

# Data Explained

---

## **Exporting, importing and Northern Ireland firm performance and strategies: Which pathway to internationalisation?**

Authors: Dr Karen Bonner, Dr Eoin Magennis, Dr Fred Booth  
Economic Policy Centre, Ulster University  
Date: March 2024

---

The overall aim of the project is to improve our understanding of the relationships, at firm-level, between firms' engagement in markets external to Northern Ireland (NI) and their performance - and indeed strategies - in entering and exiting markets. This evidence base adds to current knowledge about the role of exporting and importing, and those firms engaged in such activities, within the NI economy.

### **Initial research questions**

Through the Business Data for Research (BDR) initiative from the Northern Ireland Statistics and Research Agency (NISRA), we propose to offer a descriptive analysis of export (and import) performance - by size, by sector and by geography - over the 2014-2020 period. We also analyse firms' internationalisation trends. Our research questions are:

1. What relationship is there between firm/sector characteristics and internationalisation (both exporting and importing) for NI firms?
2. What can we say about the contributions of extensive (numbers of market sold to or imported from) and intensive (increasing sales/purchases) margins to overall export growth for the NI economy?

## Datasets and variables used

The project uses the BDR dataset which contains an extract of data from the Northern Ireland Annual Business Inquiry (NIABI) 2014–2020, which includes Broad Economy Sales and Exports Statistics (BESES).

**Table 1: Breakdown of the datasets and their variables**

<b>ABI/BESES, BRES or EES</b>	<b>Year/s</b>	<b>Variable name</b>
ABI/BESES	2014-20	Anonymised_RURef
ABI/BESES	2014-20	WQ110
ABI/BESES	2014-20	WQ111
ABI/BESES	2014-20	WQ112
ABI/BESES	2014-20	WQ113
ABI/BESES	2014-20	WQ114
ABI/BESES	2014-20	WQ115
ABI/BESES	2014-20	WQ116
ABI/BESES	2014-20	WQ117
ABI/BESES	2014-20	WQ120
ABI/BESES	2014-20	WQ121
ABI/BESES	2014-20	WQ122
ABI/BESES	2014-20	WQ123
ABI/BESES	2014-20	WQ124
ABI/BESES	2014-20	WQ125
ABI/BESES	2014-20	WQ126
ABI/BESES	2014-20	WQ127
ABI/BESES	2014-20	WQ163
ABI/BESES	2014-20	WQ164
ABI/BESES	2014-20	WQ1600
ABI/BESES	2014-20	WQ1601
ABI/BESES	2014-20	WQ1602
ABI/BESES	2014-20	WQ1603
ABI/BESES	2014-20	WQ1604
ABI/BESES	2014-20	WQ1605
ABI/BESES	2014-20	MF5006
ABI/BESES	2014-20	MF5012
ABI/BESES	2014-20	MF5000
ABI/BESES	2014-20	MF5001
ABI/BESES	2014-20	MF5002
ABI/BESES	2014-20	MF5003
ABI/BESES	2014-20	MF5004
ABI/BESES	2014-20	MF5005
ABI/BESES	2014-20	MF5008
ABI/BESES	2014-20	MF5009
ABI/BESES	2014-20	MF5010

ABI/BESES	2014-20	MF5013
ABI/BESES	2014-20	WQ052
ABI/BESES	2014-20	WQ055
ABI/BESES	2014-20	WQ058
ABI/BESES	2014-20	WQ346
ABI/BESES	2014-20	WQ399
ABI/BESES	2014-20	WQ550
ABI/BESES	2014-20	WQ450
ABI/BESES	2014-20	WQ690
ABI/BESES	2014-20	WQ710
ABI/BESES	2014-20	WQ730
ABI/BESES	2014-20	WQ2010
ABI/BESES	2014-20	WQ2020
ABI/BESES	2014-20	WQ2030
ABI/BESES	2014-20	psic2
ABI/BESES	2014-20	psic3
ABI/BESES	2014-20	psic4
ABI/BESES	2014-20	sic074
ABI/BESES	2014-20	rusection
ABI/BESES	2014-20	AWEIGHT
ABI/BESES	2014-20	GWEIGHT
ABI/BESES	2014-20	CALWGHT

## Data limitations and coverage

### Sample size and coverage

The BDR sample, based on the NIABI, currently does not include all sectors. In terms of the 10X strategy, the omission of the Financial and Insurance sector means that analysis of Fintech/Financial Services is restricted.

