United Nations ECE/CES/GE.20/2024/16

Economic and Social Council

Distr.: General
26 March 2024

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on National Accounts

Twenty-third session
Geneva, 23-25 April 2024
Item 4 of the provisional agenda
Subnational and regional accounts

Disaggregating UK annual Gross Value Added (GVA) to lower levels of geography: 1998 to 2021

Prepared by Office for National Statistics, United Kingdom[1]

Summary

There is growing demand for subnational statistics, and the Office for National Statistics is responding to this demand by implementing an ambitious plan to establish a framework for developing methods for producing granular economic statistics. This paper discusses fundamental requirements for successful production of subnational statistics of sufficient quality. It uses the development of a method to produce UK granular gross value added statistics as illustration. It emphasises the flexibility of using small building blocks to produce bespoke geographical areas for analysis.

[1] Prepared by Blessing Chiripanhura and Trevor Fenton.

I. Overview of subnational statistics development

1. Statistical users in the UK have increasingly focused on statistics and data at an ever more detailed lower-level geography for research, and for targeting and monitoring the impact of local policy making. In economic development circles, it is known that aggregated nature of national statistics masks local-level differences, hence the push for subnational statistics (Smits and Permanyer, 2019). At global level, the existence of within-country and (within-) regional disparities influenced the United Nations' development agenda to include Sustainable Development Goal number 10, which focuses on reducing inequalities between and within countries (United Nations, 2023).

2. In the UK, the interest in granular statistics has been rapidly increasing. For example, the 2016 review of UK economic statistics (Bean, 2016) recommended that the Office for National Statistics (ONS) should aim to provide greater granularity of economic statistics in terms of the amount of detailed information provided and the levels of geographic areas covered by the statistics. Although there are subnational institutions that produce subnational statistics for their own use, the statutory responsibility for producing statistics lies with the ONS. The ONS also has the resource base to produce UK-wide subnational statistics, which subnational producers of statistics do not have.

3. The ONS already publishes several statistical indicators at subnational level, including employment, gross domestic product, and gross disposable household income. The most common geographical level at which the subnational indicators are produced is local authority level. There are 379 local authority districts and equivalent administrations in the UK. There are indicators that go below the local authority level, like the ONS's small area income estimates. Further, the current Government Statistical Service subnational data strategy seeks to produce granular economic statistics, targeting sub-regional geographical levels. Since 2020, the ONS has been developing and refining a method for producing granular gross value added (GVA) estimates at lower-layer super output area and equivalent geographical levels, which we call the 'building blocks' (see Section 2.5). The latest experimental statistics were published at the end of January 2023. One of the major strengths of the building blocks datasets if that they offer users the flexibility to build their own geographical areas for analysis. However, we discourage users from comparing individual building blocks series because at that level of granularity, the timeseries are volatile. We recommend that they must use the building blocks to build larger geographical areas for analysis.

4. This paper discusses the basic requirements for successful subnational statistics development and uses the method and data sources used to produce the UK granular GVA statistics as illustration. It highlights future developments for granular economic indicators. It also highlights other geographical areas for which GVA estimates can be produced using the building blocks.

II. Basic principles for producing granular data

5. There are basic principles and requirements that are necessary to produce granular economic statistics. We discuss these and how they apply to the UK GVA statistics at lower levels of geography.

A. The need for a pre-existing national framework

6. The consideration for the development of low-level geography statistics started with an understanding that we already had national GVA statistics for the UK, the devolved administrations, and down to local authorities. The subnational GVA statistics sum up to the total UK figure. This means we have a pre-existing national framework within which we are constrained, which makes the statistics development process simpler than starting from nothing.

7. Since the existing framework is at national level, the subnational statistics development process is based on a top-down approach. This means we break down the local authority level GVA to lower-layer super output area and equivalent geographical levels. This approach is most suitable because the coverage of survey data become progressively thinner as we go down the geographical hierarchy. We can therefore develop innovative methods that use administrative and proxy data to break down GVA to lower levels of geography. By combining different data sources, we derive proxy datasets with adequate coverage of industry-level GVA. We do not produce industry-level granular GVA data series because they are disclosive.

