the RPI in 2010/11 as a result of the measures implemented during that period.

Further information

www.statistics.gov.uk/CCI/nugget.asp?ID=336

Contact

darren.morgan@ons.gsi.gov.uk

Full–time working driving up mothers' employment rate

A growth in full–time working has driven the increase in the percentage of mothers in work, new analysis pf the Labour Force Survey by the Office for National Statistics has shown. In 1996, 23 per cent of mothers worked full–time, rising to 29 per cent by the final quarter of 2010.

A higher proportion of mothers work part–time than work full–time, but this figure has been fairly stable over time: 38 per cent in 1996, 37 per cent in 2010. The gap between part–time and full–time employment rates for mothers has therefore halved.

There has also been a large narrowing of the gap between employment rates for mothers and for women who do not have dependent children. In 1996 the employment rate was 61 per cent for mothers and 67 per cent for women without children; by the end of 2010 the gap had almost gone; with 66 per cent of mothers and 67 per cent of women without children in employment. Since the onset of the recession the employment rate for women without children has fallen back from its peak of 70 per cent reached in 2006, driven mainly by a fall in employment for those aged 16 to 24.

Employment rates vary strongly between mothers who were living with a partner and lone mothers, especially those with younger children. For mothers whose youngest child is aged 0–4, there is a 27 percentage point difference in the employment rate (63 per cent for those with partners, 36 per cent for lone mothers). This difference virtually disappears where the youngest child was aged 16–18, an 81 per cent employment rate for women with partners compared with 78 per cent for lone mothers.

Further information
Further analysis is contained in a topic–based summary at www.statistics.gov.uk/cci/nugget.asp?id=2124

Contact
labour.market@ons.gsi.gov.uk

NHS productivity: new estimates for 1995 to 2009

Spending on healthcare and the output it generated both rose substantially between 1995 and 2009. Quality also rose but productivity fell slightly over the same period. These are the key findings in an article Public service output, inputs and productivity: healthcare, published on 31 March 2011 by the Office for National Statistics.

The article explains that:

- between 1995 and 2009, public expenditure on healthcare in the UK rose from £38 billion to £111 billion, 7.9 per cent a year on average
- stripping out the effects of pay and price changes, the amount of inputs (doctors and nurses, but also clinical supplies, electricity and water etc) went up by 88 per cent, an average of 4.4 per cent a year
- adjustments for quality added an average of 0.5 percentage points to annual output growth
- the increase in inputs was a little faster than the increase in outputs. So productivity (output per unit of input) fell slightly, by 2.7 per cent, an annual average fall of 0.2 per cent
- however, in the most recent year, 2009, productivity increased by 0.7 per cent

The article also includes external evidence that:

- smoking prevalence decreased by a quarter for males and a fifth for females between 1998 and 2009
- female weekly alcohol intake over the recommended units increased by one–fifth since 1998 while male consumption remained fairly constant
- emergence re–admissions increased by a quarter for children and a third for adults between 1999–00 and 2008–09

Further information

The article Public service output, inputs and productivity: healthcare can be found at www.statistics.gov.uk/cci/article.asp?ID=2666

Contact
ukcemga@ons.gsi.gov.uk

Graduates earn £12,000 a year more than non–graduates

Degree holders earned an average of £12,000 a year more than non–graduates over the past decade, new analysis from the Office for National Statistics shows. After adjusting to allow for increases in earnings over the period, the data show that graduates aged 22 to 64 had median salaries of £30,000 compared with £18,000 for non–degree holders.

Looking at how earnings differed between age, earnings were similar for those aged 22, around £15,000 regardless of whether they had a degree or not. For those without a degree, earnings initially increased for each year of age but levelled off around the age of 30 and peaked at the age of 34, where it stood at £19,400. For those with a degree, earnings increased faster for each year of age. They also increased for longer, levelling off around the age of 35 and peaking at the age of 51 at £34,000.

When comparing men's and women's earnings, it is better to use hourly earnings, because of differences in the number of hours worked between the sexes. Earnings for women, for both those with and those without a degree, levelled off at a younger age than for men with the same level of qualification. Over the last decade, a male graduate could expect to earn on average 20 per cent more than a female graduate – however the gap was marginally wider for non–degree holders at 23 per cent.

In 2010, around one in three female graduates had a degree in either health–related studies or education, compared with only one in 11 male graduates, while almost one in two graduates had degrees in business and finance, sciences or engineering compared with only one in five female graduates. Average earnings for those with a degree in the predominantly male banking and finance industry were £37,300 compared with £27,600 in the predominantly female public administration, education and health sector.

Further information
Further analysis is contained in a topic–based summary at
www.statistics.gov.uk/cci/nugget.asp?id=1166

Contact
labour.market@ons.gsi.gov.uk

Updates

Updates to statistics on www.statistics.gov.uk

11-Mar

Producer prices

Factory gate inflation rises 5.3%

www.statistics.gov.uk/cci/nugget.asp?id=248

16-Mar

Average weekly earnings

Regular pay growth decreases

www.statistics.gov.uk/cci/nugget.asp?id=10

Unemployment

Rate rises to 8.0%

www.statistics.gov.uk/cci/nugget.asp?id=12

Public sector employment

Employment decreases in Q4 2010

www.statistics.gov.uk/cci/nugget.asp?id=407

22-Mar

Inflation

CPI inflation 4.4%, RPI inflation 5.5%

www.statistics.gov.uk/cci/nugget.asp?id=19

Public sector finances

£11.8 billion net borrowing

www.statistics.gov.uk/cci/nugget.asp?id=206

24-Mar

Retail sales

Mixed picture in February

www.statistics.gov.uk/cci/nugget.asp?id=256

25-Mar

Institutional investment

Net investment £18.4 billion

www.statistics.gov.uk/cci/nugget.asp?id=396

29-Mar

GDP growth

Contracts by 0.5% in Q4 2010

www.statistics.gov.uk/cci/nugget.asp?id=192

Balance of payments

Current account deficit widened

www.statistics.gov.uk/cci/nugget.asp?id=194

Business investment

No growth in Q4 2010

www.statistics.gov.uk/cci/nugget.asp?id=258

Consumer spending

Expenditure fell by 0.3% in Q4 2010

www.statistics.gov.uk/cci/nugget.asp?id=11

30-Mar

Index of services

2.2% annual rise into January

www.statistics.gov.uk/cci/nugget.asp?id=558

Disposable household income

Growth slowed cross UK in 2009

www.statistics.gov.uk/cci/nugget.asp?id=1552

CPI and the budget

Estimated impact on inflation

www.statistics.gov.uk/cci/nugget.asp?id=336

31-Mar

UK government debt and deficit

Deficit 10.2% of GDP

www.statistics.gov.uk/cci/nugget.asp?id=277

Mothers in the labour market

More mothers working than before

www.statistics.gov.uk/cci/nugget.asp?id=2124

06-Apr

Graduate earnings

Graduates earn £12,000 a year more

www.statistics.gov.uk/cci/nugget.asp?id=1166

Corporate profitability

12.2% in Q4 2010

www.statistics.gov.uk/cci/nugget.asp?id=196

Index of production

Production: 2.45 annual increase

www.statistics.gov.uk/cci/nugget.asp?id=198

07-Apr

Labour productivity

Growh inproductivity Q4 2010

www.statistics.gov.uk/cci/nugget.asp?id=133

Forthcoming releases

Future statistical releases on www.statistics.gov.uk
08-Apr
Producer price index – March 2011
Output in the construction industry – February 2011
12-Apr
Consumer price indices – March 2011
UK Trade - January 2011
Quality adjusted labour input (experimental)
13-Apr
Labour market statistics – April 2011
Average weekly earnings – February 2011
14-Apr
Social Trends – Households and families
Social Trends – Population
Overseas travel and tourism – February 2011
How indirect taxes can be regressive and progressive – 2011
19-Apr
Turnover and orders in production and services industries – February 2011
20-Apr
Pension Trends – Chapter 10: Saving for retirement
Pension Trends – Chapter 9: Pension scheme funding and investment
Pension Trends – Chapter 14: Pensions and the National Accounts
21-Apr
Public sector finance – March 2011
Retail sales – March 2011
Index of labour costs per hour (experimental) – Q4 2010
27-Apr
Gross domestic product preliminary estimate – Q4 2010
Index of services – February 2011

Economic Indicators

PRICES AND INFLATION	Value	Period	Monthly change	Annual change	Release date
Consumer Prices Index (CPI) (2005=100)	117.8	Feb-11	0.7	4.4	22-Mar-11
Retail Prices Index (all items) (Jan 1987=100)	231.3	Feb-11	1.0	5.5	22-Mar-11
RPI excluding mortgage interest (RPIX) (Jan 1987=100)	230.5	Feb-11	1.0	5.5	22-Mar-11
Producer Prices Index - Output (2005=100)	121.9	Feb-11	0.5	5.3	11-Mar-11
Producer Prices Index - Input prices (materials and fuel) (2005=100)	161.2	Feb-11	1.1	14.6	11-Mar-11

LABOUR MARKET	Value	Period	Change on 3 months	Change on 1 year	Release date
Employment rate (%)	70.5	Nov-Jan 11	-0.1	0.1	16-Mar-11
Unemployment rate (%)	8.0	Nov-Jan 11	0.1	0.2	16-Mar-11
Average Weekly Earnings - total pay (%)	2.3	Nov-Jan 11	-0.2	1.3	16-Mar-11
Average Weekly Earnings - regular pay (%)	2.2	Nov-Jan 11	-0.1	0.8	16-Mar-11
Claimant count (Jobseeker's Allowance) (Thousands) (2005=100)	1,448.6	Feb-11	-12.1	-128.1	16-Mar-11
Vacancies (Thousands)	496	Dec-Feb 11	24	15	16-Mar-11

NATIONAL ACCOUNTS ECONOMIC ACTIVITY	Value	Period	Quarterly change	Change on 1 year[3]	Release date
UK Gross Domestic Product (chained volume measure £ billion)	329.1	Q4 10	-0.5	1.5	29-Mar-11
Private Non-Financial Corporations Net Lending (£ billion)	22.9	Q4 10			29-Mar-11
Household Saving Ratio (%)	5.4	Q4 10			29-Mar-11
Public Sector current budget (£ billion)	-6.5	Feb-11			22-Mar-11
Public Sector net debt as a % of GDP	58.0	Nov-10			21-Dec-10
Public Sector net borrowing (£ billion)	10.3	Oct-10			18-Nov-10
Public Sector net cash requirement (£ billion)	-7.0	Feb-11			22-Mar-11
Public sector net borrowing (excluding financial interventions) (£ billion)	11.8	Feb-11			22-Mar-11

Public sector net debt as a % of GDP (excluding financial interventions)	58.0	Feb-11			22-Mar-11

BALANCE OF PAYMENTS AND TRADE	Value	Period	Change on 3 months	Change on 1 year	Release date
UK's trade balance (£ billion)	-£3.0	Jan-11			9-Mar-11
Balance of Payments current account - (£ billion)	-£10.5	Q4 10			29-Mar-11
of which: EU	-£13.9				
non-EU	£3.4				
Goods export volumes - excluding oil and erratics (2006=100)	95.1	Jan-11			9-Mar-11
Goods import volumes - excluding oil and erratics (2006=100)	98.0	Jan-11			9-Mar-11

SHORT TERM INDICATORS	Value	Period	Change on 3 months[1]	Change on 1 year[2]	Release date
Retail Sales (2006=100) (chained volume, seasonally adjusted)	107.6	Feb-11	0.1	1.7	24-Mar-11
Index of Manufacturing (2006=100)	92.7	Feb-11	1.1	5.3	6-Apr-11
Index of Production (2006=100)	89.6	Feb-11	0.8	3.4	6-Apr-11
Productivity - Whole economy (2005=100)	99.2	Q4 10	-0.3	0.8	7-Apr-11
Productivity - Manufacturing (2005=100)	106.7	Q4 10	1.4	7.5	7-Apr-11
Index of Services (2006=100)	102.4	Jan-11	-0.5	1.0	30-Mar-11

Notes:

1. Three months on previous three months
2. Three months on corresponding period one year ago
3. Quarter on corresponding period one year ago

Using administrative data held by DWP as an alternative sampling frame for the Family Resources Survey

Antonia Simon
Department for Work and Pensions

Summary

The Family Resources Survey (FRS) is the leading UK household survey for the collection of private household income. The Department for Work and Pensions (DWP) is currently considering the extent to which their administrative data can be used to develop an alternative sampling frame for the FRS. This article summarises the major findings of this research, comparing the current sampling frame for the FRS, using the small–users Postcode Address File (PAF), with suitable administrative data sources. The findings offer important considerations for how UK household surveys might be sampled in the future using administrative data sources.

Background

The use of administrative data for statistical research purposes has been of growing interest in the UK (Jenkins et al 2007). There are a number of perceived benefits with administrative data, including a reduction in costs associated with conducting large–scale surveys and an improvement in data quality. The use of administrative data may also help to diminish respondent burden and the public opposition to completing questionnaires, including the census (Calderwood and Lessof 2006; Jones and Elias 2006; Wroe 1998). However, relative to other countries, such as Finland and Denmark, the UK has so far made little use of administrative data as a sampling frame for surveys.

Recognising the wider potential benefits of better utilising its administrative data, the Department for Work and Pensions (DWP) is considering the extent to which such data can be used for sampling purposes. To investigate this further, the DWP commissioned research to investigate the feasibility of using DWP administrative data as an alternative sampling frame for the Family Resources Survey (FRS). This research compared the current sampling frame for the FRS with two main DWP administrative data sources: the Address Hierarchy (AH) File, which forms the address file for the Work and Pensions Longitudinal Study (WPLS), and the Customer Information

System (CIS) which represents a more complete list of all individuals who have been allocated a National Insurance Number (NINO). The AH file contains an address history for individuals who have had some contact with DWP through claiming benefits or their state pension. It excludes individuals not known to the DWP. The CIS, in contrast, includes all NINOs and therefore potentially covers all of the population aged 16 years and over (and not just DWP customers). For the analysis reported in this paper, the AH file was reduced to represent one address per household – with only current addresses included (assumed to be the last registered address per individual).

Findings

The FRS and its current sampling frame

The FRS has collected household income data at a national level in its current form since 1993/4. It achieves an interviewed sample size of approximately 25,000 households a year. FRS data are used to assess progress towards achieving key Departmental Objectives to reduce income poverty among household types identified to be at risk, including pensioners and people with disabilities. The FRS, along with many other national household surveys in the UK, uses the small–users Postcode Address File (PAF) as the sampling frame for Great Britain (GB). This is now the most commonly used sampling frame for large-scale surveys of the general population (Tipping and Nicolaas 2006). Compiled by Royal Mail, this is a list of all addresses in the UK receiving fewer than 50 Items of post daily. While the PAF provides full address detail, it does not include any further information that might be useful for sampling, such as household composition or age structure of households.

The contractors for the FRS (currently the National Centre for Social Research and the Office for National Statistics) aim to draw a statistically representative sample from the PAF using a combination of sampling techniques such as stratification based on the latest census, systematic randomisation, and clustering (ONS and NatCen 2007). The selection process also ensures a particular household is not oversampled so as to avoid complaints from members of the public being approached more than once within a short time period to participate in government surveys (Wilson and Elliott, 1987). In addition, any multi–households in the sample are treated to take account of the differing probabilities of selection this introduces (ONS and NatCen 2007).

The geographical coverage of DWP administrative data relative to the PAF

Postcodes, and then addresses from the PAF were matched separately to the AH file and the CIS. As part of the matching process, postcodes were validated using British Standards 7666 ‘BS 7666’, which describes postcodes as being ‘valid’ if they are alphanumeric and between five and eight characters long (Cabinet Office, 2009). There are just under 1.5 million unique postcodes recorded on the small–users PAF, just over 1.7 million unique postcodes on the CIS, and just over 1.8 million unique postcodes on the AH file. The AH file has a larger number of postcodes than the CIS because it includes a historical record of addresses based on an individual’s history of claiming a

state benefit or state pension through the Department. Therefore, it is common to identify more than one address per individual in this file.

For postcodes, the 'match rate' for the PAF to the CIS and AH files was quite high – with 98 per cent of PAF's postcodes found on the CIS and 97 per cent found on the AH file (**Table 1**). For addresses, the match rate of the PAF to the CIS was 84 percent and to the AH file was 75 per cent. The match rate for addresses was lower than for postcodes because it was not possible to match addresses without street numbers (for example, those starting with a house name like 'Rose Cottage'). The lower match rate for addresses also reflects the less reliable quality of addresses in the administrative data. For example, addresses on the CIS and AH files are more likely to have errors and /or different formats. The most common issues are:

- information for flats being dropped
- suffixes on addresses being dropped (for example 41B becomes 41)
- flat and house numbers becoming confused (for example flat 4, 7 street becomes flat 7, 7 street); and
- addresses have no street / house number

The match rate for addresses for the AH file was much lower than for the CIS (75 per cent compared with 84 per cent for the CIS). This reflects the lower geographical coverage of the AH file relative to the CIS. While the CIS and AH files contain a similar number of higher level unique postcodes, the CIS offers greater coverage of lower level addresses within each unique postcode. There was no geographical bias in the match–rate for postcodes or addresses, in that the match rate was fairly similar across postcode letters on the CIS (ranging from 83 per cent to 90 per cent). The exception to this pattern was addresses with 'B' postcodes, with fewer matching to the CIS and AH files because 'B' postcodes include Belfast (but actually cover all of Northern Ireland). Those 'B' postcodes which cover Belfast (and Northern Ireland) are excluded from the PAF frame for the FRS.

Table 1 Matching CIS and AH administrative data by postcode to the small–users PAF

Percentage of postcodes on PAF matched to CIS and AH file

	CIS	AH file on WPLS
PAF unique postcodes matched	98.3% (1,446,849)	97.7% (1,439,262)
PAF unique postcodes not matched	1.7% (24,972)	2.21% (32,559)
Total unique postcodes on PAF	*100%* *(1,471,821)*	*100%* *(1,471,821)*

Analysing the extent of match between the PAF and administrative sources

While we would expect most CIS postcodes to find a match on the PAF, approximately 330,000 failed to do so. Of these, 90 per cent (300,000) had 'valid' Royal Mail (RM) postcode combinations (Table 1), as verified using the British Standards by which postcode formats are judged ('BS 7666': see previous section). Three reasons were hypothesised to explain why some postcodes from the CIS did not find matches on the PAF:

- the postcodes could belong to a deceased or 'dead letter office' (DLO) flagged individual – such a flag given where post is returned undelivered, or information is otherwise received that the person does not live at that address, or
- to a person with a foreign address (representing individuals now living abroad who for some reason had been allocated a NINO on the CIS), or
- be deleted postcodes no longer in use by Royal Mail (RM) because they may have belonged to a now demolished building

An investigation of the cases on the CIS that did not match the PAF, showed approximately half of these postcodes did not match the PAF because they were from 'old' addresses. These old addresses were flagged in the administrative data as deceased or DLO flagged. While this status is not in itself a problem for use in a sampling frame, so long as the address itself is valid, these addresses failed to find a match in the PAF because either they had been entered incorrectly in the administrative data or were addresses/postcodes that were no longer in use by Royal Mail. In addition, only a small minority (0.5 percent) of all cases not matched to PAF were found to be foreign addresses.

The 'valid, non–match' postcodes between the CIS and PAF were also compared to the National Statistics Postcode Directory (NSPD). The NSPD is an ONS database which records all postcodes that are currently, or have in the past, been in Royal Mail circulation. A 'date deleted' flag in the NSPD was used to indicate whether postcodes have been deleted. Approximately 31 per cent of these ('valid non–match') postcodes were found to have been deleted (see **Table 2**). The rest of these cases were 'dormant' (32 per cent), describing postcodes formerly in circulation but which are not currently in use (but not flagged as deleted), or 'large-users' (37 per cent) - 'live' or 'current' postcodes but excluded from the small-users PAF as they belong to large (expected to be business) addresses.

Table 2 Reasons why postcodes with 'valid' RM formats on CIS do not always match the small–users PAF

Reason for 'valid' non–matched cases between CIS and PAF and between small–users PAF and NSPD	% non–matched postcodes
Deleted postcodes: found on CIS but flagged as deleted on NSPD	31%
Dormant postcodes: found on CIS, not flagged on NSPD as deleted but not identified as 'live'	32%
Large user postcodes: found on CIS and identified as 'live' on NSPD	37%
Total on CIS not matching PAF but with 'valid' postcode format	*100% (328,378)*

Although Table 2 shows the reasons for 'valid non–match' postcodes were overall fairly evenly distributed between 'deleted', 'dormant' and 'large–user', there was actually some variation in these proportions when examined by postcode letter (for example the first letter in the postcode SW10 2PP would examine the 'S's). For example, 71 per cent of 'A's were identified as not matching and '*dormant'* on the NSPD, letters 'B', 'I', and 'J' had high proportions not matching and *'large users'* (71 per cent, 59 per cent and 96 per cent respectively), and letters 'Z', 'T', E' and 'F' had high proportions not matching and '*deleted'* (60 per cent, 45 per cent, 46% per cent and 44 per cent respectively).

Technical considerations with using administrative data for sampling

While it is always important to consider ethical impacts, there are likely to be few ethical concerns posed if administrative data are used to generate an address list for statistical sampling purposes. This is because the information generated for a sampling frame would approximate (in terms of level of detail) to that available in the PAF and which is currently widely used for sampling purposes. However, there are a number of key technical considerations in creating a reliable FRS sampling frame, including:

- reliably dealing with addresses that do not match because of data quality – work is currently underway elsewhere in DWP to 'clean' administrative data such as the CIS so that it holds more complete and reliable data and better matches the full PAF
- reducing administrative data to unique addresses so that each represents a household – currently administrative data, such as the AH file can hold several records per individual according to their benefit or pension claim histories
- excluding foreign addresses and GB addresses which do not represent private households

Conclusions

This article has outlined the potential for using administrative sources for developing alternative sources of sampling frames for the FRS. It has investigated the geographical coverage of administrative data relative to the current sampling frame for the FRS, using the small–users PAF.

The findings discussed here indicate that administrative data have the potential to provide a good alternative for sampling the FRS. This is because the match of PAF postcodes to the administrative data examined in this research is quite high (possibly even higher when non–private households are excluded from the CIS and AH files). The match rate of addresses, particularly using the CIS, also seems to offer good coverage of the GB population. However, further work is needed to make using administrative data as an alternative sampling frame for the FRS a viable option. This includes cleaning addresses on the CIS to improve the data quality and carrying out a full assessment of any bias. For example, unacceptable bias may be introduced if all addresses without house numbers or with house names are never included in a sampling frame.

While this research has concentrated on comparing the CIS and AH files with the FRS, they demonstrate the real potential of using administrative data for developing sampling frames for

other household surveys, such as the Labour Force Survey. This would also tally with wider national initiatives to make better use of administrative data held across government.

Acknowledgements

This article derives from a project jointly funded by the DWP and the Economic Social Research Council (grant number: RES-173-27-0204). The author is very grateful to John Snow, at the DWP, for his technical guidance and support with using and analysing DWP administrative data; also for the support given with this research from Valerie Christian, at the DWP, and Charlie Owen, at the Institute of Education (IOE). Thanks also to the IOE ethics committee for reviewing the work reported here and for granting full ethical approval for this research.

Further information

Further details about the Family Resources Survey can be obtained from the DWP's website: http://research.dwp.gov.uk/asd/frs/

Contact

The author of this paper, Antonia Simon, is an academic researcher at the Institute of Education, University of London. She was seconded to the DWP (February 2010 to March 2011) to carry out the research reported in this paper. Further information about this work can be obtained from her at: a.simon@ioe.ac.uk or the team manager at the DWP, Valerie Christian at: Valerie.Christian@dwp.gsi.gov.uk .

References

Cabinet Office (2009) UK postcode data formats. Retrieved from Cabinet Office: http://interim.cabinetoffice.gov.uk/govtalk/schemasstandards/e-gif/datastandards/address/postcode.aspx

Calderwood L and Lessof C (2006) 'Enhancing longitudinal surveys by linking administrative data'. Paper presented at the MOLS2006 Conference, 12–14 July 2006, University of Essex

Jenkins S P, Lynn P, Jäckle A, and Sala E (2007) 'The feasibility of linking household survey and administrative record data: new evidence for Britain', *International Journal of Social Research Methodology*, *11(1)*, 29–43

Jones P, and Elias P (2006) *Administrative data as a research resource: A selected audit.* ESRC National Centre for Research Methods (Unpublished). Retrieved from: http://eprints.ncrm.ac.uk/452/1/Admin%2520Data%2520selected%2520audit%2520Dec06.pdf

Office for National Statistics & National Centre for Social Research (2007). *Family Resources Survey. Annual technical report 2007–8*. Prepared for Department of Work and Pensions

Tipping S and Nicolaas G (2006) 'In search of a population sampling frame for UK postal surveys', *Survey Methodology Bulletin*, 58: August 2006

Wilson P R and Elliott D J (1987) 'An evaluation of the Postcode Address File as a sampling frame and its use within OPCS', *Journal of the Royal Statistical Society A*, 150 part 3: 230–240

Wroe D (1998) *Beyond 2001. Alternative to the Census. Study for Office for National Statistics. Volume 1: Report*. Office for National Statistics, Crown Copyright. Retrieved from: www.statistics.gov.uk/downloads/theme_population/previous_uk.pdf

Quality–adjusted labour input: new quarterly estimates for 1993 to 2009 and annual estimates from 1970

Jean Acheson
Office for National Statistics

Summary

Quality–adjusted labour input (QALI) is a measure of labour input to economic production which takes account of the composition of the workforce as well as volume of hours worked. It provides a more complete picture of the input of labour to the production process than traditional measures, which focus only on the quantity of labour input, and therefore provides a broader perspective in assessing productivity performance. Along with the volume index of capital services, QALI is a key input to multi–factor productivity and growth accounting analyses. This article presents new estimates of QALI for 1993 to 2009. The series has been extended by an additional year and provides a fuller picture of QALI over the recession. An additional section backcasts QALI to 1970 using annual 'labour services' estimates from the EU KLEMS dataset[1].

Introduction

Standard labour productivity measures express growth in output with respect to the volume of labour input, either in terms of employment, jobs or hours worked[2]. The implicit assumption underlying this approach is that labour is homogeneous, as it does not take into account the composition, or quality, of the workforce.

'As a result an hour worked by a highly experienced surgeon and an hour worked by a newly hired teenager at a fast food restaurant are treated as equal amounts of labour'. OECD (2001)

However, labour is far from homogeneous, and the 'value' of an hour worked, or marginal productivity, varies significantly between workers. The quality–adjusted labour input (QALI) series attempts to address this, providing a measure which explicitly recognises the heterogeneity of labour by adjusting the volume of hours worked according to certain characteristics – qualifications, age, gender and industry – which may be indicative of a worker's quality.

QALI is therefore a conceptually stronger method for use in productivity and growth accounting analyses, and is a useful tool for assessing the evolution of human capital over time. It is used alongside experimental estimates of capital services (such as Appleton 2011) to produce multi–factor productivity (MFP) estimates.

This article presents new estimates of QALI for 1993 to 2009. Data are presented for the whole economy, the market sector and for ten industries. QALI is also analysed by education, age and gender. A separate piece of analysis backcasts the QALI time series to 1970 using the EU KLEMS dataset, which produces quality adjusted labour series called 'labour services' that are conceptually similar to QALI.

Methodology

To perform the quality adjustment, hours worked are differentiated into *n* types of worker (h_1 to h_n) determined by their characteristics: age, educational attainment, industry and gender. The reasons for using these particular characteristics are explained in **Box 1**. The hours worked by these different worker types contribute to total labour input *L* through a function *g*.

$$L = g(h_1, h_2, ..., h_n)$$

Economic theory states that in competitive markets with constant returns to scale, labour will be hired until its marginal cost (wage rate) equals its marginal revenue product, or marginal productivity. Therefore, when measuring labour input, using income data as a weight takes into account the relative productivity (or 'quality') of workers as well as the quantity of hours worked.

Following the OECD (2001) recommended methodology, the growth in quality–adjusted hours is represented as a Törnqvist index:

$$\Delta \ln L_t = \ln\left(\frac{L_t}{L_{t-1}}\right) = \sum_i \left[\left(\frac{w_{i,t} + w_{i,t-1}}{2}\right) \ln\left(\frac{h_{i,t}}{h_{i,t-1}}\right)\right]$$

where $w_{i,t}$ is the share of total labour income paid to group *i* in period *t*, the weight used is the average of $w_{i,t}$ and $w_{i,t-1}$, and the income shares sum to one. The logged changes are then used to create an index where the first quarter of the time series, Q1 1993, equals 100. The use of data from the current and previous period to weight the change in hours is a feature of Törnqvist indices, making them more current or representative measures. The index is also a widely used form in economic analysis, particularly with regard to quality–adjusted labour measures (Bell *et al* 2005). A more detailed discussion of index numbers can be found in Goodridge (2007).

The difference between the Törnqvist index and an unadjusted index of hours is referred to throughout the article as 'labour composition'.

Box 1 Labour characteristics

The choice of labour characteristics involves a trade–off between parsimony and data availability, and the objective of capturing significant developments in labour inputs to production. None of the following characteristics represent labour quality in and of themselves, but only as dimensions of the income–share weights.

Age

Age is included as a proxy for work experience. Although imperfect, as it takes no account of periods of unemployment or inactivity, the assumption is that older workers tend be more productive due to their greater experience, and therefore receive greater compensation for their labour. Alternatively, it has been suggested that younger workers may be more dynamic and innovative than their older counterparts (Bell *et al* 2005). However, if this is true in some cases, then provided labour markets are competitive, these workers will be paid their marginal product and growth in hours will be weighted accordingly.

Gender

Gender is chosen because of the persistent pay differential that exists between males and females, even after holding other factors constant. Although not a driver of quality change itself, it may represent hidden characteristics such as an increased tendency to take career breaks or to fulfil part–time posts that are not as well paid. Therefore, this complements, or improves, the use of age as a proxy for work experience, as well as helping to explain the pay differential. However, if the pay differential instead reflects discrimination, then the assumption that workers are paid their marginal product is violated, resulting in hours growth being weighted incorrectly and the quality adjustment carrying a downward bias. This is a weakness of the model.

Education

This is measured as the highest qualification attained and used as a proxy for skills. Qualifications either act as a signal of ability to employers or they provide the knowledge for specific job requirements. This characteristic is the primary driver of the index. Due to the increasing prevalence of higher degrees and their growing association with higher pay, they are included as a stand–alone category.

Industry

Although primarily included for the observation of industry trends and the use of QALI in industry–level MFP, this category also helps capture inherent differences in skill and productivity that exist between industries. The industry categories chosen are broad partly because industry is self–reported in the LFS, leading to inaccuracy of response, and also because of small sample sizes for some sectors.

Data source

The main data source for QALI is the Labour Force Survey (LFS). The LFS is a continuous household survey that covers approximately 53,000 households every quarter. It provides data on the volume of hours worked, and contains series for educational attainment, industry, gender and age, plus pay data, which are used to carry out the quality adjustment. Although the LFS became quarterly in 1992, questions about the respondent's income were not asked until 1993, so the QALI series begins in the first quarter of 1993.

Scaling

QALI's consistency with other data series, particularly the National Accounts and ONS headline productivity measures, is important for its use in other types of productivity calculation such as growth accounting. To achieve this external consistency, various components of QALI are scaled to ONS aggregates. Specifically:

- gross weekly pay is scaled to National Accounts industry level 'Compensation of Employees' (CoE)
- actual hours worked are scaled to industry level productivity hours[3]
- total jobs are scaled to industry level productivity jobs[4]

The first adjustment improves the consistency of the LFS–based data with National Accounts income measures. While the LFS only provides information on wages and salaries, CoE also includes bonuses and income–in–kind and, as such, is a more complete indicator of total remuneration. Additionally, as with other household surveys, LFS micro–data include proxy responses, missing responses and inaccurate data. Respondents have particular difficulty recalling their pre–tax income or bonuses accurately; scaling the data helps overcome these issues. It should be noted that the scaling up of gross weekly pay does not impact on QALI for individual industries, but the relative weights of industries will change and so will whole economy and market sector QALI as a result.

CoE only covers remuneration of employees; the earnings of the self–employed are recorded as mixed income, which accounts for both the returns to capital and labour. Ideally, QALI would split mixed income to give labour compensation for the self–employed. However, it is not possible to do this at present. Aggregating LFS employees' wages with LFS imputed self–employed wages and scaling to CoE is the next best methodology available under current data constraints.

Scaling the jobs and hours data to the headline labour productivity jobs and hours series, which use superior business survey–based industry breakdowns, improves consistency and also helps to overcome partly the inaccuracies in the LFS industrial breakdown.

Data issues

The inclusion of the self–employed poses an issue as wages for the self–employed are not recorded in the LFS or any other survey. This is because self–employed people remunerate themselves for a combination of labour and entrepreneurial effort, without distinguishing between the two. In producing QALI, LFS data on the compositional characteristics of the self–employed are used, and the wages of employees with similar characteristics are used as an approximation for the labour income of the self–employed. This is likely to be an over–estimate, but the method used is the most appropriate for such a detailed dataset (Turvey 2009).

To measure labour's true input to production as accurately as possible, no restrictions have been placed on outliers, and actual hours rather than usual hours are used because, conceptually, it is the former that need to be measured.

Comparison with previous release

When performing the quality adjustment there is an inevitable trade–off between the different categories due to constraints imposed by the sample size of the LFS. In particular, a judgement needs to be made whether to prioritise the main compositional categories (education, age and gender) or focus on the industrial breakdown. This release uses the same compositional breakdown initiated by the previous release (Turvey, Goodridge and Franklin 2010).

The compositional and industrial breakdowns used are summarised in **Table 1**.

Table 1 Quality adjustment categories

Education	Age	Gender	Industry*	Industry description
Masters and doctorates (NVQ6)	16-29	Male	ABCE	Agriculture, hunting and forestry; fishing; mining and quarrying; utilities
First and other degrees (NVQ5)	30-49	Female	D	Manufacturing
Certificates of education or equivalent (NVQ4)	50+		F	Construction
A levels or trade apprenticeships (NVQ3)			G	Wholesale and retail trade
GCSEs or equivalent (NVQ2, NVQ1, other qualifications)			H	Hotels and restaurants
No qualifications			I	Transport, storage and communication
			J	Financial intermediation
			K	Real estate, renting and business activities
			LMN	Public administration and defence; education; health and social work
			OPQ	Other social and personal services

* Standard Industrial Classification 2003

By producing results for ten industries, other compositional categories are necessarily more limited. For example, there are only three age categories. Despite the expected pay differential between, say, workers aged 16–19 and workers aged 26–29, this limitation is necessary if enhanced industrial detail is sought (particularly for growth accounting analysis).

Table 2 Relative remuneration per hour, 2009

Whole economy = 100

Quality-adjustment categories	Relative remuneration per hour
Industry	
ABCE: Agriculture etc; mining and quarrying; utilities	101.2
D: Manufacturing	98.7
F: Construction	98.6
G: Wholesale and retail trade	73.4
H: Hotels and restaurants	59.3
I: Transport, storage and communications	95.8
J: Financial intermediation	135.5
K: Real estate, renting and business activities	119.7
LMN: Public admin and defence; education; health and social work	108.0
OPQ: Other social and personal services	89.9
Gender	
Females	91.8
Males	107.5
Age	
16-29 years	70.4
30-49 years	108.7
50+	103.2
Education	
Higher degrees	157.8
First or other degrees	131.8
Certificates of education or equivalent	110.4
A levels or trade apprenticeships	91.1
GCSEs or equivalent	77.6
No qualifications	62.9

Source: Labour Force Survey

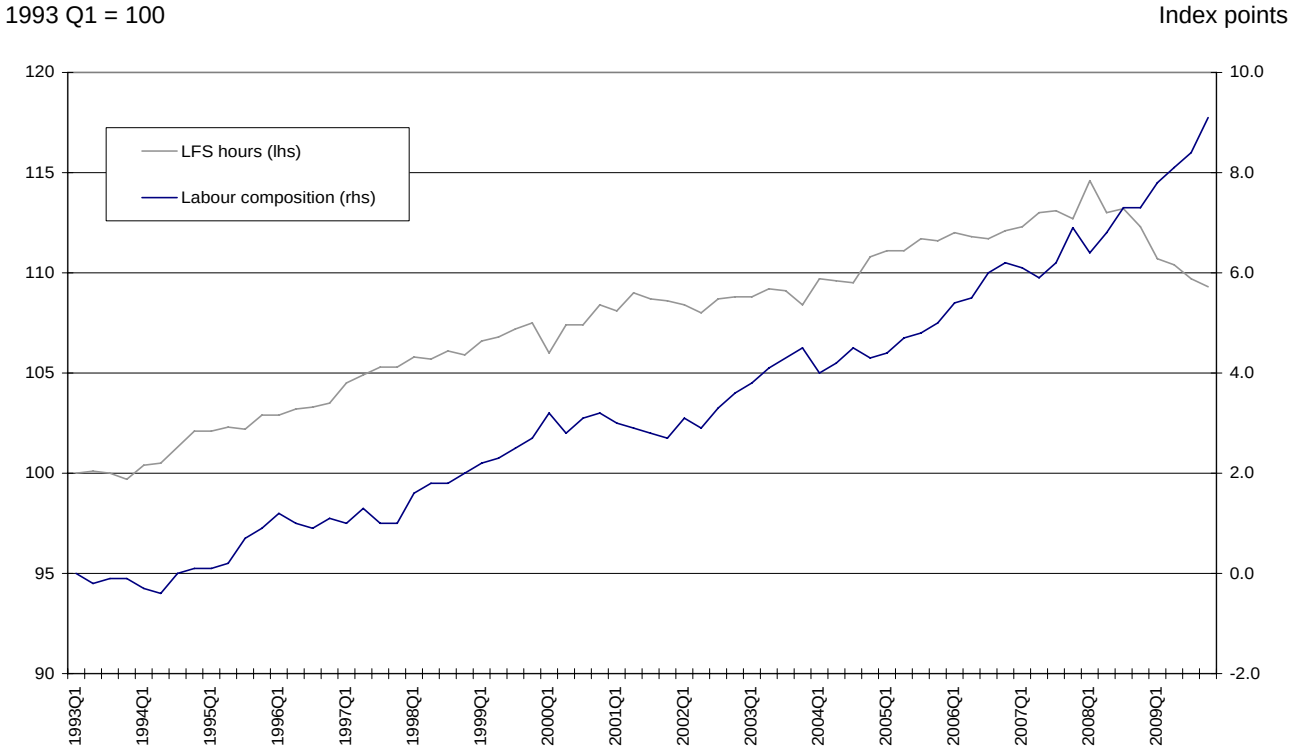
Table 2 gives an indication of the relative remuneration per hour in 2009 of the compositional categories recorded in Table 1[5]. This is extracted from the LFS micro–data and is not scaled to any National Accounts or other productivity series. The table highlights the variance in pay across the quality–adjustment categories. Not surprisingly, some QALI categories, like working in financial intermediation or having a higher degree, provide more returns to labour input than others, on average.

The table does not necessarily foreshadow the QALI results; for example, remuneration per hour in hotels and restaurants is very low relative to the whole economy average, but this does not necessarily imply that QALI comparisons between the two will yield similar results. For example, QALI for whole economy will change if labour supply moves between hotels and restaurants and other industries. And QALI for hotels and restaurants will change if the share of labour input in hotels and restaurants moves between gender or age group or mix of qualifications. The table simply shows the importance of accounting for different compositional differences when calculating quality–adjusted labour inputs.

Results

Figure 1 compares an index of unadjusted hours worked across the whole economy, based on LFS micro–data, with labour composition. These are the two components that make up QALI[6].

Figure 1 Whole economy QALI



Source: Office for National Statistics

Labour composition has been rising very steadily over the period. Hours have also been rising for most of the period but since Q1 2008 they have fallen back sharply. With labour composition increasing and less hours being worked this implies that the employed workforce is moving toward higher remunerated QALI categories, on average. The increase in labour composition combined with falling hours over the recession suggests that low–skilled workers suffered proportionately more than those with higher skill levels.

Quality–adjusted labour input by industry

Figures 2 to 11 show results for the ten industries outlined in Table 1. In each case, an unadjusted hours index is presented alongside labour composition (the difference between QALI and the unadjusted series).

Unadjusted hours in agriculture, forestry, fishing, mining and utilities (**Figure 2**) fell between 1993 and 2002 but began to rise, albeit erratically, since this period. Labour composition has shown no upward trend since 2004. There appears to be little or no adverse impact from the recession on this industry; in fact, hours increased over 2008 and 2009.

In the manufacturing sector (**Figure 3**), unadjusted hours have fallen sharply since 1998, with the rate of decline notably increasing over the recession. Labour composition has shown consistent growth over the period although it dipped slightly over the recession. Outside of the service sector, construction (**Figure 4**) has experienced the fastest growth in hours, but the slowest growth in labour composition. This may reflect limited scope to substitute towards higher skilled workers in the construction sector. In 2009, construction experienced the sharpest fall in unadjusted hours of all industries under analysis, an indication of the severe impact of the downturn on this industry. By contrast, labour composition increased substantially over the same period, suggesting that high skilled workers in this industry have fared the recession better than their low skilled counterparts.

Unadjusted hours in wholesale and retail trade (**Figure 5)** were at close to 1993 levels by the end of 2009. Labour composition, on the other hand, grew consistently over the same time period. Such a pattern highlights the importance of quality adjusting labour: a simple hours index (the blue line) hides the notable increase in labour quality in this industry over the time period in general and over the recession in particular.

Unadjusted hours for hotels and restaurants (**Figure 6**) grew by 30 per cent since 1993, the second fastest growth for all industries under analysis. However, hours in this industry have fallen by 13 percentage points between 2008 Q1 and 2009 Q4. Until the beginning of 2006, labour composition was in fact negative, indicating a weakening skill profile relative to the base period. Labour composition grew strongly between 2006 and 2008, before levelling off. The significant jump in labour composition between 2006 and 2008 is not explained by one single composition characteristic, but is a combination of age and gender interactions; the shift, however, has little to do with higher qualification levels.

The trend in unadjusted hours and labour composition in transport, storage and communication (**Figure 7**) is similar to that of wholesale and retail trade (Figure 5). Relative to 1993, there was a pure ‘quality’ effect of nine index points in 2009.