The sampling frame also means that the smallest (micro) firms are most likely to have gaps in their records, reducing the scope for longitudinal analysis. These firms represent the majority of the registered business population and therefore would represent an important cohort for time-series analysis.

“NIABI samples all businesses with 50+ employees, or 20+ employees and more than one local unit and currently includes all manufacturing businesses with 6 or more employees and those with a turnover of £10 million or larger. Businesses falling below the threshold of complete enumeration are selected on a stratified random basis.”<sup>1</sup> This means that a business dropping below the threshold for one year may not be sampled, creating a gap in data.

Currently, the BDR dataset does not include:<sup>2</sup>

<sup>1</sup> <https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/NI-Annual-Business-Inquiry-Reporting-Unit-2021.pdf>

<sup>2</sup> As indicated in: <https://www.nisra.gov.uk/statistics/annual-business-inquiry/abi-sample-coverage> and <https://www.nisra.gov.uk/system/files/statistics/NI-Annual-Business-Inquiry-Reporting-Unit-2021.pdf>.

- Public sector activities (for the most part)
- Public administration and defence (section O)
- Farming (groups 01.1, 01.2, 01.3, 01.4 and 01.5 within section A)
- Local authority and central government bodies in education (section P)
- Human health and social work activities (section Q)
- Medical and dental practice activities (Section Q, 86.2)
- Financial and insurance activities (section K)
- Businesses that are not registered for either Pay As You Earn (PAYE) or Value Added Tax (VAT).

### Gaps

The dataset is presented with no missing data. Gaps in the data would be expected due to incomplete returns or non-response. The dataset should reflect the survey with gaps, providing researchers the ability to see where companies have not responded either to a particular question or to the whole survey. This would allow researchers to impute data if required. The current policy of populating all null fields with zeroes is also problematic, as it is not then clear whether the contributor is reporting that this figure is actually zero or has not responded. It would be preferable if there was an alternative marker for non-response, e.g. -9999.

All data is presented as actual returned data with no indication of where data has been imputed by NISRA. For transparency, imputed data should be clearly marked. Whilst recent bulletins do state that “for non-returns above a selected employment threshold, data were imputed using a methodology that takes account of previous returned data and the performance of other similar businesses”<sup>3</sup>, publication of this methodology would be useful. It would also be helpful to include details of whether it is the same methodology as the one used by Office for National Statistics for the Annual Business Survey.

### Zeros

There are several instances where variables show zero numbers potentially raising issues:

#### *Zero employment*

Investigations have shown that most companies returning zero employment on the NIABI form were shown to have employment on the Inter Departmental Business Register (IDBR) at time of selection. Of these companies, approximately half also returned zero for turnover. These companies may well have become dormant or ceased trading and will probably not affect any analysis.

The other half of companies, namely those recording turnover along with zero employment, has the potential to affect any analysis performed using a ‘per employee’ measure.

---

<sup>3</sup> <https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/NI-Annual-Business-Inquiry-Reporting-Unit-2021.pdf>

### *Zero turnover*

A small number of companies return employment data and associated employment costs data but also record zero turnover. A small number also return capital expenditure with zero turnover. It is possible that these returns indicate that the company is in startup mode and has not yet generated turnover, though in some cases this occurs over multiple years.

### *Zero GVA*

A small number of companies with zero GVA show either turnover or purchase figures making a zero GVA outcome unlikely. We did not request all the underlying variables used to calculate GVA, therefore without access to these it is impossible to determine whether these are errors.

### **Weights**

Weights are pre-calculated and included within the dataset with no accompanying documentation. Weights are also only provided at the NI level. More clarity on the weights and how outliers are treated would be useful within any accompanying documentation.

### **Sub-regional analysis**

Sub-regional analysis with any further level of disaggregation (size, sector etc.) is difficult given the current level of data disclosure. Data provided by Local Government District precludes effective analysis of the effects of border regions on exporting. As the data relates to reporting unit information only, any regional analysis is limited without access to local unit data.

### **'Unknown' location**

Within the dataset those observations that have the Local Government District marked as 'unknown' are lost to any regional analysis. Whilst the majority of these observations relate to companies, there is also a number of sole proprietors, public corporations, central government and non-profits included.

## **Necessary modifications to initial research questions or research design**

Due to the six-month time period and the data not being accessible on the grant start date, it was difficult to complete the full proposed analysis within the timeframe. Quite a bit of time was spent at the outset undertaking exploratory analysis as this was the first time using the data. As a result, the second research question, which is based on the panel dataset, is part of the follow-up study as the analysis was not able to be completed within the timeframe.