B. Securing access to administrative and other proxy data

8. Surveys are often not large enough to allow the development of statistics at low levels of geography. Increasing sample sizes is costly and may over-burden respondents. This calls for other data sources to be considered, of which administrative sources are a key candidate. Securing regular access to administrative data sources will provide sufficient coverage of smaller businesses to help ensure that economic statistics for lower levels of geography are of sufficient quality.

9. In the UK, to produce GVA statistics at low level geography, we need access to Value Added Tax (VAT) turnover data. This is essential for the apportionment of industry GVA to local business units. The allocation of VAT turnover itself is based on local unit share of employment.

10. We also require other datasets to apportion the GVA of industries that do not have VAT turnover data, that for the public sector industries, the household sector, and for the imputed rent component of real estate services.

C. Maintaining an up-to-date business register

11. Breaking down local authority-level GVA by industry to low level geographies means we must understand the location where the economic activity is taking place. We must have a comprehensive picture of the structure and distribution of both private and public businesses' economic activities. For this, we use the national business register, Inter-Departmental Business Register (IDBR). The IDBR is a comprehensive list of UK businesses used by government for statistical purposes. It provides the main sampling frame for surveys of businesses conducted by the Office for National Statistics (ONS) and other government departments.

12. The IDBR is populated and regularly updated by administrative data, principally from company tax returns, plus additional information gathered by an annual Business Register and Employment Survey (BRES). We match the IDBR with the VAT returns data to create separate datasets with separate records for local business operations. The IDBR holds information on the postcode, the geographical structure of businesses, employment and the main economic activity conducted at each site.

D. Dealing with complex business operations operating across multiple sites

13. One of the challenges of dealing with business data is that larger businesses tend to operate across multiple sites, which makes breaking down GVA at industry level complex. Such large businesses are fewer than small and medium-sized businesses, but they often represent much of the economic activity in particular industries. We must quantify the amount of their activity taking place at each site.

14. It is normal for data collected by both surveys and administrative sources to only have information for the whole enterprise (although sometimes arrangements are made for a company to provide data split into several principal activities, or industries, if they are sufficiently large and diverse). Often, it is likely that the company itself may have no

conception of how to allocate its economic activity the various sites. This is because some variables, like profits, are conceptually difficult to allocate across sites. For these reasons, a national business register like the IDBR plays a vital role in providing information that can be used to facilitate such allocation, making it possible to derive lower-level GVA statistics.

E. Selecting the appropriate geographical level

15. In deciding to produce lower-level geography statistics, we must decide on the exact geographical level to target. We know that the demand for subnational statistics is driven by the needs of local administrators, policy makers and researchers. Central government also requires these statistics to monitor localised social and economic development across the country. This requires the existence of a hierarchy of statistical geographies with some degree of commonality at each level.

16. In the UK, we use the International Territorial Levels (ITL) framework, which is based on the European Nomenclature of Territorial Units (NUTS) statistics framework, and comprises a hierarchy of regions, sub-regions, and local areas, with areas at each level having resident population within certain specified bounds. There are 12 ITL 1 areas (formerly NUTS 1 areas) in the UK. In addition, we have seven other groups of statistical geographies namely postal, administrative, electoral, health, census, other geographies, and statistical building blocks (that is, middle-layer super output areas, lower-layer super output areas, and their equivalent geographies across the UK nations).

17. The UK currently produces regular economic statistics for local authorities to meet the needs of local government policy makers. The ONS is currently taking the provision of subnational statistics to the next level, breaking down our leading economic measures to lower-layer super output areas and equivalent geographies. GVA is the first statistic to be broken down to this level for the whole UK. The programme provides a set of small area statistical building blocks, which people can use to construct any area of interest, no matter how oddly shaped, or to delve into the detailed make-up of other larger areas, to identify areas in need of intervention or development.

Structure of the building blocks

18. The smaller building blocks, based on 2011 census geography codes, consist of lower-layer super output areas (LSOA) in England and Wales, data zones (DZ) in Scotland, and super output areas (SOA) in Northern Ireland. The 2021 census geography codes for Northern Ireland changed SOA to data zones.