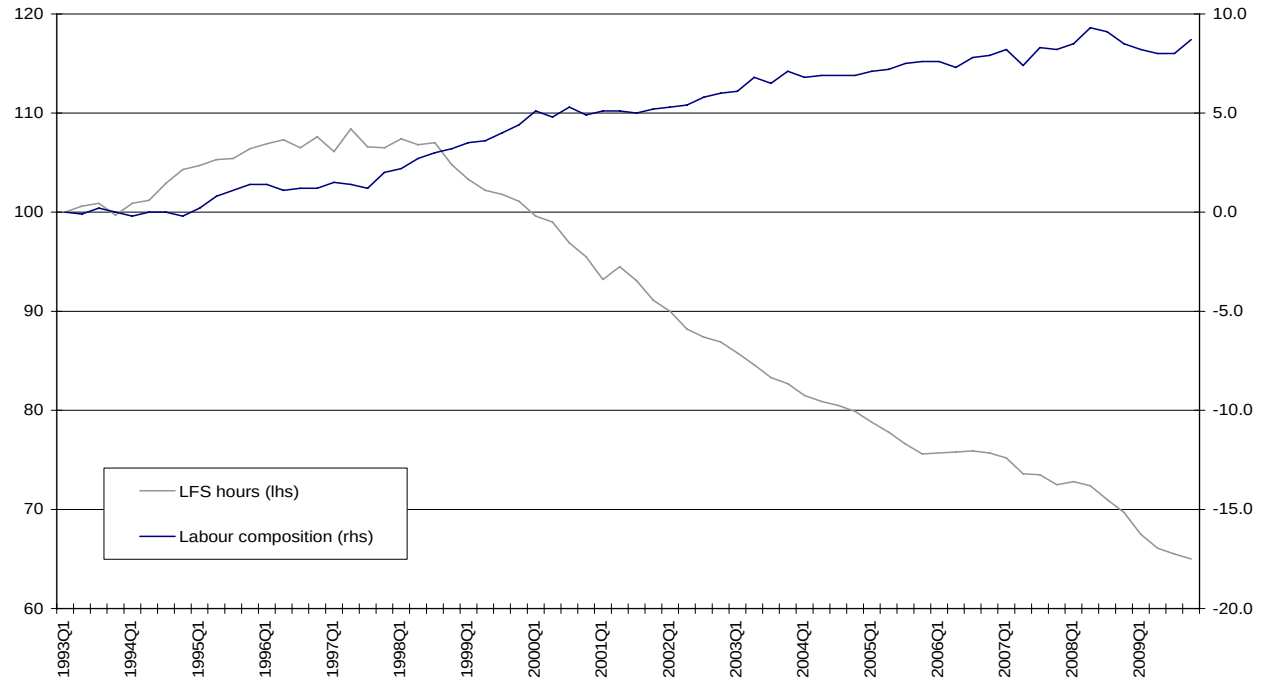
Figure 2 Agriculture, forestry, fishing; mining and quarrying; utilities (ABCE)

1993 Q1 = 100 Index points

Source: Office for National Statistics

Figure 3 Manufacturing (D)

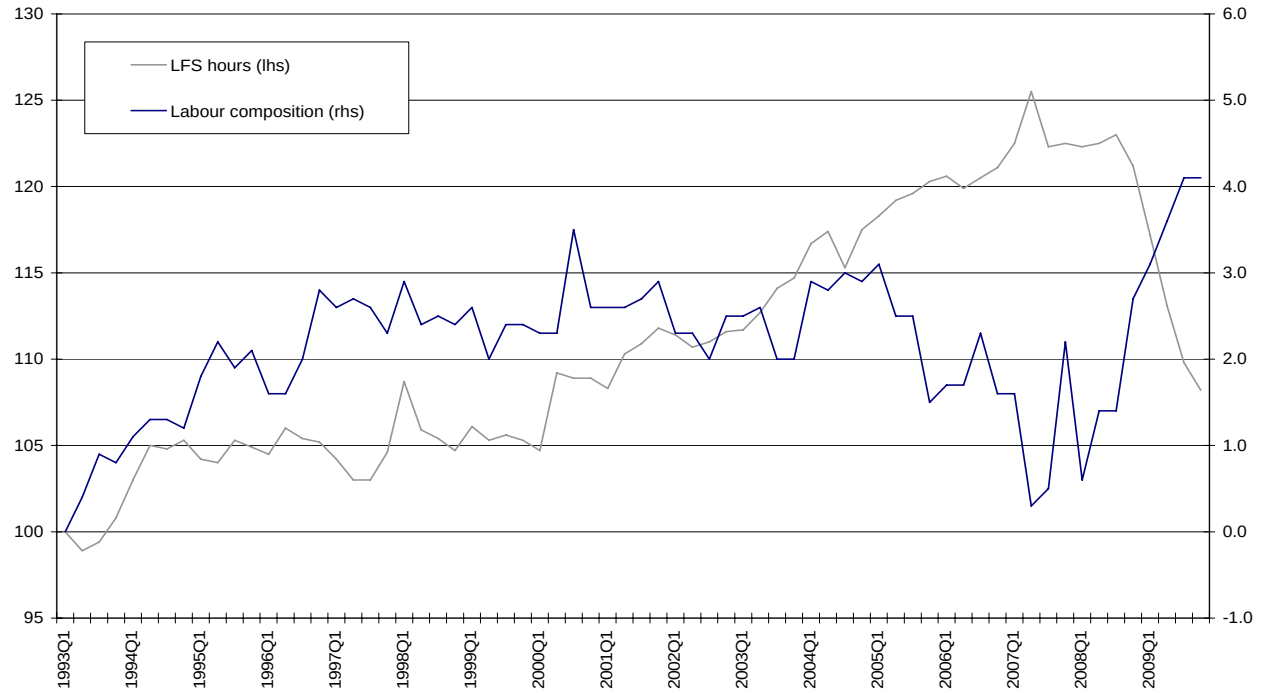
1993 Q1 = 100 Index points



Source: Office for National Statistics

Figure 4 Construction (F)

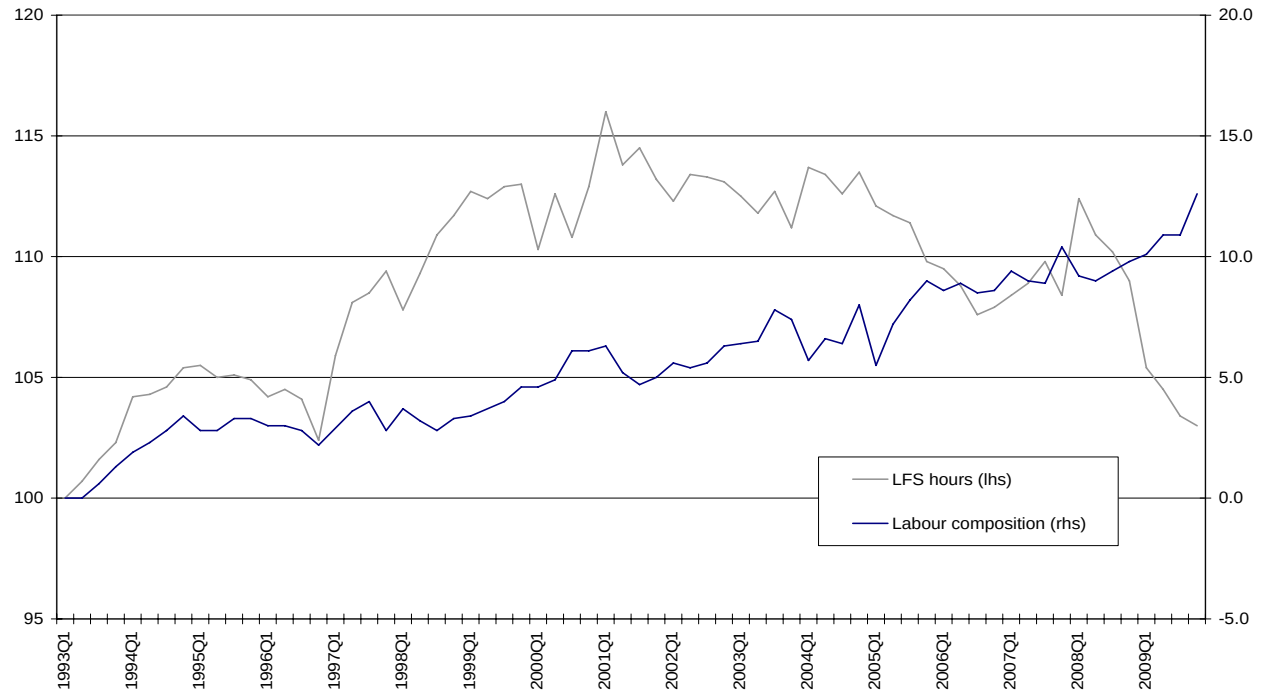
1993 Q1 = 100 Index points



Source: Office for National Statistics

Figure 5 Wholesale and retail (G)

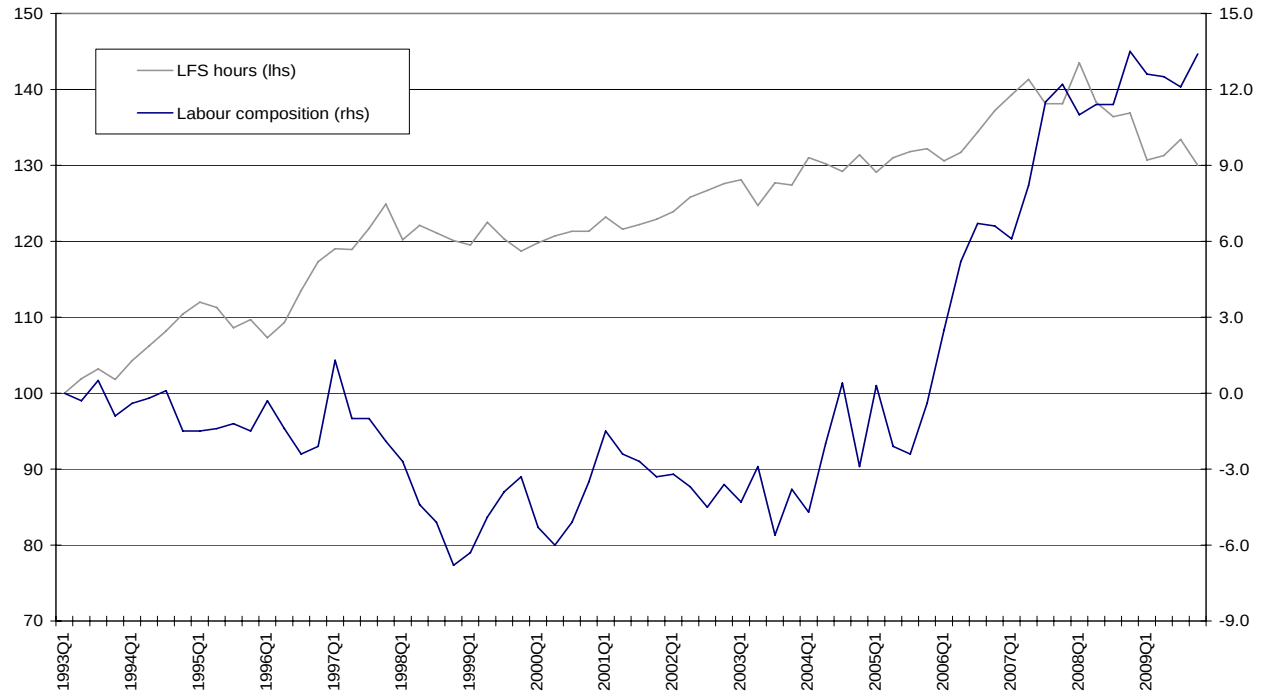
1993 Q1 = 100 Index points



Source: Office for National Statistics

Figure 6 Hotels and restaurants (H)

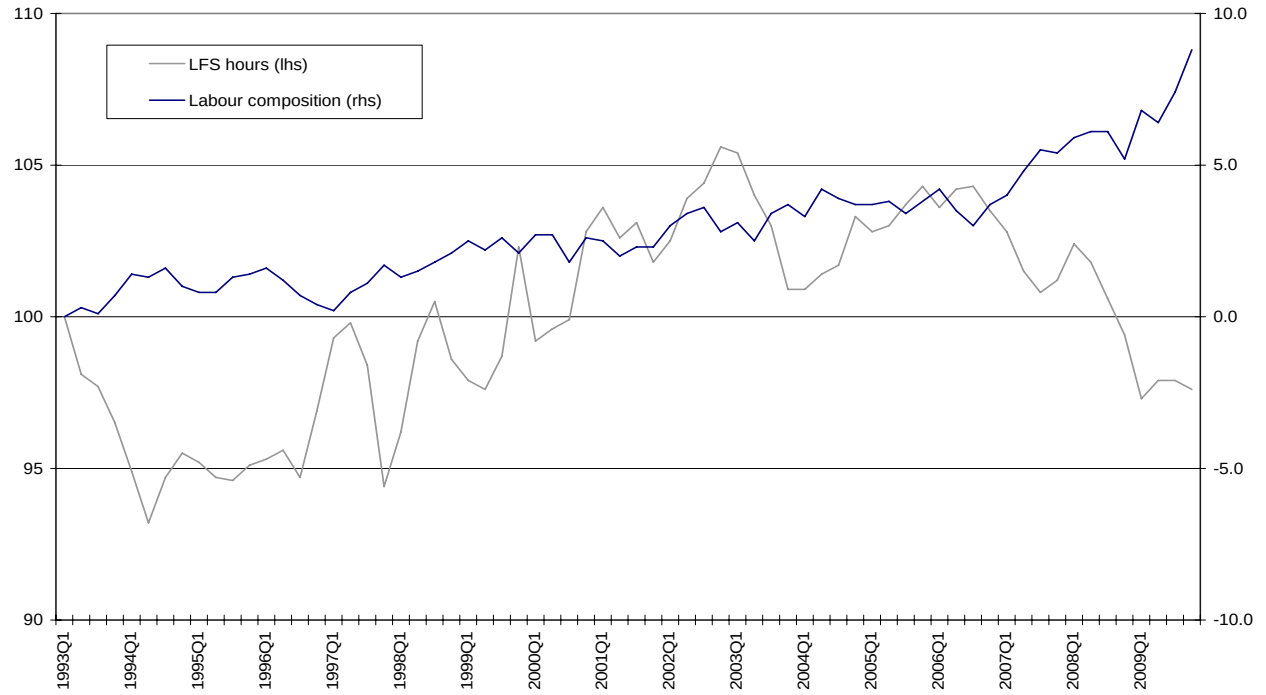
1993 Q1 = 100 Index points



Source: Office for National Statistics

Figure 7 Transport, storage and communication (I)

1993 Q1 = 100 Index points



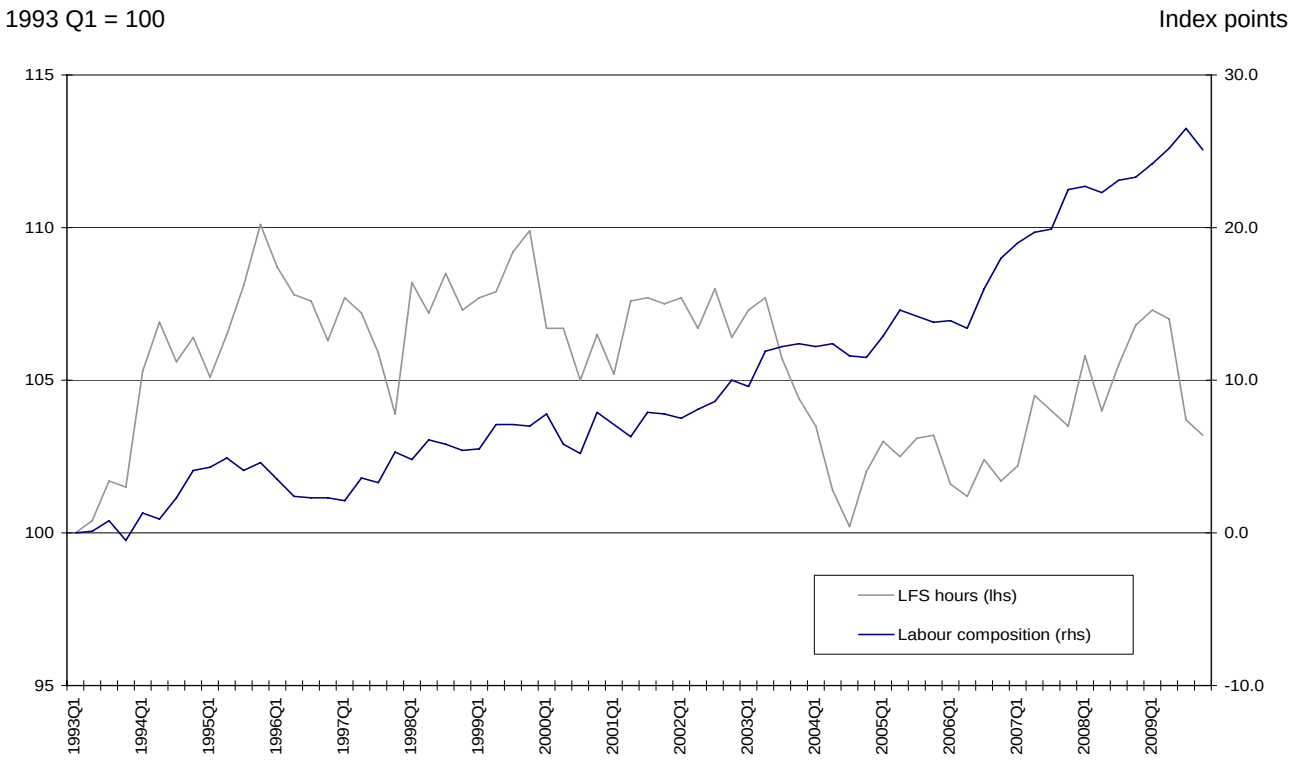
Source: Office for National Statistics

Unadjusted hours in financial intermediation (**Figure 8**) actually grew during the first half of the recession, before falling back sharply in 2009. Considering the importance of this industry to the wider economy, there has not been significant growth in hours since 1993, and it is perhaps surprising that hours have not fallen over the recession. Labour composition, however, has grown faster over the entire period than in any other industry. The proportion of QALI (obtained by adding labour composition to the unadjusted hours series) that is driven by quality is the highest for any industry. Labour composition remained on an upward trend over the recession except for a slight dip in the last quarter of 2009.

Growth in unadjusted hours for real estate, renting and business activities (**Figure 9**) was greater than in any other section: it grew by 68 percentage points, relative to the first quarter of 1993. However, labour composition has only been positive since 2002, that is to say the quality–adjustment has only led to a higher overall labour input since that time. In 2009 Q4 labour composition rose to its highest since the series began: 8.5 index points. The recession does not appear to have impacted on labour composition (although its contribution to QALI is quite small anyway).

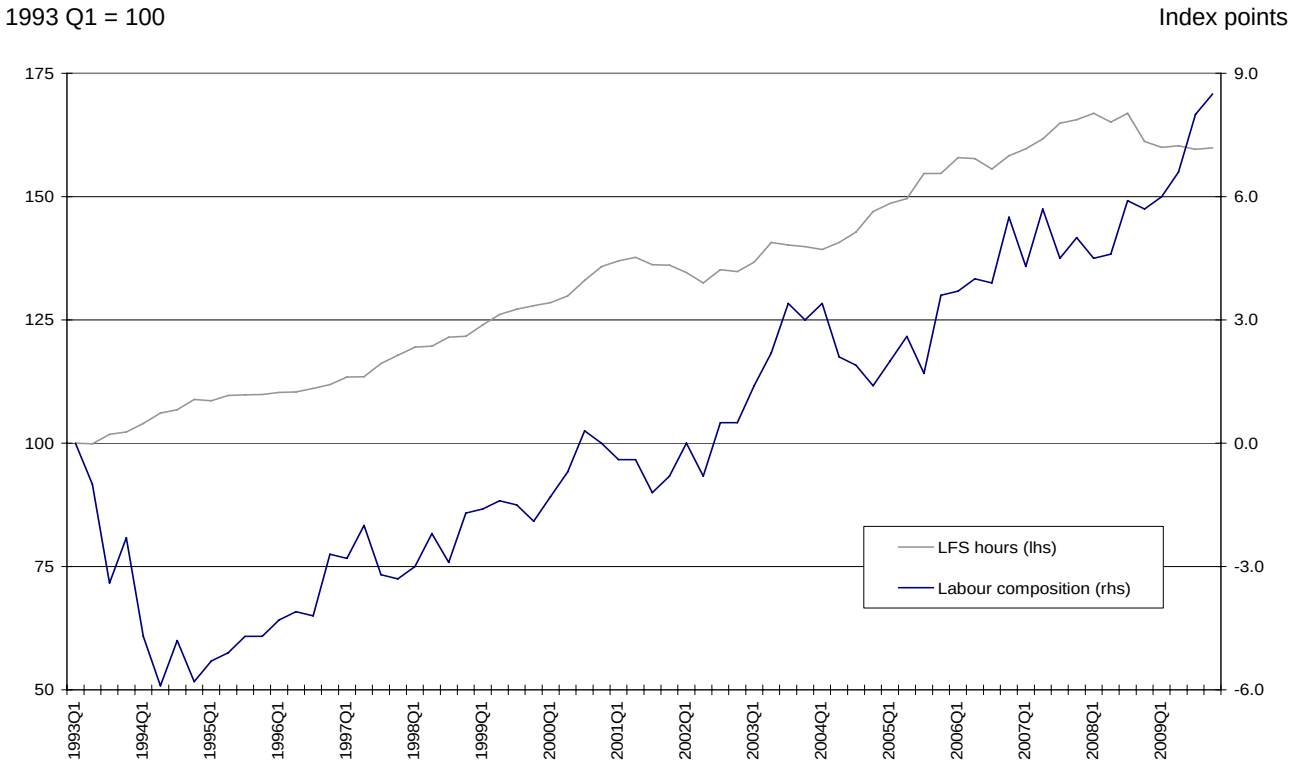
The trends in both industries are reflective of their importance to the wider economy over this period. In real estate, renting and business activities, hours worked grew very quickly as the volume of labour supply increased to meet demand, while the financial sector attracted many highly–skilled workers in order to add value. Both suffered over the recession but, from a labour input perspective, real estate, renting and business activities is recovering more quickly.

Figure 8 **Financial intermediation (J)**



Source: Office for National Statistics

Figure 9 Real estate, renting and business activities (K)



Source: Office for National Statistics

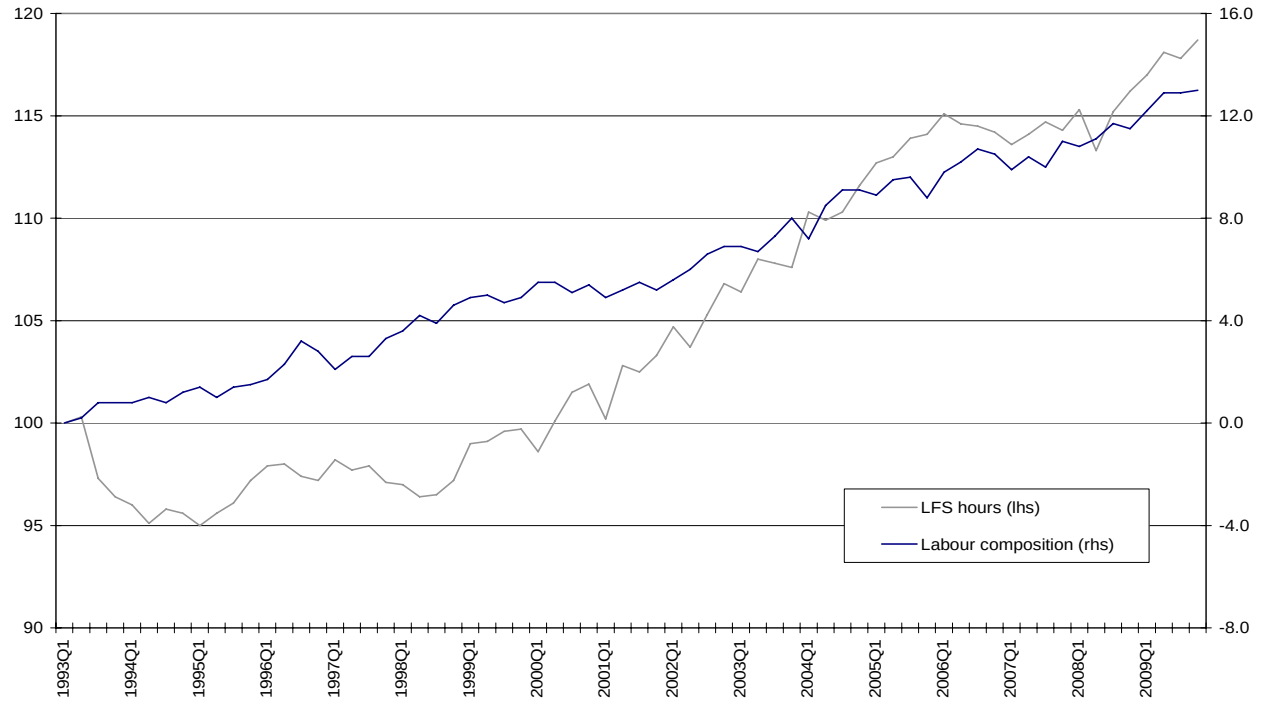
Public administration, education and health (**Figure 10**) has seen a strong increase in both labour composition and unadjusted hours over the recession. This result is unique to this industry. Other personal and social services (**Figure 11**) also exhibited strong growth in labour composition over the recession. Over the period as a whole, personal and social services experienced an increase in labour composition that was only slightly below that of financial intermediation. The output of these services (in gross value added terms) is, however, smaller than that of financial intermediation services.

Figure 12 presents labour composition and an index of unadjusted hours for the market sector. These series begin in the first quarter of 1994, the earliest date the variables used to filter out non–market sector responses can be found in the LFS. The trends are very similar to those of the whole economy (Figure 1). However, labour composition for the whole economy is slightly higher than the market sector throughout the entire period. Considering the impact of quality–adjusting labour in the predominantly non–market public services (Figure 10), this result is not surprising. Market sector hours have fallen by more over the recession than whole economy, again reflecting the impact of public services on the whole economy aggregate.

The above analysis by industry focuses on labour composition by gender, age and qualification; however it is possible to run the analysis by focusing on one of these compositional elements and weighting it by the others and by industry of employment. For example, analysis by qualification computes QALI weighted by gender, age and industry of employment; QALI by age weights by qualification, gender and industry; and QALI by gender weights by qualification, age and industry.

Figure 10 Public services (LMN)

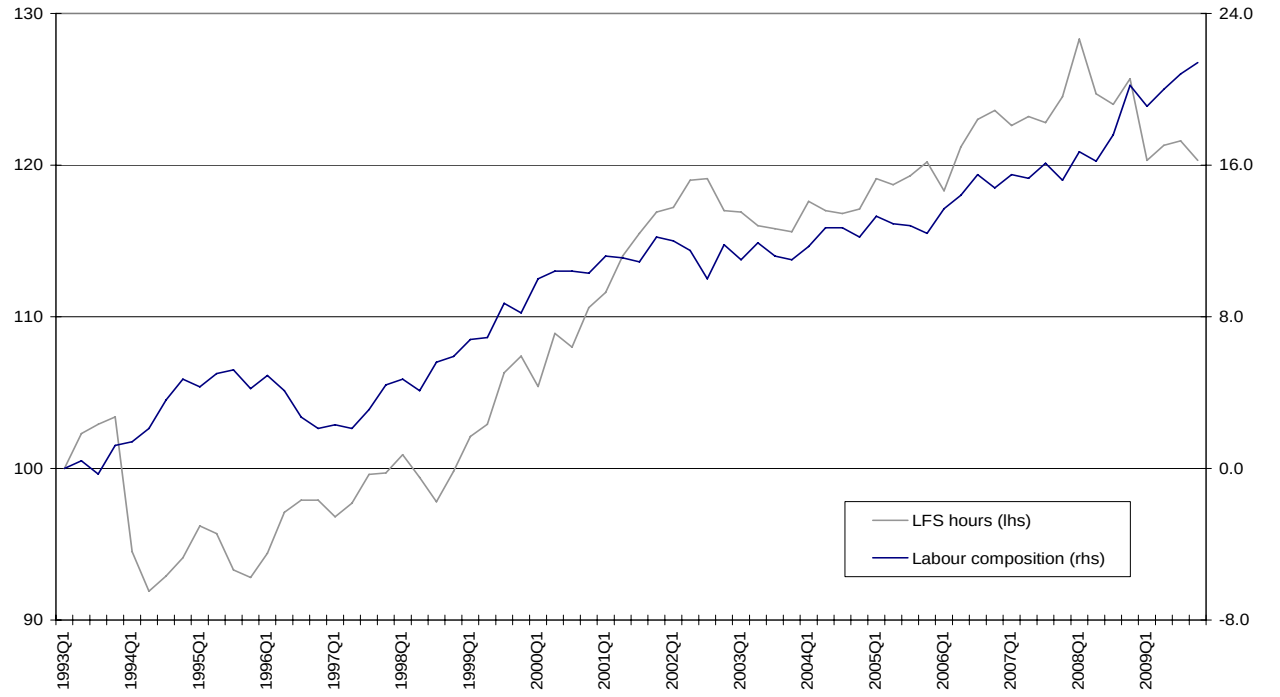
1993 Q1 = 100 Index points



Source: Office for National Statistics

Figure 11 Other services (OPQ)

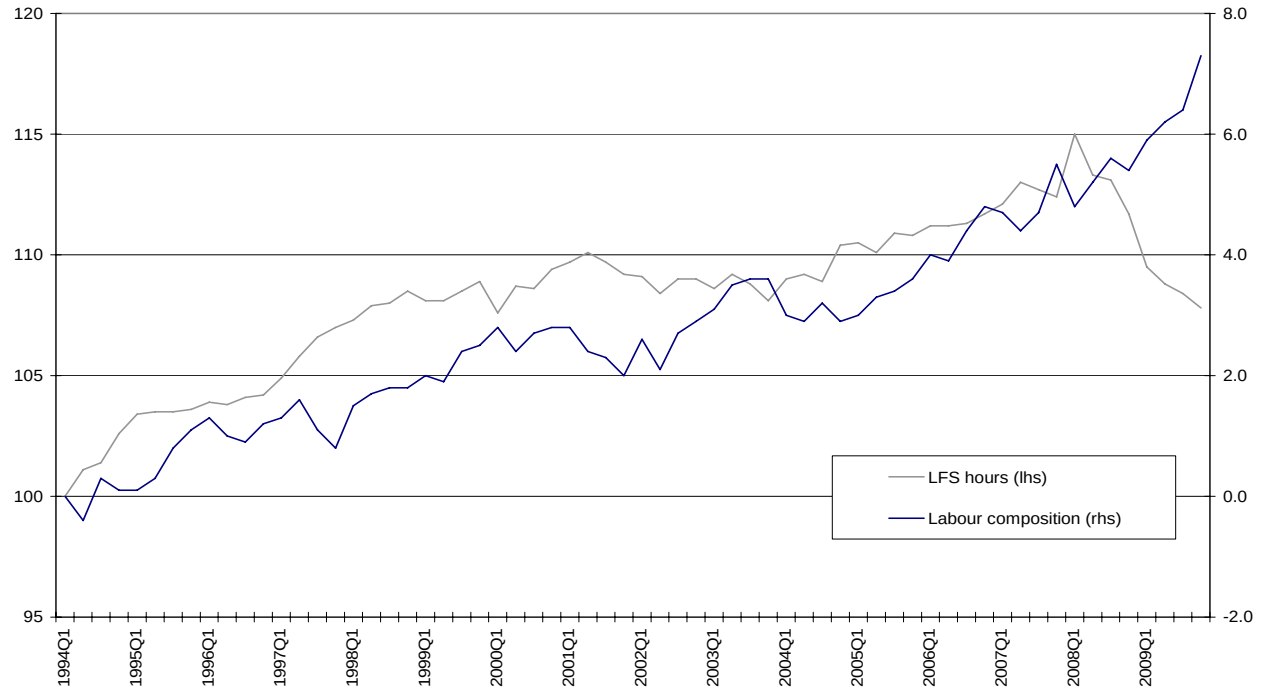
1993 Q1 = 100 Index points



Source: Office for National Statistics

Figure 12 **Market sector**

1994 Q1 = 100 Index points



Source: Office for National Statistics

Quality–adjusted labour input by education

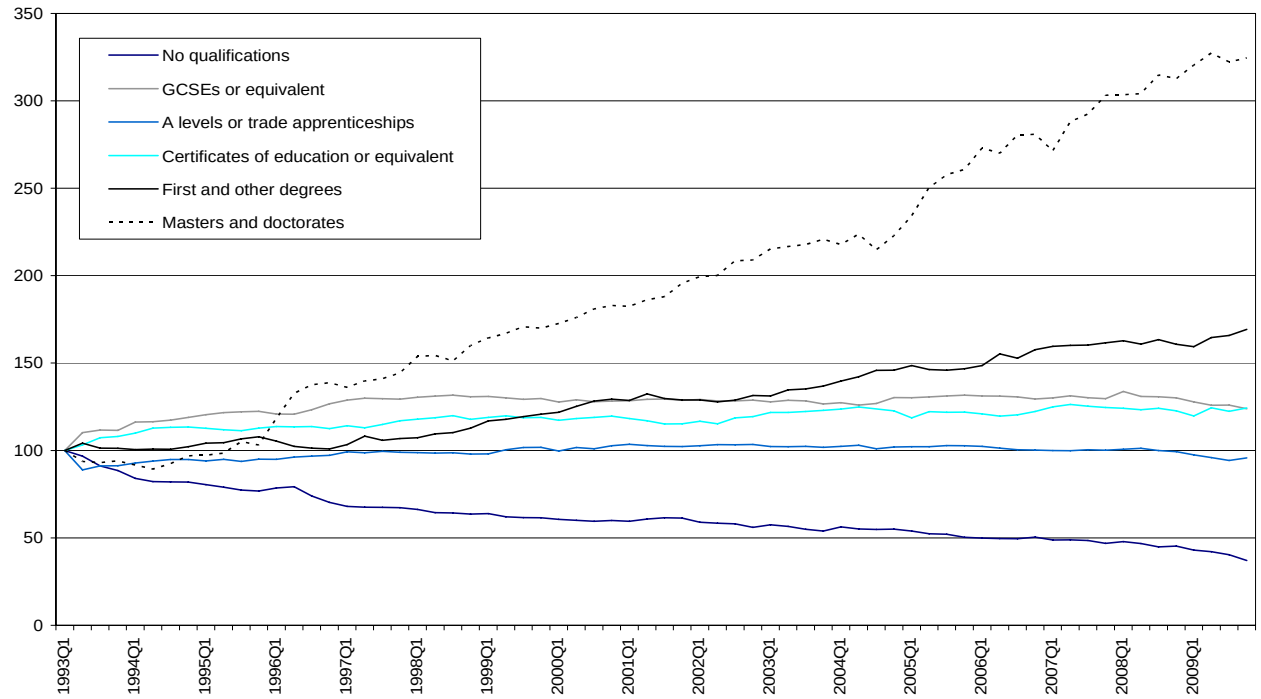
Figures 13 and 14 show estimates of unadjusted hours and labour composition by highest qualification[7].

Unadjusted hours for masters and doctorates grew at a much faster rate than any other qualification level (**Figure 13**): this series increased by 300 percentage points over the entire period. But the same cohort also saw rapid decline in their labour composition: it decreased by 19 percentage points between 1993 and 2009 (**Figure 14**). This suggests those with higher qualifications do not automatically take up high–paying jobs. It could also be related to the other labour characteristics of workers with these qualifications; for example their age, gender and industry of employment could result in lower earnings, which would impact on the labour composition result. Table 2 demonstrates that although relative remuneration for this cohort is high, with such high growth in the volume of labour it is inevitable that at the margin there will be some workers earning less, particularly if this cohort has a young age profile or increasingly works in low–paying industries.

Combining the two results from Figures 13 and 14, the rapid increase in the volume of labour taken with the overall fall in labour composition for this group indicates an over–qualification of labour input for these workers. However, over the recession, labour composition for masters and doctorates increased, suggesting that higher skill levels offer, on average, the best protection during economic downturns.

Figure 13 Unadjusted hours by qualification level

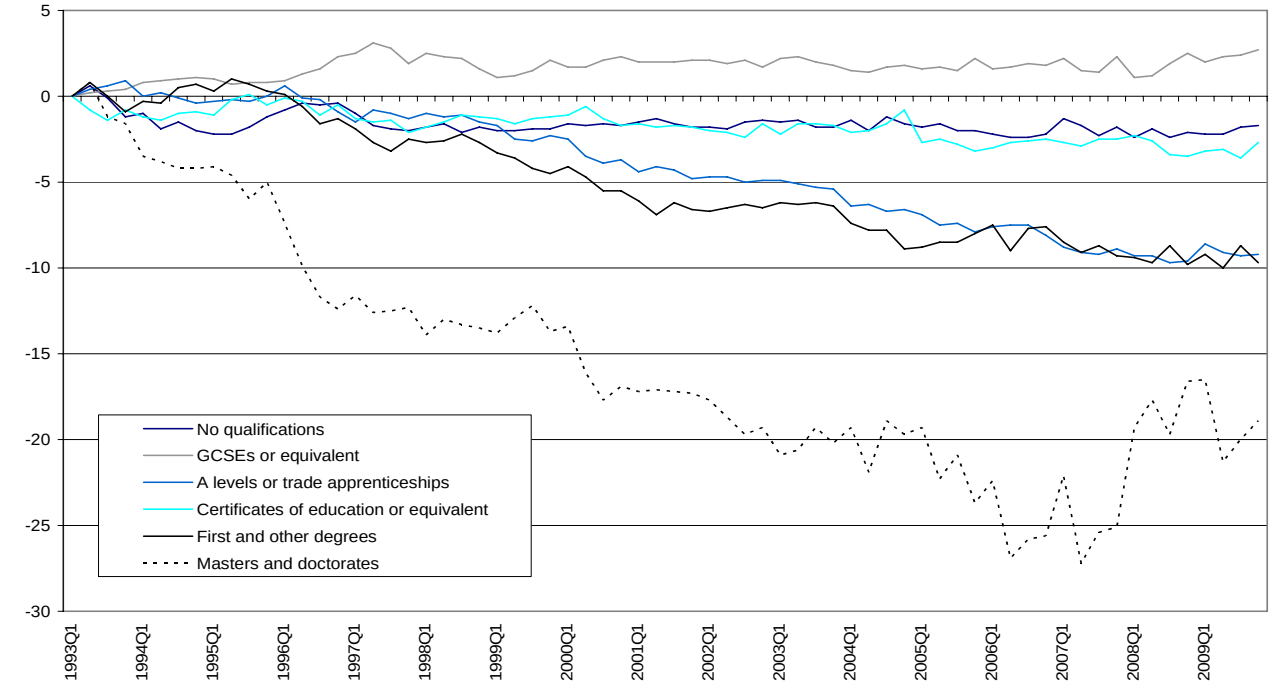
1993 Q1 = 100



Source: Office for National Statistics

Figure 14 Labour composition by qualification level

Index points



Source: Office for National Statistics

Only those with no qualifications saw their hours decline over the entire period, with the recession not appearing to accentuate the overall trend of decline by much. Those with GSCEs or equivalent were the only cohort to see positive labour composition throughout the period; this could be related to their age profile as well as the industry they work in. Those with first or other degrees saw their hours increase slightly and labour composition stabilise over 2008 and 2009. Similarly to those with masters and doctorates, this suggests that the higher qualified are better protected during recessions.

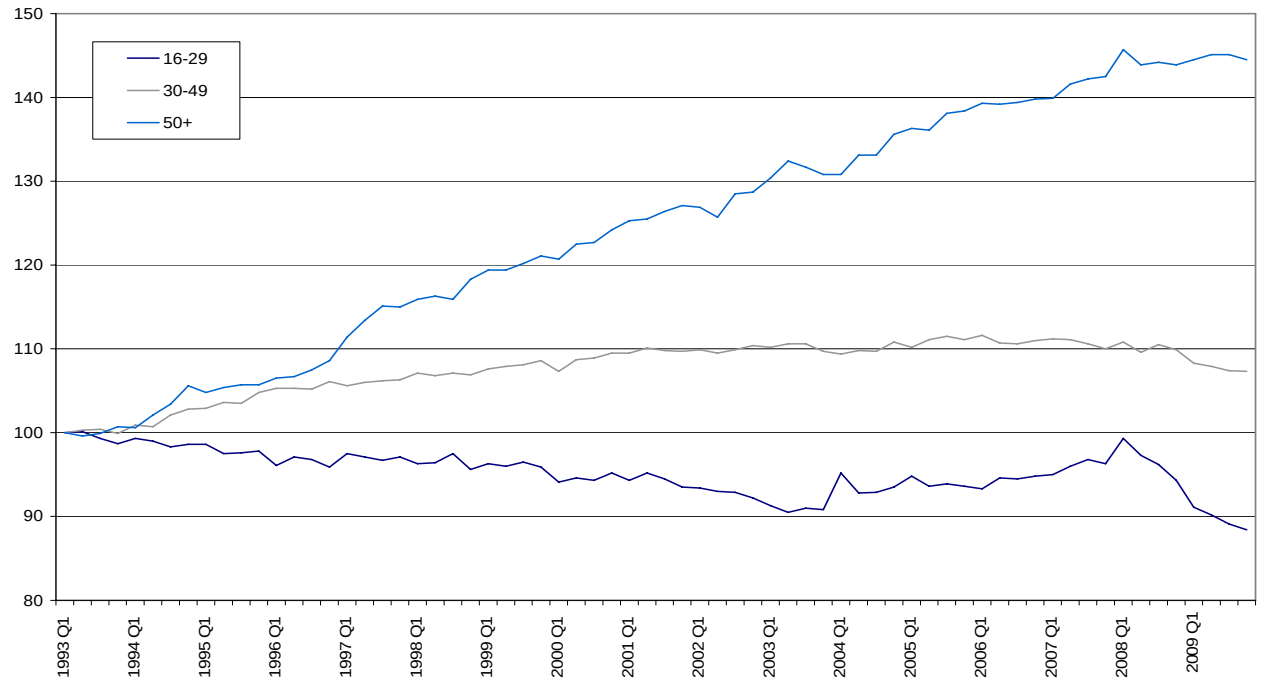
Quality–adjusted labour input by age

Figures 15 and 16 present indices of unadjusted hours and labour composition by age group. The unadjusted hours series increased by the most for the over–50s, reflecting their increased labour-force participation between 1993 and 2009[8]. Labour composition growth for this group easily exceeded that of the other two age groups. It is clear from looking at both indices that the over–50s were not so adversely affected by the economic downturn.

On the eve of the recession, unadjusted hours for workers aged 16–29 were back at 1993 levels, but an 11 point fall was recorded from Q1 2008 to Q4 2009, which highlights how much the recession disproportionately impacted on this cohort[9]. Labour composition rose over the period, but not by a substantial amount.