Disclosure restrictions also limited the extent to which some interesting findings could be discussed. These small numbers of specific cases may have provided some interesting case studies but did not meet the threshold for disclosure.

---

## Necessary modifications to the data

### Transformations for regressions

As zeros and negative values were present, data used in regressions was transformed using the hyperbolic sine function.

### Priority sectors

The Department for the Economy 10X vision for Northern Ireland set out priority areas of interest within the NI economy. As 10x companies are not currently identified in any documentation, we used the following definition of 10X sectors mapped into 2-digit SIC codes (from the Ulster University Economic Policy Centre Skills Barometer) to identify priority sectors:

**Digital and Creative Industries** is defined as the following sectors (SIC, 2-digit):

- 61 Telecommunications;
- 62 Computer programming, consultancy and related activities;
- 63 Information service activities;
- 58 Publishing activities;
- 59 Motion picture, video and television programme production, sound recording and music publishing activities;
- 60 Programming and broadcasting activities;
- 71 Architectural and engineering activities; technical testing and analysis;
- 73 Advertising and market research;
- 74 Other professional, scientific and technical activities;
- 90 Creative, arts and entertainment activities;
- 91 Libraries, archives, museums and other cultural activities.

**Agri-tech** is defined as the following sectors (SIC, 2-digit):

- 10 Manufacture of food products;
- 11 Manufacture of beverages;
- 12 Manufacture of tobacco products.

**Fintech/Financial Services** is defined as the following sectors (SIC, 2-digit):

- 64 Financial service activities, except insurance and pension funding;
- 65 Insurance, reinsurance pension funding, except compulsory social security;
- 66 Activities auxiliary to financial services and insurance activities.

Plus 20% of the following Tech sectors:

- 61 Telecommunications;
- 62 Computer programming, consultancy and related activities;
- 63 Information service activities.

**Advanced Manufacturing and Engineering** is defined as the following sectors (SIC, 2-digit):

- 20 Manufacture of chemicals and chemical products;
- 22 Manufacture of rubber and plastic products;
- 23 Manufacture of other non-metallic mineral products;



- 25 Manufacture of fabricated metal products, except machinery and equipment;
- 26 Manufacture of computer, electronic and optical products;
- 27 Manufacture of electrical equipment;
- 28 Manufacture of machinery and equipment n.e.c.;
- 29 Manufacture of motor vehicles, trailers and semi-trailers;
- 30 Manufacture of other transport equipment;
- 32 Other manufacturing.

**Life and Health Sciences** is defined as the following sectors (SIC, 2-digit):

- 21 Manufacture of basic pharmaceutical products and pharmaceutical preparations;
- 72 Scientific research and development.

Due to disclosure restrictions these had to be further amalgamated resulting in three classifications: Non 10x, Digital and Creative Sectors, and Other 10x.

### NI regions

As publication at Local Government District level would be disclosive under current guidelines, we created City Deal Regions for the geographic analysis. These regions were:

- Belfast City – encompassing Belfast Local District Council
- Belfast City Region – encompassing Antrim & Newtownabbey, Ards & North Down, Lisburn & Castlereagh, Mid & East Antrim and Newry Mourne and Down District Council areas
- Derry City & Strabane & Causeway Coast – encompassing Derry City & Strabane and Causeway Coast & Glens District Council areas
- Mid Southwest – encompassing Armagh City, Banbridge & Craigavon, Fermanagh & Omagh, and Mid Ulster District Council areas.

## Recommendations to data owners

To undertake more in-depth research into the Northern Ireland economy, our recommendations are:

### Sample size

Increase the sample size in general and include sectors not currently sampled.

Once qualifying for inclusion, these companies should be sampled each year (even if they subsequently fail to meet the qualifying criteria) to provide a better understanding of trends. Similarly, the sample should include a higher number of micro and smaller companies.

Include all selected companies in the dataset, whether responding or not, and also aim to achieve higher response rates to help with data completeness.

### Data provision

Without prior knowledge of the dataset it is difficult to predict which variables will be needed, particularly without knowing how complete the variable is (no. of zeros and non-zeros). The

provision of the full dataset with all available variables would enable a more in-depth analysis, and allow for the creation of proxy variables if needed.

Export/import data relating to all individual countries would also allow more scope for analysis than just the five broad markets currently available.

Data provided within the ONS Secure Research Service (SRS) in a number of formats would also be useful, e.g. SPSS, STATA, CSV.

### **Variables**

Clarify the weighting. It would also be useful to create sub-regional weights so that sub-regional analysis can be undertaken.