19. The building blocks are designed to divide the UK based on the number of households and include any businesses operating in the same area. They are population-based and hence comparable. We have smaller and larger building blocks, described as follows:

Smaller building blocks

- Lower-layer super output areas have a population of 1,000 to 3,000 people (400 to 1,200 households).
- Data zones have a population of 500 to 1,000 household residents.
- Super output areas have a population of 300 to 6,000 people.

Larger building blocks

- Middle-layer super output areas (MSOA) in England and Wales have a population of 5,000 to 15,000 people (2,000 to 6,000 households).
- Intermediate Zone (IZ) in Scotland have a population of 2,500 to 6,000 household residents.

(Office for National Statistics, 2021; Scottish Government, 2021; Northern Ireland Statistics and Research Agency Geography, 2019)

20. There are no larger building blocks geography in Northern Ireland. We produce GVA statistics for smaller building blocks. These can be used to produce the larger building blocks and other geographical areas. We publish lookup tables alongside the GVA statistics to make it possible for users to build their own areas.

III. Method and data for producing low-level geography GVA statistics

21. The production of GVA statistics at building blocks level is based on an apportionment method. We start with GVA at local authority level, which must be apportioned to LSOA and equivalent geographies (that is, LSOA, DZ and SOA). We use VAT Turnover data, employment, population, and dwelling stock to apportion GVA to LSOA and equivalent geographical levels, which we call the 'building blocks' in the method description.

22. The Digital Economy Act (2017) has enabled access to an administrative source of data collected through VAT returns, which contains records for most businesses operating in the UK (all those registered for VAT) and includes variables for the company's turnover and expenditure.

23. By matching these records to the IDBR, we have created a dataset that contains a separate record for each physical site (local unit) where a business operates, including information on the address, employment, and main activity. We use this information to allocate VAT turnover to business sites in any geographic area, even small ones.

24. The VAT Turnover information is for the whole company. It is allocated between the sites of a business according to the relative share of the total workforce located at each site within each building block. This approach assumes workers across all sites contribute equally to a company's output, which assumes equal productivity. Although this may not be the case because of differences in skills and capabilities, and in economic activities at different sites, this is a reasonable assumption suitable for the purpose of apportioning GVA to small geographical areas. From the steps described here, we generate VAT turnover by section for each building block. The sections are described in the International Standard Industrial Classification of all economic activities publication (United Nations, 2008)[2].

25. There are no VAT turnover records for households with employees and own account production, nor for imputed rental of owner-occupied dwellings and non-market activities of public sector industries. We apportion the GVA for these industries using alternative data sources at building block level, as follows:

- Population is used to apportion the GVA of Section T (activities of households as employers; undifferentiated goods- and services-producing activities of households for own use).
- Dwelling stock is used to apportion the GVA of part of Section L (68.2IMP) (that is, renting and operating of own or leased real estate).
- Employment is used to apportion the GVA of public sector sections O, P and Q (public administration and defence, and compulsory social security; education; and human health and social work activities).

26. The apportioning datasets have sufficient coverage of all areas. We therefore derive a composite apportioning dataset and fill up gaps by rolling back the first existing data point of each variable. The composite apportioning dataset is used to apportion the GVA statistics for the period 1998 to 2021.

[2] https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf

A. Apportioning local authority (LA) level GVA to building blocks level

27. We start with data engineering to allocate VAT turnover to enterprises' local units. This is achieved by matching VAT turnover records to the Inter-Departmental Business Register (IDBR) to create a new dataset that is used to allocate VAT turnover to business sites.

28. Next, we apportion the largest number of sections using the new VAT turnover data.

a) We apportion GVA for all sections except O, P, Q, T, and part of L using VAT turnover data:

$$\frac{\text{Building block Section } i \text{ VAT turnover}}{\text{LA SUM of Section } i \text{ VAT turnover}} \text{ x LA Section } i \text{ GVA } = \text{ Building block Section } i \text{ GVA}$$

where i = all other sections except O, P, Q, T, and part of L (68.2IMP)

29. This gives the building blocks GVA for all sections with VAT turnover data in a local authority.

30. Next, we apportion the GVA of sections O, P, Q, T, and part of L.

b) Sections O, P and Q:

$$\frac{\text{Building block Section } i \text{ employment}}{\text{LA SUM of Section } i \text{ employment}} \text{ x LA Section } i \text{ GVA } = \text{ Building block Section } i \text{ GVA}$$

where i = Sections O, P and Q.