Figure 15 Unadjusted hours by age

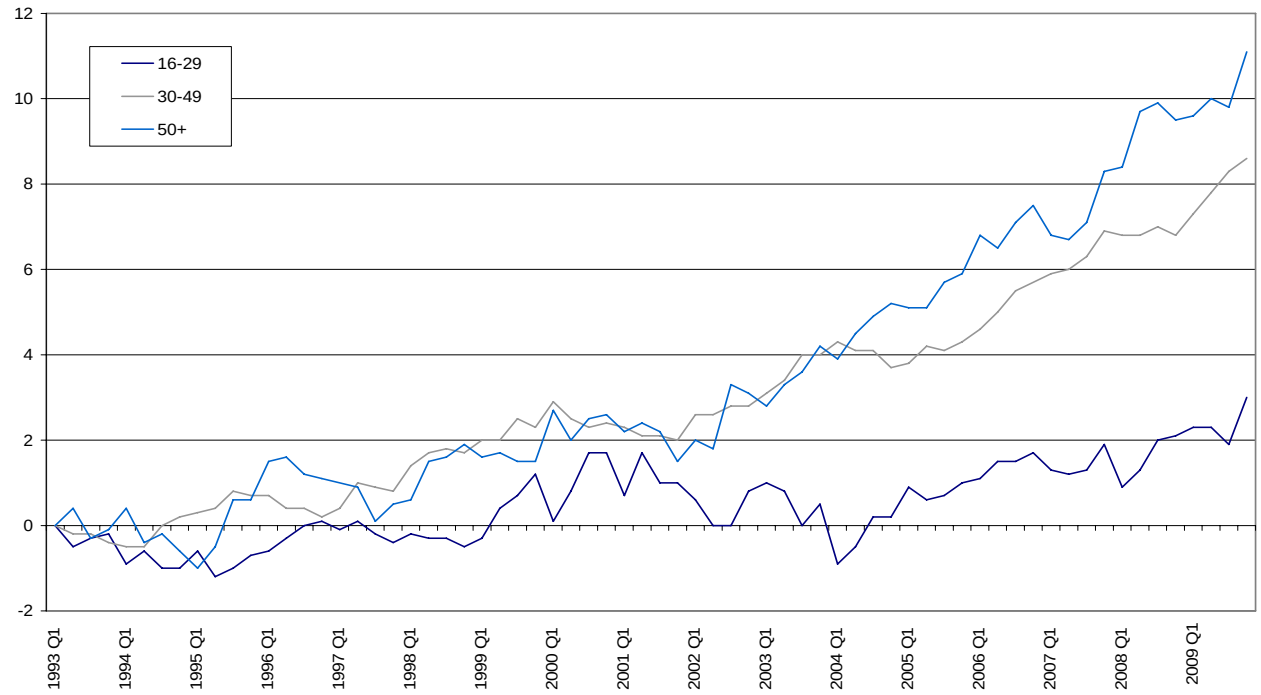
1993 Q1 = 100



Source: Office for National Statistics

Figure 16 Labour composition by age

Index points



Source: Office for National Statistics

Labour composition increased steadily for workers aged 30–49, and unadjusted hours also displayed an upward trend until 2006, although it was far weaker than that experienced by the over–50s. It has declined since then, with the recession having only a marginal impact on this rate of decline.

The increase in labour composition for all age groups over the recessions suggests a substitution away from the less skilled (as reflected in relative remuneration) in all age groups, but especially for those under 30 years.

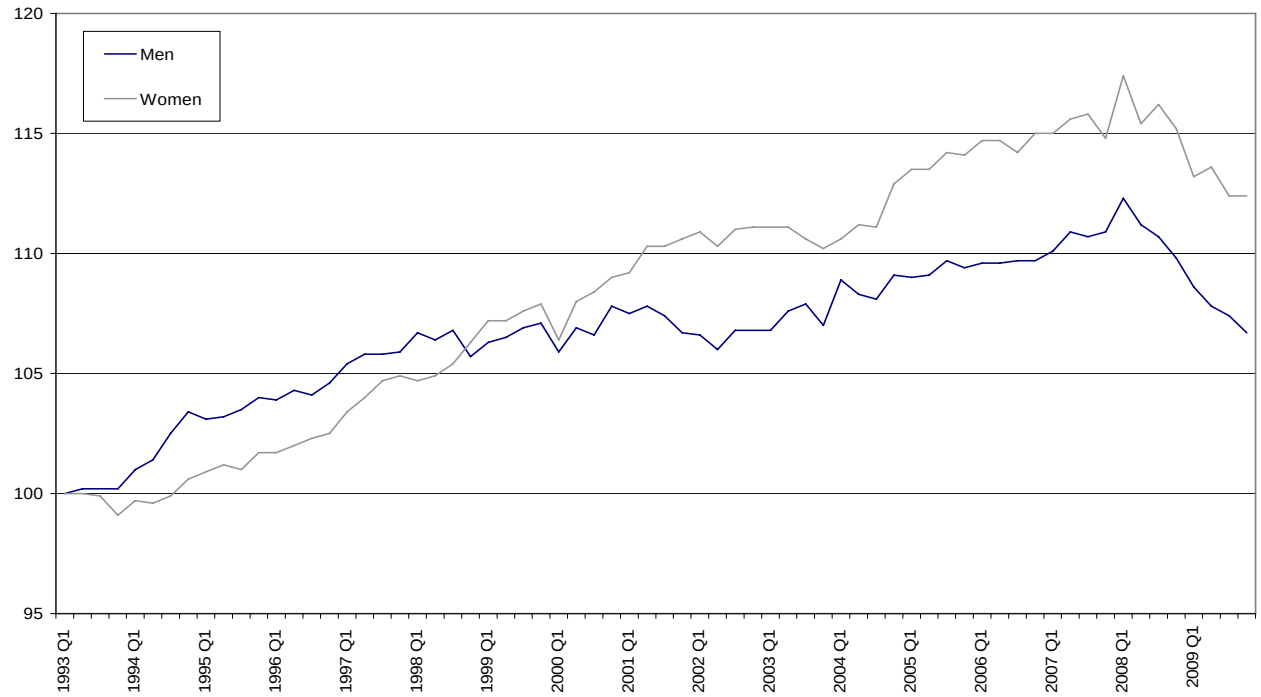
Quality–adjusted labour input by gender

Figures 17 and 18 compare estimates of unadjusted hours and labour composition for men and women. The gender variable is included in QALI in order to compensate for the fact that age is only a rough proxy for experience; using a gender variable compensates for this as empirically men and women have different working patterns, with women showing a greater tendency for career breaks and part–time work.

The unadjusted hours series (**Figure 17**) highlights the pure volume effect of increased participation by women in the labour market. While the graph indicates that, relative to 1993, both men and women have been similarly affected by the recession, the headline labour market statistics highlight that, on a levels basis, men were affected more by the recession[10].

Figure 17 Unadjusted hours by gender

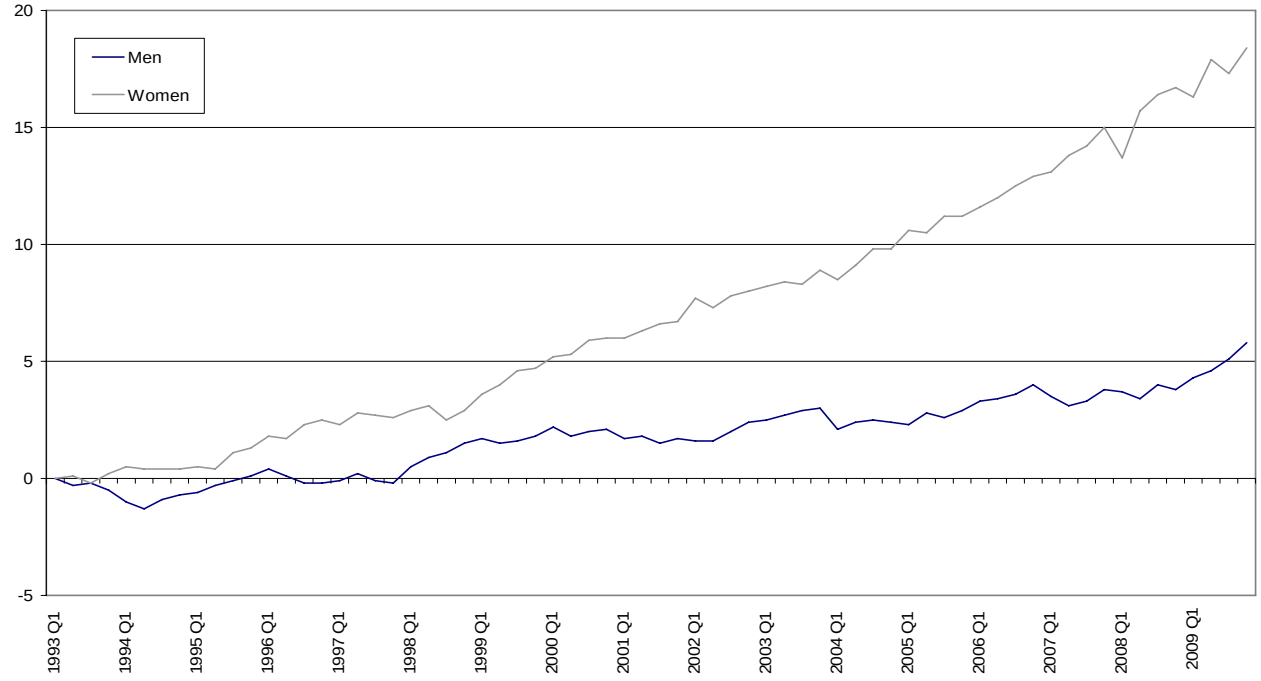
1993 Q1 = 100



Source: Office for National Statistics

Figure 18 Labour composition by gender

Index points



Source: Office for National Statistics

For labour composition (**Figure 18**), growth for women compared to men is striking. Reasons for the difference include the increased participation of women in the labour market and improving educational attainment for women relative to men. It is also possible that it picks up some effects of equal pay legislation mandating a convergence between male and female compensation. To the extent that it shows the existence (but weakening) of discrimination in the labour market, it is demonstrating the weakness of using pay as a proxy for marginal productivity.

EU KLEMS backcasting

One advantage of the current compositional breakdown is that it aligns QALI quite closely with the 'labour services' series generated by the EU KLEMS project[11]. The quality–adjustment in EU KLEMS has identical age and gender categories to QALI but fewer education categories: it uses three categories compared to six in QALI. This allows the EU KLEMS data to be produced at a greater industrial breakdown. Like QALI, EU KLEMS assumes that, other things equal, the compensation per hour of a self–employed person is equal to that of an employee. EU KLEMS also employs the Törnqvist index, but on an annual basis.

It is possible to map the EU KLEMS industries to QALI industries because the former is produced at a more disaggregated level. QALI industries, that is the ten industries listed in Table 1, can be obtained by aggregating particular EU KLEMS industries together.

Conceptually 'labour services' and QALI are similar, and in their derivation at industry level it is only the educational categories that differ significantly. As the two are so similar, it is reasonable to use the growth profile of the 'labour services' series, which extends back to 1970, to backcast QALI on an annual basis. Furthermore, the educational differences can be assumed to narrow the further into the past we go. 1993 is taken as the base year, so that a comparison can be drawn with the quarterly analysis in the previous section.

Table 3 examines the differences between the QALI system and the EU KLEMS system by looking at average annual growth rates between the years of overlap, 1993 to 2007. The difference is greatest for unadjusted hours in agriculture, forestry, fishing, mining and utilities (ABCE) and other services (OPQ). In general, the differences in labour composition are smaller than the differences in hours[12]. For the majority of industries the differences are small and suggest the results produced by the two systems are similar enough to make the backcasting a plausible exercise.

At the whole economy level (**Figure 19**), the pattern in hours growth prior to 1993 is quite different to that from 1993 onward, with the former showing a smooth upward trend until the most recent recession. It can clearly be seen that larger falls in hours were recorded for previous recessions in the early 1980s and early 1990s. Labour composition, however, has risen quite steadily between 1970 and 2009, despite the changing economic cycles, with the average annual growth rate increasing since 1993.

Table 3 QALI and EU KLEMS: Average annual growth rates, 1993–2007

Percentages

	Labour Composition			Unadjusted Hours		
	QALI	EU KLEMS	Difference	QALI	EU KLEMS	Difference
Whole economy	0.4	0.6	**-0.2**	0.9	0.8	**0.1**
ABCE: Agriculture etc; mining and quarrying; utilities	0.5	0.2	**0.3**	-1.1	-2.9	**1.8**
D: Manufacturing	0.7	0.9	**-0.2**	-2.1	-2.1	**0.0**
F: Construction	0.0	0.3	**-0.3**	1.5	0.8	**0.7**
G: Wholesale and retail trade	0.6	0.5	**0.0**	0.5	0.4	**0.2**
H: Hotels and restaurants	0.5	0.7	**-0.2**	2.3	1.6	**0.7**
I: Transport, storage and communications	0.3	0.4	**-0.1**	0.3	0.4	**-0.1**
J: Financial intermediation	1.3	1.2	**0.1**	0.2	0.8	**-0.6**
K: Real estate, renting and business activities	0.3	0.8	**-0.4**	3.5	3.6	**-0.1**
LMN: Public admin and defence; education; health and social work	0.6	0.7	**-0.1**	1.1	1.3	**-0.2**
OPQ: Other social and personal services	0.8	0.8	**0.0**	1.4	2.9	**-1.5**

Source: Office for National Statistics

Figure 19 Whole economy backcasting

Source: Office for National Statistics

In **Table 4**, the annual growth rates for the periods 1970–1993 and 1993–2009 are averaged in order to provide a picture of movements in unadjusted hours and labour composition across the entire period.

Table 4 Average annual growth rates, 1970–1993 and 1993–2009

Percentages

	1970-1993		1993-2009	
	Hours	**Labour Composition**	**Hours**	**Labour Composition**
Whole economy	-0.2	0.3	0.5	0.5
ABCE: Agriculture etc; mining and quarrying; utilities	-3.0	-0.6	-1.0	0.2
D: Manufacturing	-2.8	0.2	-2.6	0.7
F: Construction	-0.5	0.2	0.6	0.1
G: Wholesale and retail trade	0.5	0.0	0.1	0.6
H: Hotels and restaurants	1.8	-0.1	1.3	0.6
I: Transport, storage and communications	-0.7	0.0	-0.1	0.4
J: Financial intermediation	2.4	0.4	0.2	1.3
K: Real estate, renting and business activities	3.1	1.4	2.8	0.4
LMN: Public admin and defence; education; health and social work	1.2	0.7	1.1	0.6
OPQ: Other social and personal services	2.0	1.4	1.3	1.0

Source: Office for National Statistics

Agriculture, forestry, fishing, mining and utilities (ABCE) has seen a turnaround in labour composition between the two periods, with positive annual growth rates, on average, from 1993 onward compared to negative average growth rates previously. The rate of decline in hours slowed substantially in the later period, possibly due to the effects of earlier privatisation in mining and utilities fading. In manufacturing (D), hours have been declining throughout the entire period but at a slower rate since 1993. Labour composition has risen at a much higher average annual rate since 1993, which suggests that employment in high–paying manufacturing jobs has increased relative to low–paying jobs, and the extension of this is that the skill profile in manufacturing has improved markedly over the time period.

With the exception of transport, storage and communication (I) and hotels and restaurants (H), the services industries all show positive average growth rates in both hours and labour composition across the entire period. An increase in the average growth rate for labour composition has been especially marked in financial intermediation (J). However, its growth rate decreased for hours, indicating that employment in this industry is becoming dominated by relatively well remunerated workers. Although other services has the second highest average growth rate for labour composition in the later period, this is still down by a noticeable amount compared to the earlier

period. Real estate, renting and business activities (K) has seen labour composition growth decline by a percentage point, the largest decline by far for all industries.

These growth rates reflect the compositional movements of the employed workforce; the age, qualification level and predominant gender of workers in these industries may explain some of the changes, or it could be related to structural change in the industry in question. However, nothing can be said about the relative strengths of these effects because the EU KLEMS dataset only reports ‘labour services’ by industry.

Conclusion

This article presented new estimates of quality–adjusted labour inputs between the years 1993 and 2009. As the adjustment categories show a high level of consistency with the EU KLEMS ‘labour services’ measure, backcasting the QALI time series on an annual basis to 1970 was also possible.

Making explicit adjustment for the quality of the workforce has the effect of increasing estimated labour input to the production process. The magnitude of the quality–adjustment, or labour composition, at the whole economy level has been increasing over time. The observed impact tends to be greatest in service industries, with financial intermediation exhibiting the largest increase in labour composition, while growth in QALI itself was highest in real estate, renting and business activities. For the majority of industries, labour composition increased over the recession whilst growth in hours declined. These results suggest that the low-skilled suffered more than the high-skilled over the downturn as the employed workforce moved to higher–remunerated QALI categories, on average.

Within adjustment groups, the overall trend has been towards growth in hours worked by those with undergraduate and postgraduate qualifications at the expense of workers with little or no qualification, and those over 50 compared to the young. Growth in both QALI and labour composition has been far stronger for women than for men over the period studied. The recession had little impact on older workers whilst it impacted quite notably on younger workers. The labour market position of those with the highest level of qualifications improved over the downturn.

Analysing labour inputs over a longer period, using the growth profile of the ‘labour services’ series, highlights how the growth of labour composition and hours, and thus QALI, have changed over time. Financial intermediation and real estate, renting and business activities, in particular, have seen notable growth rate changes since the 1970s.

Notes

1. The EU KLEMS project, which ran from 2003 to 2008, created a database on measures of economic growth, productivity, employment creation, capital formation and technological change at the industry level for all European Union member states from 1970 onwards. Its equivalent of QALI is referred to as ‘labour services’. Further details available at: http://www.euklems.net/

2. Employment is a head–count of all participants in the labour market. A jobs measure differs from head–count measures like employment as it takes account of workers with second jobs.

3. Productivity hours and productivity jobs are series used in the calculation of headline ONS labour productivity measures, available at www.statistics.gov.uk/STATBASE/Product.asp?vlnk=7476. They provide conceptually consistent measures of the volume of labour input for productivity purposes as they are produced using more reliable industry breakdowns, from both short–term and annual business surveys, which are constrained to LFS aggregates.

4. Productivity jobs and productivity hours were both extracted from the Q3 2010 Labour Productivity Statistical Bulletin for use in this article. However, there have been revisions to both series since that release, which are available in the Q4 2010 bulletin.

5. The calculation is performed by dividing gross annual pay by total hours for each QALI category, with both variables obtained from the LFS micro–data.

6. QALI is a Törnqvist weighted index of hours, where the weights are labour income shares. The difference between this index and an unadjusted index of hours from the LFS represents 'labour composition', expressed in index points. Where QALI is greater than hours, this implies that labour composition is moving towards QALI categories with higher wages, on average.

7. These refer to National Vocational Qualifications (NVQ), which are recorded in the LFS microdata. For ease of interpretation, descriptive categories are used, but these refer to NVQ levels.

8. See 'Older people in the labour market' for further details: www.statistics.gov.uk/cci/nugget.asp?id=2648

9. See 'Graduates in the labour market' for further details: www.statistics.gov.uk/CCI/nugget.asp?ID=1162

10. See the Labour Marked Statistical Bulletins for further details: www.statistics.gov.uk/STATBASE/Product.asp?vlnk=1944

11. See O'Mahony and Timmer (2009) for further details on EU KLEMS and Dunn, Goodridge and Turvey (2010) for an evaluation of EU KLEMS.

12. This suggests an alternative methodology for backcasting may be possible that involves using EU KLEMS labour composition combined with an ONS–produced hours series.

Acknowledgements

The author wishes to thank Mark Franklin and Peter Goodridge for their help in producing this article.

Contact

elmr@ons.gov.uk

References

Appleton (2011) ‘Volume index of capital services: new annual and quarterly estimates for 1950 to 2009’, *Economic & Labour Market Review* forthcoming

Bell V, Burriel–Llombart P and Jones J (2005) ‘A quality–adjusted labour input series for the United Kingdom (1975-2002)’, *Bank of England Working Paper* No. 280

Dunn R, Goodridge P and Turvey A (2010) ‘Evaluation of the EU KLEMS analytical database – Labour Accounts and Intermediate Inputs’. Report compiled for Eurostat.

EU KLEMS (2007a) ‘EU KLEMS Growth and Productivity Accounts, Part 1: Methodology’

EU KLEMS (2007b) ‘EU KLEMS Growth and Productivity Accounts, Part 2: Sources by Country’

Goodridge P (2007) ‘Methods Explained: Index Numbers’, *Economic and Labour Market Review* 1(3), pp 54-57 and at www.statistics.gov.uk/cci/article.asp?ID=1743

Goodridge P (2009) ‘Quality–adjusted labour input: estimates for 1997 to 2007’, *Economic and Labour Market Review* 3(1), pp 62-66 and at www.statistics.gov.uk/cci/article.asp?ID=2102

O’Mahony M and Timmer M.P (2009) ‘Output, input and productivity measures at the industry level: the EU KLEMS database’, *The Economic Journal* 119 (June), pp 374–403.

Organisation for Economic Co–operation and Development (2001) ‘Measuring Productivity – OECD Manual’.

Turvey A (2009) ‘Developing a unit labour costs indicator for the UK’, *Economic and Labour Market Review* 3(6), pp 51-56 and at www.statistics.gov.uk/cci/article.asp?id=2218

Turvey A, Goodridge P and Franklin M (2010) 'Quality adjusted labour input: new estimates for 1993 to 2008', Economic & Labour Market Review 4(10), pp 113–131 and at www.statistics.gov.uk/cci/article.asp?ID=2596

Business Population Estimates for the UK and Regions

Introducing improved statistics on the UK enterprise population

Julian Shaw
Department for Business, Innovation and Skills

Summary

In May 2011 the Department for Business, Innovation and Skills will publish *Business Population Estimates for the UK and Regions 2010*. This National Statistics release is a continuation of the series in *Small and Medium Enterprise Statistics for the UK and Regions* but uses a new methodology to improve the quality of estimates of the number of enterprises[1]. This article sets out the detail of three major methodological changes and other key changes, and explains how they impact on the estimates by size, legal status, sector and region. To illustrate the combined impact of these methodological changes, the estimate of the number of enterprises at the start of 2009 is presented as a case study. Under the new methodology, there were an estimated 4.4 million enterprises at the start of 2009, compared to 4.8 million reported in *Small and Medium Enterprise Statistics 2009*.

Introduction

The Department for Business, Innovation and Skills (BIS) has previously published estimates of the total number of enterprises in the publication *Small and Medium Enterprise Statistics for the UK and Regions (SME Statistics)*. This National Statistics series ran from 1994 to 2009, containing data on the number of enterprises at the start of the year with their associated employment and turnover[2].

In December 2008 BIS launched a public consultation on *SME Statistics*, a key component of which was a set of questions on proposals to change the methodology (see BIS 2008). Respondents on the whole welcomed the ideas BIS put forward for improving the methodology. In June 2009 BIS published the *Government Response to the Consultation on Small and Medium Enterprise Statistics for the UK and Regions* (see BIS 2009). In this response BIS committed to further investigate the options for improving the methodology and implementing changes where appropriate.

BIS identified a number of areas where improvements could be made to the estimate of the number of enterprises. One key area is the over–count in the self–employment estimate used to produce *SME Statistics*. It is only recently that a change to the Labour Force Survey questionnaire has allowed a robust methodology to be developed for estimating the 'true' level of self–employment.

Other improvements were identified, such as more closely following ONS practice when selecting enterprises from the Inter–Departmental Business Register (IDBR) and using HMRC self–assessment data for the first time. Recent improvements to the timing of employment estimates on the IDBR have presented BIS with the opportunity to use an earlier IDBR data extract and produce estimates sooner than before. For the first time all the constituent data sources used in producing *SME Statistics* are available in the Standard Industrial Classification 2007 (SIC2007), allowing the publication to be based on this format.

In order to minimise the number of discontinuities in the series, all of the improvements to the methodology will be applied at the same time. *SME Statistics for the UK and Regions* has been renamed *Business Population Estimates for the UK and Regions* to ensure users are aware of the significant change in methodology between the 2009 and 2010 estimates.

Overview

This article describes the methodology used in *SME Statistics* and *Business Population Estimates*. The first section highlights where the three major methodology changes described in this article fit into the overall methodology.

The second section covers each of the three major methodology changes in detail. In summary these are:

- a change to the Labour Force Survey self–employment data used in producing the estimates. This change in isolation results in a 421,000 (9 per cent) decrease in the estimated overall number of enterprises at the start of 2009 (compared with *SME Statistics 2009*)
- a change to using HMRC self–assessment data in producing the estimates. This change in isolation results in a 71,000 (1 per cent) increase in the estimated overall number of enterprises at the start of 2009 (compared with *SME Statistics 2009*)
- a change to the Inter–Departmental Business Register (IDBR) data used in producing the estimates. This change in isolation results in a 30,000 (1 per cent) fall in the estimated overall number of enterprises at the start of 2009 (compared with *SME Statistics 2009*)

When all three of these major methodology changes are applied at the same, the overall estimate for the business population at the start of 2009 decreases from the published 4.8 million to 4.4 million (or by around 398,000 and 8 per cent). The third section describes this overall impact of the changes in more detail, showing results by size, legal status, industry sector and UK country/region.

The fourth section describes other changes that will be introduced with the publication of *Business Population Estimates*.

Summary of the estimation methodology

The methodology used in producing *SME Statistics* is described in this section. Although the broad methodology is the same for both publications, the methodology used to derive the underlying source data has changed. This is discussed in subsequent sections.

There is no single database in the UK which contains details on every active business. Further information on the different sources of enterprise population data is shown in **Box 1**.

Box 1 Sources of data on the enterprise population

There are several sources of information on the enterprise population. The three main official sources are:

- *UK Business: Activity, Size and Location* (ONS) provides data on all enterprises registered for either VAT and/or PAYE in the UK. These data refer to the number of enterprises active on a particular day (see ONS 2010a).
- *Business Demography* (ONS) provides data on all enterprises registered for either VAT and/or PAYE in the UK which are active and listed on the IDBR at any time during the full calendar year (see ONS 2010b).
- *Small and Medium Enterprise Statistics* (BIS) provides the only estimate of the total number of enterprises in the UK. Unlike the ONS publications *SME Statistics* incorporates an estimate of the un–registered enterprise population. These data show the estimated number of enterprises active on 1 January each year.

The total number of enterprises therefore has to be estimated using a number of different data sources. The *SME Statistics* methodology takes data from the approximately 2.2 million enterprises which are on the government business register, the Inter Departmental Business Register (IDBR). The IDBR contains all VAT and/or PAYE registered enterprises[3]. It is assumed that all active companies[4] are on the IDBR. The IDBR is used to count how many self–employed people are running registered businesses, either alone or in partnership. The methodology then estimates how many very small unregistered enterprises there are to create a total business population for the UK (see **Figure 1**).

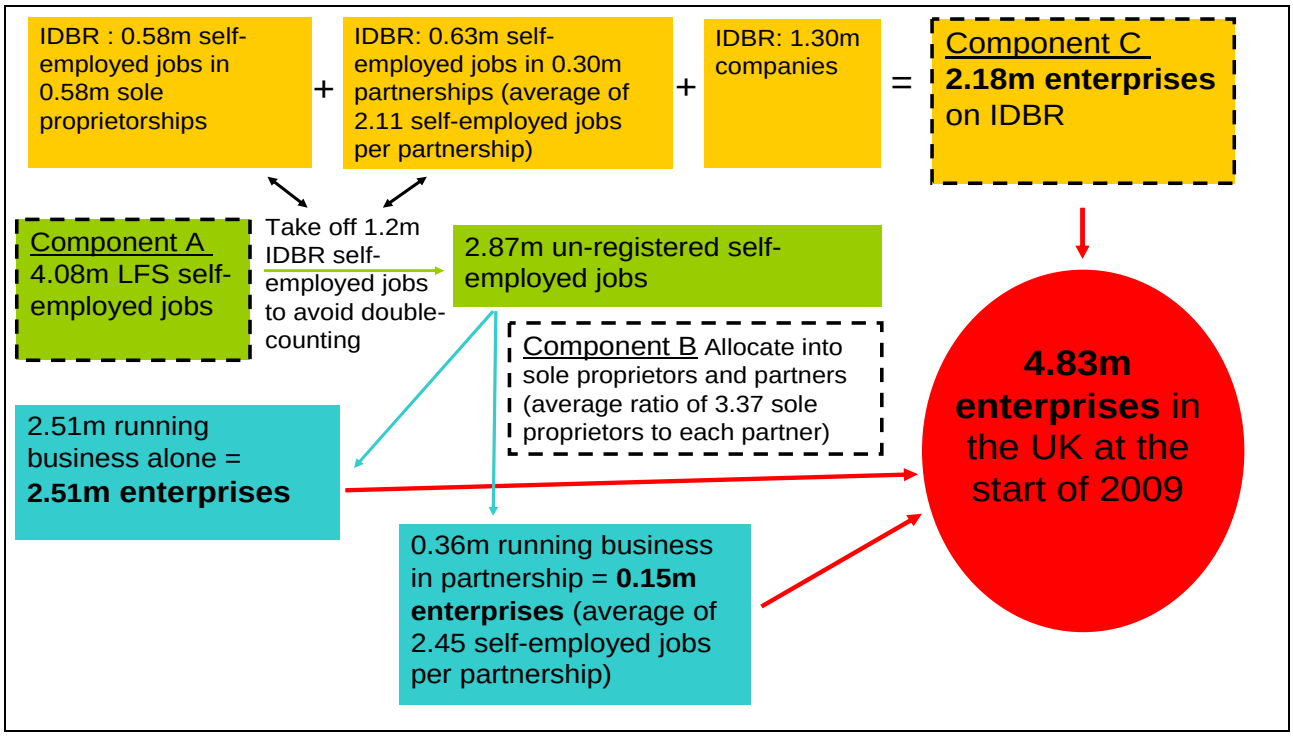
The ONS Labour Force Survey (LFS) is used to get estimates of all self–employed jobs. In the *SME Statistics* methodology this involves adding together all main job and all second job self–employed (the LFS does not ask about subsequent jobs). By subtracting the IDBR total self–employed jobs (in sole proprietorships or partnerships) from the LFS estimate of self–employed jobs, it is possible to estimate how many self–employed people there are running non–registered enterprises.

HMRC data on the ratio of self–employed people working alone or in partnership and the average number of partners per partnership is used to allocate these working proprietors into enterprises[5]. These are added to the IDBR registered enterprises to get the estimated total enterprise population

in the UK. A more detailed description of the methodology used is available in the methodology note for *SME Statistics* and in the forthcoming methodology note for *Business Population Estimates* (see BIS 2010a and BIS 2010b for more details).

Figure 1 **Flowchart to illustrate the estimation methodology, with figures representing data used in *SME Statistics 2009***

Elements of the methodology impacted by changes are shown in boxes with dashed borders.



Component A: the number of self-employed jobs, taken from the LFS

Component B: the ratio of sole proprietors to partners, in SME Statistics this is taken from the HMRC Survey of Personal Incomes

Component C: the number of enterprises registered on the IDBR

The methods used to calculate each of the three components highlighted in Figure 1 will change in the *Business Population Estimates* methodology. These three major changes are described in more detail in the following section.

Description of each major methodology change

New methodology for calculating Labour Force Survey self–employment

– This relates to Component A in Figure 1

The ONS Review of Employment and Jobs Statistics (see ONS 2006) recommended that comparisons between estimates of jobs produced from household and business surveys should be made on a quarterly basis. Following on from this ONS made a commitment to reconcile the

differences between the estimates of the number of jobs published in the Workforce Jobs Series (WFJ) and the estimates published in the Labour Force Survey.

The Review of Employment and Jobs Statistics identified approximately 30 reasons why the LFS and WFJ estimates of jobs differ from each other, one of which was the double–counting of some self–employment jobs.

Some employees sampled for the employer surveys may state they are self–employed when surveyed as part of the LFS. As self–employment data in WFJ is sourced from the LFS, the estimates in WFJ potentially double–count jobs in a way that the LFS does not.

The reasons for the over-counting of self–employed are discussed in more detail in the article 'Comparison of Statistics on Jobs: September 2007' published in *Economic & Labour Market Review*, March 2008 (see Machin 2008).

Reclassifying main job activity

To help quantify this over–counting, the ONS has included a new question on the Labour Force Survey since Q1 2007 which asks self–employed people (in their main job), and employees not paid a wage or salary by an employer, who it was that pays their National Insurance and Income Tax.

This new question (the 'NITax' variable) when combined with the question asking respondents who paid their salary and wage (the 'SELF' variable), allowed the ONS for the first time to produce a better estimate of the 'true' self–employed population.

Individuals who pay their own National Insurance and Income Tax are legally considered to be self–employed, and should be registered as self–employed with HM Revenue and Customs. Sole directors of limited companies should be classed as employees rather than self–employed[6].

ONS use this methodology to reconcile information on the number of jobs in the economy from the LFS with that from the Work Force Jobs (WFJ) series[7]. BIS Enterprise Directorate can now use the ONS methodology for generating improved estimates of the self–employed population[8].

Under this new methodology individuals are considered to be self–employed if:

- they pay either their own Income Tax and/or National Insurance, **and** they are not a sole director of a limited company; or
- they do not give an answer to the question about Income Tax and National Insurance, **and** they do not give an answer to the question about who pays their salary or wage (on the basis that when there is no additional information the self–reported status has to be used).

Those individuals stating that they are the sole director of a limited company will in all cases be classed as employees. Similarly those individuals stating that their National Insurance and Tax are deducted by an organisation are classed as employees in all cases.

The NITax and SELF questions are asked to all people that self–report they are self–employed. However for self–reported employees a filter question (PDWAGE) is applied before these questions. PDWAGE asks whether the individual gets paid a salary or wage by an employer. Those that say 'no' are then asked the NITax and SELF questions, and can potentially be reclassified as self–employed.

The worked examples in **Figures 2** and **3** use averaged LFS data from Q4 2008 and Q1 2009. These two quarters were used in the production of *SME Statistics 2009*, and therefore allow comparisons to be made with the data that was used in the publication.

These figures illustrate that of the 3.8 million people who identify themselves as self–employed in their main job, only 3.3 million are to be considered as 'true' self–employed, with approximately 500,000 being reclassified as employees.

On the employee side approximately 22,000 individuals are reclassified as self–employed. **Box 2** displays the detail of this reclassification.

Box 2 Detail on the new estimates for main and second job self-employment

Main job self–employment reclassification in Q4 2008/Q1 2009

- 3,826,000 self–reported self-employment in their main job
- In the new methodology 476,000 are reclassified as employees
- And 22,000 employees are reclassified as self–employed
- The estimate for 'true' self–employed in main job is 3,371,000
- After exclusions for industry sectors outside the scope of the publication a figure of 3,277,000 main job self–employed is used in the estimation process

Second job self-employment estimation in Q4 2008/Q1 2009

- 374,000 self-reported self–employment in their second job
- Once excluding industry sectors outside the scope of the methodology this figure is 355,000
- The proportion of self–reported self–employment that is true self–employment in the main job is 88 per cent
- This proportion is calculated for each 2–digit SIC, and within each region and broad industry grouping. The proportions are applied to the second job self-employed numbers in each of these categories
- When each of these categories is summed up to the UK total (for industry sectors included in the scope of the methodology) they equal 315,000 individuals who are estimated to be 'true' self–employed in their second job

This results in a total self–employment estimate of 3,592,000

Figure 2 Methodology for reclassification of the self–employed in main job

Self-employed reclassification flow chart
(LFS population average Q4 2008/Q1 2009)
Weighted using PWT09

Self-employed (INECAC05=2)
3,826,000

NITax Payment of NI and Tax	**SELF(1-4)** Other methods of payment aside from receiving a salary or wage direct from an employer				
	Paid a salary or wage by an agency (1)	Sole director of a limited company (2)	Both (1) and (2)	Neither or not answered	Total
Pay own tax and NI	55,000	292,000	11,000	3,086,000	3,444,000
Pay own NI and tax but not both	5,000	8,000	*	171,000	185,000
NI and tax deducted by organisation	19,000	13,000	2,000	127,000	160,000
No answer	2,000	2,000	*	32,000	36,000
Total	81,000	315,000	13,000	3,417,000	3,826,000

True self-employed	3,349,000
Self-employed reclassified as employees	476,000

* Sample size too small to give a reliable estimate.

Figure 3 Methodology for reclassification of employees in main job

Employee reclassification flow chart
(LFS population average Q4 2008/Q1 2009)
Weighted using PWT09

Employee (INECAC05=1)
25,237,000

Paid a salary or wage by employer? (PDWAGE)

Yes (PDWAGE=1)

No (PDWAGE=2)
108,000

SELF(1-4) Other methods of payment aside from receiving a salary or wage direct from an employer

NITax Payment of NI and Tax	Paid a salary or wage by an agency (1)	Sole director of a limited company (2)	Both (1) and (2)	Neither or not answered	Total
Pay own tax and NI	6,000	3,000	*	8,000	17,000
Pay own NI and tax but not both	1,000	*	*	2,000	3,000
NI and tax deducted by organisation	71,000	*	*	10,000	81,000
No answer	1,000	1,000	*	5,000	7,000
Total	79,000	4,000	*	25,000	108,000

Employees reclassified as Self-employed	22,000
True employees (after PDWAGE=2 filter)	87,000

* Sample size too small to give a reliable estimate.

Estimating second job self–employment

Both *SME Statistics* and *Business Population Estimates* use information on individuals who are self–employed in their main and/or second jobs, to more fully capture self–employment activity in the UK. This allows a more complete estimate of the total number of enterprises.

The variables used in estimating 'true' self–employment in the respondents' main job are not available for second jobs. Therefore it is necessary to develop a way to better estimate 'true' second job self–employment in the absence of the additional questions in the LFS.

Analysis conducted by BIS found there was no evidence that main job self–employed individuals with a second job were any more likely to get their self–employment status correct when self–reporting in their main job than the average. Most second job self–employed had a main job that was not in self–employment.

Furthermore there was no bias in the distribution of second job self–employed towards sectors or occupations where, in the main job, they got the self–classification correct. Taking this into account, the method adopted for estimating second job self–employment was to take the ratio of self–reported self–employment to 'true' self–employment in main job, and to apply that ratio to second job self–employment. Box 2 displays the detail of this method for the LFS data in Q4 2008/Q1 2009.

Summary of the effects of the methodology on main and second self–employment jobs

The preceding analysis gives new estimates for main and second job self–employed. This is the methodology that will be used in the production of *Business Population Estimates*.

Table 1 brings together the data to show the overall difference between the self–reported self–employment levels and levels based on the new methodology.

Table 1 Self–employment jobs, UK, average Q4 2008 and Q1 2009

Not seasonally adjusted

	Self–reported	Self–reported for industry sectors in scope*	New methodology estimating 'true' self–employment	New methodology for industry sectors in scope*
Main job	3,826,000	3,721,000	3,371,000	3,277,000
Second job	374,000	355,000	330,000	315,000
Total	4,200,000	4,075,000	3,701,000	3,592,000

* Industry sectors that are used in the estimation process, see the methodology note for SME Statistics for further information (BIS 2010a)

Source: BIS analysis of ONS Labour Force Survey

As can be seen in Table 1 the number of self–employment jobs in the UK is 4.2 million when based on self–reporting, and 4.1 million for those sectors in the scope of the methodology for *SME Statistics* and *Business Population Estimates*. The figure of 4.1 million was used in the estimation of unregistered businesses for *SME Statistics 2009*.

Using the new methodology, and excluding industry sectors outside the scope of the estimation process, the number is 3.6 million self–employment jobs.

This new figure of 3.6 million self–employment jobs, when used in the estimation process for the total number of enterprises in the UK, produces an overall estimate of 4.4 million enterprises. This represents a decrease of 0.4 million (9 per cent) on the estimated 4.8 million enterprises reported in *SME Statistics 2009*.

Changes to the HMRC data used to calculate the ratio of sole proprietors to partners

– This relates to Component B in Figure 1

In order to estimate the number of unregistered businesses, the number of self–employed jobs associated with unregistered businesses is first estimated. The number of businesses accounted for by these jobs is then estimated.

Under the *SME Statistics* methodology the starting point for these estimates is the number of people in the LFS stating that they are self–employed. The number of self–employed individuals working in IDBR registered enterprises is then removed. In order to estimate the remaining number of enterprises, information on sole proprietorships (one enterprise, one person) and partnerships (one enterprise, more than one self–employed person) is required. Previously, HMRC's *Survey of Personal Incomes* (SPI) has been used to do this. This provided, for each SIC 2–digit classification, an estimate of the ratio of partners to sole proprietors. This was used to convert the number of self–employed people in each classification into the number of enterprises.

One of the methodological changes being introduced improves the estimate of these ratios. *Business Population Estimates* will use HMRC self–assessment (SA) data to carry out the same process. SA data is not perfect, for example, it includes approximately 15 per cent of cases where industry data is missing. However it is based on the whole self–assessment record, almost a census of the self–employed[9]. As well as having more data, a further benefit comes from additional information about individuals. In particular, in the SPI, the ratios of sole proprietors to partners was derived from all enterprises, whereas SA data allows us to focus only on individuals working within unregistered enterprises, by excluding enterprises registered for VAT.

The impact of this specific improvement is an increase in the estimates of the number of unregistered enterprises. This occurs because of the exclusion of registered enterprises from the ratio calculation and the fact that partnerships with more individuals are more likely to be registered enterprises. By excluding these enterprises, the remaining data contains relatively fewer partners than sole proprietors, which results in a higher ratio of businesses per self–employed person. Using 2008/9 self–assessment data instead of 2007/8 SPI data in the estimation of the number of enterprises produced a figure of 4.9 million enterprises at the start of 2009. This represents an increase of 0.1 million (1 per cent) on the figure published in *SME Statistics 2009* of 4.8 million.

Changes to the Inter–Departmental Business Register data

– This relates to Component C in Figure 1

In addition to the changes in the LFS self–employment data and HMRC data, *Business Population Estimates* will also include IDBR data with new selection criteria.

The IDBR contains a number of selection markers which allow the most appropriate units to be selected in relation to any specific survey requirement. IDBR data currently used in *SME Statistics* includes enterprise units from all of these selection marker categories.

However this is not in line with ONS practice on excluding enterprises in a number of categories. In the *Business Population Estimates* methodology, BIS is moving closer to ONS practice by excluding a number of categories (such as enterprises with no UK activity and enterprises with little evidence of trading activity).

Additionally a new inclusion statement will be used for *Business Population Estimates*. Inclusion statements are a check to remove larger enterprises with only limited evidence for their existence. In line with ONS practice the inclusion statement is being changed for enterprises with 20 or more employment with the result that some of these enterprises that were previously excluded from *SME Statistics* will be included in *Business Population Estimates*.

As a result of this change in the IDBR extract criteria some enterprises have been taken out while other enterprises have been added in. As can be seen in **Table 2** the change in criteria has the overall impact of reducing the number of enterprises by around 30,000 (-1 per cent). However in some of the larger size categories the number of enterprises has increased due to this change of methodology. The 20–49 employees size category increased by almost 5,000 (9 per cent), using the new IDBR criteria.