Include variables for company status (closed/dormant/non-responder) – with empty cells left for non-responders.

Leave missing cells blank, or use an indicator for missing (e.g. -9999) so that zeros represent an actual returned value of 0.

Identify 10x companies – once a definition for 10x companies is agreed, these should be marked as such on the dataset.

Access to company location at Super Output Area would allow analysis of trade in border regions, whilst access to local unit information would also enable more accurate geographic analysis.

Reduce the number of 'unknown' Local Government District markers and/or indicate the proportion of returned data to be attributed to different Local Government Districts, or ideally Super Output Areas could be identified.

### **Disclosure**

During accreditation training, it is demonstrated that in reality, at least three companies' data needs to be included to prevent disclosure. However, current guidelines state that there must be at least 10 companies before figures can be published, and that anything less is potentially disclosive. Even if there are more than 10 observations, further dominance checks need to be completed to ensure that (a) the largest entity is >43.75% of the total figure and/or (b) all entities except the largest two represent <12.5% of the largest entity.

It is arguable that if the priority of these tests were reversed, with dominance being checked first and discounted as not being an issue, then the observation count could be reduced (to five for example). This would allow a better understanding of the NI economy whilst preserving the anonymity of individual contributors. Whilst the current guidelines may allow meaningful results for large economies such as England, Northern Ireland has a smaller business population, and the current restrictions (of a cell count of 10) stifles meaningful analysis.



### **Research team**

Whilst the provision of support in the form of explanatory documents relating to aspects of the survey would be invaluable, provision of a NISRA representative to be included in the project would be beneficial. Currently, researchers cannot discuss specific observations with NISRA due to disclosure restrictions. Having a NISRA representative named as part of a project would allow researchers to share information within the SRS. This would allow researchers to formulate specific questions on the data and provide NISRA with specific information on possible problems or errors found within the dataset. This information could then be used to either 'fix' the data or form the basis of a 'known issues' document to be provided alongside the data for future research projects.

### **Additional data which would help to further develop the research**

The main restriction to the analysis has been that the sales and purchases data has been provided only at the level of the five broad markets. In order to properly examine internationalisation strategies, it would be beneficial to have access to data on all individual markets. This would allow for an analysis into whether firms start with the nearest markets and gradually expand further afield. It would also allow for a better analysis of whether firms export with more intensity to the same markets over time and/or whether they diversify into other export markets.

The above recommendations to the data owners also suggest some further additions to the variables.

More generally, if the data were to be provided in the SRS, it would be useful to include the original IDBR enterprise reference number so that linking could be undertaken with other NI datasets.

### **Please include code files used in your analysis**

Please contact NISRA Research Support Unit: [rsu@nisra.gov.uk](mailto:rsu@nisra.gov.uk) for code files.

### **Feedback on metadata, synthetic data and other documentation provided**

Documentation should be available within the SRS to help with the analysis. This should include the survey questionnaires and detail on the survey methodology, sample size, etc.

### **Any other comments**

It was beneficial to have access to the BDR dataset to undertake the analysis. We would welcome its further access to researchers via the SRS as it will enable a more in-depth understanding of the NI economy.

## Acknowledgements

Administrative Data Research Northern Ireland (ADR NI) takes privacy protection very seriously. All information that directly identifies individuals will be removed from the datasets by trusted third parties, before researchers get to see it. All researchers are trained and accredited to use sensitive data safely and ethically, they will only access the data via a secure environment, and all of their findings will be vetted to ensure they adhere to the strictest confidentiality standards. The help provided by the staff of Administrative Data Research Centre Northern Ireland (ADRC NI) and the Northern Ireland Statistics and Research Agency (NISRA) Research Support Unit is acknowledged. ADR NI is funded by the Economic and Research Council (ESRC). The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of the ADR NI. NISRA's data has been supplied for the sole purpose of this project.

## About ADR UK

ADR UK (Administrative Data Research UK) is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives. By linking together data held by different parts of government and facilitating safe and secure access for accredited researchers to these newly joined-up and de-identified data sets, ADR UK is creating a sustainable body of knowledge about how our society and economy function – tailored to give decision makers the answers they need to solve important policy questions. ADR UK is made up of four national partnerships (ADR England, ADR Northern Ireland, ADR Scotland and ADR Wales), and the Office for National Statistics (ONS), which ensures data provided by UK Government bodies is accessed by researchers in a safe and secure form with minimal risk to data holders or the public.

## Contact

[Ulster University Economic Policy Centre](#)

---