c) Section T:

$$\frac{\text{Building block population}}{\text{LA total population}} \text{ x LA Section T GVA = Building block Section T GVA}$$

d) Part of Section L (68.2IMP):

$$\frac{\text{Building bock dwelling stock}}{\text{LA total dwelling stock}} \text{ x LA Section L: 68.2IMP GVA } = \text{ Building block Section L: 68.2IMP GVA}$$

31. After apportioning all industries to building blocks level, we calculate the building block section i total GVA by summing across all section i values in the building block. We can sum section i GVA across all building blocks in a local authority to get local authority section i GVA.

$$LA \text{ Section } i \text{ total GVA } = \sum \sum \text{ Building block Section } i \text{ GVA } in\ LA$$

where i = Sections A to T.

32. The internal summation gives the total GVA of section i in a building block. The outside summation gives the total apportioned Section i GVA in all building blocks in a local authority. The result must equal to the starting value of local authority section i GVA. The equality must hold, otherwise the apportionment process will be inaccurate.

33. The sum of GVA of all sections in all building blocks GVA in a local authority must equal the local authority GVA. This is a global check that we have effectively constrained the apportioned data to the local authority totals.

$$\sum \text{Total building block GVA in LA } = \text{ Total LA GVA}$$

34. We apply this method for the years 1998 to 2021 to produce GVA time series for each building block.

B. Dealing with the risk of statistical disclosure

35. Apportioning GVA data to small geographical areas comes with the perceived risk that people with local knowledge may try to estimate the GVA of local dominant businesses, which may jeopardise ONS business data collection processes. This is only a perceived risk because GVA is an economic concept, and no one knows the true GVA values. The businesses themselves may not know their true GVA. Despite all this, we must assure businesses that we are taking precautions and applying the necessary statistical controls and procedures to protect their data. For example, we do not publish local area industry GVA data because it may be disclosive. The risk is further mitigated by the fact that all the GVA figures are estimates compiled using auxiliary variables to break down figures from the UK total (thus ensuring that our estimates sum to all ITL), meaning that there is no way to derive precise company values from our published data.

36. We refer to the Government Statistical Service's disclosure control guidance for tables produced from surveys (PDF, 250KB) to address perceived risk of disclosure. We have guarded against disclosive data by ensuring that the industry level data we use for apportionment includes a minimum of at least four separate business enterprises.

37. We calculate the level of business and industry dominance to identify LSOA and equivalent geographies with elevated risk of statistical disclosure. We set aside these building blocks for disclosure treatment in line with organisational guidance. Our treatment method does not suppress disclosive building blocks because the strength of the dataset is in maintaining all the building blocks for users.

38. Our approach is to average a disclosive building block with a neighbouring non-disclosive one within the same middle-layer super output area, intermediate zone, or district electoral area (which we call the middle-level geographical areas). The selection of a suitable building block to average with is based on a condition that the partner building block must have a minimum size that will allow the disclosive building block to be masked to a minimum of 20% of its current GVA.

39. Next, we determine whether a middle-level geographical area contains one or multiple disclosive building blocks. If it has multiple disclosive building blocks, we average all the building blocks for the whole timeseries.

40. If a middle-level geographical area has one disclosive building block, we use formulas to calculate bounds which help to choose the most suitable partner building block to average with across the entire timeseries. To avoid the potential problem of a partner being matched with multiple building blocks multiple times, we select the partner with the highest match counts over time as the most suitable to average with across the whole timeseries. If there is no potential partner with the highest count, we select a partner that is closest to the appropriate bound parameter.

C. Testing for and treating outliers

41. After disclosure treatment, we check the data for outliers. Outliers may come from the underlying apportionment datasets, or from the disclosure treatment. We reduce the influence of outliers checking and adjusting the building blocks time series.

42. We generate adjustments for each local authority and re-apportion the net adjustments to all building blocks in the local authority that have not been adjusted.

43. The re-apportionment ensures that the sum of building blocks data in each local authority remains equivalent to the local authority total GVA. It is these series that we publish for all building blocks and for different geographical areas. In the next section we demonstrate how the granular data can be used to for analysis.