Key results from applying the *Business Population Estimates* methodology

Summary of the impact of each methodology change on the estimate of the number of enterprises at the start of 2009

As outlined before self–employment data is used to estimate the number of unregistered enterprises, which when added to the number of VAT and/or PAYE registered enterprises (taken from the Inter Departmental Business Register) provides the BIS estimate of the total number of UK enterprises.

Table 3 shows the impact of making each methodology change (new LFS data, new IDBR criteria, using HMRC self–assessment data) and the combined impact of making all these changes on the estimates of the number of enterprises at the start of 2009.

Table 2 Impact of the change in IDBR criteria on the estimate of the number of businesses

Start 2009, United Kingdom

	As published in SME Stats 2009	Calculated with the new IDBR data	Difference (number)	Difference (per cent)
All enterprises	4,834,045	4,803,840	-30,205	*-1*
All employers	1,220,070	1,223,625	3,555	*0*
With no employees*	3,613,975	3,580,215	-33,760	*-1*
1	189,120	188,115	-1,005	*-1*
2-4	606,485	605,590	-895	*0*
5-9	224,000	223,800	-200	*0*
10-19	113,620	113,710	90	*0*
20-49	54,050	58,855	4,805	*9*
50-99	17,770	18,360	590	*3*
100-199	7,665	7,790	125	*2*
200-249	1,470	1,485	15	*1*
250-499	3,005	3,025	20	*1*
500 or more	2,885	2,895	10	*0*

*'With no employees' comprises sole proprietorships and partnerships comprising only the self-employed owner-manager(s), and companies

Source: BIS estimates

Table 3 Impact of each methodology change on the estimate of the number of enterprises

Start 2009, United Kingdom

	Data published in *SME Statistics 2009*	Calculated using LFS self-employment data on the new methodology*	Calculated using HMRC self-assessment data*	Calculated using IDBR data based on the new criteria*	Calculated using all new data sources - the *Business Population Estimates* methodology
Number	4,834,045	4,413,160	4,905,205	4,803,840	4,436,050
Change on *SME Statistics 2009* (number)	-	-420,885	71,160	-30,205	-397,995
Change on *SME Statistics 2009* (per cent)	-	*-9*	*1*	*-1*	*-8*

* In each case the other data sources used in the calculation were the same as those used in producing SME Statistics 2009.

Source: BIS estimates

Table 3 shows that using the new *Business Population Estimates* methodology, the estimate for the business population at the start of 2009 is reduced from the published 4.8 million to 4.4 million (or by around 398,000 and 8 per cent).

These results detail the impact of each key methodological change in isolation, alongside the overall impact of applying all of the changes together. As the interaction between these changes is complex the overall impact is different to the sum of each of the changes.

As can be seen the largest impact is from the change in the LFS self–employment data, which reduces the overall number of enterprises by nine per cent. The change in the IDBR data in isolation has a smaller impact on count, reducing the estimate of the total number of enterprises by around one per cent.

The change to using HMRC self–assessment data in isolation leads to a one per cent increase in the total number of enterprises.

The impact of the new methodology on the number of enterprises at the start of 2009 by size, legal status, industry sector and UK region

The composite results from the new methodology are shown in more detail in **Table 4** and compared against figures published in SME Statistics 2009. These data show that the change in the methodology mostly impacts on the number of enterprises with zero employees (and the 'all enterprises' total). This is to be expected as the self–employment data, which has the largest impact on the estimate, is only used to estimate unregistered enterprises, all of which have no employees.

Changes to the number of enterprises in the larger size bands are caused only by a change to the IDBR data in these size bands (see Table 3).

SME Statistics and *Business Population Estimates* are comprised of enterprises from three types of legal status. These are companies (including public corporations and nationalised bodies) in which the working directors are counted as employees; partnerships, run by two or more self–employed people; and sole proprietorships, run by one self–employed person.

Table 5 shows that partnerships saw the largest difference under the new methodology. The estimated number of partnerships at the start of 2009 would have been 80,000 (18 per cent) less under the new methodology. Partnerships saw more of a difference than sole–proprietorships due to the change to the HMRC data used to calculate the ratio of sole proprietors to partners. The self-assessment data used in the new methodology focussed only on unregistered enterprises, which contain fewer partnerships than identified in the SPI data used in *SME Statistics*.

The estimate of the number of companies was least affected by the change in methodology. The assumption that all companies are registered on the IDBR means there is no need to estimate unregistered enterprises in this category. The two per cent fall in companies under the new methodology is entirely due to using IDBR data based on the new criteria.

Table 4 Number of enterprises by size: comparison of the results from SME Statistics 2009 and data based on the Business Population Estimates methodology

Start 2009, United Kingdom

	Data published in *SME Statistics 2009*	Calculated using the *BPE* methodology, 2009	Difference, number	Difference, per cent
All enterprises	4,834,045	4,436,050	-397,995	*-8*
All employers	1,220,070	1,223,625	3,555	*0*
With no employees*	3,613,975	3,212,425	-401,550	*-11*
1	189,120	188,115	-1,005	*-1*
2-4	606,485	605,590	-895	*0*
5-9	224,000	223,800	-200	*0*
10-19	113,620	113,710	90	*0*
20-49	54,050	58,855	4,805	*9*
50-99	17,770	18,360	590	*3*
100-199	7,665	7,790	125	*2*
200-249	1,470	1,485	15	*1*
250-499	3,005	3,025	20	*1*
500 or more	2,885	2,895	10	*0*

* 'With no employees' comprises sole proprietorships and partnerships comprising only the self-employed owner-manager(s).

Source: BIS estimates

Table 5 Number of enterprises by legal status: comparison of the results from SME Statistics 2009 and data based on the Business Population Estimates methodology

Start 2009, United Kingdom

	Data published in *SME Statistics 2009*	Calculated using the *BPE* methodology, 2009	Difference, number	Difference, per cent
All enterprises	4,834,045	4,436,050	-397,995	*-8*
Companies	1,300,390	1,270,350	-30,040	*-2*
Partnerships	444,250	364,230	-80,020	*-18*
Sole proprietorships	3,089,405	2,801,470	-287,935	*-9*

Source: BIS estimates

Table 6 shows that although the overall number of enterprises decreased by eight per cent under the new methodology the level of change varied greatly by industry section, but almost all sections saw a fall in level. Mining and quarrying; electricity, gas and water supply (-26 per cent) and financial intermediation (-26 per cent) saw the largest percentage change in estimated number of enterprises.

Table 6 Number of enterprises by industry section*: comparison of the results from SME Statistics 2009 and data based on the Business Population Estimates methodology

Start 2009, United Kingdom

	Data published in *SME Statistics 2009*	Calculated using the *BPE* methodology, 2009	Difference, number	Difference, per cent
All enterprises	4,834,045	4,436,050	-397,995	*-8*
A, B - Agriculture, Hunting and Forestry; Fishing	195,480	184,850	-10,630	*-5*
C, E - Mining and Quarrying; Electricity, Gas and Water Supply	15,165	11,170	-3,995	*-26*
D - Manufacturing	303,245	277,005	-26,240	*-9*
F - Construction	1,017,210	914,785	-102,425	*-10*
G - Wholesale and Retail Trade; Repairs	562,815	535,410	-27,405	*-5*
H - Hotels and Restaurants	156,470	153,695	-2,775	*-2*
I - Transport, Storage and Communication	315,020	298,245	-16,775	*-5*
J - Financial Intermediation	83,890	62,245	-21,645	*-26*
K - Real Estate, Renting and Business Activities	1,195,825	1,068,570	-127,255	*-11*
M - Education	180,825	162,660	-18,165	*-10*
N - Health and Social work	279,560	281,010	1,450	*1*
O - Other Community, Social and Personal Service Activities	528,540	486,405	-42,135	*-8*

* SIC2003

Source: BIS estimates

In contrast health and social work was the only section to see a small increase in the number of enterprises (1 per cent). Although changing the LFS and IDBR source data caused a fall in the estimate of the number of enterprises in the health and social work sector, this was more than off–

set by the impact of using the self-assessment data, leading to an overall increase when all three major methodology changes were applied together.

Table 7 shows that under the new methodology there would have been lower estimates for the number of enterprises in every English region and UK country at the start of 2009. London showed the largest fall in the number of enterprises under the new methodology (-14 per cent). This can be explained by London having a higher proportion of unregistered enterprises than other regions, meaning the overall estimate for enterprises in London is disproportionately affected by the change in LFS methodology.

Table 7 Number of enterprises by English region and UK country: comparison of the results from SME Statistics 2009 and data based on the Business Population Estimates methodology

Start 2009

	Data published in *SME Statistics 2009*	Calculated using the *BPE* methodology	Difference, number	Difference, per cent
United Kingdom	4,834,045	4,436,050	-397,995	*-8*
North East	139,750	129,900	-9,850	*-7*
North West	482,220	438,120	-44,100	*-9*
Yorkshire and the Humber	345,565	315,660	-29,905	*-9*
East Midlands	307,335	289,980	-17,355	*-6*
West Midlands	363,375	339,255	-24,120	*-7*
East of England	501,655	473,490	-28,165	*-6*
London	837,605	721,850	-115,755	*-14*
South East	743,620	695,895	-47,725	*-6*
South West	464,735	435,685	-29,050	*-6*
England	4,185,860	3,839,835	-346,025	*-8*
Wales	200,180	186,985	-13,195	*-7*
Scotland	324,110	293,870	-30,240	*-9*
Northern Ireland	123,895	115,360	-8,535	*-7*

Source: BIS estimates

Other changes being implemented in *Business Population Estimates*

Changing the IDBR extract date and publication schedule

Business Population Estimates 2010 will include a further significant change to the way the IDBR extract is used to produce the statistics. This change will impact on the employment and turnover information reported and also on the timing of publication of subsequent editions.

SME Statistics is produced using analysis from an extract taken 15 months after the reference point. So for example the *SME Statistics 2009* publication, representing data at the start of 2009, uses data extracted from the IDBR in March 2010. The data on the extract is analysed based only on enterprises live at the start of 2009.

With *Business Population Estimates for the UK and Regions 2010* the IDBR extract used will also be taken in March 2010. This means that there will only be a lag of three months between the reference point and the extract being taken from the IDBR. As with *SME Statistics* the extract is analysed in a way that only enterprises live at the start of 2010 are included.

The change to using a three–month lag in *Business Population Estimates*, will not impact on the count of enterprises, as only enterprises active at the start of 2010 will be selected. However it will have an impact on the employment and turnover data reported.

Employment and turnover data on the IDBR extract reflect the latest position reported for any particular enterprise. So for example although the enterprises live at the start of 2009 can be selected, the precise employment and turnover from the start of 2009 cannot be identified. The data for employment and turnover on the register will reflect the latest data as reported.

The expectation is that the impact on the quality of the employment data will be different to the impact on the turnover data. For the very smallest units on the IDBR employment information is derived from PAYE jobs or employment imputed from VAT turnover. However the majority of employment information comes from the *Business Register and Employment Survey (BRES),* which first collected information in September 2009. Previously this has been sourced from the *Business Register Survey* (BRS). In addition employment data on the IDBR is now updated on a monthly rather than an annual basis. This is because information captured by the 81,000 enterprises surveyed by BRES is updated as completed forms are returned from respondents. A proportion of this more up–to–date employment data is available on the IDBR before the extract is taken, whereas previously it was updated the July after extract was taken. At the time the March 2010 extract was taken 81 per cent of responses to the September 2009 BRES had been received and used to update the IDBR.

By moving to a three–month lag, more of the employment data will be relevant to the September before the reference period, representing an improvement compared to *SME Statistics*. So for *Business Population Estimates 2010* employment data will mostly relate to that reported for September 2009.

However no changes have been made to the way turnover data is updated on the IDBR. In the March 2010 extract the turnover data provided to the ONS for the majority of enterprises is based on VAT returns for a 12 month period ending in December 2008, or January/February 2009,

according to the reporting pattern of the enterprise. For other enterprises turnover may relate to an earlier period. For those which are only registered for PAYE turnover is imputed from employment. This means that turnover data on the IDBR tends to be less timely than both enterprise counts and employment data.

The change in the IDBR extract used means that *Business Population Estimates* will be produced much more quickly after the reference date than *SME Statistics*. Although due to methodological work *Business Population Estimates 2010* has been produced on a slightly extended timescale, the expectation is that future publications will be available around 10 months after the reference period. **Table 8** shows the current and expected publication schedule.

Table 8 Publication schedule for SME Statistics and Business Population Estimates

Publication	Reference point	IDBR Extract	Publication schedule (provisional)	Lag between reference point and publication month
SME Stats 2008	Start 2008	March 2009	October 2009	22 months
SME Stats 2009	Start 2009	March 2010	October 2010	22 months
BPE 2010	Start 2010	March 2010	May 2011	17 months
BPE 2011	Start 2011	March 2011	October 2011	10 months
BPE 2012	Start 2012	March 2012	October 2012	10 months

Source: BIS

Standard Industrial Classification (SIC) 2007

Business Population Estimates for the UK and Regions 2010 will be the first time estimates of the total number of enterprises have been produced using the Standard Industrial Classification 2007 (SIC2007). Estimates published in *SME Statistics* were produced using Standard Industrial Classification 2003 (SIC2003).

This move presents a challenge to users as for any particular industry it will not be possible to identify whether any changes are due to moving to the SIC2007 classification or due to the change in methodology.

To help users BIS will publish full tables in the SIC2003 format for *Business Population Estimates 2010*. These tables will be available on the web site within a few weeks of the main publication release.

Conclusions

Small and Medium Enterprise Statistics for the UK and Regions will no longer be produced. Instead a new publication, called *Business Population Estimates,* will be published by BIS in May 2011 using an improved methodology for estimating the total enterprise population.

The improved methodology uses better estimates of the self–employed population from the Labour Force Survey and more comprehensive data from HMRC to calculate more accurately how many unregistered enterprises there are. It also better captures the most appropriate enterprises from the IDBR.

This improved methodology provides more accurate estimates of the total number of enterprises in the UK. It estimates 4.4 million enterprises at the start of 2009, eight per cent fewer than previously estimated under the *SME Statistics* methodology.

Business Population Estimates for the UK and the Regions 2010 will use this improved methodology for the first time in a National Statistics publication. It will also have a more timely release than *SME Statistics*. For example *Business Population Estimates 2011* is expected to be published 12 months earlier than *SME Statistics* 2011 would have been.

Business Population Estimates for the UK and Regions 2010 will be released May 2011 and will be available on the National Statistics Publication Hub and at http://stats.bis.gov.uk/ed/bpe

Contact

julian.shaw@bis.gsi.gov.uk

Notes

1. *Business Population Estimates* uses the same broad methodology as *SME Statistics*. It therefore represents a continuation of the SME Statistics series and can be classed as National Statistics. The National Statistics status of *Business Population Estimates* will be formally assessed by the UK Statistics Authority in October 2011.

2. For the years 1997–1999, 2001, 2003, 2005, 2007–2009 SME Statistics also contained data for each English region and UK country. Data is available here: http://stats.bis.gov.uk/ed/sme

3. Enterprises are required to register for VAT if their turnover reaches the compulsory registration threshold (£70,000 in 2010/11). However some enterprises will register voluntarily below this threshold. Enterprises are not required to register for PAYE if they have no employees or when their employees are paid below the PAYE threshold (£540 a month in 2010/11). See ONS 2009 for a guide to the IDBR.

4. Enterprises are classified as the following legal forms – companies/public corporations and businesses run by self–employed people, which are either partnerships (if two or more self–employed working proprietors) or sole proprietorships (one working proprietor).

5. Data used to calculate the ratio of sole proprietors to partners and average number of partners in a partnership comes from analysis of HMRC Survey of Personal Incomes data and HMRC Self–Assessment data.

6. For further information see http://www.hmrc.gov.uk/leaflets/es-fs1.pdf

7. An example of the reconciliation can be found in Annex 1 here: www.statistics.gov.uk/downloads/theme_labour/LMS_Q&A.pdf

8. The official measure of self-employment is reported in Labour Market Statistics, (see Table 3, Labour Market Statistical Bulletin www.statistics.gov.uk/statbase/Product.asp?vlnk=1944). To adhere to Eurostat requirements, the official measure of self-employment is based entirely on the self-reported status of individuals in their main job.

9. Further information on the self–assessment form can be found here: http://search2.hmrc.gov.uk/kbroker/hmrc/forms/viewform.jsp?formId=3813

References

Department for Business Innovation and Skills (2008) Consultation on Small and Medium Enterprise Statistics for the UK and Regions at www.berr.gov.uk/files/file49351.pdf

Department for Business Innovation and Skills (2009) Government Response to Consultation on Small and Medium Enterprise Statistics for the UK and Regions at www.berr.gov.uk/files/file51821.pdf

Department for Business Innovation and Skills (2010a) Small and Medium Enterprise Statistics for the UK and Regions 2009 Methodology Note at http://stats.bis.gov.uk/ed/sme/smestats2009-meth.pdf

Department for Business Innovation and Skills (2010b) Small and Medium Enterprise Statistics for the UK and Regions 2009 at http://stats.bis.gov.uk/ed/sme

Machin A (2008) ‘Comparison of statistics on jobs: September 2007’, *Economic & Labour Market Review* 2 (3) pp 18-23 and at: www.statistics.gov.uk/elmr/03_08/downloads/ELMR_Mar08_Machin.pdf

Office for National Statistics (2006) ‘Review or Employment and Job Statistics’, National Statistics Quality Review Series, Report Number 44 at www.ons.gov.uk/about-statistics/methodology-and-quality/quality/nat-stats-qual-revs/qual-revs-by-theme/labour-market/ann-of-emp-and-jobs.pdf

Office for National Statistics (2009) A Guide to the IDBR, Inter-Departmental Business Register; a key survey source at www.statistics.gov.uk/CCI/nugget.asp?ID=195

Office for National Statistics (2010a) UK Business: Activity, Size and Location 2010 at www.statistics.gov.uk/StatBase/Product.asp?vlnk=933

Office for National Statistics (2010b) Business Demography 2009 at www.statistics.gov.uk/StatBase/Product.asp?vlnk=15186

Investigating the characteristics of patents and the businesses which hold them

Andrew Thomas
Office for National Statistics

Summary

This article describes how patent data from the UK Intellectual Property Office has been matched at the firm–level with the business surveys administered by the Office for National Statistics. The resulting dataset should provide a useful resource for those undertaking research on innovation in the UK. Some initial analysis of the matched dataset has been presented here. Results show that businesses hold more patents than individuals; large businesses hold more patents than small businesses on average; and businesses patent inventions which would be intuitively associated with their industry.

Introduction

Intellectual property, along with other knowledge assets, has become increasingly important to economies across the world. Recent estimates suggest that firms' investment in intangible assets has overtaken that of tangible capital (OECD 2009). Consequently, policy interest in this area has been growing, and the desire for policy evidence has been driving research in the innovation field. Patent statistics have provided one source of data for this research.

This article describes how patent and business datasets have been matched at the firm–level and provides some initial results from analysis of these data. The concept of combining patent and business databases is not novel[1] but the analysis in this article is based on patent data from the UK Intellectual Property Office (IPO) which have not previously been made available for research. The matched data also provide a first step towards being able to associate innovative activity to a particular enterprise and location. Previous research has been able to identify whether a particular firm may exploit the intellectual property associated with a patent, for example as part of an international group that owns the patent, but until now it has been difficult to determine if a particular business at a particular location actually applied for the patent. The reason for this is the use of 'name only' matching of patent and business data sources. Matching by name and address can reduce the raw counts of matched observations, but has the added refinement of ensuring that

the branch of the corporation which applied for the patent is the business with which a match is found.

The name and address matching in this instance is not directly with a business database. Rather, the names and addresses from the patent data are matched to the Inter–Departmental Business Register (IDBR) held by the Office for National Statistics (ONS). The IDBR is a live updating register of businesses in the UK, and is used as a sampling frame for the business surveys undertaken by ONS. The Enterprise Reference Numbers from the IDBR were attached to the patent data when a match was found. These reference numbers are common to all of the ONS's business surveys and the research possibilities arising from this are wide ranging as the patent data could be matched to any of these surveys.

The IPO is particularly interested in obtaining information about the characteristics of businesses that hold patents. Using the newly matched database, this article presents some initial findings on:

- the distributions of patents by age and size of their respective patentees
- the age and size distributions of the patentees themselves
- trends in the total stock of patents over time
- differences in patenting by businesses and individuals, and
- the relationship between the industry in which a business operates and the types of inventions being patented (based on International Patent Classification (IPC) codes)

The results suggest that the stock of patents is increasing over time, along with the number of patentees. In general, larger more established businesses hold more patents than smaller younger businesses. An investigation in to the types of inventions being patented reveals that businesses tend to patent inventions which have some logical link to the industry in which they operate. Finally, businesses hold significantly more patents than individuals.

The rest of this article proceeds as follows. The next section describes the sources for patent and business data. The third section explains the matching process. The fourth section presents some results from analysing the matched data set, and finally some brief conclusions are offered.

Data sources

Patent data

Two sources of patent data have been used in this work. The UK IPO provided a data extract from their Optics system, and use of the PATSTAT database containing information on patents applied for at the European Patent Office (EPO) was granted. Together these two data sources contain information on all patents applied for at the IPO or the EPO. However, as the aim of the matching exercise is to link UK business information with the patents held by those businesses, only the patents which have been applied for from a UK address are considered.

Optics

Optics is the live register used by the IPO for the management of patent information. This database is updated every time a patent is applied for, published, amended or granted. To use this data for research purposes an extract has been taken which includes published patents that were applied for after 1990 and also all patents which have been granted since 1990. Naturally there is considerable crossover between the two, but there are a number of patents that were applied for before 1990 but granted afterwards. The data from 1986 to 2008 contains at least 2,000 observations per year. There are limited numbers of observations from 1977 to 1985 and in 2009. There is a time lag between when a patent is applied for and when it is published. The small number of observations prior to 1986 represent a longer than usual time lag, whilst the small number of observations in 2009 represent patents which have been published much more quickly than usual.

Optics is a live updating system so any new information is recorded in the same place as the existing information. For example, if a patent is renewed, the renewal date and annuity date are simply changed for that particular patent. As a result, all of the information extracted from Optics is attached to each observation. Aside from matching names and addresses to IDBR reference numbers, there was very little manipulation of the Optics data required to make it suitable for research.

PATSTAT

PATSTAT is a database of international patent statistics administered by the EPO and has been specifically designed for research purposes. PATSTAT data contains information on patents that were applied for between 1986 and 2009. Between 1986 and 2007 there are at least 3,000 observations per year, whilst there are fewer observations for 2008 to 2009. As with Optics, this drop in the number of observations in recent years comes as a result of the time lag between the application and publication of a patent.

Despite being designed for research purposes, a certain amount of data manipulation was required before PATSTAT data could be analysed. As PATSTAT is not a live updating register, new information is not simply captured in the existing observation for a patent. For instance, a patent application and publication come in separate observations resulting in multiple observations per patent. It is known that the patents in the PATSTAT data have been published, and the information in the separate observations is otherwise identical. Therefore it was decided to simply keep the application observation for each patent.

The IPC codes for PATSTAT come in a separate file. As this information has one observation per IPC code rather than one observation per patent there are multiple observations per patent because each patent can have multiple IPC codes. To deal with this issue, the intention was to obtain the primary IPC code at the three character level of aggregation (the first three characters of the primary IPC code). To determine the primary IPC when there are multiple codes per patent, a marker was used indicating ‘F’ or ‘L’. This marker is present in less than half of the IPC dataset. The author was initially informed that this code indicated the first and last IPC for the patent, and

attached the code marked ‘F’ to the patent data. When this marker was unavailable, a new variable containing only the first three characters of the IPC codes was created.

In the cases where the 3 character IPC code was the same for all observations relating to a particular patent, this IPC code was attached to the patent data. For the remaining cases where there was uncertainty regarding the primary IPC code, the IPC variable was left as a missing value in the patent data. Subsequent to this process and the production of the results presented here, the author was informed that the ‘F’ marker does not carry the meaning which was initially given. The consequence of which is that the results using IPC codes in this article may be spurious to a certain extent. As these results are presented at the 3 character level, and only a proportion of the PATSTAT data which has been affected, it is believed these consequences to be relatively minor but any future research on this database should bear in mind that the IPC codes for the PATSTAT information need further cleaning.

The patent information which is present in both the Optics and PATSTAT data and is used in this report is:

- ***Priority Year*** – the initial year that the patent was applied for at any intellectual property office
- ***Grant Year*** – the year in which the patent is granted
- ***Renewal Year/Annuity Year*** – the year that the patent was last renewed and the year that it is next due to be renewed
- ***International Patent Classification (IPC)*** – the classification representing the type of invention being patented

Business data

Matching patentee names and addresses to the IDBR results in a patent dataset which, in principle, can be matched to any of ONS’s business surveys. For this work, the Business Structure Database was used to provide the basic characteristics of the businesses holding patents.

The Business Structure Database (BSD)

The Business Structure Database is constructed from an extract of the IDBR[2], and as such has the same breadth of coverage. An extract from the IDBR is taken annually to create a database of yearly files which can be used for research purposes. Although this database is constructed for research, its origin is an administrative source which was not intended for the same purpose. As such it should be treated with a certain amount of caution.

The administrative sources used to capture business information on the IDBR are primarily VAT registrations (for turnover) and PAYE registrations (for employment). This gives a very wide coverage as any business which is registered for VAT, or has an employee on the PAYE system, is present on the IDBR. However, if only one of these sources is present, the business information which arises from the absent source is imputed. This is particularly the case for new and small businesses.

As the coverage of the BSD is synonymous with that of the IDBR (at the point in time that the extraction was made) a high match rate can be achieved between the BSD and the patent data which have enterprise references. Such a match allows the following information about patentees to be attached to the patent data:

- ***Employment/Employees*** – the number of employees: a) including self–employed and b) excluding self–employed
- ***Standard Industrial Classification (SIC)*** – the industry in which the business operates
- ***Birth Year*** – the year that the firm first appeared on the IDBR[3]. This is used to calculate firm age

Data matching

To facilitate the matching of patent data to business data in the Virtual Microdata Laboratory (VML – this is a facility at the ONS which allows secure access to confidential micro–data), the names and addresses of applicants first need to be matched to the IDBR via a string comparison process. This allows the extraction of Enterprise Reference Numbers from the IDBR, and attaches these reference numbers to the patent data in place of the names and addresses. The Enterprise Reference Numbers are used as unique identifiers for businesses in the ONS datasets in the VML. They are therefore used as the linking field for matching the patent and business data via a direct match.

Matching to the IDBR

The first step of the matching process was to link the patent data names and addresses to enterprises on the IDBR. Some preparatory steps are necessary as there are specific formatting norms which facilitate the matching. Names and addresses need to be subjected to character cleaning before matching. For instance, in the patent data the character '¬' was frequently used in place of a space, and the phrase 'Incorporated in the United Kingdom' preceded a large number of business names. It was also necessary to separate the postcode from the rest of the address and, because the IDBR is a live register of UK businesses, only UK addresses were selected.

Once the initial data cleaning was complete, a list of names and addresses was sent to the Business Register Strategy and Outputs (BRSO) branch at ONS, who provide a service for matching names and addresses to the IDBR. The data sent to the BRSO branch contained only unique names and addresses, along with a unique identifier number. The BRSO branch then returned three separate files; one containing definite matches, one containing multiple matches, and one containing no matches. A definite match is simply when a name and address has been matched to one enterprise on the IDBR. Multiple matches are when a name and address has matched to a number of enterprises on the IDBR. The names and addresses that have found no matches on the IDBR are also returned in a separate file.

A name and address may return multiple matches from the IDBR because businesses can often have a number of separate enterprises on the IDBR. For example pension funds and holdings

companies may exist alongside the company actually undertaking the business activity. In these cases it is possible to manually identify which enterprise the name and address pertain to. This choice of enterprise can also be checked against the BSD in terms of employment and turnover. Given the number of multiple matches that returned from the IDBR matching process, undertaking this manual matching is not feasible for the whole dataset. However, it was possible to use the unique identifier number to rank the multiple matched names and addresses by the number of patents each holds. Having done so, the manual matching process on those multiple matched names and addresses that hold the most patents was done which increased the definitely matched dataset by a large number of patents for a relatively small number of names and addresses.

Through this process some of the multiple matches were moved into the definite matches file. It was then possible to match these names and addresses back to the original patent data using the unique identifier number. All of the definite matches, multiple matches and no matches were linked back to the patent data in this manner. A marker for each category of match was also added to the data in order to easily calculate the number of patents applicable for each set of matched names and addresses. The matching results in terms of patents are displayed in **Table 1**.

Table 1 Matching patents to enterprises

	Optics		PATSTAT	
	Number of patents	Per cent of total*	Number of patents	Per cent of total*
Definite Matches	50,696	44	51,494	56
Multiple Matches	12,663	11	16,612	18
No matches	52,209	45	23,140	25

* Numbers may not sum to 100 due to rounding

As sources of patent data are usually from an administrative system, the data has not specifically been constructed for research purposes. One of the challenges often faced by researchers using patent statistics is name harmonisation. The patent statistics databases created from administrative systems often contain many different entries for a single firm. Researchers must therefore harmonise a number of versions of one name into a single version before beginning a name matching process. This need can arise from different spellings, naming conventions, or simply upper and lower case entries[4]. The IDBR matching process overcomes this name harmonization issue automatically. The fuzzy matching process allocates a match score to each case and applies a threshold level to determine if a match is found[5]. For example, the following names and addresses could be sent for matching:

- ADT Ltd 1 Astreet Road, London
- A D T LTD 1 Astreet Road, London
- A.D.T PLC 1 Astreet Road, London
- A. T. Limited 1 Astreet Road, London

Assuming that there is one enterprise on the IDBR called 'A D T Limited' at this address, these options will all match to the same Enterprise Reference Number and be definite matches. As the Enterprise Reference Number will be the same for all entries, these different versions of the same name have been harmonised by the matching process itself. The other reason for taking this approach is because the IDBR is used as a sampling frame for ONS's business surveys, and the Enterprise Reference Numbers are a common unique identifier across these surveys. This facilitates matching across a number of data sources and greatly increases the breadth of research possibilities.

Another contribution which this matching process makes is in increasing the accuracy with which it is possible to associate a patent with its source. Name matching only exercises may produce a higher match rate, but whilst this is in itself desirable, it is only possible to determine that a business has access to the intellectual property associated with the patent. Using name only matching, it is not possible to determine where the intellectual property was produced. By matching patent applicants by both their name and address is a step closer to being able to make that distinction. There is still likely to be a 'head office' effect in a number of cases. An invention may occur at a particular address, and then the patent may be applied for from the business' head office. In these cases, the patent would be matched to the head office, possibly introducing a bias to any spatial analyses which these matched data could be used for.

Whilst there are advantages to matching patent applicants to the IDBR, there are also limitations to this approach. The IDBR is a live register of businesses in the UK, and is updated daily. When a name and address matching exercise is done, it is the IDBR at that point in time which is being used. If a local unit has changed name, but is undertaking the same activity in the same location, then this local unit will be updated on the IDBR to reflect the name change. Subsequently matching a patent application made under the old name is therefore likely to produce a no match result.

Matching patent and business data

The patent datasets which were uploaded to the VML had the names and addresses of applicants removed for data security purposes. However, upon completion of the name and address matching to the IDBR these patent datasets had Enterprise Reference Numbers in place of these names and addresses when a definite match was found. This leaves an Optics dataset which includes all patents applied for from a UK address to the UK IPO; and a PATSTAT dataset which includes all of the patents applied for from a UK address to the EPO.

For the purposes of this work, these datasets are appended together then matched to the BSD using the Enterprise Reference Number when it is available. Yearly BSD files are used in combination with the application year information from the patent data[6]. The resulting matched dataset includes all of the patent information from Optics and PATSTAT along with business information from the BSD for the applicant, when the applicant found a definite match on the IDBR. The use of yearly BSD files for matching ensures that the business information is relevant at the year of application. The cross sectional results presented in this article were produced using this dataset. Yearly matched datasets were also created to form a time series so that trends in the stock of live patents could be investigated.

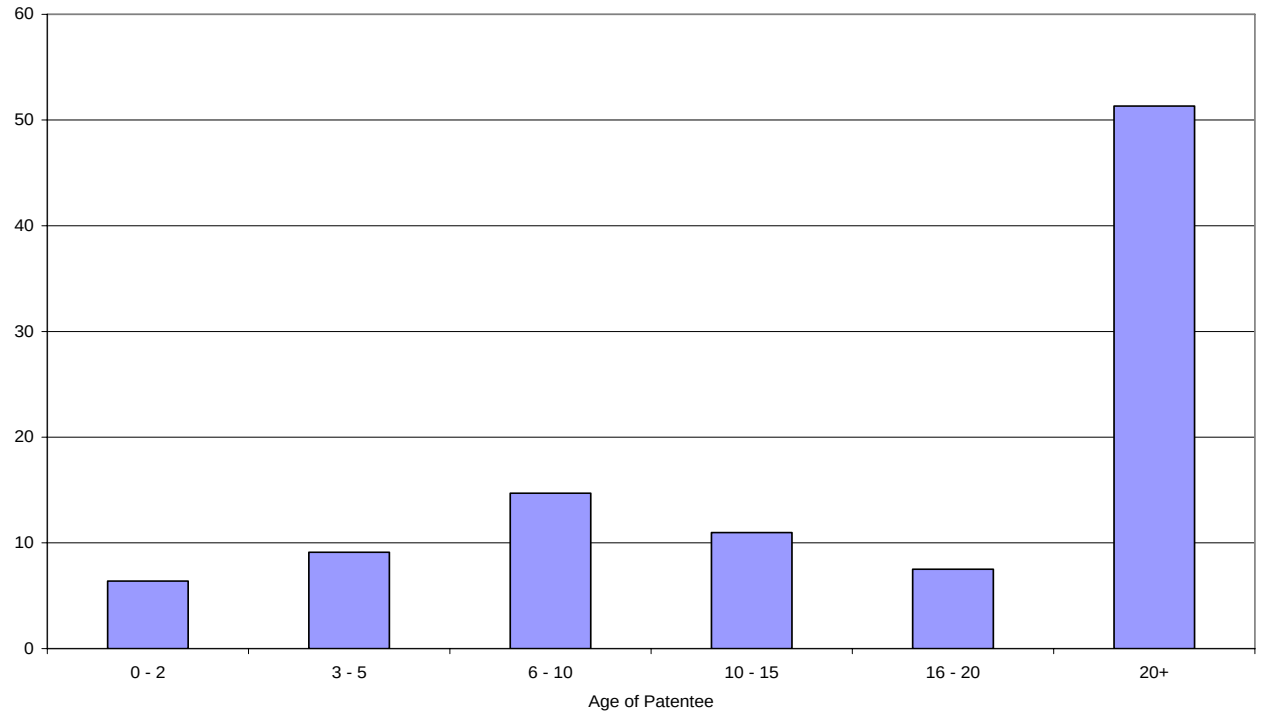
Results

Distributions of patents and patentees by patentee age

The distribution of patents by patentee age presented in **Figure 1a** can be interpreted as representing the number of patents for which the patentee business is in a particular age band. For **Figure 1b**, the number of patents was not taken in to account. This figure simply represents the age distribution of firms in patent data at the year of application (Optics and PATSTAT)[7]. Both of these distributions display a pronounced spike at the top coded end of the distribution. More than 50 per cent of patents are owned by businesses greater than 20 years old. Correspondingly, more than 45 per cent of patentees are greater than 20 years old. A feature of the BSD, which was used to calculate the patentee age, is that there are a large number of enterprises which have a birth year of 1973. This arises from the history of the IDBR. These enterprises do not have an accurately known birth year, and this has been allocated as 1973. This means that analysis of age can lose some meaningfulness at the upper end of the age distribution, and these results should be treated with caution.

Figure 1a **Distribution of patents by patentee age**

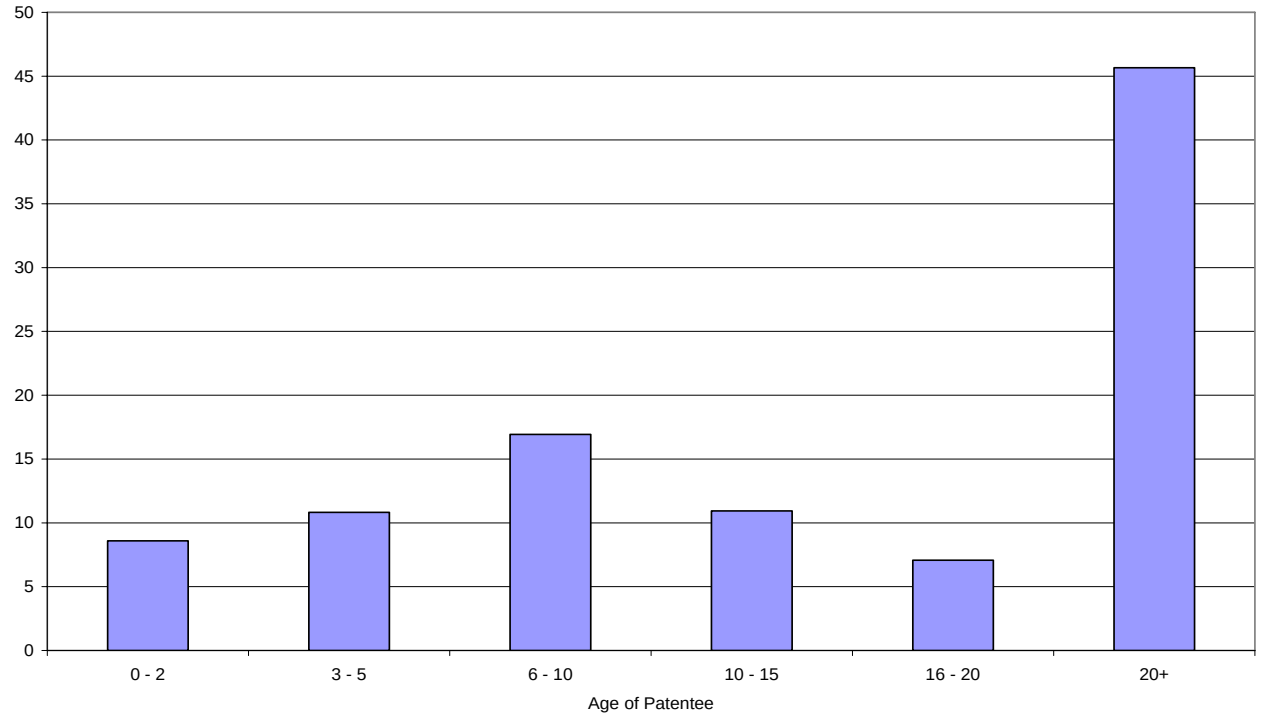
Per cent of patents



Source: Intellectual Property Office and Office for National Statistics

Figure 1b Distribution of patentees by age

Per cent of patentees



Source: Intellectual Property Office and Office for National Statistics

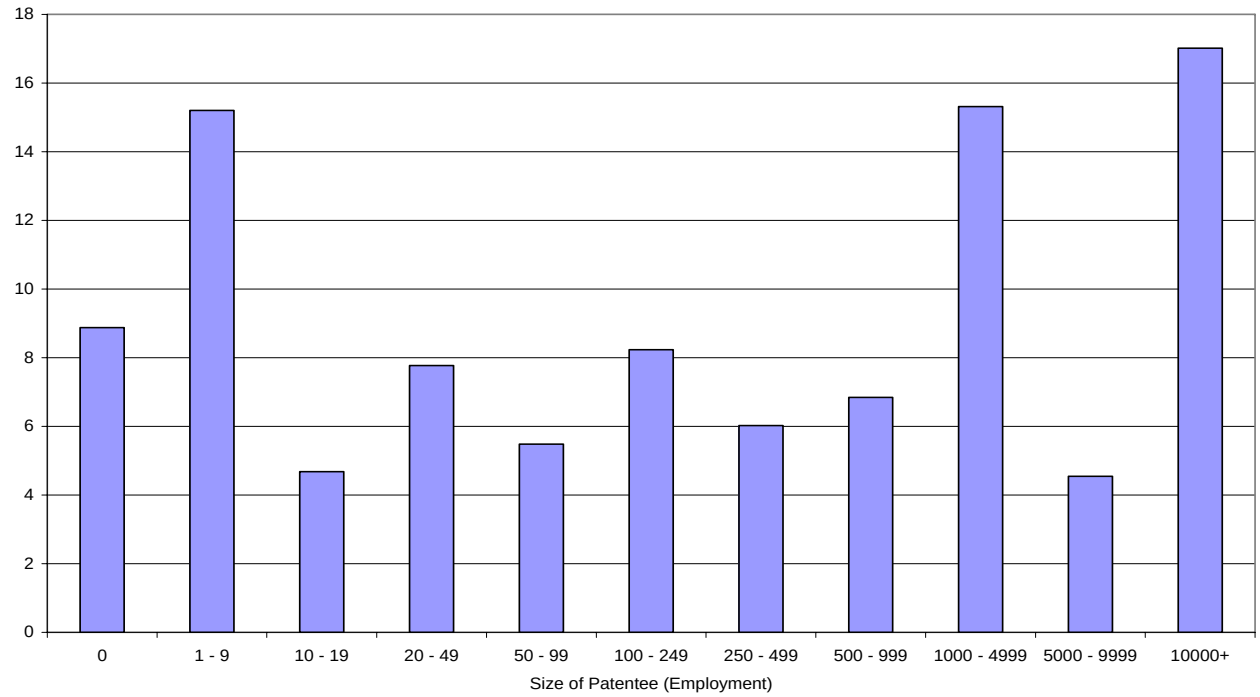
When considering these distributions excluding the top category, both display a similar pattern. Each of the distributions now has a shape which is at least superficially similar to that of a normal distribution. The 6–10 year age band of patentees hold the most patents and also represent the largest group of patentees. The distributions then reduce towards each tail. Note that the figures presented here were produced using the whole dataset. This procedure was repeated for a number of different time periods within the data and the results did not change materially from those presented.

Distributions of patents and patentees by patentee size

The size distributions presented in **Figures 2a** and **2b** show a contrast between the distribution of patents by their owner's size and the distribution of the patentees themselves. The number of patents held by firms of different sizes is relatively even, with the majority of size bands having 5 per cent to 10 per cent of patents. Even the size band with the largest share, the 10,000 employees or more size band, holds only 17 per cent of patents. When considering only the number of firms which patent, the distribution is markedly different. The dominant category in this case is firms with 1 to 9 employees, making up over 30 per cent of patentees. This can be used as an indication of the average number of patents held by firms in different size bands. Firms with 1 to 9 employees dominate the patentees, but on average they hold fewer patents. Conversely, firms with 1,000 to 4,999 employees make up just 2 per cent of patentees but own more than 15 per cent of patents[8].

Figure 2a Distribution of patents by patentee size

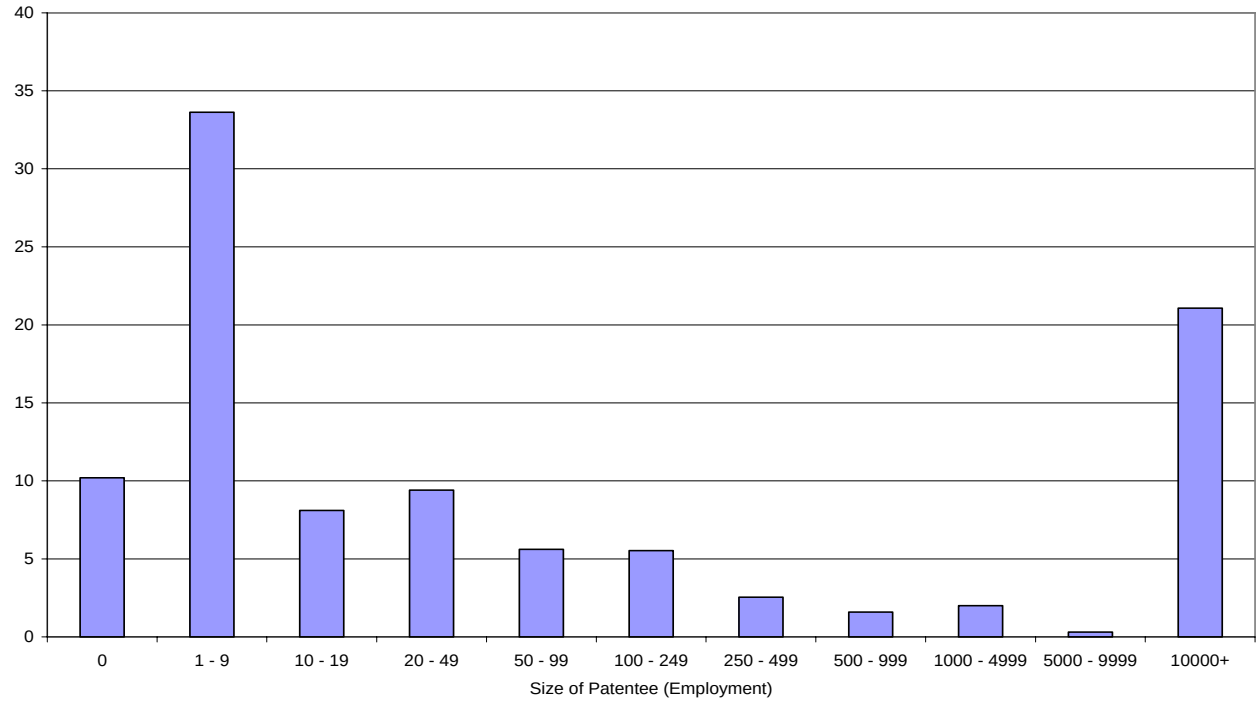
Per cent of patents



Source: Intellectual Property Office and Office for National Statistics

Figure 2b Distribution of patentee size

Per cent of patentees



Source: Intellectual Property Office and Office for National Statistics

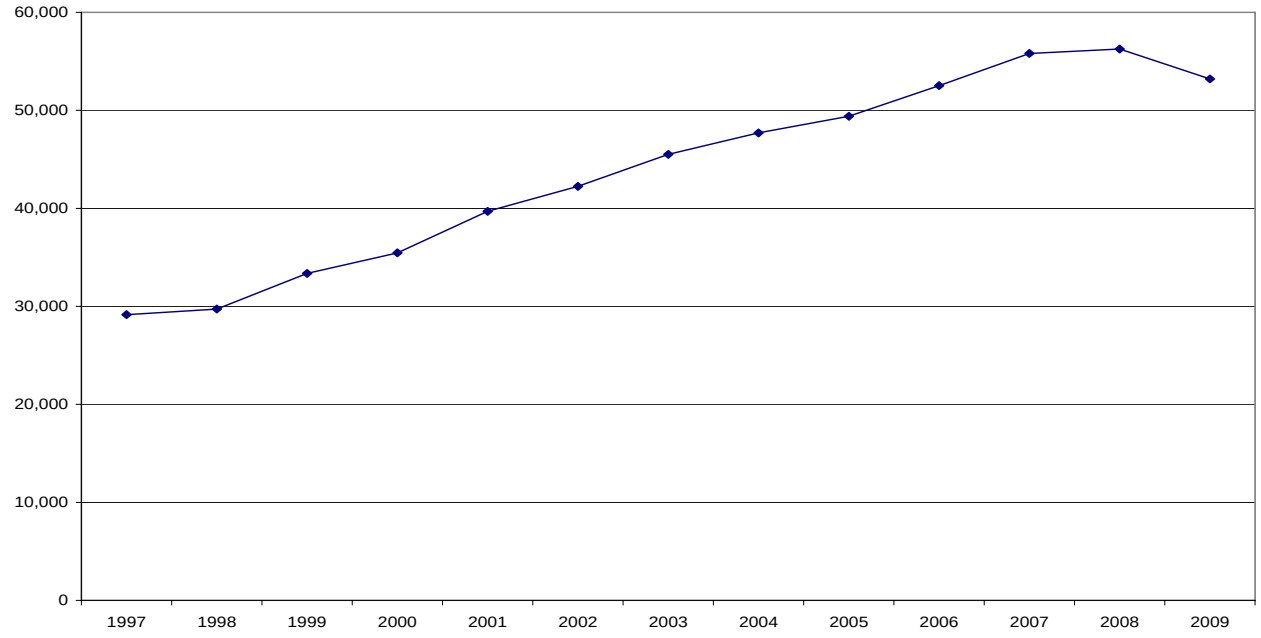
Stock of patents

To define a patent as live, the renewal information from the patent data is used in conjunction with the yearly BSD files. For a given year in the dataset, each patent can be in one of four states: applied for and granted; applied for and not yet granted; applied for, granted and subsequently lapsed; and not yet applied for. This last category occurs because we are considering the whole patent dataset, and patents applied for in more recent years are obviously not live in the years before application. For example, a patent in our dataset which was applied for in 2003 will not be considered a live patent in 1997. Patents which have lapsed by a certain year are no longer live in that year. Both of the first two categories are taken as live patents in this analysis. The first being obvious, these are the patents which have been granted and are enforceable. However, the patents which have been applied for and published but not granted still have some force of protection. This arises from the fact that a patent pending is likely to discourage competitors from copying the invention. If a competitor was to do so, they run the risk of the patent being granted and finding themselves liable to pay the inventor any gains from the copied invention.

Figure 3a indicates that the stock of patents which are live in a given year is steadily increasing over time. There is a turning point at the end of the time period, but this is likely to represent the lower number of published patents we have data for in 2008 and 2009. It seems unlikely to represent a genuine reduction in the stock of patents. Similarly, the number of businesses holding live patents is increasing over the time (**Figure 3b**).

Figure 3a **Stock of live patents**

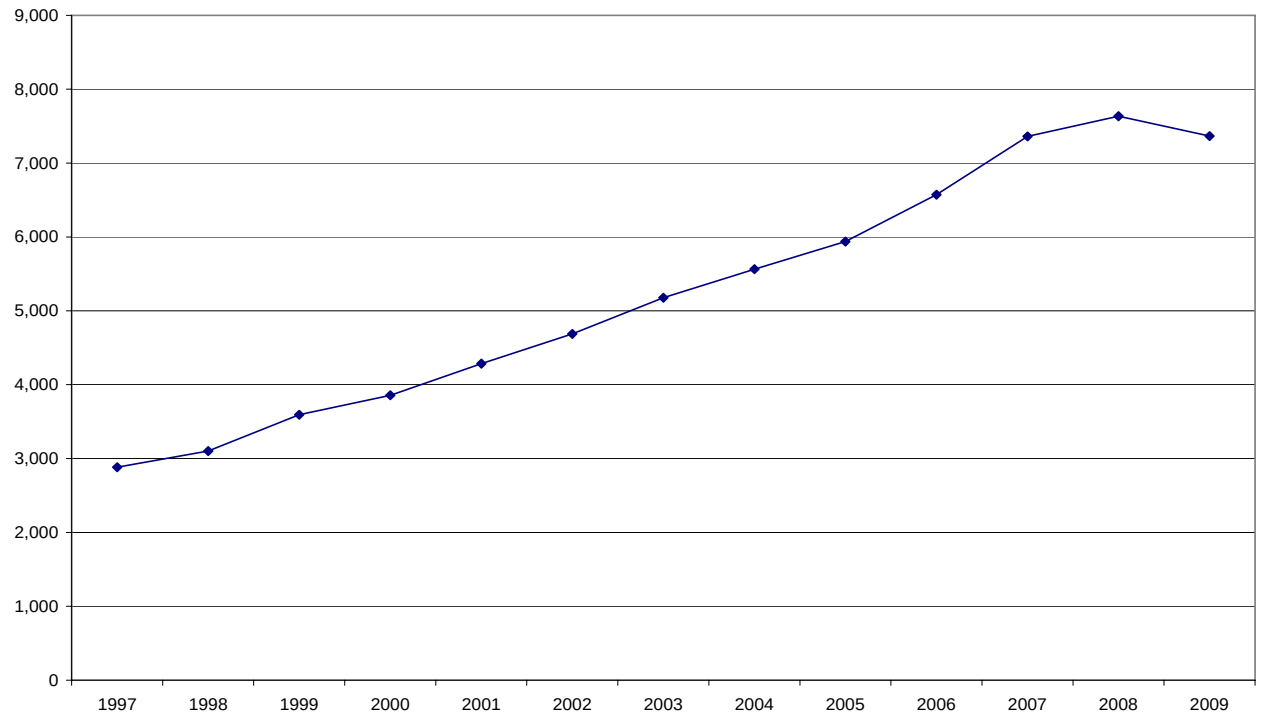
Live patents



Source: Intellectual Property Office and Office for National Statistics

Figure 3b **Patentees holding live patents**

Number of patentees



Source: Intellectual Property Office and Office for National Statistics

Businesses, non–businesses and IPC codes

The matching process distinguishes businesses from individuals to some extent. Any name and address that found either a definite match or multiple matches on the IDBR must be a business. For the analysis in this article, we also identify a number of businesses from the no match file. This is achieved by searching for certain strings within a name variable on the data. For example, any name with 'LTD', 'Ltd', 'Plc' and so on in it was defined as a business. At this point, there are some remaining businesses in the no match file which have not been possible to identify. For the purposes of this analysis, these are considered individuals. When interpreting the results in **Table 2**, it should be noted that there are no individuals defined as businesses but it is likely that there are a small number of businesses defined as non–businesses.

Overall, Table 2 shows the number of patents owned by businesses to be 2.6 times the number of patents owned by individuals. A breakdown of these results by the type of invention (IPC code) tells broadly the same story, with some specific variations represented by the highlighted rows. ‘Sports, games and amusements’ is the only IPC category in which the number of patents held by individuals is higher than the number of patents held by business. Patents in all other categories of invention are held mostly by businesses. In a couple of categories, this result is particularly pronounced (IPC C07 – Organic chemistry and IPC C11 – Oils, fats, detergents and so on). Whilst it may be intuitive for this to occur in the chemical classifications, these results also relate to relatively few observations.

Table 2 Patents held by businesses and non–businesses

IPC Code (3 character)	Business	Non-Business	Business/ Non-Business
A01 - Agriculture; Forestry; Animal Husbandry; Hunting; Trapping; Fishing	2,788	2,268	1.2
A23 - Foods or Foodstuffs; Their Treatment	1,524	267	5.7
A47 - Furniture; Domestic Articles or Appliances; Coffee Mills; Spice Mills; Suction Cleaners	3,353	3.263	1.0
A61 - Medical or Veterinary Science; Hygiene	11,542	3,670	3.1
A63 - Sports; Games; Amusements	1,370	2,536	0.5
B01 - Physical or Chemical Processes or Apparatus	2,485	520	4.8
B60 - Vehicles in General	3,941	2,368	1.7
B62 - Land Vehicles other than on Rails	1,165	1,119	1.0
B65 - Conveying; Packing; Storing; Handling Thin or Filamentary Material	5,720	2,399	2.4
C07 - Organic Chemistry	5,607	321	17.5
C08 - Organic Macromolecular Compounds	1,794	239	7.5
C11 - Animal or Vegetable Oils, Fats, Fatty Substances or Waxes; Fatty Acids Therefrom; Detergents; Candles	1,563	49	31.9
C12 - Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation or Genetic Engineering	2,769	382	7.2
E04 - Building	3,204	2,269	1.4
E05 - Locks; Keys; Windowor Door Fittings: Safes	1.955	796	2.5
E06 - Doors; Windows; Shutters; Or Roller Blinds in General; Ladders	1,357	1,008	1.3
F16 - Engineering Elements or Units	6,732	2,056	3.3
G01 - Measuring; Testing	10,595	2,676	4.0
G02 - Optics	2,314	687	3.4
G06 - Computing; Calculating; Counting	6,332	1,308	4.8
G08 - Signalling	997	811	1.2
G09 - Educating; Cryptography; Display; Advertising; Seals	1,393	1,117	1.2
H01 - Basic Electric Elements	6,213	1,415	4.4
H02 - Generation, Conversion, or Distribution of Electric Power	1,894	645	2.9
H03 - Basic Electronic Circuitry	1,730	234	7.4
H04 - Electric Communication Technique	9,905	1,720	5.8
H05 - Electric Techniques not Otherwise Provided for	1,248	304	4.1
Other	37,674	16,352	2.3
Total	**139,164**	**52,799**	**2.6**

Grey rows represent businesses holding more than three times the patents of individuals. Blue rows represent individuals holding more patents than business.

Source: Intellectual Property Office and Office for National Statistics

Industry and IPC code

The list of industries shown in this article is obviously not extensive. A few industries which appear to be particularly relevant for patent statistics have been analysed to provide an indication of the link that may exist between industry and the type of innovation being patented[9]. Standard Industrial Classification (SIC) codes have been grouped to provide a subset of industries, and a breakdown of patents by IPC code (at the 3 character level) is presented for each industry. A subset of IPC codes is also selected for each industry. However this selection is empirically driven, the classifications chosen for each industry were simply those with any significant proportion of patents (3 per cent or more).

IPC codes which account for less significant amounts of patents are grouped together in a catch–all category named 'Other'. Broadly the results of this analysis, shown in **Table 3**, seem intuitive. Businesses in the chemical and pharmaceutical industry predominantly hold patents classified in chemicals and medical science; businesses in the motor vehicle industry tend to hold patents in vehicles, engines and engineering; and businesses in the computer industry mostly hold patents in computing.

The breakdown for the manufacture of electrical and optical equipment industry is much more evenly distributed. Though measuring and electrical elements are the inventions which are patented most, there are a larger number of other types of inventions being patented by businesses in this industry.

The research and development industry patents inventions predominantly in a few key areas. Medical science; biochemistry and microbiology; organic chemistry; and measuring and testing make up nearly half of the patents by this industry. The remaining patents held by research and development businesses are spread across the many remaining IPC codes.

The industry classified as other business activities comes from the services section of the SIC codes. This subset of SIC codes contains business management services; holdings companies; legal services; and a number of other business administration related activities. One phenomenon that was observed when completing the matching process and manually checking some of these matches, was enterprises changing from a manufacturing SIC code to other business activities. This may represent growth in the business, whereby the original entity becomes a holdings company which is associated with a number of subsidiary enterprises. It is difficult to tell the significance of this phenomenon – partly because this information only exists from 1997 onwards so any business which had made this change prior to 1997 would appear as a holdings company for the whole time period. The types of patents held by these companies is somewhat varied, with medical science being the largest category.

Table 3 Innovations being patented by select industries

Manufacture of chemicals and pharmaceuticals (SIC 24000–24999)

IPC	Patents	Per cent	Businesses
A61 - Medical or Veterinary Science; Hygiene	312	14.6	46
B01 - Physical or Chemical Processes or Apparatus	142	6.7	19
C07 - Organic Chemistry	510	23.9	29
C08 - Organic Macromolecular Compounds	226	10.6	27
Other	942	44.2	323
Total	**2,132**		

Manufacture of electrical and optical equipment (SIC 30000–33999)

IPC	Patents	Per cent	Businesses
A61 - Medical or Veterinary Science; Hygiene	301	5.2	78
B60 - Vehicles in General	200	3.5	20
F16 - Engineering Elements or Units	170	2.9	54
G01 - Measuring; Testing	902	15.6	194
G02 - Optics	198	3.4	41
G06 - Computing; Calculating; Counting	317	5.5	59
H01 - Basic Electric Elements	873	15.1	151
H02 - Generation, Conversion, or Distribution of Electric Power	188	3.3	69
H03 - Basic Electronic Circuitry	350	6.1	40
H04 - Electric Communication Technique	966	16.7	91
Other	1,315	22.8	573
Total	**5,780**		

Manufacture of motor vehicles; other transport equipment (SIC 34000–34999)

IPC	Patents	Per cent	Businesses
B60 - Vehicles in General	718	17.1	67
B62 - Land Vehicles other than on Rails	131	3.1	27
E05 - Locks; Keys; Windowor Door Fittings: Safes	117	2.8	13
F01 - Machines or Engines in General	291	6.9	11
F02 - Combustion Engines	422	10.1	18
F16 - Engineering Elements or Units	692	16.5	44
G01 - Measuring; Testing	363	8.6	23
H01 - Basic Electric Elements	151	3.6	17
Other	1,314	31.3	310
Total	**4,199**		

Computer and related activities (SIC 72000–72999)

IPC	Patents	Per cent	Businesses
G01 - Measuring; Testing	25	11.4	18
G06 - Computing; Calculating; Counting	67	30.6	30
H04 - Electric Communication Technique	27	12.3	20
Other	100	45.7	84
Total	**219**		

Research and development (SIC 73000–73999)

IPC	Patents	Per cent	Businesses
A61 - Medical or Veterinary Science; Hygiene	214	15.1	39
C07 - Organic Chemistry	169	11.9	20
C12 - Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation or Genetic Engineering	158	11.1	32
G01 - Measuring; Testing	135	9.5	41
Other	744	52.4	267
Total	**1,420**		

Other business activities (SIC 74000–74999)

IPC	Patents	Per cent	Businesses
A23 - Foods or Foodstuffs; Their Treatment	398	5.7	22
A61 - Medical or Veterinary Science; Hygiene	1,082	15.6	58
B65 - Conveying; Packing; Storing; Handling Thin or Filamentary Material	378	5.4	92
C07 - Organic Chemistry	635	9.1	19
F16 - Engineering Elements or Units	256	3.7	109
G01 - Measuring; Testing	471	6.8	130
Other	3,721	53.6	1,366
Total	**6,941**		

Source: Intellectual Property Office and Office for National Statistics

Concluding remarks

This article provides a description of patent data matching, along with some initial results from the matched dataset. The data sources used in this paper are very rich and will be used for a wide range of further research. Having access to the information contained in Optics has provided a unique opportunity to begin research on previously unavailable data. Along with this, using the

IDBR for matching names and addresses has given rise to the potential for exploiting these data alongside large amounts of business information which is present in ONS surveys.

Trends which have become apparent from the initial descriptive results seem predominantly intuitive. The findings suggest that there are many small businesses that own patents, but each of these owns fewer than larger businesses. Businesses hold more patents than individuals in total, and this holds in the vast majority of cases when different types of patents are taken into account. There does seem to be a link between the industry the business is classified in and the type of inventions the business patents. Whilst businesses may not patent one type of invention exclusively, the industries as a whole appear to be patenting the types of inventions that would be expected.

Whilst it should be accepted that any data matching exercise has its limitations, these results can be considered along with other quality checks. Taken as a whole, the results provide a quality assurance of the data matching process. It appears that the final matched datasets are producing intuitive results, and that these datasets are an effective research tool.

Notes

1. See for example Thoma et al. (2010), Motohashi (2009), Helmers and Rogers (2008)

2. www.statistics.gov.uk/idbr/idbr.asp

3. Note: this is not necessarily when the business started trading, a business may have existed prior to registering for VAT or employing a worker on PAYE.

4. See Thoma et al (2010) for a detailed discussion.

5. Please contact maus@ons.gsi.gov.uk for a document explaining this process.

6. All applications prior to 1997 have been matched to BSD 1997, as this is the earliest year in the BSD.

7. This approach is follows OECD 2010. This is also the case for the size distributions presented later.

8. The figures presented were produced using the whole dataset. This procedure was repeated for a number of different time periods within the data and the results do not change materially from those presented.

9. The IPO Facts and Figures 2008/2009 document was used as an indication of which industries are most relevant.

Disclaimer

Acknowledgements

The author is grateful to the UK Intellectual Property Office for providing the funding for this work. Thanks go to Khalid Khan, Eric Scheffel and Felix Ritchie from the Office for National Statistics for helpful support and advice and also to Peter Evans, Benjamin Mitra–Kahn, and Tony Clayton from the Intellectual Property Office for comments on an earlier version of the article. For support with the data matching process, the author appreciates advice from Verity Sweeney and Mike Villars of the Business Register Strategy and Outputs branch at ONS.

Contact

elmr@ons.gov.uk

References

Balasubramanian N and Sivadasan J (2009) ‘What Happens when Firms Patent? New Evidence from US Economic Census Data’ *Working Papers 08–03,* Centre for Economic Studies, U.S. Census Bureau

Evans P and Khan M K N (2009) ‘The Characteristics of Patenters’ *Economic & Labour Market Review,* Vol 3, No 12, pp 20–27

Griliches Z (1990) ‘Patent Statistics as Economic Indicators: A Survey’, *Journal of Economic Literature,* Vol XXVIII, pp 1661–1707

Hall B H, Thoma G and Torrisi S (2007) ‘The Market Value of Patents and R&D: Evidence from European Firms’ *NBER Working Paper Series,* No 13426

Helmers C and Rogers M (2008) ‘Innovation and the Survival of New Firms Across British Regions’ *Economics Series Working Papers, University of Oxford, Department of Economics,* No 416.

Intellectual Property Office (2010) ‘Facts and Figures 2008–2009’ *http://www.ipo.gov.uk/about-facts0809.pdf*

Kerr W and Fu S (2006) ‘The Industry R&D Survey: Patent Database Link Project’, *Working Papers 06–28*, Centre for Economic Studies, U.S. Census Bureau

Magerman T, Van Looy B, Song X (2006) 'Data Production Methods for Harmonized Patent Statistics: Patentee Name Harmonization', *K .U .Leuven FETEW MSI Research Report 0605,* Leuven

Motohashi K (2009) 'Software Patent and its Impact on Software Innovation in Japan', *Discussion papers 09038*, Research Institute of Economy, Trade and Industry (RIETI)

OECD (2008) 'Compendium of Patent Statistics 2008'

OECD (2009) 'An Agenda for Policy Action on Innovation' *2009 Interim Report on the OECD Innovation Strategy*

OECD (2010) 'Matching Patent and Business Data to Investigate Firms' Innovativeness' *Working Party on Industrial Analysis*

Ozman M (2007) 'Breadth and Depth of Main Technology Fields: An empirical investigation using patent data,' *STPS Working Papers 0701*, STPS – Science and Technology Policy Studies Center, Middle East Technical University

Thoma G, Torrisi S, Gambardella A, Gullec D, Hall B H and Harhoff D (2010) 'Harmonizing and Combining Large Datasets – An Application to Firm–Level Patent and Accounting Data', *NBER Working Paper Series,* No 15851

Content of Blue Book 2011

Glenn Everett
Office for National Statistics

Summary

The Quarterly National Accounts consistent with Blue Book 2011 (BB11) and Pink Book 2011 (PB11) will be published on 5 October 2011. Electronic publication of the BB11 and PB11 will be on 1 November 2011. In line with ONS policy, this article informs users about the content of these publications, including changes from previous years.

Introduction

BB11 is a major step forward, and also a foundation on which to build future improvements. This article confirms the initial plans for content of Blue Book 2011 set out in the April 2010 edition of *Economic and Labour Market Review* (see Everett 2010). The Pink Book will be largely unchanged from previous years. The Quarterly National Accounts consistent with Blue Book 2011 (BB11) will be a significant achievement:

- it will be the first time that the National Accounts will be published using the revised Standard Industrial Classification 2007 (SIC07), providing an expanded coverage of the service sector
- similarly, it will be the first time that the National Accounts will be published using the revised Classification of Product by Activity 2008 (CPA08)
- improved deflation methodology will be incorporated
- the National Accounts will be published using an integrated production system on a new IT platform; and
- the reference year and base year will both be 2008, bringing them back to normal practice and completing the catching up of the lag which had opened up in earlier years

Scope and contents of Blue Book 2011

The full scope of the Quarterly National Accounts consistent with the 2011 Blue Book dataset that will be published on 5 October 2011 will be:

- an update of the current price balancing approach used in Blue book 2010 to reconcile and balance fully the three measures of GDP for 2007, 2008 and 2009. The years 1992 to 2006 will be reconciled and rebalanced using conversion factors reflecting changes to the new industrial classification – SIC07

- the reference year will be 2008=100
- the last base year for the chained volume measure of GDP will be 2008
- all ‘industry’ outputs will be classified according to the Standard Industrial Classification 2007 (SIC07)
- all ‘product’ outputs will be classified according to the Classification of Products by Activity 2008 (CPA08)
- Consumer Price Indexes will replace the Retail Price Indexes for the purposes of deflation (that is adjusting current price estimates of GDP onto a ‘volume’ basis); and
- some improvements to Financial Services statistics will be introduced

The contents of BB11 will include:

- Part 1 – Main aggregates and summary accounts
- Part 2 – Industrial analyses
- Part 3 – the sector accounts, including Non–financial corporations, Financial corporations, General government, Households and non–profit institutions serving households (NPISH) and Rest of the world
- Part 4 – other analyses and derived statistics
- Part 5 – UK Environmental Accounts

The UK Balance of Payments Pink Book for 2011 will also be released electronically on 1 November 2011.

A related development will be the production of current price Input Output Analytical (I/O) tables. These were last produced in 2006 for the year 1995. New tables, for 2005, are now being compiled and will be published in May 2011. These tables will still classify industries according to the current SIC03.

System changes

The methods used to produce past Blue Books largely reflected the independent development of systems and processes for estimating gross domestic product (GDP), short–term indicators and supply and use tables. The three stages of the GDP process, short–term indicators, quarterly GDP estimates and annual supply and use estimates, were not fully integrated into the same framework.

ONS has been working for some time on system changes to improve the quality of the National Accounts. The current development programme (Effective National Accounts and Blue Book to measure the economy – 'ENABLE') will complete, by the end of March 2011, the integration of the National Accounts production systems, and their development and migration to a new IT platform.

The IT platform, CORD, uses SAS and Oracle and has been designed by ONS for use with time series. It will provide improvements in quality control, efficiency of work processes and business resilience, as well as enabling the methods changes set out in the following sections of this article.

Methodology changes

The annual Blue Book provides an opportunity to review methods and introduce improved methodology. In BB11, two significant changes will be introduced, a new industrial classification and an improved method of deflating current price data. There will also be some other improvements, largely in improved methods for the Financial Services. These are now described in more detail.

Industrial classification

BB11 will be published for the first time using the revised Standard Industrial Classification 2007 (SIC07) in keeping with EU regulations. These revisions are motivated by the need to adapt the classifications to changes in the structure of the economy. Some outputs have already moved to the new industrial classification, for example the Retail Sales Index has been on SIC07 since January 2010.

Key changes in SIC07 include a number of new sections giving more service sector detail. Under the new SIC, manufacturing is significantly reduced as a proportion of economic activity, which reflects the move towards more services–based economies over the past 20 years. Some of the significant changes are:

- **Business activities**: this broad heading was part of section K, 'Real estate, renting and business activities', under the previous classification, but now moves to several different areas including Section M, 'Professional, scientific and technical activities'; Section N, 'Administration and support services'; Section S, 'Other service activities'; and Section J, 'Information and communication'.
- **Retail sale of automotive fuel**: until now, this activity has been considered part of the motor trade but in today's circumstances, the sale of fuel is better considered as part of retail activity. This is reflected by its new classification to Group 47.3 (Retail sale of automotive fuel in specialised stores).
- **Recycling**: the significant increase in recycling sees it move from Section D, 'Manufacturing' in SIC 2003 to Section E in SIC07, 'Water supply, sewerage, waste management and remediation activities'.
- **Construction**: the new wider definition of construction includes development activity.

To incorporate these changes and to provide users a time series on the new SIC07 basis, data on the previously published SIC03 basis are being converted using factors provided by the ONS's Business Register, where businesses are classified on the old and new basis. The main indicator used for these changes is turnover. This is used as a starting point for the adjustments, with some additional manual adjustments made using available additional information. This will provide a time series back to 1992 on the new SIC07 basis – the data from 2008 onwards are 'real' SIC07 data (collected using the new classification).

ONS's Business Registers Unit has produced a cross categorisation between SIC 2003 and SIC 2007 which can be accessed via the ONS website. See Hughes (2008), Hughes and Brook (2009) and Hughes et al (2009) for details of related SIC07 articles.

Deflators

Until now, some National Accounts series have been deflated using the Retail Prices Index (RPI). BB11 will change to using to the conceptually better Consumer Prices Index (CPI) to meet best practice and in line with Eurostat's Price and Volume Handbook. The CPI is designed to be consistent with the European System of National Accounts. This should ensure greater consistency across all National Accounts compiler areas, thus improving the coherence of the National Accounts.

The basic theory and expected impact of changing from a Paasche index (used in the RPI) to a Laspeyres index (in the CPI) is that it will raise the level of GDP. A separate article detailing this change and expected impact will be published prior to the publication of BB11.

Financial services improvements

BB11 will contain some improvement to financial services data. These improvements stemmed from the *Better financial statistics for policy* project (see Walker 2011 for the latest progress report). Generally the improvements replace existing data sources and methodology with more reliable ones but have little overall impact on top level GDP. These include adjustments for the Bank of England's Special Purpose Vehicles (SPVs), net spread earnings and debt securities.

Adjustments

The production of the National Accounts necessarily involves the application of data adjustments and balancing items, as recognised in international rules and guidance (for example SNA 2008, paragraph 22.77). There are many reasons for applying data adjustments, ranging from low level error corrections, to adjustments used to aid the alignment of high level series or balance aggregate economic series according to agreed practices.

The application of adjustments in the processes of production is not new. However, ONS has taken the opportunity, synchronised with the development and use of new systems for the production of the UK's National Accounts, to review the categorisation and application of data adjustments. The new adjustment process will be used for the first time in the production of the 2011 Blue Book and the datasets linked and associated with this product.

The new categorisation provides the benefit of clearer and more standardised recording of adjustments. The classification has also been aligned more directly with the adjustments required in the Gross National Income (GNI) process tables that the UK must provide periodically to

Eurostat. These tables break down the data in the accounts into sources of data and types of adjustment, so the new process should simplify the production of these tables. Internal analysis and interpretation of the accounts should also be aided by the new categorisation.

There have also been changes to the stage at which some adjustments are applied, in line with the changes to production introduced with the new systems, and to resolve some current production difficulties.

Other

Some other minor revisions will be incorporated, for example where new data are available to replace buffers or estimated/forecast data. They cover, for example, areas such as child trust funds and social security administration.

Revisions

The introduction of new data and methodological changes into the National Accounts is carefully managed. Revisions for the most recent three years reflect a range of routine annual revisions, including the annual benchmarking exercise to incorporate new and revised current price data as part of the regular supply and use tables.

In Blue Book 2011, 2009 will be balanced through the input–output supply and use framework for the first time and 2007 and 2008 will be fully rebalanced. The years 1992–2006 will be revised using relevant conversion factors for the changes to the new industrial classification.

Regional Accounts

The Regional Accounts statistical bulletin 'Regional, Sub–regional and Local Gross Value Added 2010' is planned for release on 14 December 2011. This will include the move to the new industrial classification SIC07. The publication will reflect the SIC07 changes, and NUTS1 (Nomenclature of Units for Territorial Statistics) estimates will be for 2010 (provisional), while NUTS2 and NUTS3 estimates will be for 2009.

Estimates will be provided at NUTS1 (regional), NUTS2 (sub–regional) and NUTS3 (local area) geographical levels; with a 20– industry breakdown at NUTS1 and NUTS2. For NUTS3, a 10–industry breakdown will be produced. Data will be presented for total GVA, GVA per head of population and GVA per head indices.

Background

Definitive estimates of GDP are produced at the Blue Book stage by reconciling and balancing annual estimates of expenditure, income and production in current prices through the input–output

framework. The input–output framework consists of the regularly produced supply and use tables and input–output analytical tables. The supply and use tables are compiled around 18 months after the year in question, when comprehensive information on expenditure, income and production becomes available.

The expenditure, income and production approaches are based on different survey and administrative data sources, and each produces estimates that, like all statistical estimates, are subject to errors and omissions. Accordingly, the three measures produce different estimates, although theoretically they should be the same. A single, definitive, GDP estimate can only emerge therefore after a process of balancing and adjustment.

Supply and use tables are constructed to show a balanced and complete picture of the flows of products in the economy and illustrate relationships between producers and consumers of goods and services. In addition, they show the interdependence between industries: what industries either purchase from one another or import to produce their output.

The supply and use tables are the basis on which input output analytical tables (IOATs) are constructed. IOATs describe how different goods and services are used in the production of other goods and services. They are used to measure how changes in demand for one product can filter through to the rest of the economy.

Blue Book 2011 production and risk management

In June last year, the publication of the Quarterly National Accounts consistent with the Blue Book was delayed for several days. In the light of that experience, steps have been taken to minimise risk and ensure delivery of BB11 to time. This has taken account of advice received from Peter van de Ven, the head of National Accounts in the Netherlands, on ways of improving procedures and risk management. .

These steps include:

- enhanced project management procedures
- more focus on identifying and managing key risks, including contingencies
- early planning and clear criteria for revision bids
- refreshing practical training for all the staff involved in compilation work
- extensive user testing and piloting the new ENABLE systems
- phased sign off dates
- additional progress review dates; and
- a ‘twin key’ quality assurance of final tables

Change to Quarterly Accounts (M2) in August 2011

To ensure compilers have sufficient time to prepare for the major changes in Blue Book 2011, ONS has decided to change the approach to Quarterly National Accounts for Q2 2011. A Month 1 estimate, based only on output data, will be published as normal in July. Then in August, this output–based estimate will be updated. This would supplement the July Preliminary GDP estimate by replacing the third month of forecast data for the Index of Production (IoP), the Index of Services (IoS) and the monthly construction output survey with actual data. This would replace the Output, Income and Expenditure statistical bulletin in August 2011. There will be no published information on the income or expenditure components in the second quarter until the Quarterly National Accounts (QNA) release in early October.

Scope of Blue Book 2012

In line with normal practice, production of the Blue Book is being managed over a two–year period – currently BB11 and BB12. Initial plans for the scope and content of the Blue Book in 2012 include:

- publication date will coincide with the release of Quarterly National Accounts for Q2 2012
- reference year will be 2009 = 100
- last base year for the chained volume measure of GDP will be 2009
- new methodology for the Insurance industry will be introduced (this will include revision to previous years)
- consistency between the Public Sector Finance statistics and the National Accounts will be improved, also incorporating the Whole Government Accounts (WGA), which are becoming available; and
- further improvements to Financial Services statistics will be included

Future developments

Over the next 2–3 years, ONS will make further improvements to the systems, methods and analysis supporting core National Accounts outputs.

One driver is the need to prepare for transition to the new European System of Accounts 2010, with the expected implementation date of 2014. The key changes here include the treatment as capital of Research and Development (R&D) activities, improvements to the treatment of pension statistics, to the treatment of goods sent abroad for processing and the delineation between market and non–market activities. There will also be changes in preparation for implementation of the Balance of Payments Manual 6, to a similar timescale. ONS will consult further on the methods and data requirements for these changes, aiming to achieve compliance with international regulations at the minimum cost, while also meeting the analytic and quality needs of UK users.

Some key additional developments are being considered for the next few years, including:

- balancing the three measures of GDP through supply and use tables in constant price (volume) terms, adding to the existing approach which involves balancing just nominal, or current price, estimates. Initially, this may be just for annual estimates, but ONS will investigate the feasibility of producing quarterly constant price supply and use tables, which would form the basis of the official estimates of GDP
- exploration of options for producing a monthly measure of GDP
- further improvements in the measurement of Financial Services
- further development of the 'satellite' accounts (including accounts for Research and Development and the Environment)
- the development of a labour accounting framework, integrated with the National Accounts

These developments will be additional to implementation of ongoing methodological improvements, which will be taken forward on a prioritised basis and subject to available resources.

ONS will also move some further areas onto CORD, offering additional benefits in terms of efficiency and quality control. Details of plans and timescales will be set out in further articles.

Additional articles planned

The following are topics of articles that are planned prior to the publication of the Blue Book. These will provide additional detail regarding methodology changes and the impact on the National Accounts:

- Deflators
- Trade in services
- Industrial classification change to SIC07

Contact

elmr@ons.gov.uk

References

Everett G (2010) 'Plans for Blue Book 2010' Economic & Labour Market Review April 2010. Available at www.statistics.gov.uk/elmr/04_10/downloads/elmr-Apr10.pdf

Hughes J C (2008), ‘SIC 2007: implementation in ONS’, Economic & Labour Market Review August 2008. Available at www.statistics.gov.uk/cci/article.asp?id=2034

Hughes J C and Brook K (2009) ‘Implementation of SIC 2007 across the Government Statistical Service (GSS)', *Economic & Labour Market Review* August 2009. Available at www.statistics.gov.uk/elmr/08_09/downloads/ELMR_Aug.pdf.

Hughes J C, James G, Evans A and Prestwood D (2009) ‘Implementation of Standard Industrial Classification 2007: December 2009 update, *Economic & Labour Market Review* December 2009. Available at www.statistics.gov.uk/elmr/12_09/downloads/ELMR_Dec09.pdf

United Nations (2008) ‘System of National Accounts 2008’ (ISBN 978-92-1-161522-7)

Walker G (2011), Developing financial statistics for policy – Progress report January 2011’, *Economic & Labour Market Review* January 2011 Available at www.statistics.gov.uk/cci/article.asp?ID=2627

Consumer Prices Index and Retail Prices Index: the 2011 basket of goods and services

Philip Gooding
Office for National Statistics

Summary

The 'shopping baskets' of items used in compiling the Consumer Prices Index (CPI) and Retail Prices Index (RPI) are reviewed each year. Some items are taken out of the baskets and some are brought in to make sure the CPI and RPI are up to date and representative of consumer spending patterns. This article describes the review process, explains how and why the various items in the CPI and RPI baskets are chosen, and discusses the main changes to the contents of the CPI and RPI baskets for 2011 from the 2010 price collection. Similar articles have been published in previous years. This article also describes one other change relating to the way in which prices of seasonal items are measured in the CPI and RPI.

The shopping basket

A convenient way of thinking about both the CPI and RPI is to imagine a 'shopping basket' containing those goods and services on which people typically spend their money. As the prices of the various items in the basket change over time, so does the total cost of the basket. Movements in the CPI and RPI represent the changing cost of this representative shopping basket.

In principle, the basket should contain all consumer goods and services purchased by households, and the prices measured in every shop or outlet that supplies them. In practice, both the CPI and RPI are calculated by collecting a sample of prices for a selection of representative goods and services in a range of UK retail locations. Currently, around 180,000 separate price quotations are used every month in compiling the indices, covering over 650 representative consumer goods and services for which prices are collected in around 150 areas throughout the UK.

Within each year, the CPI and RPI represent the changing cost of a basket of goods and services of fixed composition, quantity and quality. In practice, this is achieved by (a) keeping constant the sample of representative goods and services; (b) applying a fixed set of weights to price changes

for each of the items such that their influence on the overall index reflects their importance in the typical household budget; and (c) taking care to ensure that replacements for brands that are no longer stocked in an individual shop are of comparable quality. In this way, changes in the CPI and RPI from month to month reflect only changes in prices, and not ongoing variations in consumer purchasing patterns.

However, the contents of the CPI and RPI baskets of goods and services and their associated expenditure weights are updated annually. This is important in helping to avoid potential biases that might otherwise develop over time – for example, due to the development of entirely new goods and services, or the tendency for consumers to move away from buying goods and services which have risen relatively rapidly in price. For example, if the price of tea rose dramatically during one year, consumers might switch their spending towards coffee, making it necessary to adjust the expenditure weights accordingly in the following year.

These procedures also help to ensure that the indices reflect longer–term trends in consumer spending patterns. For example, the proportion of household expenditure devoted to household services has broadly risen overall over the last 25 years. This is reflected both in an increasing weight for this component in the CPI and RPI, and the addition of new items in the basket to improve measurement of price changes in this area: examples include internet subscriptions, playgroup and nanny fees.

For the RPI, changes to the items and weights are introduced in the February index each year, but prices are collected for both old and new items in January. This means that the figures for each year can be 'chain linked' together to form a long–run price index spanning many years. In other words, price changes into January are based on the old basket and weights, and price changes between January and February, and beyond, are based on the new basket and weights. This procedure ensures that the annual changes to the basket and weights have no impact on estimated changes in prices as measured by the indices. The same basic approach is adopted in the CPI although, for technical reasons, it is necessary to chain link the published index twice each year rather than only once as in the RPI[1].

ONS (2010a) '*Consumer Price Indices – a Brief Guide*' provides a helpful introduction to the concepts and procedures underpinning the compilation of the CPI and RPI indices. These are described in much greater detail in ONS (2010b) '*Consumer Price Indices – Technical Manual*'.

Representative items

There are some individual goods and services where typical household spending is so large that they merit inclusion in the basket in their own right: examples include petrol, and electricity and gas supply. However, it would be both impractical and unnecessary to measure price changes of every item bought by every household in compiling the CPI and RPI.

More commonly, a sample of specific goods and services has to be selected that gives a reliable measure of price movements for a broader range of similar items. For example, price changes for

garden spades might be considered representative of price changes for other garden tools. The selection of these representative items is judgmental since the significant difficulties involved in defining an adequate sampling frame (that is, a list of all the individual goods and services bought by households) restrict the use of traditional random sampling methods when choosing the representative items.

For each product grouping, a number of items are selected whose price movements, when taken together, provide a good estimate of the overall change in prices for the group. For example, there are around 20 representative items in the CPI 'furniture and furnishings' class, from bedroom wardrobes to kitchen units, whose prices are used to calculate an overall estimate of price changes for all furniture products.

The prices collected for each product group are then combined to produce the overall CPI and RPI, with weights proportional to total expenditure on the entire product group. So the weight given to 'furniture and furnishings' in the CPI shopping basket, or 'furniture' in the RPI basket, reflects average household spending on all furniture products as opposed to spending on the basket items only. Similarly, the weight of garden spades would be derived from all spending on garden tools.