IV. Application: The West Midlands Metro Region

44. This section demonstrates the flexibility of using the GVA building blocks data to analyse a selected geographical area. In this case, the selected geographical area is a metro line in the West Midlands. This case is interesting because it illustrates the geographical flexibility offered by our subnational GVA data. It shows how we can fit the analysis in an area with no initially defined boundary.

45. The results of this application do not allow for conclusions regarding cause and effect. They are for illustrative purposes only, and further investigation into the causes of economic change are beyond the scope of this paper.

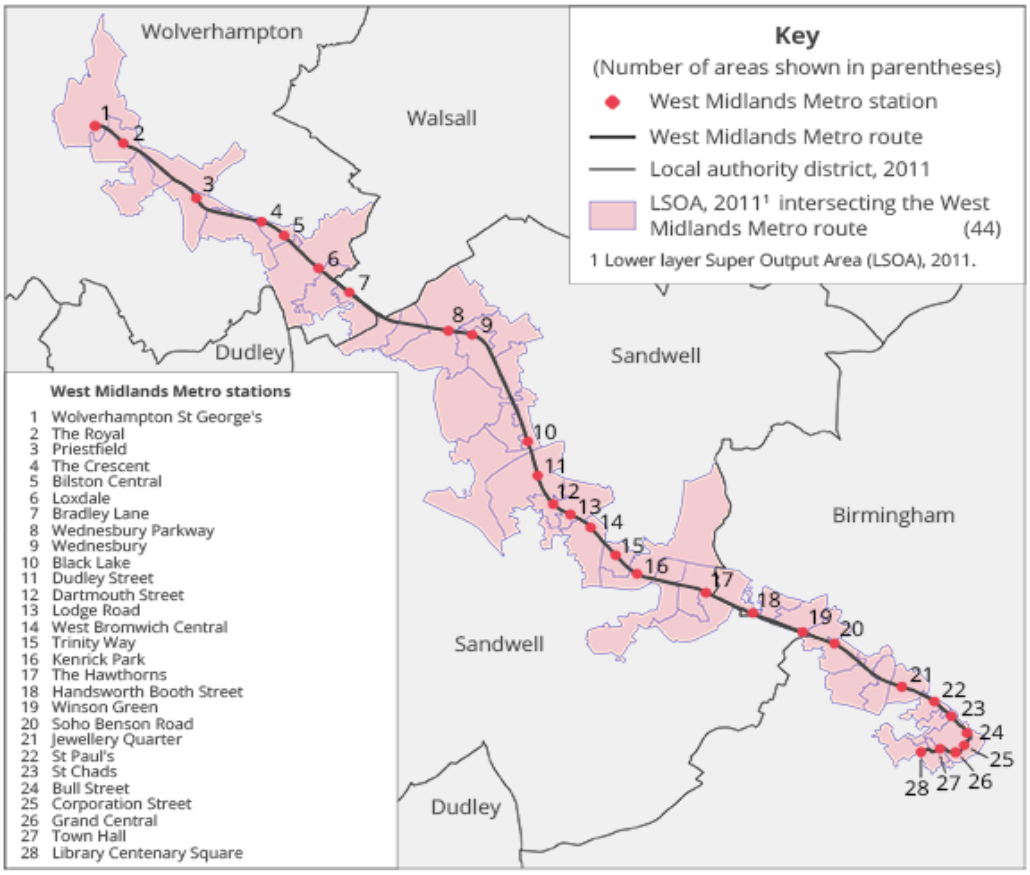
46. Case: The West Midlands Metro, illustrated in Figure in Figure 1, is a light-rail system operating between Wolverhampton St. George's and Birmingham Bull Street. The construction of the rail line started in 1995, and the first section was opened in 1999. Extensions of the line were opened in 2015 and 2019, respectively.

47. We can analyse the changes in local output over this period using the GVA time series. We want to understand how the establishment and extension of the rail line influenced economic development and outcomes in the surrounding areas.

48. Mapping the area: We select all LSOA through which the metro line passes. We also include all LSOA whose boundaries intersect the metro line.