Table 1 High level RPI weights since 1987

Weights are specified as parts per 1000 of the all items RPI

RPI GROUP	1987	1992	1997	2002	2007	2008	2009	2010	2011
Food	167	152	136	114	105	111	118	112	118
Catering	46	47	49	52	47	47	50	47	47
Alcohol	76	80	80	68	66	59	63	64	60
Tobacco	38	36	34	31	29	27	27	27	28
Housing	157	172	186[1]	199	238	254	236	237	238
Fuel and light	61	47	41	31	39	33	49	40	42
Household goods	73	77	72	73	66	66	70	67	65
Household services	44	48	52	60	65	64	61	59	63
Clothing and footwear	74	59	56	51	44	42	39	40	44
Personal goods and services	38	40	40	43	39	41	41	41	38
Motoring expenditure	127	143	128	141	133	133	121	144	137
Fares and other travel costs	22	20	20	20	20	20	20	20	20
Leisure goods	47	47	47	48	41	38	38	37	36
Leisure services	30	32	59[2]	69	68	65	67	65	64

1. Depreciation costs were added to the housing group in 1995

2. Foreign holiday costs were added to the leisure services group in 1993, followed by UK holidays a year later.

Source: Retail Prices Index

These expenditure weights are also updated annually so that the indices reflect current spending patterns. In line with usual practice, CPI class weights were updated with effect from the January 2011 index, and RPI section weights will be revised with effect from the February index, at which point the weights for the more detailed (unpublished) item indices will also be revised. A brief comparison of high level RPI weights since 1987 is shown in **Table 1**, including the new weights for 2011. The table illustrates that over the period there are some clear shifts in expenditure. Broadly speaking, weights for services have increased while those for goods have decreased. A more detailed article on changes to the published CPI and RPI weights will be published on the National Statistics website in April 2011.

The contents of the CPI and RPI baskets are very similar, although the precise weights attached to the individual items in each index differ[2]. There are, however, some differences between the baskets. For example, the RPI basket includes a number of items chosen to represent owner-occupier housing costs, including mortgage interest payments and depreciation costs, all of which are excluded from the CPI. These differences are summarised in ONS (2010c) *'Differences between the RPI and CPI Measures of Inflation'* and described in greater detail in Roe and Fenwick (2004*), 'The New Inflation target: the Statistical Perspective*'.

Selecting the representative items

A number of factors need to be taken into account when choosing representative items. Of course, the items must be easy to find by price collectors, so ensuring that estimates of price change are based on an adequate number of price quotes collected throughout the UK. Since the CPI and RPI are based on the cost of a fixed in–year basket of goods and services, ideally they should also be available for purchase throughout the year. However, availability of some food and clothing items is clearly seasonal, and so these goods require a slightly different treatment in the indices.

The number of items chosen to represent each product group within the CPI and RPI depends both on the weight (expenditure) of the group and also the variability of price changes between the various items that could be selected to represent the group (reflecting, for example, the diversity of products available). Intuitively, it makes sense to choose more items in product groups where spending is high; this helps to minimise sampling variability in the estimate of price change for high–weighted groups, and therefore in the overall price index. However, if price movements of all possible items in the group are very similar, it is sufficient to collect prices for only a few[3]. By contrast, if price movements of all the possible items are very different, prices will be needed for many representative items to get a reliable overall estimate of price change for the group.

Based on this, the allocation of items to broad commodity groups can be analysed, as shown for the 12 divisions of the CPI in **Table 2**, and the balance used as an anchor for the annual review of the basket. The significant allocation of items to the food division relative to its index weight, for example, is partly explained by the relatively high variation in observed price changes between the individual goods in this area. Conversely, a smaller proportion of items relative to index weight is allocated to the restaurants and hotels division, reflecting greater similarity in observed price changes. In some cases, such as transport and housing, apparent low allocations of items are explained by the presence of some dominant individual items (for example car purchase and motor

fuels, and housing rents respectively). Here, the case for adding further items to improve coverage of these divisions' remaining index weights is much weaker – instead, it is far more important to ensure that the sampling of prices for these heavily weighted items is as comprehensive as possible.

Table 2 Allocation of items to CPI divisions in 2011

		CPI weight (per cent)	Observed variation in price changes[1]	Representative items[2] (per cent of total)
1	Food & non-alcoholic beverages	11.8	High	23
2	Alcohol & tobacco	4.2	Low	3
3	Clothing & footwear	6.2	High	11
4	Housing & household services	12.9	High	5
5	Furniture & household goods	6.1	Medium	11
6	Health	2.4	Low	3
7	Transport	15.9	High	6
8	Communication	2.6	High	1
9	Recreation & culture	14.7	High	17
10	Education	1.8	High	1
11	Restaurants & hotels	12.0	Low	8
12	Miscellaneous goods & services	9.4	High	11

1. Based on an analysis of variation in price changes between the individual items chosen to represent each division in the period 2004-2008.

2. These figures should be treated as providing only a broad indication of the allocation of items to the 12 CPI divisions. For example, the sample of prices underpinning an existing item might easily be stratified in some way to form two or more distinct items; conversely, items could be merged to form a single item representing a wider, more heterogeneous, spending category. A specific example of this is the item 'University tuition fees'. This is classified as one item but the index takes into account prices for several hundred courses, including undergraduate, postgraduate, part–time etc.

The analysis also helps to highlight those areas of the index which might benefit most from improved coverage, such as miscellaneous goods and services. The current allocation of items to the division is broadly comparable to its index weight but variation in price changes appears relatively high, possibly reflecting the diversity of goods and services covered by this division. As discussed later, this type of analysis has motivated some of the changes to the basket introduced in 2011. Conversely, it also helps to highlight areas where there is scope to remove items from the basket without any significant loss of precision in the indices. It is important that growth in the overall size of the basket is limited each year so that production costs and processing times are contained.

Such analysis cannot tell us *which* items should be priced, and so choosing a particular set of items to represent each area remains a matter of judgement. CPI and RPI commodity groupings are regularly reviewed with the aim that all significant items or distinct markets where consumers' expenditure exceeds around £400 million annually are explicitly represented in the basket, except where those items are judged to be adequately represented by other items in the basket[4]. Conversely, where spending on items falls below the £100 million mark, there should be good reason for their continuing inclusion in the basket. For example, while spending on acoustic guitars and power drills is relatively low, both are included in the basket to represent wider markets (musical instruments and electrical tools respectively) that would otherwise not be covered explicitly. Trends in expenditure, as well as the latest available figures, help to inform the decisions in all cases.

This focus on expenditures in determining the contents of the basket partly reflects the data that are available describing household spending patterns. One major source of information comes from the diaries and questionnaires filled in by people taking part in the ONS Living Costs and Food Survey, a continuous survey of around 6,000 households each year. This is supplemented by detailed analyses of trends presented by market research companies, trade journals and in press reports. Changes in the retail environment are also reported to ONS by the price collectors, and together these various sources of information help to ensure that the goods and services that the average household spends its money on are appropriately represented in the CPI and RPI baskets.

It is very important to note that the contents of the basket and, in particular, changes from one year to the next should not be given significance beyond their purpose as representative items used in estimating retail price changes. Changes to the basket will reflect evolving consumer tastes, but only over a long run of years. In any particular year, changes to the basket will reflect a range of considerations such as practical experience in collecting prices, the desire to improve coverage in high spending areas, or analysis that suggests that estimated price changes could be improved at the margin by varying the number or type of representative items collected. Indeed, within each product grouping there is usually a point at which the exact number, choice of items and the precise weights attached to them becomes a matter of relatively fine judgement. At this detailed level, it is unlikely that such choices would have any significant impact on the CPI and RPI indices. For example, a selection of specific household appliances has been chosen to represent spending on small electrical goods, including irons and kettles. However, other representations would clearly be possible and equally valid.

It should also be noted that the vast majority of the 650 or so representative items remain unchanged in 2011.

In summary, selection of representative items is based on a number of factors, including:

- ease of finding and pricing the product
- availability throughout the year
- variability of prices within a class
- analysis of balance across the basket
- amount spent on a particular item or the group of items

Changes to the basket in 2011

Changes to the basket of goods and services this year are being introduced with the February 2011 consumer price indices published on 22 March 2011; that is, monthly changes in prices from February 2011 to January 2012 inclusive are estimated with reference to the updated basket. The basket will be updated again at the same time next year.

New additions to the basket in 2011 and those items removed are set out in **Tables 3** and **4**, together with a brief summary of the motivation for these changes. As the tables make clear, these motivations are diverse. As in previous years, changes to the basket in 2011 certainly should not be viewed as a simple indicator of those products or services whose popularity has either grown or fallen significantly over the past year. Note that all of the changes to the basket this year affect both the CPI and RPI indices.

The bullet points below give a brief summary and explanation of the themes behind the changes to the basket for 2011:

- A number of new items are introduced to represent specific markets where consumer spending is significant, and existing items in the basket may not adequately represent price changes for such goods. For example, craft kits are being introduced to represent a distinct sector not previously covered in the games, toys and hobbies class. Oven–ready joints are also included for the first time to reflect a longer–term move to prepared food and away from traditional joints of meat.

- Two high–technology goods are being introduced this year that can be seen as representing evolving trends: smart phone handsets and mobile phone applications are introduced to capture price changes in these rapidly expanding markets.

- In addition to introducing items to represent distinct sectors or markets, a number of items have been introduced to diversify the range of products collected for already established groupings, usually where spending is significant. For example, hardback fiction books widen coverage of the books sector. Similarly the number of television items in the basket has been increased from two to three, reflecting the greater diversity in size now available and, in particular, to better represent the upper end of the range where they form part of home entertainment systems.

- Analysis of the broad balance of the existing sample of representative items across the CPI highlighted a need to improve coverage of price changes for a number of CPI classes. These areas include:
  - Appliances and products for personal care (12.1.2/3), with a new item covering hair conditioner to supplement existing toiletries such as toothpaste and deodorant;
  - Other services not elsewhere covered (12.7.0), where dating agency fees have been added to broaden coverage in a particularly diverse grouping.

Table 3 Additions to the basket in 2011

CPI Class	RPI Section	New Item	Notes
01.1.2 Meat	2110 Other Meat	Oven-ready joint	Replaces pork shoulder joint reflecting a longer-term movement to prepared food and replacing an item which was sometimes difficult to collect since joints are sometimes only available towards the end of the week and on weekends.
01.1.6 Fruit	2128 Processed Fruit	Dried fruit	New item. Introduced to improve coverage of fruit products which have been identified as an under-represented area of the basket. It also reflects a distinct market and widens the coverage of processed fruit in the basket.
02.1.2 Wine	3104 Wines & Spirits Off Sales	Sparkling wine	New item. Introduced to reflect spending on this type of item.
03.1.2 Garments	5101 Men's Outerwear	Men's casual shirt, long/short sleeve	Replaces men's casual long-sleeved shirt with the aim of improving coverage during the summer months when the item was difficult to collect.
03.1.2 Garments	5102 Women's Outerwear	Women's blouse/shirt	New item. A second blouse/shirt has been added to the basket reflecting the expenditure. This second item is aimed at the casual end of the market.
03.1.2 Garments	5103 Children's Outerwear	Girl's jacket	Replaces girls' winter & girls' summer jacket. Introduced to improve coverage of the item across the year and to remove seasonal items whose prices have to be imputed during months when they are not available.
03.2.0 Footwear incl Repairs	5105 Footwear	Women's high heeled shoes	Replaces women's party shoes and women's high heeled sensible shoes due to footwear being over-represented in the basket.
04.3.1 Materials for Maintenance and Repair	4106 DIY Materials	Medium density fibreboard (MDF)	Replaces hardboard as more representative of this area of the market and reflecting the relative spending on the two items.
06.1.2/3 Other Medical and Therapeutic Equipment	5201 Personal Articles	Spectacle frames with single vision lens	Replaces spectacle frames without lens reflecting the way in which outlets increasingly do not price frames separately from lens.
08.2.0 Telephone and Telefax Equipment and Services	4404 Telephone Charges	Mobile phone applications	Replaces mobile phone downloads. Introduced to capture price changes in this new expanding technology.
08.2.0 Telephone and Telefax Equipment and Services	4303 Electrical Appliances	Smart phone handsets	New item. Introduced to capture price changes in this new expanding technology.
09.1.1 Equipment for the Reception and Reproduction of Sound and Pictures	6301 Audio-visual Equipment	Flat panel TV 14"-22"	New item. Introduced to capture the differing price movements from the wider range of screen sizes available, in particular the top end where they form part of home entertainment systems.
09.1.1 Equipment for the Reception and Reproduction of Sound and Pictures	6301 Audio-visual Equipment	Flat panel TV 23"-32"	New item. Introduced to capture the differing price movements from the wider range of screen sizes available, in particular the top end where they form part of home entertainment systems.
09.1.1 Equipment for the Reception and Reproduction of Sound and Pictures	6301 Audio-visual Equipment	Flat panel TV 33"+	New item. Introduced to capture the differing price movements from the wider range of screen sizes available, in particular the top end where they form part of home entertainment systems.
09.3.1 Games, Toys and Hobbies	6303 Toys, Photographic and Sports Goods	Craft kit	New item. Introduced to improve coverage of this under-represented area of the basket. It also represents a distinct sector not previously covered within the class.
09.5.1 Books	6304 Books and Newspapers	Hardback book - fiction	New item. Introduced to improve coverage of books which has been identified as an under-represented area of the basket.
09.5.2 Newspapers and Periodicals	6304 Books and Newspapers	Provincial newspaper	Replaces morning & evening provincial newspapers where the distinction has blurred over time.
12.1.2/3 Appliances and Products for Personal Care	5202 Chemists Goods	Hair conditioner	New item. Introduced to improve coverage of this under-represented area of the basket and reflects spending on hair conditioners.
12.7.0 Other Services not elsewhere covered	4402 Fees and Subscriptions	Dating agency fees	New item. Prices of services in this residual category vary greatly and dating agency fees have been introduced to improve coverage of this under-represented class.

Table 4 Items removed from the basket in 2011

(Note: 'low weighted' denotes an item with a CPI weight of less than 0.5 parts per thousand in 2010)

CPI Class	RPI Section	Dropped Item	Notes
01.1.2 Meat	2107 Pork	Pork shoulder	Replaced by oven-ready joint reflecting a longer-term movement to prepared food. Also aimed at improving coverage since the pork shoulder was often only available towards the end of the week and on weekends.
02.1.3 Beer	3102 Beer Off Sales	Lager 4 cans	Removed. The pack of four cans of lager has been removed to reduce the number of premium strength beers and rebalance the basket in this area.
02.2.0 Tobacco	3201 Cigarettes	Vending machine cigarettes	Removed. Spending on vending machine cigarettes has been declining and current legislation will see these withdrawn during 2011.
03.1.2 Garments	5101 Men's Outerwear	Men's casual shirt, long sleeved	Replaced by men's casual long or short sleeved shirt in order to improve coverage during the summer months.
03.1.2 Garments	5102 Women's Outerwear	Women's casual outer jacket (e.g fleece)	Removed. Women's jackets remain represented in the basket by a separate women's casual outer jacket and a women's formal jacket.
03.1.2 Garments	5102 Women's Outerwear	Women's premium branded dress	Removed. This reflects the reduced weight of dresses in the basket. Dresses remain represented by a women's casual or formal dress.
03.1.2 Garments	5103 Children's Outerwear	Girls' summer jacket	Combined with girls' winter jacket to improve coverage of the item across the year and to remove a seasonal item whose price has to be imputed during months when it is not available.
03.1.2 Garments	5103 Children's Outerwear	Girls' winter jacket	Combined with girls' summer jacket to improve coverage of the item across the year and to remove a seasonal item whose price has to be imputed during months when it is not available.
03.2.0 Footwear including Repairs	5105 Footwear	Women's high heeled sensible shoes	Combined with women's high heeled party shoes since footwear is over-represented in the basket.
03.2.0 Footwear including Repairs	5105 Footwear	Women's high heeled party shoe	Combined with women's high heeled sensible shoes since footwear is over-represented in the basket.
04.3.1 Materials for Maintenance and Repair	4106 DIY Materials	Hardboard	Replaced by medium density fibreboard (MDF) reflecting the market share of the two items.
06.1.2/3 Other Medical and Therapeutic Equipment	5201 Personal Articles	Spectacle frames (without lenses)	Replaced by spectacle frames with single vision lens since outlets increasingly do not price frames without lens.
08.2.0 Telephone and Telefax Equipment and Services	4404 Telephone Charges	Mobile phone downloads	Replaced by mobile phone applications.
09.1.1 Equipment for the Reception and Reproduction of Sound and Pictures	6301 Audio-visual Equipment	Flat panel TV 14"-25"	Replaced by three sizebands of flat panel televisions to better measure differing price movements across the variety of television sizes available, particularly the top end where they form part of home entertainment systems.
09.1.1 Equipment for the Reception and Reproduction of Sound and Pictures	6301 Audio-visual Equipment	Flat panel TV 26"-42"	Replaced by three sizebands of flat panel televisions to better measure differing price movements across the variety of television sizes available, particularly the top end where they form part of home entertainment systems.
09.3.3 Gardens, Plants and Flowers	6305 Gardening Products	Rosebush	Removed. A low weighted item in a section which is over-represented.
09.3.4/5 Pets, Related Products and Services	4306 Pet Care	Vet fees for spaying a kitten	Removed. This section of the basket is over-represented and in some cases cat spaying may be free. Vet fees remain represented by annual booster injections.
09.5.2 Newspapers and Periodicals	6304 Books and Newspapers	Morning provincial newspaper	Combined with the evening provincial newspaper as the distinction has become blurred over time.
09.5.2 Newspapers and Periodicals	6304 Books and Newspapers	Evening provincial newspaper	Combined with the morning provincial newspaper as the distinction has become blurred over time.

- In other cases, the new items are direct replacements for similar products that leave the basket in 2011. For example, medium density fibreboard (MDF) replaces hardboard with the new product representing an increasing proportion of the market compared with the product it replaces. Another example here is a provincial newspapers item replacing separate morning and evening provincial newspapers with the main driver being the blurring of the distinction between the two separate items. Collection issues can also have an effect. For example, spectacle frames with single vision lens replace frames without lens as an increasing number of outlets do not price frames separate from the lens.

- It is important that the review of the basket considers not just the list of items to be priced, but also where the prices are collected. For example existing items are being collected from an increasing number of large supermarkets as they widen their product bases. These items include Blu–ray discs, computer games and reference books.

- The seasonality of some items in the basket is also reviewed. This year, separate items for girls' winter and summer jackets have been merged into one girls' jacket item. This should improve coverage across the year and remove two seasonal items for which prices have to be imputed during months when they are not available.

- Finally, as always, minor changes are made to the definitions in the item list caused by the availability of products. The main example this year is foreign exchange which is priced based on commission charges and changes in exchange rates. Previously the RPI item was based only on the cost of buying foreign currency but from 2011, it will be based on selling currency as in the CPI.

As noted earlier, it is important that growth in the overall size of the basket is limited each year so that production costs and processing times may be contained. A number of items therefore have been removed from the basket in 2011 to make room for the new additions. Some reflect reduced expenditure such as vending machine cigarettes which are set to be withdrawn during 2011 due to legislation. In other cases, the removal of items from the basket does not necessarily imply that the markets for these goods and services are very small or are declining significantly.

- Some items have been removed to make way for new additions to the basket within the same product grouping. For example, spectacles without lens have been removed in 2011 to be replaced by spectacles with lens, although both products represent spectacles. Similarly hardboard has been replaced by MDF. In other cases, items have been removed so that new items may be introduced covering distinct markets not previously represented explicitly within the product grouping. For example, the pork shoulder has been removed so that oven–ready joints can be represented.

- In some cases a product will still remain represented in the basket even if there is no longer an explicit item. For example, a women's premium branded dress has been removed but it is still

covered by a more broadly defined casual and formal dress item. Similarly a women's casual outer jacket (for example fleece) is still covered by other casual and formal jacket items.

- Elsewhere, analysis suggested that there was scope to remove items from certain product groupings without any significant loss of precision in estimates of price changes overall. Within these groupings, those items with relatively low index weights or those items which are variants of other items have typically been chosen; examples include rose bushes and vet fees for spaying a kitten. In each case, it is judged that price changes for these goods remain adequately represented by those items that remain in the basket. The removal of items in such cases therefore represents a rebalancing of the basket, helping to offset the expansion of coverage in other product areas.

Other changes

One other change has already been introduced this year, with the January index published in February. This relates to the method used to measure prices of seasonal items. These are products likely to be unavailable to price for certain months of the year.

Previously the price index calculated in the last month that the product was available was carried forward until it became available again. The new method involves imputing the price forward each month using the average price movement of 'in–season' products. The 'in-season' products used are ones in the same classification group as the 'out of season' product but they are available to price when the 'out of season' product is not. The new method is internationally recognised as an improvement over the previous approach. This change is described in greater detail in a consultation document and a response to that consultation, *'Measurement of Seasonal Items within the Consumer Prices Index and Retail Prices Index'*[5].

Notes

1. CPI indices are chain–linked first each January, when weights for CPI classes and higher level aggregates are updated, and again in February when changes to the basket are introduced and hence weights for individual item indices are reviewed.

2. RPI weights are based primarily on household spending estimates derived from the Living Costs and Food Survey, and relate to expenditures by private households only, excluding the top 4 per cent of households by income and those pensioner households mainly dependent on state benefits. CPI weights are based on National Accounts estimates of household final consumption consistent with the wider CPI population coverage (that is, all private households, residents of institutional households and foreign visitors to the UK).

3. At the extreme, if price changes for all the possible items that could be selected in a particular group were identical each month, it would be necessary to select only one of the items for inclusion

in the basket. Price changes for this one item would be perfectly representative of price changes for the group as a whole.

4. Under CPI regulations, items should be included in the CPI where estimated consumers' expenditure is 1 part per thousand or more of all expenditure covered by the CPI; based on household final consumption data underpinning calculation of the 2011 CPI weights, this is equivalent to around £730 million.

5. www.ons.gov.uk/about/consultations/closed-consultations/measurement-of-seasonal-items-within-the-cpi-and-rpi/measurement-of-seasonal-items-within-the-cpi-and-rpi.html

Contact

Public enquiries: 0845 601 3034
CPI/RPI enquiries: 01633 456900
Email: cpi@ons.gov.uk

References

Roe D and Fenwick D (2004) 'The New Inflation Target: the Statistical Perspective', *Economic Trends* No 602. Available at www.statistics.gov.uk/articles/economic_trends/ET602Roe.pdf

ONS (2010a) 'Consumer Price Indices – a Brief Guide'. Available at www.statistics.gov.uk/downloads/theme_economy/CPI-briefguide.pdf

ONS (2010b) 'Consumer Price Indices – Technical Manual'. Available at www.statistics.gov.uk/downloads/theme_economy/CPI_Technical_manual-2010.pdf

ONS (2010c) 'Differences between the RPI and CPI Measures of Inflation'. Available at www.statistics.gov.uk/downloads/theme_economy/cpi-rpi-information-note.pdf

Profits from UK Foreign Direct Investment

A view from the micro–data

Tullio Buccellato, Eric Scheffel and Andrew Thomas
Office for National Statistics

Summary

This article uses the Annual Foreign Direct Investment survey to investigate the changing patterns and the determinants of profits from the foreign subsidiaries of UK–based multinational enterprises. The results show that these profits are greatest in countries which are closely related to the UK historically, linguistically and geographically – particularly the USA and the Netherlands. However, low tax economies and the BRICs (Brazil, Russia, India and China) are also becoming more significant. Econometric analysis shows that factors such as openness to trade, GDP per capita, population, human capital and unit labour costs help to explain the level of these profits.

Introduction

Foreign direct investment (FDI) is a hugely important economic phenomenon. The flows of investment income that it generates form a significant part of the Balance of Payments in open economies like the UK. International flows of capital and ideas also play a fundamental role in stimulating growth around the world and is particularly crucial for developing and emerging market economies (see Lucas 1990). A rich literature has grown up around the subject looking at the various determinants of FDI.

This article presents some analysis of firm–level data reported in the Annual Foreign Direct Investment (AFDI) survey. This is an important source of information for the production of FDI and trade statistics frequently published by the Office for National Statistics and also annually in the Pink Book[1]. The majority of previous studies have confined their analysis to cross–country data of aggregated FDI flows and earnings by country. The great advantage of studying micro–data for one particular country is in exploring the effects of individual firm (or enterprise) characteristics on these flows, which might otherwise be difficult to discern from aggregated data. It is a well–known problem in econometric analysis that aggregation can introduce biases in estimated relationships between variables.

This article focuses on the profits earned by the foreign subsidiaries of UK–based companies which are or have been recipients of FDI flows emanating from the UK. It is important to note that responses collected through the AFDI survey do not make it easy to distinguish between firms which are UK–based and UK–owned multinational enterprises (MNEs) and firms which are UK–based and foreign–owned MNEs. To investigate this dimension is interesting in its own right and has been the attention of a number of other studies using the AFDI survey micro–data set. These typically employ some form of identification scheme in order to be able to tell the two types of firms apart (see Breinlich and Criscuolo 2011). In this study all UK firms which have been engaged in FDI are not in any further way disaggregated.

This article reports on three main aspects of profits from UK FDI:

- Using simple descriptive statistics to provide a detailed view of aggregate FDI profits by country and region and how these have changed over time.
- Looking at trends in the number and location of foreign subsidiaries.
- Using econometric analysis to investigate some of the determinants of FDI profits for UK firms. This is achieved by using firm–level data and also a number additional country–specific variables such as population size, corruption, stock of human capital and also unit labour costs among others.

Profits from foreign subsidiaries owned by UK–based multinational firms

Table 1 summarises the main sources of profits from the foreign subsidiaries of UK–based multinational companies between 1997 and 2008. Three–year averages are used in order to smooth out the more volatile year–to–year movements (1997–1999; 2000–2002; 2003–2005 and 2006–2008). The countries occupying the Top 10 positions tend not to change their ranking positions very much and are usually among the more advanced group of economies. This stability often reflects that the UK's external stock of FDI has been built up over many years. This picture changes markedly when considering the Top 20 and Top 30 positions. Here, different countries change their positions more often throughout the years in question including some countries entering and leaving the Top 30.

The Top 10 countries in terms of profits earned from the overseas subsidiaries of UK–based multinational companies tend to share either a linguistic, historical or geographic proximity to the UK. The two largest countries – the U.S. and the Netherlands – together account for around half of all profits earned by the Top 10 countries. South Africa is a fairly new entrant to the Top 10 appearing in the final two time categories. Tax–friendly countries such as Switzerland and Luxembourg have also been prominent fixtures in the Top 10.

Table 1 Profits from UK outward FDI – the Top 30 countries

£ millions

	1997-1999		2000-2002		2003-2005		2006-2008	
1	United States	11,520	United States	11,746	United States	13,504	United States	16,094
2	Netherlands	6,728	Netherlands	8,508	Netherlands	10,480	Netherlands	7,280
3	Australia	2,124	Ireland	1,787	Germany	2,312	Luxemburg	5,099
4	France	1,762	France	1,642	Ireland	2,223	Hong Kong	2,822
5	Germany	1,611	Germany	1,634	France	2,111	Switzerland	2,656
6	Ireland	1,325	Switzerland	1,453	Luxemburg	2,013	South Africa	2,613
7	Hong Kong	1,206	Australia	1,444	South Africa	1,975	Australia	2,396
8	Singapore	1,161	Luxemburg	1,427	Australia	1,906	France	2,376
9	Canada	972	Canada	1,082	Switzerland	1,639	Germany	2,167
10	Switzerland	799	Singapore	931	Singapore	1,328	Ireland	2,132
11	Spain	735	South Africa	805	Hong Kong	1,254	Jersey	1,800
12	Brazil	679	Sweden	745	Canada	1,205	Singapore	1,729
13	South Africa	631	Spain	695	Bermuda	1,185	Canada	1,510
14	Bermuda	627	Jersey	685	Sweden	1,167	Sweden	1,189
15	Belgium	597	Hong Kong	645	Jersey	783	Bermuda	1,165
16	Jersey	476	Italy	582	Spain	757	Guernsey	963
17	Italy	468	Belgium	504	Guernsey	678	Spain	923
18	Malaysia	447	Bermuda	475	Italy	587	Belgium	841
19	New Zealand	366	Japan	433	Malaysia	560	British Virgin Islands	683
20	Japan	331	Brazil	376	Belgium	551	Chile	674
21	Denmark	290	Malaysia	353	Brazil	487	Italy	617
22	Chile	289	Guernsey	344	Chile	433	Brazil	552
23	Sweden	266	New Zealand	294	Trinidad and Tobago	429	Cayman Islands	502
24	United Arab Emirates	242	India	283	Canada	391	Canada	439
25	Norway	216	Puerto Rico	274	India	364	Vatican City State	435
26	Luxembourg	189	Trinidad and Tobago	261	Mexico	347	Mexico	422
27	Portugal	182	Denmark	253	New Zealand	319	Malaysia	421
28	India	176	Norway	238	Cayman Islands	299	Senegal	406
29	Mexico	166	Mexico	214	Japan	297	India	395
30	Panama	156	Canada	211	Poland	282	Isle of Man	394
Total		**36,738**		**40,323**		**51,864**		**61,697**

Source: Annual Foreign Direct Investment survey

The list of countries ranked 10–20 in the list is more variable with a total of 17 different countries appearing in these positions between 1997 and 2008 (compared to 12 different countries making up the Top 10 positions over the same time period). Low–tax economies such as Jersey, Guernsey, Bermuda and the British Virgin Islands have become increasingly prominent. Chile has also moved strongly up the rankings, perhaps a reflection of the strong increase in grade A copper prices for which Chile produced a third of total global supply.

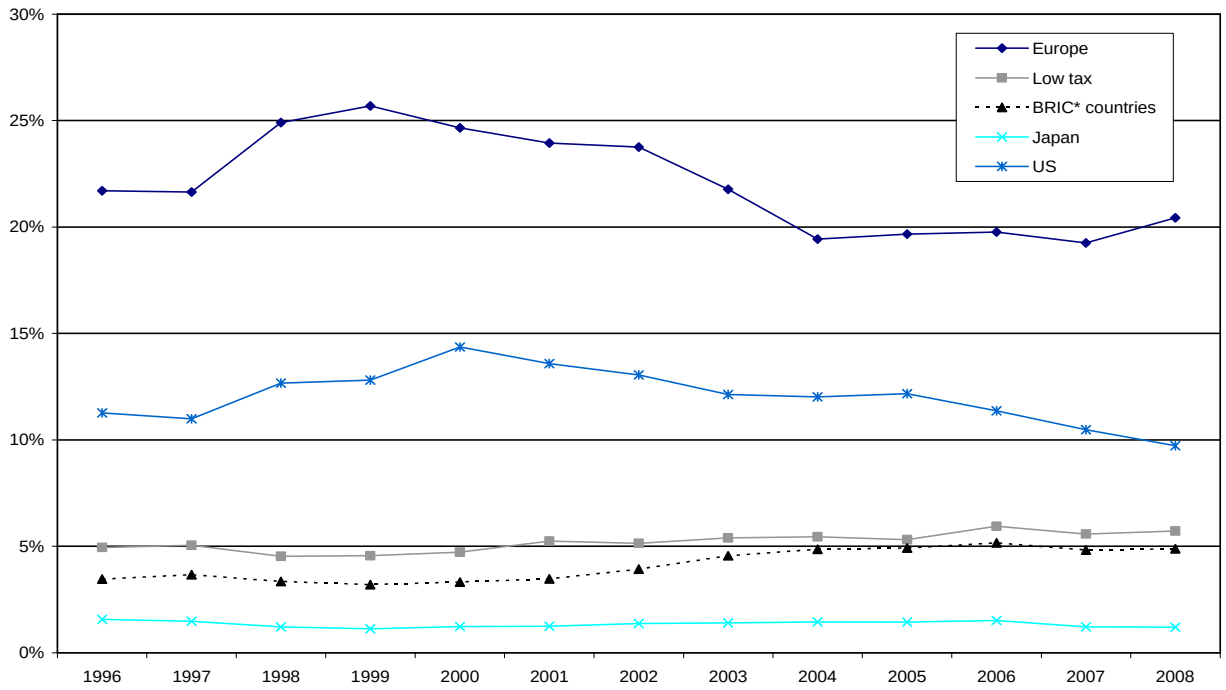
The list of countries making up positions 20–30 is even more variable – a total of 24 different countries. Some of the more recent entrants are the Cayman Islands, Senegal and Vatican State all of which moved into the Top 30 in 2003–2005 and 2006–2008.

Trends in the location of foreign subsidiaries

Figure 1 shows how the relative location of foreign subsidiaries owned by UK MNEs have changed between 1996 and 2008. The proportions shown in the figure reflect, for each region, the count of foreign subsidiaries owned as a proportion of all foreign subsidiaries owned by UK–based firms. This provides a measure of the relative positioning of each UK–based MNE in terms of the number of subsidiaries they operate in different countries across the globe and how this changes through time.

Figure 1 Locations of UK–owned foreign subsidiaries

Average proportion



* Brazil, Russia, India and China

Source: Annual Foreign Direct Investment Survey

As Figure 1 shows, Europe and the U.S. are traditionally the main locations of foreign subsidiaries owned by UK–based firms. In the late 1990s both regions exhibited falls in their respective shares although there was a slight upturn for Europe in the last year. In contrast, the relative shares accounted for by a group of low–tax economies[2] and the BRIC countries have increased since the early 2000s, while the trend for Japan has remained relatively flat. Although the shares for Europe and the U.S. are still very large in direct comparison with – say – the BRIC countries (for which this share rise from roughly 3 per cent to just under 5 per cent), the substantial fall in this trend of the U.S. from close to 15 per cent in 2000 to less than 10 per cent in 2008 and for Europe from 25 per cent to 20 per cent over the same period could be some evidence of a structural shift in the pattern of FDI.

Of course, an important consideration is that modern technology, especially in providing new ways of distributing products and services (through the internet, for example) may partially have led to a reduction of physical subsidiaries and a concentration of such activities in only a few centres. This may be particularly pervasive in the case of service–sector firms, which typically serve the more mature and developed markets of Europe and North America.

Econometric analysis on the determinants of the FDI profits of UK–based multinational enterprises

Data sources

The main source of data is the AFDI which provides micro–data on FDI profits earned by UK–based multinational companies. This is augmented by using the Business Structure Database[3] to add the employment size of each business to the AFDI dataset and by a number of macroeconomic control variables which are listed in **Table 2**. These are country–specific factors that may well have a bearing on profits from FDI such as openness to trade, corruption perception, GDP per capita, population, unit labour cost, human capital measures such as secondary school gross enrolment, market capitalisation as proportion of GDP as well as a measure of purchasing power parity.

Table 2 Data sources and their expected influence on FDI profits

Data	Source	Expected Sign
FDI profits	AFDI micro–data set	*None*, independent variable
Firm–level employment of UK MNE headquarter	BSD micro–data set	*Positive*, if size implies greater market power
Openness to trade	World Bank	*Positive*, openness to trade may increase competitiveness and innovation which may result in higher profits
Secondary school gross enrolment	World Bank	*Positive*, better stock of human capital may also spur growth and innovation leading to higher profits.
Corruption perception index	Transparency International	*Uncertain*, corruption could be bad for UK-controlled foreign subsidiaries if this disfavours them vis-à-vis local firms. But it could also result in higher profits if corruption favours them. (Important: the corruption index is inverted in the sense that a higher index implies less corruption)
Unit labour cost	OECD	*Negative*, if labour cost is very low then profits could potentially be higher.
Population	Penn World Tables	*Positive*, The greater the population the greater the potential consumer market and the greater the potential for firms to benefit from economies of scale.
Market capitalisation as a proportion of GDP	Penn World Tables	*Positive*, Market capitalisation of listed companies as percentage of GDP. Higher financial sophistication should alleviate financing constraints and help profits.
Purchasing power parity (PPP)	Penn World Tables	*Uncertain*, depends on the nature of the subsidiary. If the subsidiary services the local market, higher (consumption goods) prices could increase revenues (if demand is relatively inelastic). If the subsidiary is an export platform and the higher prices pass through to intermediate input factors, then this could diminish profits.

Regression results

This section summarizes the results obtained from running a number of cross–section and panel regressions at the individual firm–level. All of the regression tables reported here take the logarithm of profits from UK–based MNE's subsidiaries in foreign countries as the dependent variable.

Table 3 presents the results obtained from running year–by–year cross–sectional regressions using simple OLS. Employment is highly statistically significant with a positive sign – as expected – and is likely to be an indicator of the scale of a particular firm's foreign operations. Market capitalisation as a percentage of GDP is also estimated with a positive sign and strong statistical significance. The partial effect of unit labour costs on profits provided a mixed picture although a clearer result is achieved in an alternative estimation framework presented later in the article.

The total size of the population within each country also contributes positively to profits from the foreign subsidiary, however only by a margin of statistical significance in the last three years of the sample. It is however noteworthy that regardless of significance, the estimate remains positive throughout all years. Finally, while openness to trade marginally appears to positively influence profits, corruption, gross enrolment in secondary education and PPP are found to be largely insignificant.

Table 3 Main variables affecting FDI profits by year[1]

Dependent variable: FDI–profits

Indicator[2]	2000	2001	2002	2003	2004	2005	2006	2007
Employment	0.337***	0.358***	0.393***	0.474***	0.454***	0.356***	0.273***	0.300***
	(0.012)	(0.012)	(0.012)	(0.016)	(0.017)	(0.012)	(0.012)	(0.011)
Openness	-0.000	0.001	0.001	0.002	0.002	0.001	0.002***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Corruption	-0.048*	-0.008	-0.028	0.002	-0.065*	-0.038	-0.060**	-0.027
	(0.026)	(0.031)	(0.027)	(0.035)	(0.034)	(0.024)	(0.026)	(0.029)
School enrolment	0.000	-0.001	-0.002	-0.001	0.008**	0.004*	-0.000	-0.000
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
GDP per capita	0.187	0.081	0.054	-0.507**	-0.503**	-0.103	0.299*	-0.245*
	(0.165)	(0.172)	(0.163)	(0.223)	(0.225)	(0.147)	(0.159)	(0.140)
Labour cost index	-0.100	0.024	-0.051	0.208	0.428**	0.076	0.039	0.342***
	(0.128)	(0.126)	(0.118)	(0.130)	(0.170)	(0.100)	(0.135)	(0.120)
Population	0.030	0.045	0.053	0.034	0.057	0.076**	0.094***	0.101***
	(0.041)	(0.042)	(0.041)	(0.043)	(0.046)	(0.031)	(0.035)	(0.033)
Market capitalisation/GDP	0.003***	0.004***	0.007***	0.004***	0.004***	0.005***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
PPP	0.000	0.001*	0.001	0.003***	0.002**	-0.002*	-0.002***	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-3.775***	-4.551***	-3.707***	-1.536	-4.305***	-3.756***	-6.405***	-4.731***
	(0.980)	(1.059)	(1.092)	(1.576)	(1.608)	(1.153)	(1.193)	(1.116)
Observations	4164	4099	4887	4221	3467	4784	4042	4816
R-squared	0.212	0.230	0.224	0.205	0.218	0.188	0.153	0.188

1. OLS regression. Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1.
2. Indicators described in Table 2. All estimations were done in logarithms.

Table 4 takes a different perspective on the data by simply running OLS year–by–year regressions on firm–level profits in foreign subsidiaries against the total ratio of subsidiaries in that particular foreign country relative to the total number of foreign subsidiaries (see Figure 1). If the ratio is high for one particular firm, it means that it possesses a high foreign country–specific market penetration for that particular foreign market. In general, this is expected to be an indicator of scale in the foreign market (whether as a distributor in the foreign market or as an acquirer of inputs) and therefore a strong positive driver of profits.

Table 4 Regressions by year and geographic area of investment[1]

Dependent variable: FDI–profits

Indicator[2]	2000	2001	2002	2003	2004	2005	2006	2007	2008
Employment	0.255***	0.260***	0.299***	0.348***	0.350***	0.296***	0.225***	0.258***	0.008***
	(0.009)	(0.008)	(0.010)	(0.012)	(0.013)	(0.010)	(0.009)	(0.008)	(0.002)
Low tax economies	1.014***	0.912***	1.152***	0.681*	1.083***	0.510**	1.072***	1.015***	-0.037*
	(0.215)	(0.163)	(0.237)	(0.354)	(0.379)	(0.212)	(0.211)	(0.185)	(0.021)
Europe	-0.276***	-0.246***	-0.401***	-0.532***	-0.255	-0.124	-0.262***	0.110	-0.039***
	(0.066)	(0.075)	(0.087)	(0.131)	(0.161)	(0.097)	(0.077)	(0.074)	(0.013)
BRIC	1.878***	2.346***	2.140***	2.704***	2.057***	1.792***	0.659***	1.113***	-0.054**
	(0.451)	(0.341)	(0.484)	(0.556)	(0.466)	(0.273)	(0.220)	(0.205)	(0.024)
Japan	1.438**	0.251	1.122**	3.256***	1.791***	0.366	-0.231	0.755**	-0.265
	(0.581)	(0.459)	(0.551)	(0.902)	(0.685)	(0.383)	(0.262)	(0.331)	(0.163)
US	-0.267***	-0.273***	-0.151	-0.962***	-0.751***	-0.287**	0.124	0.212**	-0.022*
	(0.069)	(0.084)	(0.098)	(0.141)	(0.163)	(0.118)	(0.084)	(0.083)	(0.013)
Constant	-2.094***	-2.165***	-2.493***	-2.860***	-2.831***	-2.322***	-1.581***	-2.102***	-0.030***
	(0.071)	(0.073)	(0.086)	(0.117)	(0.125)	(0.079)	(0.069)	(0.058)	(0.007)
Observations	6172	6490	7399	6815	5686	7216	6097	7586	7592
R-squared	0.186	0.190	0.196	0.196	0.198	0.162	0.118	0.149	0.005

1. OLS regression. Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1.
2. Indicators described in Figure 1. All estimations were done in logarithms.

The sign and statistical significance of the estimated coefficient for the BRIC countries is particularly marked. For this group the level of profits appears to be higher at firm–level when compared to other countries/regions (with the exception of the crisis year of 2008). In contrast, partial regression coefficients for the U.S. and Europe more often result in negative estimates, albeit often statistically insignificant ones. Countries which are typically considered to be 'low tax' have also been separately identified and the estimated coefficient is generally statistically significant and positive (again, except for the last year of 2008). It goes without saying that the final year of 2008 may have been affected by globally–driven factors not captured in this regression which have led to a depression in FDI profits more generally.

Table 5 provides results on a panel estimation (so includes both the time series and cross–sectional elements simultaneously) of the factors investigated in Tables 3 and 4. To make the results as robust as possible regressions are run using both fixed effects (FE) and random effects (RE) models[4].

Table 5 Panel estimation results on the determinants of FDI profits[1]

Dependent variable: FDI–profits

Indicator	Random effects	Fixed effects	Random effects	Fixed effects	Random effects	Fixed effects	Random effects	Fixed effects
Employment[2]	0.273***	0.051***	0.272***	0.051***	0.184***	0.038***	0.247***	0.048***
	(0.005)	(0.010)	(0.005)	(0.010)	(0.004)	(0.007)	(0.005)	(0.010)
Openness[2]	0.001***	0.012***	0.001***	0.011***			0.002***	0.012***
	(0.000)	(0.002)	(0.000)	(0.002)			(0.000)	(0.002)
Corruption[2]	-0.103***	-0.082***	-0.099***	-0.082***			-0.052***	-0.080***
	(0.012)	(0.024)	(0.012)	(0.024)			(0.012)	(0.024)
School enrolment[2]	0.002***	0.004***	0.002***	0.004***			0.002**	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)			(0.001)	(0.001)
GDP[2]	0.994***	0.600***	1.005***	0.537***			0.784***	0.542***
	(0.067)	(0.172)	(0.071)	(0.188)			(0.071)	(0.188)
Unit labour cost[2]	-0.196***	0.062	-0.161**	0.140			-0.102	0.052
	(0.069)	(0.219)	(0.072)	(0.240)			(0.070)	(0.239)
Population[2]	0.062***	5.088***	0.061***	5.133***			0.195***	4.749***
	(0.019)	(0.723)	(0.019)	(0.727)			(0.020)	(0.726)
Market capitalisation/GDP[2]			-0.000	0.000			0.000	0.000
			(0.000)	(0.000)			(0.000)	(0.000)
PPP[2]			0.003***	0.001			0.000	0.002
			(0.001)	(0.004)			(0.001)	(0.004)
Low tax[3]					0.522***	-0.156	1.664***	-0.025
					(0.073)	(0.134)	(0.131)	(0.191)
Europe[3]					-0.503***	-0.439***	-0.634***	-0.502***
					(0.035)	(0.067)	(0.047)	(0.082)
BRIC[3]					1.333***	0.408***	2.456***	-0.137
					(0.099)	(0.158)	(0.174)	(0.225)
Japan[3]					1.364***	1.965***	0.919***	1.516***
					(0.172)	(0.296)	(0.217)	(0.352)
USA[3]					-0.627***	-0.940***	-1.018***	-0.872***
					(0.040)	(0.074)	(0.056)	(0.089)
Constant	-10.807***	-60.777***	-11.293***	-61.464***	-1.611***	-0.379***	-11.041***	-56.369***
	(0.423)	(6.321)	(0.446)	(6.384)	(0.030)	(0.057)	(0.439)	(6.385)
Observations	38400	38400	38400	38400	67058	67058	38400	38400
R-squared		0.031		0.031		0.006		0.036

1. Panel estimation on all years. Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1.
2. See Table 2
3. See Figure 1

Table 5, which presents more robust panel estimations on a number of indicator and country variables, provides different results to those obtained by simple OLS methods and reported in Tables 3 and 4.

First, the openness to trade, corruption and human capital–related gross enrolment of secondary employment are all now found to be statistically significant. While it is not surprising to see openness and stocks of human capital in the foreign country positively affecting profits of subsidiaries located there, the result for corruption is interesting. This is a difficult variable to interpret and may well be acting as a proxy for other factors – for example countries which have higher levels of corruption may have a poorer legal framework resulting in less competitive markets.

The remaining variables display coefficient estimates which are approximately in line with those expected and estimated using the simple cross–sectional OLS approach. In particular the BRIC countries stand out favourably, but market capitalisation and purchasing power parity are now insignificant. It is also noteworthy that the partial effect ascribed to the unit labour cost coefficients now almost always displays a statistically significant negative sign across a number of estimation runs as would be expected.

Conclusions

This study uses micro–data from the AFDI survey to investigate the main determinants of the profits earned by the foreign subsidiaries of UK–based MNEs. Looking at simple descriptive statistics shows that BRIC countries are becoming more attractive locations although profit levels and the proportion of all foreign subsidiaries located in this group of countries is still relatively low, especially when compared to the United States, Europe, and other historically, geographically and linguistically close nations to the UK such as many (former) Commonwealth countries.

Regression analysis finds that openness to trade, human capital, GDP per capita and the total size of the foreign country's population all contribute positively to profits in foreign subsidiaries. Panel estimation also shows unit labour costs are inversely related to profits. The regression analysis also supports the growing importance of BRIC nations to firm–level profits – these being particularly fast growing markets.

Notes

1. www.statistics.gov.uk/StatBase/Product.asp?vlnk=1140

2. The 17 countries making up the low tax group are: Bermuda, Cyprus, Seychelles, Switzerland, Trinidad & Tobago, Bahamas, Cayman Islands, Dominica, Luxembourg, Isle of Man, Liechtenstein, Channel Islands, Solomon Islands, Fiji, US Virgin Islands, British Virgin Islands and St. Lucia.

3. The Business Structure Database (BSD) is an annual snapshot of the Inter–Departmental Business Register (IDBR). This provides a live update of all businesses in the UK registered for

VAT and for which employees are registered for PAYE. The BSD provides data on business turnover and employment.

4. There are two common assumptions made about individual specific effects, the random effects (RE) assumption and the fixed effects (FE) assumption. The random effects assumption is that the individual specific effects are uncorrelated with the independent variables. The fixed effect assumption is that the individual specific effect is correlated with the independent variables.

Contact

elmr@ons.gov.uk

References

Lucas R (1990), Why Doesn't Capital Flow from Rich to Poor Countries?, *The American Economic Review*, Vol. 80, No. 2, May 1990.

Breinlich H and Criscuolo C (2011), *Matching ARD and AFDI*, Ceriba Data Guide.

Assessing the accuracy of business–level forecasts

Pete Brodie, Tullio Buccellato and Eric Scheffel
Office for National Statistics

Summary

This article presents original work on some aspects of forecasting at the individual business– or firm–level. In particular, two ways are suggested for assessing the accuracy of these forecasts based on the calculation of average percentage errors and the construction of a 95 per cent confidence interval. It is found that the quality of forecasts tend to become increasingly unreliable after two years and that the decay in forecast quality is inversely related to the frequency of the time series – that is the less often a time series is updated/published the faster the deterioration in forecast quality.

Introduction

Economists routinely generate forecasts of economic variables such as gross domestic product (GDP) and its components. These are important for helping policy–makers and organisations to make decisions about the future. However, forecasts are usually based on published aggregates while, so far, there has been little investigation of the potential for using micro– or business–level data that underlie these aggregates to forecast. This article aims to do just that by using simple time series methods to generate forecasts of business–level data on turnover and capital expenditure as well as suggesting how the accuracy of these forecasts can be assessed.

The main purpose for collecting business–level or micro–data is for the production of economic statistics. However, once these statistics have been published, the remaining value in the micro–data is mainly in their use for research. The Virtual Microdata Laboratory (VML) at ONS allows secure access to this data, enabling researchers from government, universities and other organisations to undertake work that focuses on the individual characteristics of businesses and households whilst maintaining data confidentiality.

This work aims to find new value in micro–data. To produce timely aggregated figures for the state of the economy it is usual practise for National Statistics Institutions to impute missing values in business registers. One method for imputing this data to extrapolate past trends using simple one

variable (or univariate) methods – a process know as *nowcasting* as the approach is simply attempting to predict the present. Forecasting just takes this one step further, investigating whether past trends in micro–data can be used to predict the future.

Being able to assess the accuracy of these forecasts is important and two ways are proposed. First, estimated points are classed as inaccurate when the margin of error exceeds 20 per cent. A second method implies that the prediction is inaccurate if it falls outside of a 95 per cent confidence interval.

The analysis is conducted on four important ONS business datasets. These are the Annual Business Inquiry (ABI), the quarterly Capital Expenditure survey (CAPEX), the Monthly Inquiry into Distribution and Services Sector (MIDSS) and the Monthly Production Inquiry (MPI).

The next section of this article outlines the formal framework used to compute forecasts at the individual firm level and how to access their accuracy. The third section presents the results of the empirical analysis and the final section concludes.

Methodological framework

This section outlines a simple time series procedure, namely the Holt–Winters process that is used to produce forecasts. Then, two methods are proposed for testing the accuracy of out–of–sample forecasts – these are based on mean and median percentage errors and also confidence intervals.

Time series procedures

A standard multiplicative Holt–Winters process is used to produce forecasts of turnover and capital expenditure for an individual business. This method is already widely–used at the Office for National Statistics (for example in various parts of the Index of Services). It works by using information on the level, trend and seasonality of the time series to form a one period ahead forecast of a variable $\hat{y}_{t+1}$ such that

$$\hat{y}_{t+1} = \left(m_t + b_t\right) c_{t-s}$$

where m_t is the level of the series at time t, b_t the gradient of the time series and c_{t-s} the multiplicative seasonal coefficient such that $k = 1, 2, ..., s$ for each of the s seasons (for example, monthly s=12 and quarterly s=4).

Each of these three coefficients is updated recursively as another observation (or forecast) becomes available, such that:

Level

$$m_t = \alpha_0 \frac{y_t}{c_{t-s}} + (1-\alpha_0)(m_{t-1} + b_{t-1})$$

Trend:

$$b_t = \alpha_1 (m_t - m_{t-1}) + (1-\alpha_1) b_{t-1}$$

Seasonal:

$$c_t = \alpha_s \frac{y_t}{m_t} + (1-\alpha_s) c_{t-s}$$

The coefficients α_0, α_1 and α_s describe how quickly the level, trend and seasonal coefficients are updated in response to new data observations. The higher each of these coefficients, a larger weight is placed on the current observation relative to the past history of the series. These coefficients can be estimated using the past time series. These updating coefficients are likely to be lower the better–behaved/less volatile the past data.

The great advantage of this method lies in its simplicity and its ability to take account of the presence of both a time trend as well as a seasonal component in the data, even when these components are experiencing changes through time. Forecasts are simply generated in a one–step–ahead fashion with the Holt–Winters equations re–estimated each time to take account of a new data observation or forecast. The process also allows a large number of forecasts to be produced fairly quickly, which is important when dealing with a large number of firm–level time series.

The Holt–Winters procedure is used for monthly turnover data from MPI and MIDSS and quarterly investment spending from CAPEX. However, for the annual turnover data collected by the ABI, where seasonality is not an issue and the number of observations is smaller, a basic time trend is used for the basis of extrapolating past data.

Applying these univariate methods requires a complete time series so that there are no gaps or missing observations. Large and small firms are treated differently in ONS's sampling frame, with large businesses typically included in every period while smaller businesses are rotated in and out of the sample, so will typically have periods of non–inclusion. This is usually to lessen the administrative burden on smaller companies. As a result, the forecasting exercise here is limited to a select number (ranging from 38 to 50) of the largest businesses in each dataset which have been continuously sampled. Larger firms also tend to be less volatile and therefore easier to forecast using these simple time series methods.

Once the forecasts at the business–level have been produced they are then aggregated.

Accuracy of forecasts

A succession of one–step–ahead forecasts $\hat{y}_{i,t+h}$ can be formed for each of the i = 1,..,N firms, but as the data moves further away from the actual reference time t (that is as *h* gets larger) the predictive power tends to fall. But how long does it actually take for data to become 'obsolete' in the sense that the quality of the predictive values falls below an acceptable level?

Two approaches are used to assess the quality of forecasts.

Mean and median absolute percentage error

First, prediction accuracy is summarised with the mean and median absolute percentage errors (MAPE and MedAPE) for the N firms.

$$MAPE_{t+h} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|y_{i,t+h} - \hat{y}_{i,t+h}\right|}{y_{i,t+h}} * 100$$

$$MedAPE_{t+h} = \frac{\left|y_{i=N/2,t+h} - \hat{y}_{i=N/2,t+h}\right|}{y_{i=N/2,t+h}} * 100$$

This allows conclusions to be drawn on accuracy in a similar way to what is usually done in National Statistics Institutions with sampling errors.

Confidence intervals

Secondly, upper and lower bounds around the forecasted values are constructed showing a 95 per cent confidence interval. A forecast is judged unreliable if it falls outside of this range. There are two ways in which the 95 per cent confidence interval can be constructed.

The first is based on fixed upper and lower boundaries calculated with the standard deviation computed at the initial point of the forecasting horizon. Confidence intervals estimated using only the in–sample data can become quite inaccurate as *h* gets relatively large (that is the forecasts extent into the future). However, results based on this measure are reported because it is based on a real information set.

The second approach updates these boundaries over the forecasting horizon, computing the standard deviation iteratively as new data points (from forecasts and observed values) emerge so uses both in–sample and out–of–sample information. The variance is expected to grow reflecting a gradual loss in forecasting accuracy over time, so it is anticipated that the confidence bands would widen.

Empirical analysis on ONS business data

This section presents results from the forecast exercises on the Annual Business Inquiry (ABI), the Capital Expenditure survey (CAPEX), the Monthly Inquiry into Distribution and Services Sector (MIDSS) and the Monthly Production Inquiry (MPI) relating to turnover and capital expenditure. In each case the Holt–Winters or trend extrapolation procedure is fitted but not to the end of the sample. This then allows a number of observations for the accuracy of forecasts to be evaluated. In the case of ABI, MPI and MIDSS turnover five years of observations at the end of the sample are reserved for out–of–sample testing. For investment spending on the quarterly CAPEX survey three years of data are used for testing the accuracy of predictions.

Absolute percentage errors

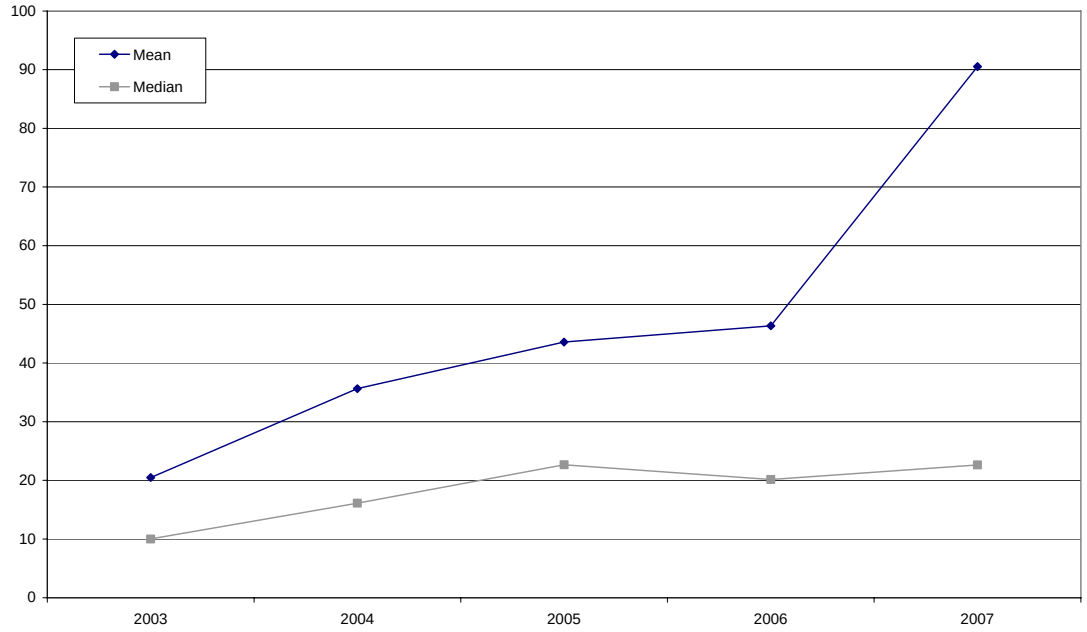
Figures 1 to 4 compute the MAPE and MedAPE for the available cross–section of businesses at any given point in time in each survey. This is an equivalent approach to computing of the conditional forecast accuracy of out–of–sample predictions of individual firms. Unsurprisingly, all of the line graphs generally exhibit a positive slope which implies a decreasing forecasting accuracy as a greater out–of–sample prediction horizon is considered. Forecasts of the CAPEX survey (Figure 2) are particularly bad, perhaps due to the inherent volatility in investment data – especially at the firm–level.

Because the MedAPE considers the central rather than the average forecast error it is more robust to outliers. Therefore, significant differences between the MAPE and MedAPE are usually accounted for by a small number of particularly bad forecasts at the individual business level. For example, consider the predictive accuracy measures computed for the last year of the ABI (Figure 1), where a large reported turnover value in a single business caused the MAPE to jump from below 50 per cent in 2006 to over 90 per cent in 2007.

Figures 3 and 4 show the conditional forecast accuracy based on the MAPE and MedAPE statistics for the MPI and MIDSS micro–data. These also shows a widening gap between the MedAPE and the MAPE in the last third of the out–of–sample forecast horizon – which is also predominantly due to the influence of one or a few businesses, whose forecasts turn out to be comparatively bad relative to the actual turnover outcome. It is also clear from the monthly forecasts for MPI and MIDSS that not all of the seasonality present in the individual business' turnover series is modelled satisfactorily, as characteristic peaks and troughs are plainly apparent at regular intervals in Figures 3 and 4.

Figure 1 Absolute percentage error – ABI (turnover)

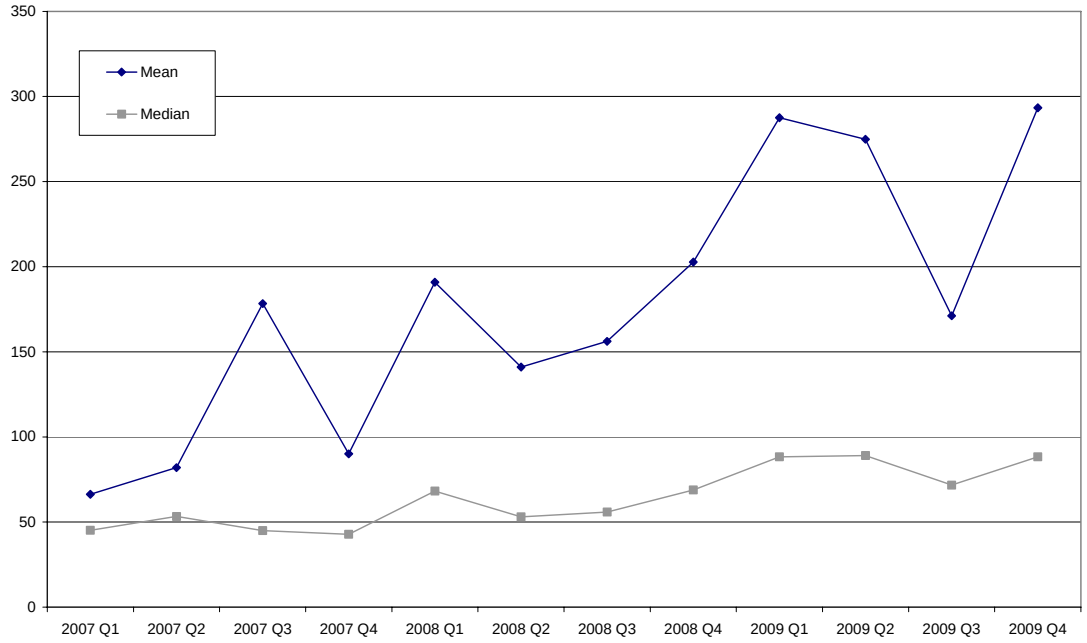
Per cent



Source: Authors' calculations

Figure 2 Absolute percentage errors – CAPEX (investment)

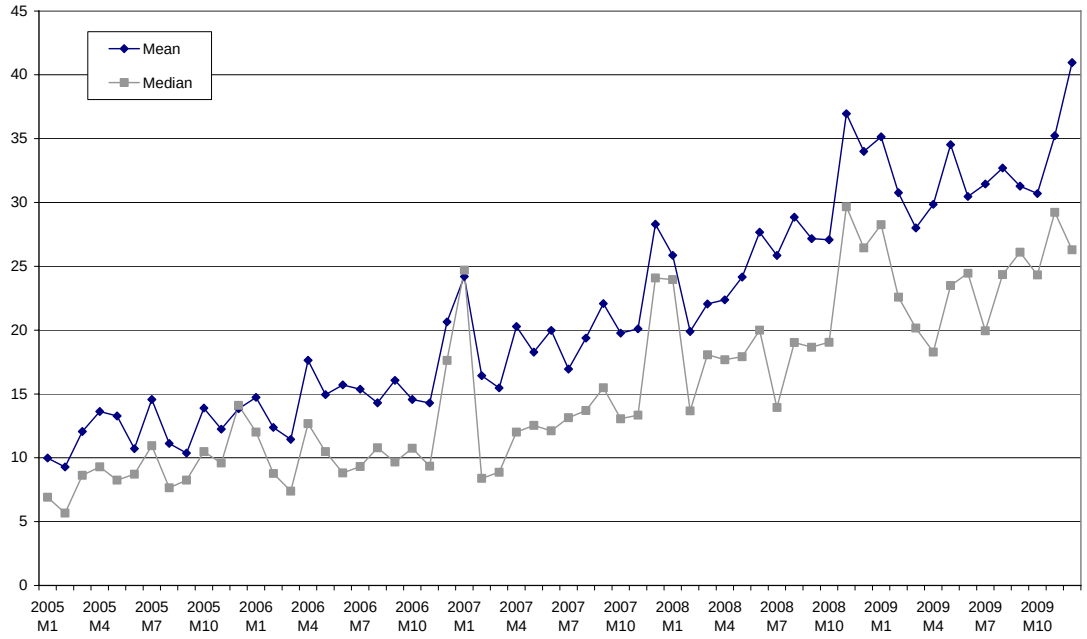
Per cent



Source: Authors' calculations

Figure 3 Absolute percentage errors – MIDSS (turnover)

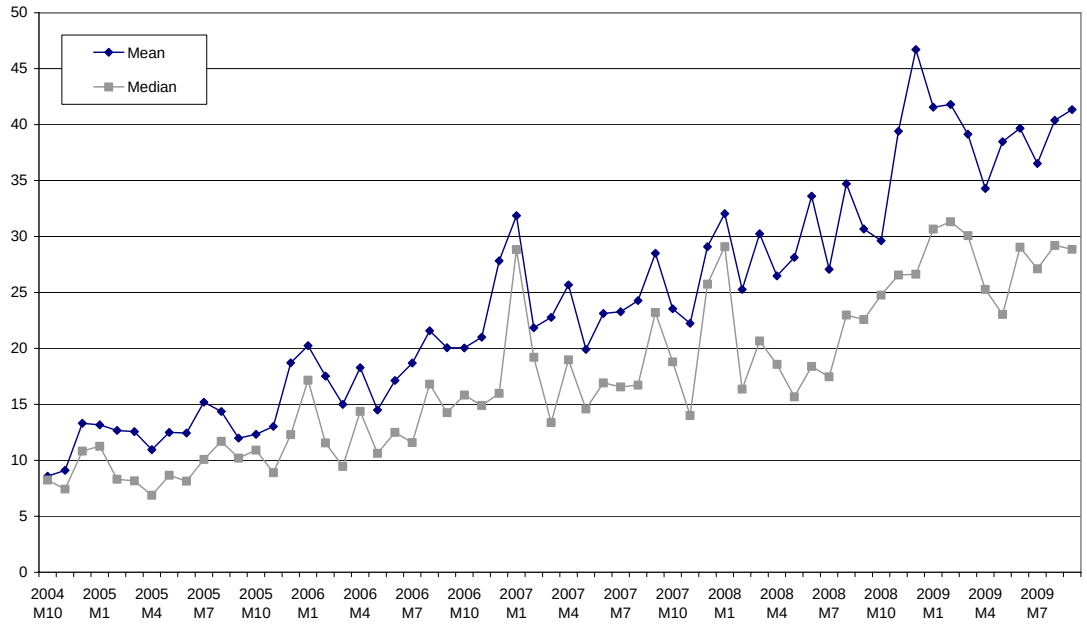
Per cent



Source: Authors' calculations

Figure 4 Absolute percentage errors – MIP (turnover)

Per cent

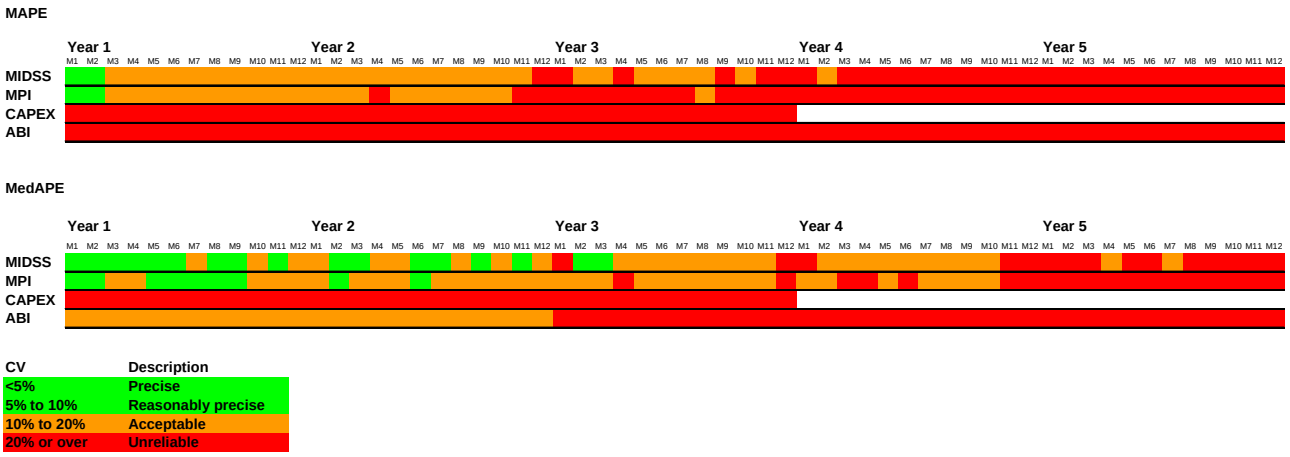


Source: Authors' calculations

In the **Annex**, **Figures A1 to A4** show the unconditional distribution of the MAPE. In each survey, these exhibit very strong right hand side tails which tends to support the argument that divergence between MAPE and MedAPE is driven by a small number of particularly bad forecasts towards the end of the out–of–sample period.

Figure 5 provides a summary of Figures 1 to 4 where a forecast is considered acceptable providing the absolute percentage error is less than 20 per cent. Based on an assessment using MAPE, MIDSS and MPI forecasts are acceptable for around two years, whereas MAPEs for CAPEX and ABI are at no point are below this 20 per cent threshold. MedAPEs, for the argument already given, are generally lower than the respective MAPEs and hence the forecasts are generally below the 20 per cent threshold for longer. The MedAPEs for both MIDSS and MPI are generally below this threshold until nearly the end of the fourth year of forecasts. Out–of–sample ABI forecasts are now acceptable for two years, however CAPEX forecasts remain unreliable throughout the evaluation period.

Figure 5 Summary of MAPE and MedAPE tests



Source: Authors' calculations

Confidence intervals

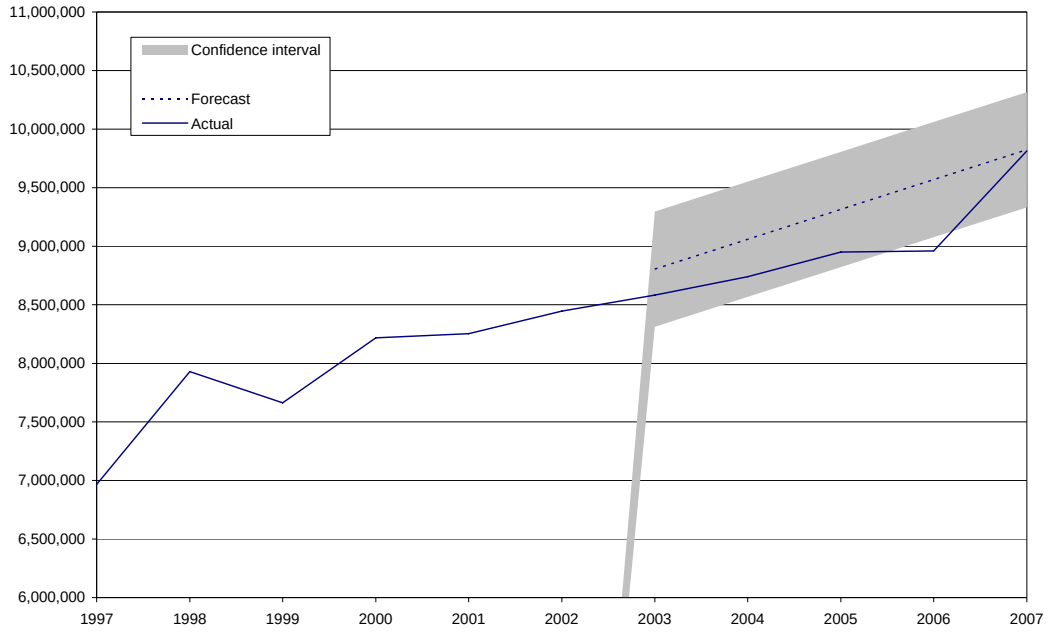
The second approach to assessing forecast accuracy is to calculate at each point in time a 95 per cent prediction interval. The confidence interval can be calculated based on in–sample variances (fixed variance approach) or also to include a rolling variance calculation over the forecast horizon which is updated each time a new forecast is evaluated against the outturn (recursive variance approach).

Figures 6 to **9** display the results. In each case, part (a) of the respective figure shows confidence intervals based on fixed variances and part (b) based on recursive variances.

Figure 6 Point and density* forecasts – ABI (turnover)

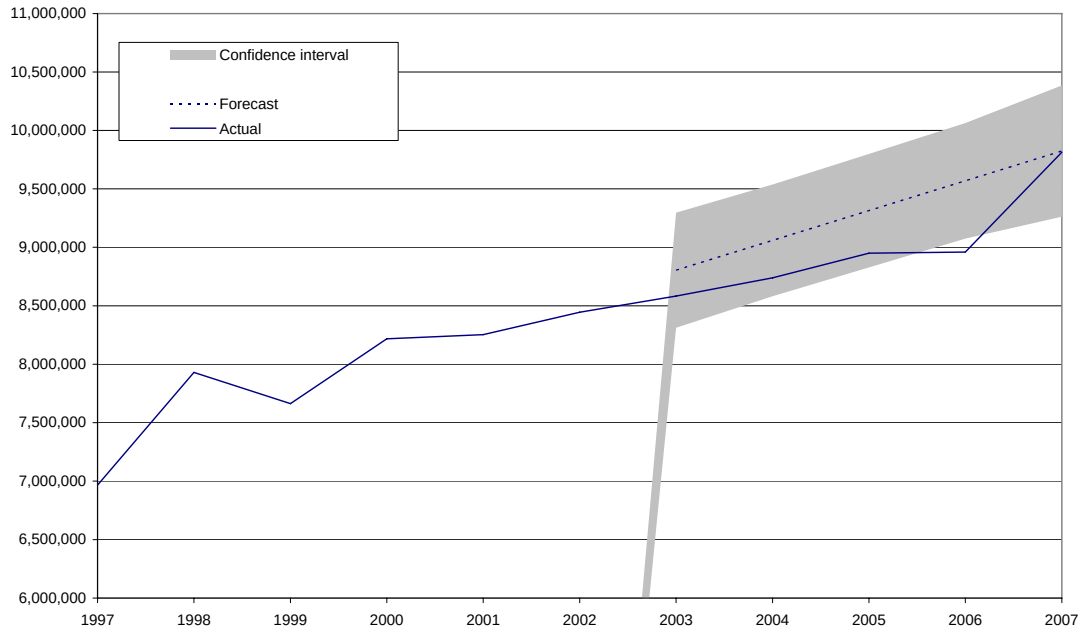
(a) Fixed variance

Value (aggregate over 50 firms)



(b) Recursive variance

Value (aggregate over 50 firms)



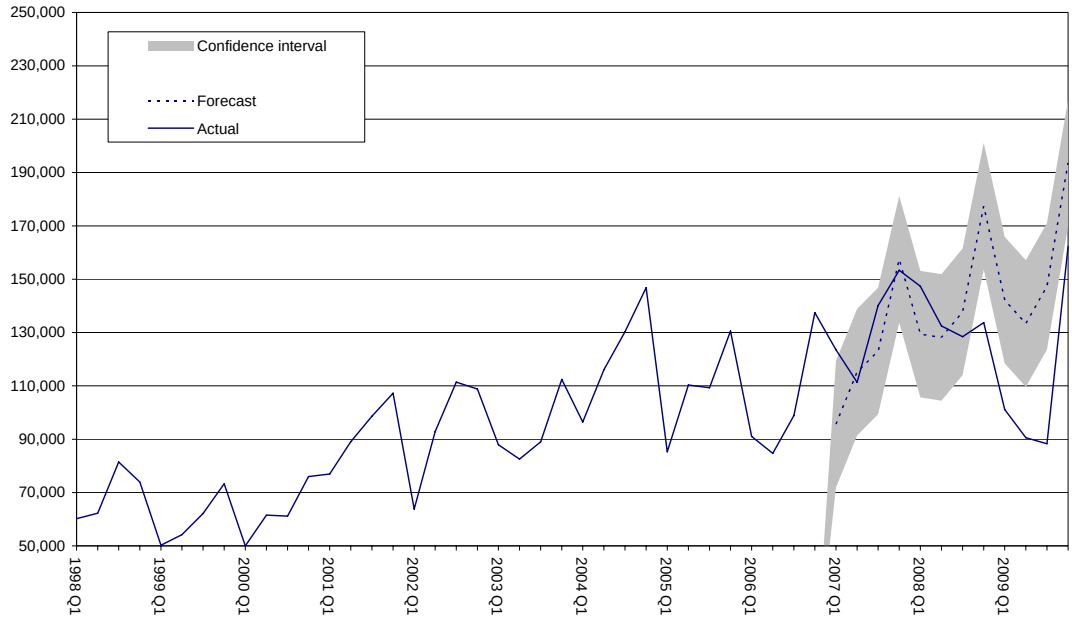
* 95% confidence interval

Source: Authors' calculations

Figure 7 Point and density* forecasts – CAPEX (investment)

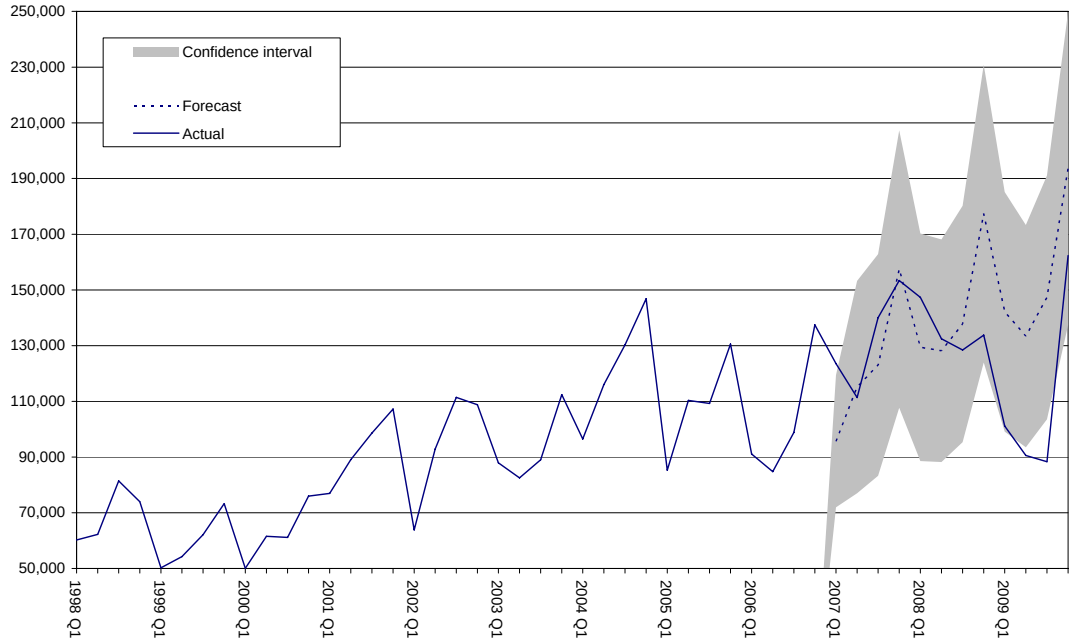
(a) Fixed variance

Value (aggregate over 50 firms)



(b) Recursive variance

Value (aggregate over 50 firms)



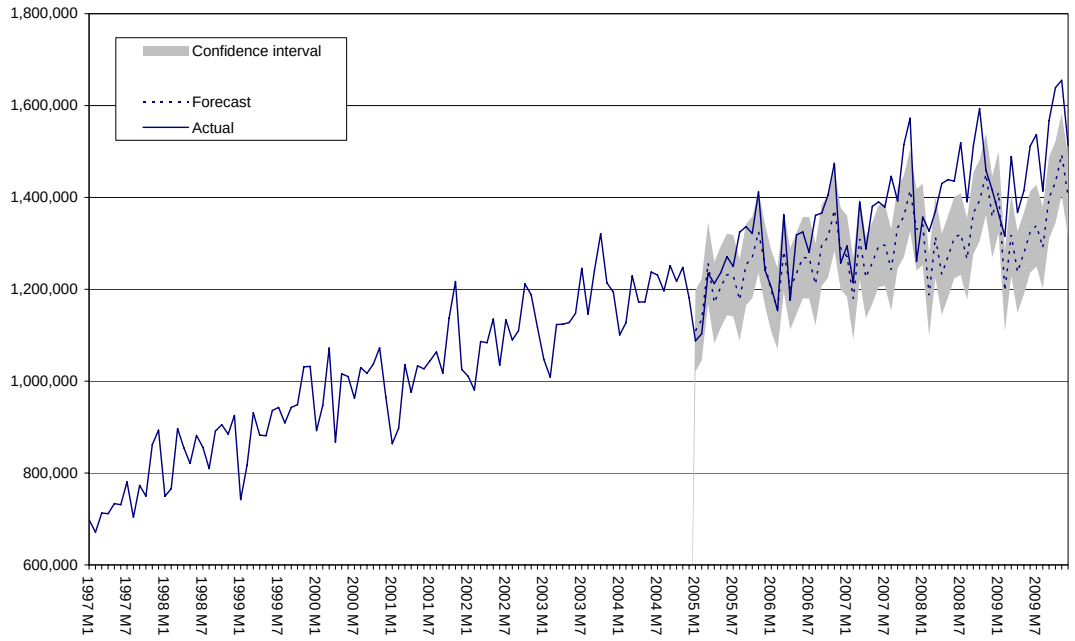
* 95% confidence interval

Source: Authors' calculations

Figure 8 Point and density* forecasts – MIDSS (turnover)

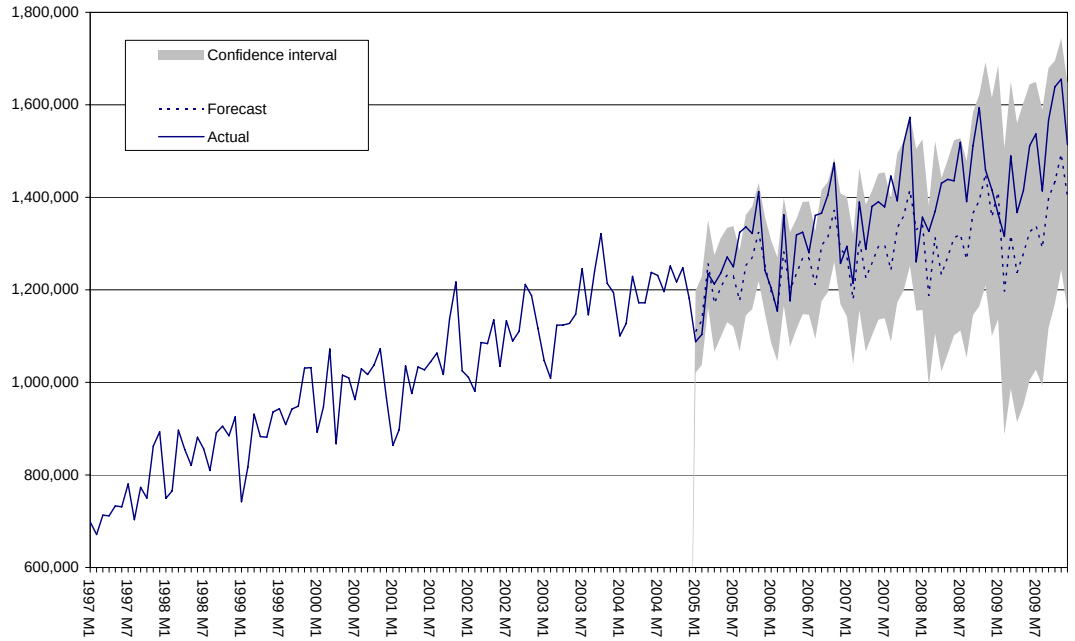
(a) Fixed variance

Value (aggregate over 50 firms)



(b) Recursive variance

Value (aggregate over 50 firms)



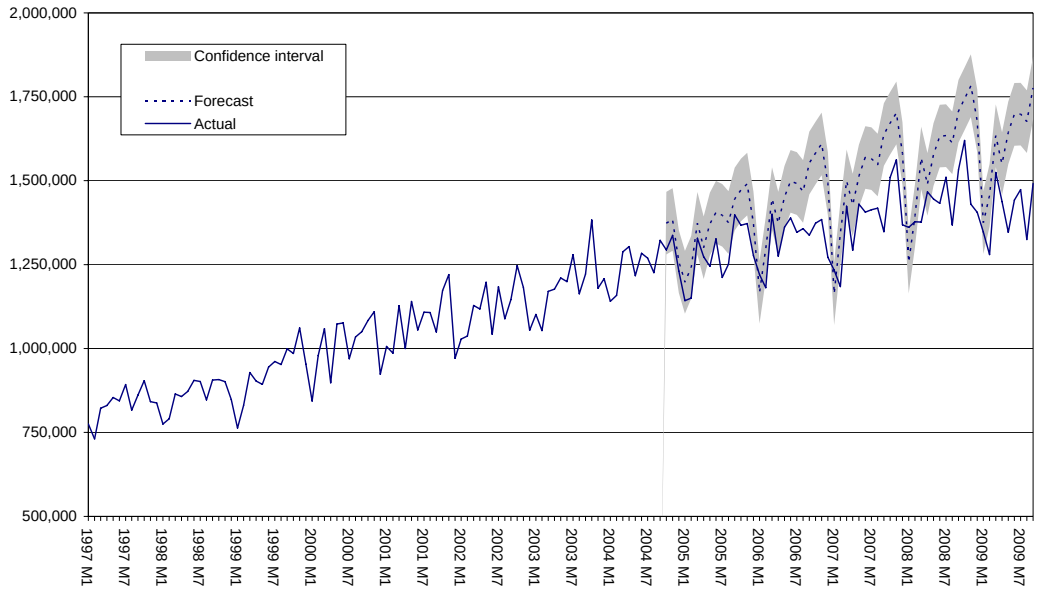
* 95% confidence interval

Source: Authors' calculations

Figure 9 Point and density* forecasts – MPI (turnover)

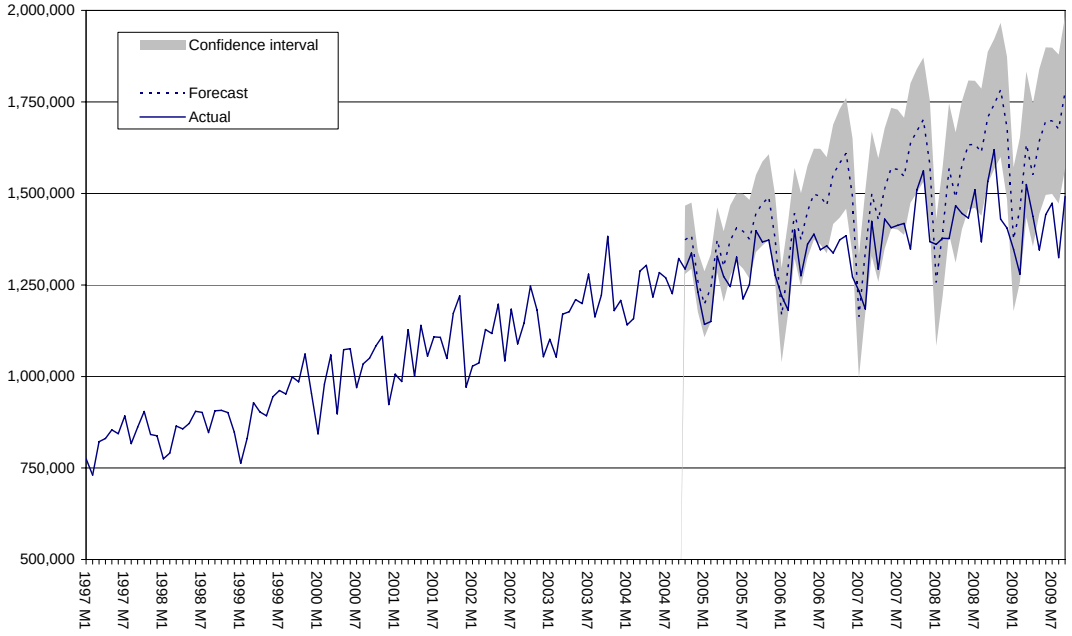
(a) Fixed variance

Value (aggregate over 50 firms)



(b) Recursive variance

Value (aggregate over 50 firms)

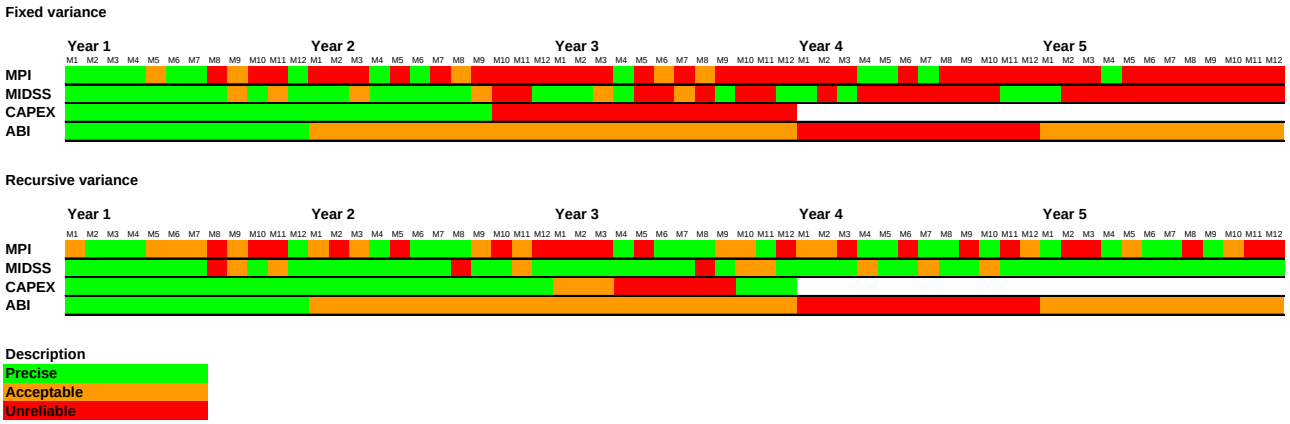


* 95% confidence interval

Source: Authors' calculations

Figure 10 provides a summary of the main results shown in Figures 6 to 9. Different colours have been used to highlight the degree of accuracy of the forecast with respect to the prediction interval – green if completely within the intervals, amber if the difference between the closest interval boundary and the standardized actual value is less than two percent and red when it is outside of the confidence interval.

Figure 10 Summary of forecast findings from Figures 6 to 9



Source: Authors' calculations

Generally, forecasts perform better on the recursive variance as the 95 confidence interval tends to widen over the forecast horizon, making it easier for the outturn to fall within the prediction interval. On the fixed variance basis, outturns for MPI, MIDSS and CAPEX start to routinely fall outside the 95 per cent confidence interval after two years and three years for ABI. The improved performance of CAPEX and ABI forecasts, relative to the absolute percentage error tests, is primarily down to the size of the confidence intervals reflecting the inherent volatility in the underlying data. For instance, quarterly CAPEX data tends to be relatively volatile, so the 95 per cent confidence interval will be calculated to be relatively wider.

Conclusions

The main objective of this work is to explore the predictive power of individual firms' turnover and capital expenditure series based on four key business surveys administered by the Office for National Statistics. These are the monthly MPI and MIDSS, the quarterly CAPEX and the annual ABI. The results suggest that the predictive power of the data starts to drop substantially after the end of the second year and the beginning of the third. However, the rate of decay is not continuous over the forecasting horizon and this seems to be particularly true for annual data (the decay is slower when referring to the rolling forecast). As a result, it is concluded that forecast precision is not independent of the frequency of the data for which such forecasts have been produced. In general, the lower the frequency the slower the deterioration in the forecasts, and vice versa.

CAPEX forecasts based on quarterly data tend to perform fairly badly and forecast precision declines much faster than the monthly MPI and MIDSS turnover series. This is probably because investment data is much more volatile or lumpy compared to the generally much more stable turnover data taken from the monthly surveys.

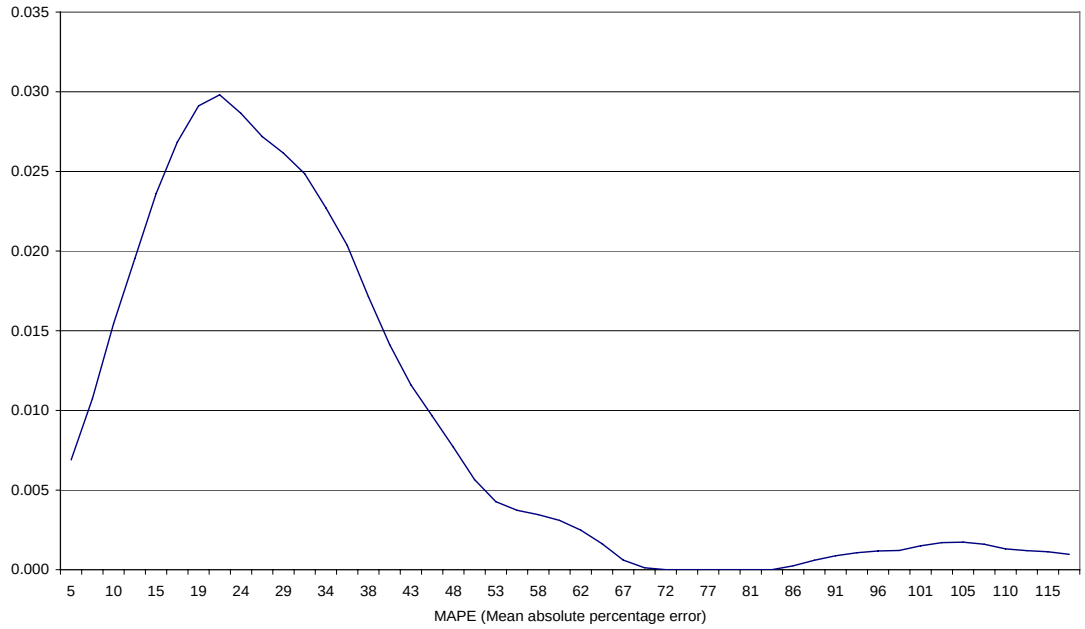
Contact

elmr@ons.gov.uk

Annex – Unconditional MAPE distributions of MPI, MIDSS, CAPEX and ABI

Figure A1 Distribution of absolute percentage errors – MPI

Relative frequency

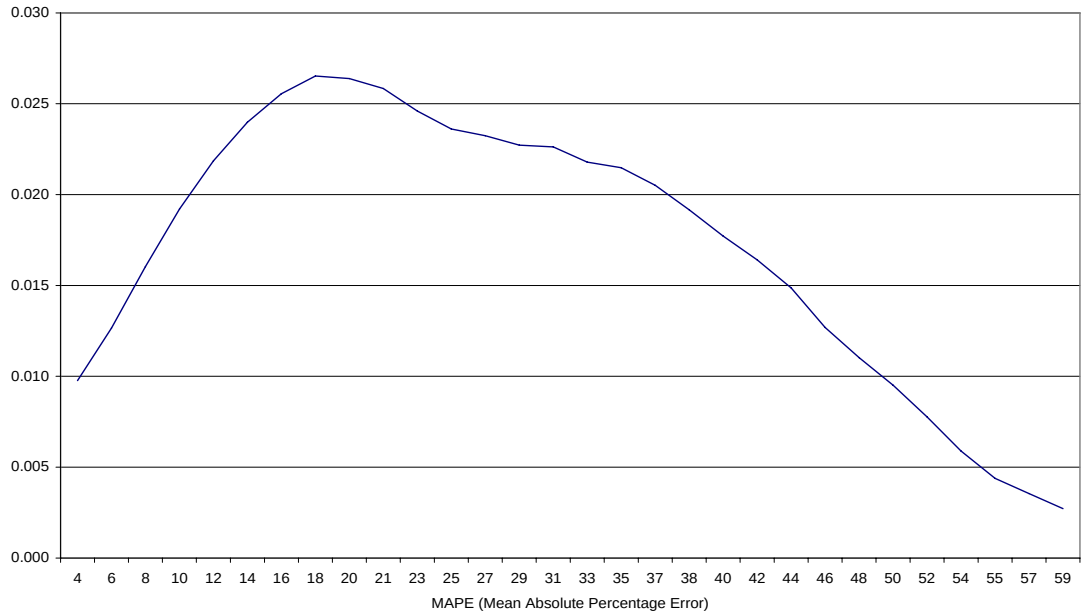


Turnover; Range: 3% – 120%; Mean – 30%; Mode – 22%

Source: Authors' calculations

Figure A2 Distribution of absolute percentage errors – MIDSS

Relative frequency

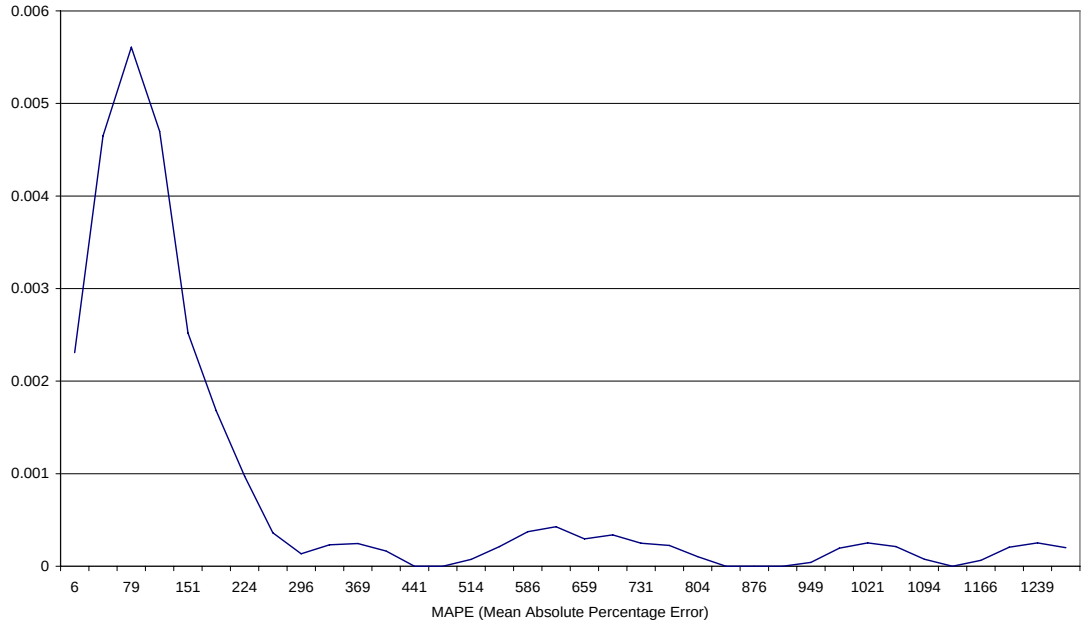


Turnover; Range: 3% – 60%; Mean – 27%; Mode – 18%

Source: Authors' calculations

Figure A3 Distribution of absolute percentage errors – CAPEX

Relative frequency

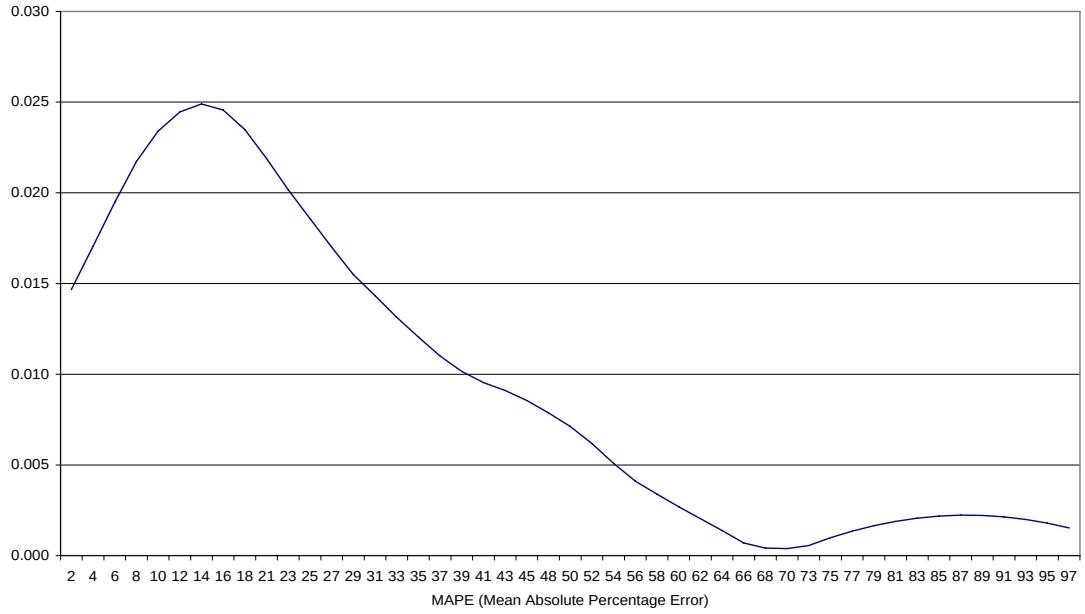


Investment; Range: 40% – 1200%; Mean – 207%; Mode – 80%

Source: Authors' calculations

Figure A4 Distribution of absolute percentage errors – ABI

Relative frequency



Turnover; Range: 2% – 100%; Mean – 20%; Mode – 14%

Source: Authors' calculations

Key time series

1. National Accounts aggregates

Last updated 29/03/11

Seasonally adjusted

	£ million		Indices (2006 = 100)						
	At current prices		Value indices at current prices		Chained volume indices			Implied deflators[3]	
	Gross domestic product (GDP) at market prices	Gross value added (GVA) at basic prices	GDP at market prices[1]	GVA at basic prices	Gross national disposable income at market prices[2]	GDP at market prices	GVA at basic prices	GDP at market prices	GVA at basic prices
	YBHA	ABML	YBEU	YBEX	YBFP	YBEZ	CGCE	YBGB	CGBV
2009	1,394,989	1,257,627	105.0	106.2	98.4	97.6	97.9	107.6	108.5
2010	1,453,616	1,296,402	109.4	109.5	100.0	98.8	99.3	110.7	110.3
2008 Q3	361,466	325,041	108.8	109.8	103.8	102.6	102.6	106.1	107.1
2008 Q4	358,848	324,009	108.1	109.5	100.9	100.5	100.5	107.5	108.9
2009 Q1	349,801	317,113	105.3	107.2	99.5	98.2	98.4	107.2	108.9
2009 Q2	344,504	311,156	103.7	105.1	96.6	97.4	97.7	106.5	107.6
2009 Q3	348,081	313,018	104.8	105.8	98.3	97.2	97.5	107.9	108.5
2009 Q4	352,603	316,340	106.2	106.9	99.4	97.6	98.0	108.8	109.0
2010 Q1	359,147	319,774	108.1	108.1	98.6	97.8	98.3	110.6	109.9
2010 Q2	361,485	322,395	108.9	108.9	100.2	98.9	99.3	110.1	109.7
2010 Q3	365,496	326,117	110.1	110.2	100.8	99.6	100.0	110.5	110.2
2010 Q4	367,488	328,116	110.7	110.9	100.4	99.1	99.5	111.7	111.4
			Percentage change, quarter on corresponding quarter of previous year						
			IHYO	ABML[4]	YBGO[4]	IHYR	ABMM[4]	IHYU	ABML/ABMM[4]
2008 Q3	2.4	3.4	2.4	3.4	0.3	-0.4	-0.6	2.9	4.0
2008 Q4	0.5	1.6	0.5	1.6	-5.2	-2.7	-2.8	3.3	4.6
2009 Q1	-3.4	-1.8	-3.4	-1.8	-7.2	-5.4	-5.4	2.1	3.8
2009 Q2	-5.2	-3.9	-5.2	-3.9	-8.2	-5.9	-5.8	0.8	2.0
2009 Q3	-3.7	-3.7	-3.7	-3.7	-5.3	-5.3	-5.0	1.7	1.4
2009 Q4	-1.7	-2.4	-1.7	-2.4	-1.6	-2.8	-2.5	1.1	0.1
2010 Q1	2.7	0.8	2.7	0.8	-0.8	-0.4	-0.1	3.1	0.9
2010 Q2	4.9	3.6	4.9	3.6	3.7	1.5	1.6	3.4	2.0
2010 Q3	5.0	4.2	5.0	4.2	2.5	2.5	2.6	2.5	1.6
2010 Q4	4.2	3.7	4.2	3.7	1.0	1.5	1.5	2.7	2.2

Notes

1. 'Money GDP'
2. This series is only updated once a quarter, in line with the full quarterly national accounts data set
3. Based on chained volume measures and current price estimates of expenditure components of GDP
4. Derived from these identification (CDID) codes.

2. Gross Domestic Product: by category of expenditure

Last updated 29/03/11

£ million, chained volume measures, reference year 2006, seasonally adjusted

	Domestic expenditure on goods and services at market prices											
	Final consumption expenditure			Gross capital formation								
	Households	Non-profit institutions1	General government	Gross fixed capital formation	Changes in inventories2	Acquisitions less disposals of valuables	Total	Exports of goods and services	Gross final expenditure	less Imports of goods and services	Statistical discrepancy (expenditure)	Gross domestic product at market prices
	ABJR	HAYO	NMRY	NPQT	CAFU	NPJR	YBIM	IKBK	ABMG	IKBL	GIXS	ABMI
2008	842,174	32,338	293,464	232,777	130	1,290	1,402,173	372,104	1,774,277	411,138	0	1,363,139
2009	814,666	32,281	296,306	196,997	-16,012	1,222	1,325,460	334,601	1,660,061	362,026	-1,346	1,296,689
2010	821,271	30,356	298,571	202,980	2,550	938	1,356,666	352,246	1,708,912	392,914	-3,087	1,312,911
2008 Q1	213,214	8,292	72,104	59,619	3,228	206	356,664	93,858	450,522	105,712	0	344,809
2008 Q2	211,525	8,183	73,334	59,779	872	440	354,134	94,284	448,418	104,550	0	343,868
2008 Q3	210,330	8,018	73,473	57,254	645	367	350,088	93,918	444,005	103,226	0	340,780
2008 Q4	207,105	7,845	74,553	56,125	-4,615	277	341,287	90,044	431,332	97,650	0	333,682
2009 Q1	204,262	8,153	73,972	51,112	-4,514	420	333,404	83,645	417,050	90,636	-156	326,257
2009 Q2	202,792	8,078	74,089	48,858	-3,796	239	330,260	82,166	412,426	88,581	-260	323,585
2009 Q3	202,828	8,026	73,958	48,878	-4,191	212	329,711	82,879	412,590	89,547	-388	322,655
2009 Q4	204,784	8,024	74,287	48,149	-3,511	351	332,085	85,911	417,995	93,262	-542	324,192
2010 Q1	204,649	7,744	74,664	50,050	-1,901	265	335,471	85,043	420,514	94,998	-670	324,846
2010 Q2	205,753	7,716	74,781	50,106	-231	368	338,493	87,592	426,085	97,026	-758	328,301
2010 Q3	205,713	7,504	74,425	51,892	1,631	131	341,298	89,034	430,331	98,857	-818	330,656
2010 Q4	205,156	7,392	74,701	50,932	3,051	174	341,404	90,577	431,982	102,033	-841	329,108
	Percentage change, quarter on corresponding quarter of previous year											
2008 Q1	2.9	0.1	0.8	-1.9			1.8	3.7	2.2	3.1		1.9
2008 Q2	1.4	-1.5	1.9	-1.4			1.2	2.7	1.5	3.1		1.0
2008 Q3	0.1	-4.1	1.2	-6.0			-1.3	0.5	-0.9	-2.5		-0.4
2008 Q4	-2.1	-6.9	2.5	-10.5			-4.4	-2.7	-4.1	-8.4		-2.7
2009 Q1	-4.2	-1.7	2.6	-14.3			-6.5	-10.9	-7.4	-14.3		-5.4
2009 Q2	-4.1	-1.3	1.0	-18.3			-6.7	-12.9	-8.0	-15.3		-5.9
2009 Q3	-3.6	0.1	0.7	-14.6			-5.8	-11.8	-7.1	-13.3		-5.3
2009 Q4	-1.1	2.3	-0.4	-14.2			-2.7	-4.6	-3.1	-4.5		-2.8
2010 Q1	0.2	-5.0	0.9	-2.1			0.6	1.7	0.8	4.8		-0.4
2010 Q2	1.5	-4.5	0.9	2.6			2.5	6.6	3.3	9.5		1.5
2010 Q3	1.4	-6.5	0.6	6.2			3.5	7.4	4.3	10.4		2.5
2010 Q4	0.2	-7.9	0.6	5.8			2.8	5.4	3.3	9.4		1.5

Notes

1. Non–profit institutions serving households
2. This series includes a quarterly alignment adjustment

3. Labour Market summary

Last updated 16/03/11

United Kingdom (thousands) seasonally adjusted

			Headline indicators					
	LFS household population[1]		Employment		Unemployment		Inactivity	
			Level	Rate[2]	Level	Rate[3]	Level	Rate[4]
	All aged 16 & over	All aged 16 to 64	All aged 16 & over	All aged 16 to 64	All aged 16 & over	All aged 16 to 64	All aged 16 & over	All aged 16 to 64
People	MGSL	LF2O	MGRZ	LF24	MGSC	MGSX	LF2M	LF2S
Nov–Jan 2009	49,260	39,691	29,349	72.2	2,065	6.6	8,994	22.7
Nov–Jan 2010	49,646	39,905	28,861	70.4	2,443	7.8	9,396	23.5
Feb–Apr 2010	49,744	39,955	28,872	70.3	2,475	7.9	9,422	23.6
May–Jul 2010	49,842	40,006	29,158	70.7	2,467	7.8	9,264	23.2
Aug–Oct 2010	49,935	40,050	29,125	70.6	2,502	7.9	9,286	23.2
Nov–Jan 2011	50,027	40,094	29,157	70.5	2,529	8.0	9,328	23.3
Change on quarter	93	44	32	-0.1	27	0.1	43	0.1
Change on quarter %	0.2	0.1	0.1		1.1		0.5	
Change on year	381	189	296	0.1	87	0.2	-67	-0.3
Change on year %	0.8	0.5	1.0		3.6		-0.7	
Men	MGSM	YBTG	MGSA	MGSV	MGSD	MGSY	YBSO	YBTM
Nov–Jan 2009	23,991	19,746	15,803	77.8	1,241	7.3	3,156	16.0
Nov–Jan 2010	24,202	19,858	15,355	75.0	1,505	8.9	3,476	17.5
Feb–Apr 2010	24,257	19,885	15,389	75.0	1,511	8.9	3,468	17.4
May–Jul 2010	24,311	19,912	15,610	75.8	1,452	8.5	3,375	17.0
Aug–Oct 2010	24,362	19,935	15,620	75.8	1,464	8.6	3,372	16.9
Nov–Jan 2011	24,413	19,959	15,632	75.7	1,488	8.7	3,384	17.0
Change on quarter	51	24	12	-0.2	24	0.1	12	0.0
Change on quarter %	0.2	0.1	0.1		1.6		0.3	
Change on year	211	101	277	0.7	-17	-0.2	-92	-0.6
Change on year %	0.9	0.5	1.8		-1.1		-2.7	
Women	MGSN	LF2P	MGSB	LF25	MGSE	MGSZ	LF2N	LF2T
Nov–Jan 2009	25,268	19,945	13,546	66.6	824	5.7	5,838	29.3
Nov–Jan 2010	25,444	20,047	13,506	65.8	938	6.5	5,920	29.5
Feb–Apr 2010	25,487	20,071	13,483	65.6	964	6.7	5,954	29.7
May–Jul 2010	25,531	20,095	13,548	65.7	1,015	7.0	5,888	29.3
Aug–Oct 2010	25,573	20,115	13,505	65.5	1,039	7.1	5,914	29.4
Nov–Jan 2011	25,614	20,135	13,525	65.3	1,042	7.2	5,945	29.5
Change on quarter	41	20	19	-0.1	3	0.0	31	0.1
Change on quarter %	0.2	0.1	0.1		0.3		0.5	
Change on year	170	88	19	-0.5	104	0.7	25	0.0
Change on year %	0.7	0.4	0.1		11.1		0.4	

Notes

1. The Labour Force Survey (LFS) is a survey of the population of private households, student halls of residence and NHS accommodation.

2. The headline employment rate is the number of people aged 16 to 64 in employment divided by the population aged 16 to 64.

3. The headline unemployment rate is the number of unemployed people (aged 16+) divided by the economically active population (aged 16+). The economically active population is defined as those in employment plus those who are unemployed.

4. The headline inactivity rate is the number of people aged 16 to 64 divided by the population aged 16 to 64.

Note on headline employment, unemployment and inactivity rates

The headline employment and inactivity rates are based on the population aged 16 to 64 but the headline unemployment rate is based on the economically active population aged 16 and over. The employment and inactivity rates for those aged 16 and over are affected by the inclusion of the retired population in the denominators and are therefore less meaningful than the rates for those aged from 16 to 64. However, for the unemployment rate for those aged 16 and over, no such effect occurs as the denominator for the unemployment rate is the economically active population which only includes people in work or actively seeking and able to work.

Note on headline employment, unemployment and inactivity levels

The headline employment and unemployment levels are for those aged 16 and over; they measure all people in work or actively seeking and able to work. However, the headline inactivity level is for those aged 16 to 64. The inactivity level for those aged 16 and over is less meaningful as it includes elderly people who have retired from the labour force.

4. Prices

Last updated 22/03/11

Percentage change over 12 months, Not seasonally adjusted

	Consumer prices						Producer prices			
	Consumer prices index (CPI)			Retail prices index (RPI)			Output prices		Input prices	
	All items	CPI excluding indirect taxes (CPIY)[1]	CPI at constant tax rates (CPI-CT)	All items	All items excluding mortgage interest payments (RPIX)	All items excluding mortgage interest payments and indirect taxes (RPIY)[2]	All manufactured products	Excluding food, beverages, tobacco and petroleum products	Materials and fuels purchased by manufacturing industry	Excluding food, beverages, tobacco and petroleum products
	D7G7	EL2S	EAD6	CZBH	CDKQ	CBZX				
2009 Jan	3.0	4.5	4.1	0.1	2.4	3.4	3.5	4.1	1.9	11.4
2009 Feb	3.2	4.6	4.2	0.0	2.5	3.5	3.1	4.0	0.5	8.9
2009 Mar	2.9	4.3	3.9	-0.4	2.2	3.2	2.2	3.6	-0.7	7.3
2009 Apr	2.3	3.8	3.4	-1.2	1.7	2.7	1.8	3.5	-6.0	2.5
2009 May	2.2	3.6	3.3	-1.1	1.6	2.6	0.7	3.0	-9.2	-0.2
2009 Jun	1.8	3.1	2.9	-1.6	1.0	1.9	0.2	2.3	-12.6	-3.5
2009 Jul	1.8	3.1	2.8	-1.4	1.2	2.1	-0.3	1.8	-12.5	-3.8
2009 Aug	1.6	2.9	2.7	-1.3	1.4	2.3	0.0	1.5	-7.9	-2.1
2009 Sep	1.1	2.2	2.1	-1.4	1.3	2.0	0.4	1.3	-6.1	-1.0
2009 Oct	1.5	2.6	2.5	-0.8	1.9	2.8	1.5	1.7	0.8	1.1
2009 Nov	1.9	3.0	2.9	0.3	2.7	3.5	2.6	1.5	3.9	0.6
2009 Dec	2.9	2.8	2.6	2.4	3.8	3.8	3.2	2.1	7.1	0.8
2010 Jan	3.5	1.9	1.7	3.7	4.6	3.3	3.5	2.0	7.4	0.8
2010 Feb	3.0	1.4	1.2	3.7	4.2	2.9	3.8	2.2	7.7	2.2
2010 Mar	3.4	1.8	1.6	4.4	4.8	3.5	4.5	2.7	10.3	4.0
2010 Apr	3.7	2.0	1.9	5.3	5.4	3.9	4.9	2.8	12.6	6.1
2010 May	3.4	1.7	1.6	5.1	5.1	3.8	5.0	3.2	11.6	7.1
2010 Jun	3.2	1.6	1.5	5.0	5.0	3.8	4.4	3.8	10.7	7.1
2010 Jul	3.1	1.4	1.3	4.8	4.8	3.5	4.2	3.8	10.7	7.5
2010 Aug	3.1	1.4	1.3	4.7	4.7	3.4	4.2	3.7	8.3	6.3
2010 Sep	3.1	1.5	1.4	4.6	4.6	3.4	3.8	3.4	8.8	5.6
2010 Oct	3.2	1.6	1.4	4.5	4.6	3.2	4.0	3.2	8.5	6.0
2010 Nov	3.3	1.6	1.5	4.7	4.7	3.4	4.1	3.3	9.2	7.0
2010 Dec	3.7	2.0	1.9	4.8	4.7	3.5	4.2	2.7	13.0	9.0
2011 Jan	4.0	2.4	2.3	5.1	5.1	3.8	5.0	3.2	14.1	10.4
2011 Feb	4.4	2.8	2.7	5.5	5.5	4.1	5.3	3.1	14.6	10.6

Notes

1 The taxes excluded are VAT, duties, insurance premium tax, air passenger duty and stamp duty on share transactions.

2 The taxes excluded are council tax, VAT, duties, vehicle excise duty, insurance premium tax and air passenger duty.

Notes to tables

Identification (CDID) codes

The four-letter identification code at the top of each data column is the ONS reference for this series of data on our time series database. Please quote the relevant code if you contact us requiring any further information about the data.

Conventions

Where figures have been rounded to the final digit, there may be an apparent slight discrepancy between the sum of the constituent items and the total as shown. Although figures may be given in unrounded form to facilitate the calculation of percentage changes, rates of change etc by users, this does not imply that the figures can be estimated to this degree of precision as they may be affected by sampling variability or imprecision in estimation methods.

The following standard symbols are used:

..	not available
–	nil or negligible (less than half the final digit shown)
P	provisional
—	break in series
R	revised
r	series revised from indicated entry onwards

Labour market statistics concepts and definitions

Labour Force Survey 'monthly' estimates

Labour Force Survey (LFS) results are three-monthly averages, so consecutive months' results overlap. Comparing estimates for overlapping three-month periods can produce more volatile results, which can be difficult to interpret.

Labour force summary table

Economically active

People aged 16 and over who are either in employment or unemployed.

Economically inactive

People who are neither in employment nor unemployed. This includes those who want a job but have not been seeking work in the last four weeks, those who want a job and are seeking work but not available to start work, and those who do not want a job.

Employment and jobs

There are two ways of looking at employment: the number of people with jobs, or the number of jobs. The two concepts are not the same as one person can have more than one job. The number of people with jobs is measured by the Labour Force Survey (LFS) and includes people aged 16 or over who do paid work (as an employee or self-employed), those who have a job that they are temporarily away from, those on government-supported training and employment programmes, and those doing unpaid family work. The number of jobs is measured by workforce jobs and is the sum of employee jobs (as measured by surveys of employers), self-employment jobs from the LFS, people in HM Forces, and government-supported trainees. Vacant jobs are not included.

Unemployment

The number of unemployed people in the UK is measured through the Labour Force Survey following the internationally agreed definition recommended by the ILO (International Labour Organisation) – an agency of the United Nations. Unemployed people:

are without a job, want a job, have actively sought work in the last four weeks and are available to start work in the next two weeks, or

are out of work, have found a job and are waiting to start it in the next two weeks

Other key indicators

Claimant count

The number of people claiming Jobseeker's Allowance benefits.

Earnings

A measure of the money people receive in return for work done, gross of tax. It includes salaries and, unless otherwise stated, bonuses but not unearned income, benefits in kind or arrears of pay.

Productivity

Whole economy output per worker is the ratio of Gross Value Added (GVA) at basic prices and Labour Force Survey (LFS) total employment. Manufacturing output per filled job is the ratio of manufacturing output (from the Index of Production) and productivity jobs for manufacturing (constrained to LFS jobs at the whole economy level).

Redundancies

The number of people who:

were not in employment during the reference week, and

reported that they had been made redundant in the month of, or the two calendar months prior to, the reference week

plus the number of people who:

were in employment during the reference week, and

started their job in the same calendar month as, or the two calendar months prior to, the reference week, and

reported that they had been made redundant in the month of, or the two calendar months prior to, the reference week

Unit wage costs

A measure of the cost of wages and salaries per unit of output.

Vacancies

The statistics are based on ONS's Vacancy Survey of businesses. The survey is designed to provide comprehensive estimates of the stock of vacancies across the economy, excluding those in agriculture, forestry and fishing. Vacancies are defined as positions for which employers are actively seeking recruits from outside their business or organisation. More information on labour market concepts, sources and methods is available in the *Guide to Labour Market Statistics* at www.statistics.gov.uk/about/data/guides/LabourMarket/default.asp

Directory of online tables

Weblink: www.statistics.gov.uk/StatBase/Product.asp?vlnk=14692

Title	Frequency of update
1. UK economic accounts	
Weblink: www.statistics.gov.uk/elmr/downloads/elmr1.pdf	
1.01 National accounts aggregates	M
1.02 Gross domestic product and gross national income	M
1.03 Gross domestic product, by category of expenditure	M
1.04 Gross domestic product, by category of income	M
1.05 Gross domestic product and shares of income and expenditure	M
1.06 Income, product and spending per head	Q
1.07 Households' disposable income and consumption	M
1.08 Household final consumption expenditure	M
1.09 Gross fixed capital formation	M
1.10 Gross value added, by category of output	M
1.11 Gross value added, by category of output: service industries	M
1.12 Summary capital accounts and net lending/net borrowing	Q
1.13 Private non-financial corporations: allocation of primary income account	Q
1.14 Private non-financial corporations: secondary distribution of income account and capital account	Q
1.15 Balance of payments: current account	M
1.16 Trade in goods (on a balance of payments basis)	M
1.17 Index of Services	M
2. Selected labour market statistics	
Weblink: www.statistics.gov.uk/elmr/downloads/elmr2.pdf	
2.01 Summary of Labour Force Survey data	M
2.02 Employment by age	M
2.03 Full-time, part-time and temporary workers	M
2.04 Public and private sector employment	Q
2.05 Workforce jobs	Q
2.06 Workforce jobs by industry	Q
2.07 Actual weekly hours of work	M
2.08 Usual weekly hours of work	M

2.09 Unemployment by age and duration	M
2.10 Claimant count levels and rates	M
2.11 Claimant count by age and duration	M
2.12 Economic activity by age	M
2.13 Economic inactivity by age	M
2.14 Economic inactivity: reasons	M
2.15 Educational status, economic activity and inactivity of young people	M
2.16 Average weekly earnings - total pay	M
2.16A Average weekly earnings - bonus pay	M
2.17 Average weekly earnings - regular pay	M
2.18 Productivity and unit wage costs	M
2.19 Regional labour market summary	M
2.20 International comparisons	M
2.21 Labour disputes	M
2.22 Vacancies by size of enterprise	M
2.23 Vacancies by industry	M
2.24 Redundancies: levels and rates	M
2.25 Redundancies: by industry	Q
2.27 Employment levels by country of birth and nationality	M
2.28 Working age employment rates by country of birth and nationality	Q
2.29 Lone parent claimants of Jobseekers Allowance by age of youngest child	M
2.30 Key out of work benefits	M
2.31 Production industry employee jobs	M
2.32 Public sector employment by industry	Q

3. Prices

Weblink: www.statistics.gov.uk/elmr/downloads/elmr3.pdf

3.01 Producer and consumer prices	M
3.02 Harmonised Indices of Consumer Prices: EU comparisons	M

4. Selected output and demand indicators

Weblink: www.statistics.gov.uk/elmr/downloads/elmr4.pdf

4.01 Output of the production industries	M
4.02 Construction output	M
4.03 Construction new orders	M
4.04 Indicators of fixed investment in dwellings	M
4.05 Number of property transactions	M
4.06 Change in inventories	Q
4.07 Retail sales and credit business	M

5. Selected financial statistics

Weblink: www.statistics.gov.uk/elmr/downloads/elmr5.pdf

5.01 Sterling exchange rates and UK reserves	M
5.02 Monetary aggregates	M
5.03 Counterparts to changes in money stock M4	M
5.04 Public sector receipts and expenditure	Q
5.05 Public sector key fiscal indicators	M
5.06 Consumer credit and other household sector borrowing	M
5.07 Analysis of MFI lending to UK residents	M
5.08 Interest rates and yields	M
5.09 A selection of asset prices	M

6. Further labour market statistics

Weblink: www.statistics.gov.uk/elmr/downloads/elmr6.pdf

6.01 Working-age households	A
6.02 Local labour market indicators by unitary and local authority	Q
6.03 Employment by occupation	Q
6.04 Workforce jobs by industry	M
6.05 Employee jobs by industry	Q
6.06 Workforce jobs by region and industry	Q
6.07 Key productivity measures by industry	Q
6.08 Total workforce hours worked per week	Q
6.09 Total workforce hours worked per week by region and industry group	Q
6.10 Job-related training received by employees	Q
6.11 Unemployment rates by previous occupation (discontinued Q4 2007)	Q
6.12 Average Earnings Index by industry: excluding and including bonuses	M
6.13 Average Earnings Index: effect of bonus payments by industry	M
6.14 Median earnings and hours by main industrial sector	A
6.15 Median earnings and hours by industry section	A
6.16 Index of wages per head: international comparisons	M
6.17 Regional Jobseeker's Allowance claimant count rates	M
6.18 Claimant count area statistics: counties, unitary and local authorities	M
6.19 Claimant count area statistics: UK parliamentary constituencies	M
6.20 Claimant count area statistics: constituencies of the Scottish Parliament	M
6.21 Jobseeker's Allowance claimant count flows	M
6.22 Number of previous Jobseeker's Allowance claims	Q
6.23 Interval between Jobseeker's Allowance claims	Q
6.24 Average duration of Jobseeker's Allowance claims by age	Q

6.25 Vacancies and unemployment	M
6.26 Redundancies: re-employment rates	Q
6.27 Redundancies by Government Office Region	Q
6.28 Redundancy rates by industry	Q
6.29 Labour disputes: summary	M
6.30 Labour disputes: stoppages in progress	M

Notes
A Annual
Q Quarterly
M Monthly

More information

- Time series are available from www.statistics.gov.uk/statbase/tsdintro.asp
- Subnational labour market data are available from www.statistics.gov.uk/statbase/Product.asp?vlnk=14160 and www.nomis.web
- Labour Force Survey tables are available from www.statistics.gov.uk/statbase/Product.asp?vlnk=11771
- Annual Survey of Hours and Earnings data are available from www.statistics.gov.uk/StatBase/Product.asp?vlnk=13101

Recent articles

October 2010

- The experimental tourism satellite account for the United Kingdom (E–UKTSA)
- A proposed methodology for nowcasting the demand and supply estimates of tourism activities
- Estimating regional exports of services trade for the UK
- Total public service output, inputs and productivity
- Quality adjusted labour input: new estimates for 1993 to 2008
- Volume of capital services: annual estimates for 1950 to 2008 and new quarterly series

November 2010

- Measuring the UK's human capital stock
- Regional economic indicators: A focus on sub–regional GVA using shift–share analysis
- Methods explained: Temporal disaggregation

December 2010

- Measuring the environmental goods and services sector
- Googling the present
- Producer prices and services producer prices: implementation of SIC 2007

January 2011

- Employment characteristics of UK tourism industries in 2008
- Developing financial statistics for policy – Progress report January 2011
- Exploring the geographical distribution of wealth using the output area classification
- Employment and intangible spending in the UK's creative industries – A view from the micro–data

February 2011

- The impact of the recession on different sized firms – A view from the micro–data
- Median and mean income analyses – Their implications for material living standards and national well–being
- Education: public service output, inputs and productivity
- Measuring the accuracy of the Retail Sales Index
- The development of the Monthly Business Survey

- Okun's Law revisited
- Regional economic indicators with a focus on the relationship between skills and productivity

March 2011

- Patterns of pay: results of the Annual Survey of Hours and Earnings 1997 to 2010
- Changes to the picture of the UK economy – impact of new SIC 2007 industry classification
- Standard Occupational Classification 2010– implementation in the Office for National Statistics
- Monitoring the coherence between ONS and PMI data – an update

Future articles

May 2011 will be the final edition of the Economic & Labour Market Review. This is a list of articles that may be included, although at this point of time it is highly provisional. Once the new website goes live in August 2011 future articles will be published directly on the ONS website and tagged to specific themes..

- Enhancing the coverage of financial sector activity
- China and the UK – patterns of international trade and investment
- Gross domestic product, national income and economic welfare
- The UK economy in 2010
- Volume index of capital services (VICS)
- Multifactor productivity estimates (MFP)