49. The map shows the West Midlands Metro route as a line from Wolverhampton in the North-West of the map to Birmingham City Centre in the South-East of the map. The line has a total of 28 stations, crossing or meeting with 44 LSOA.

Figure 1
West Midlands Metro Region as of December 2019

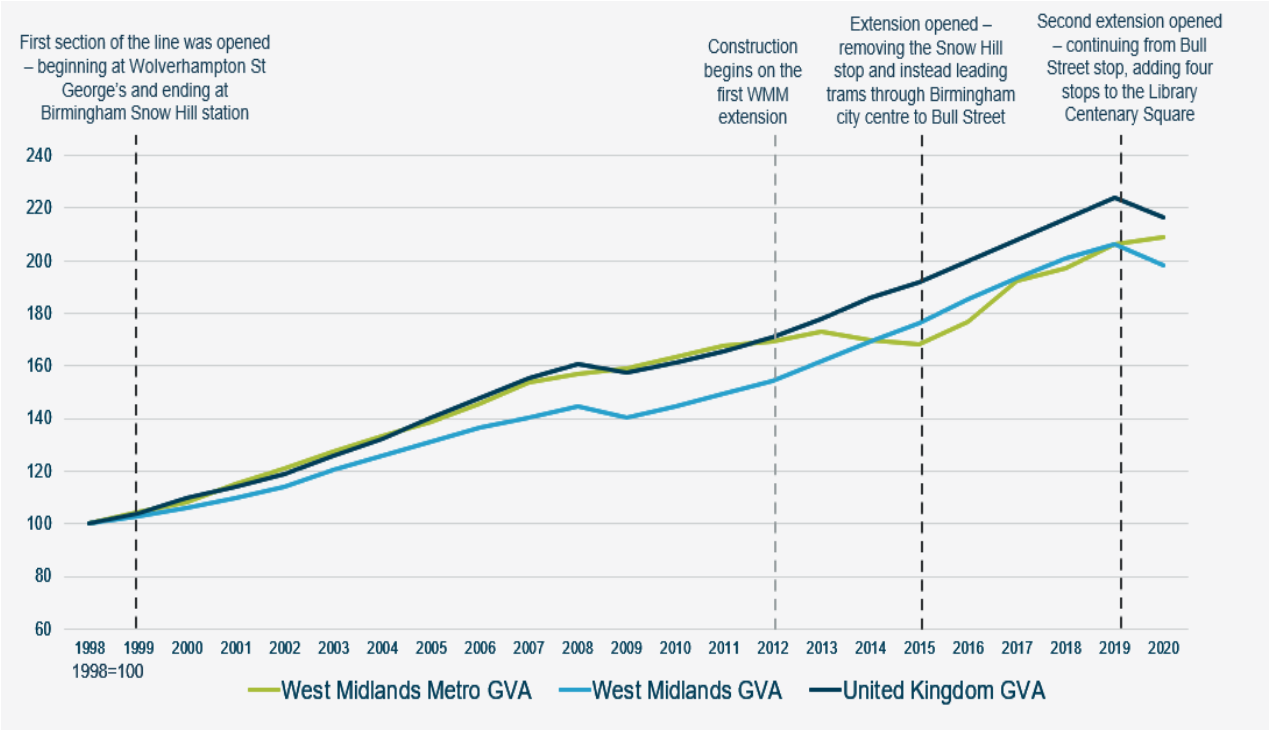


Source: OS Open Zoomstack, Ordnance Survey; West Midlands Metro; Wikipedia; Office for National Statistics licensed under the Open Government Licence v.3.0.; Contains OS data © Crown copyright 2023; Graphic created by ONS Geography

50. In 2020, the region had a total GVA of £9,845 million, more than doubling from £4,717 million in 1998, all measured in current prices.

51. We can enhance our analysis of the West Midlands Metro region (WMMR) by plotting the region's GVA against the UK and the West Midlands region GVA. Figure 2 shows that the three geographical areas had similar rates of growth across the period, but the West Midlands Metro region had stronger growth than the West Midlands region until 2014. The WMMR GVA grew slightly more sharply in 2015 following the opening of the first extension to the metro line, which then ended in Birmingham City Centre. The GVA ticked up again in 2019 when the line was expanded.

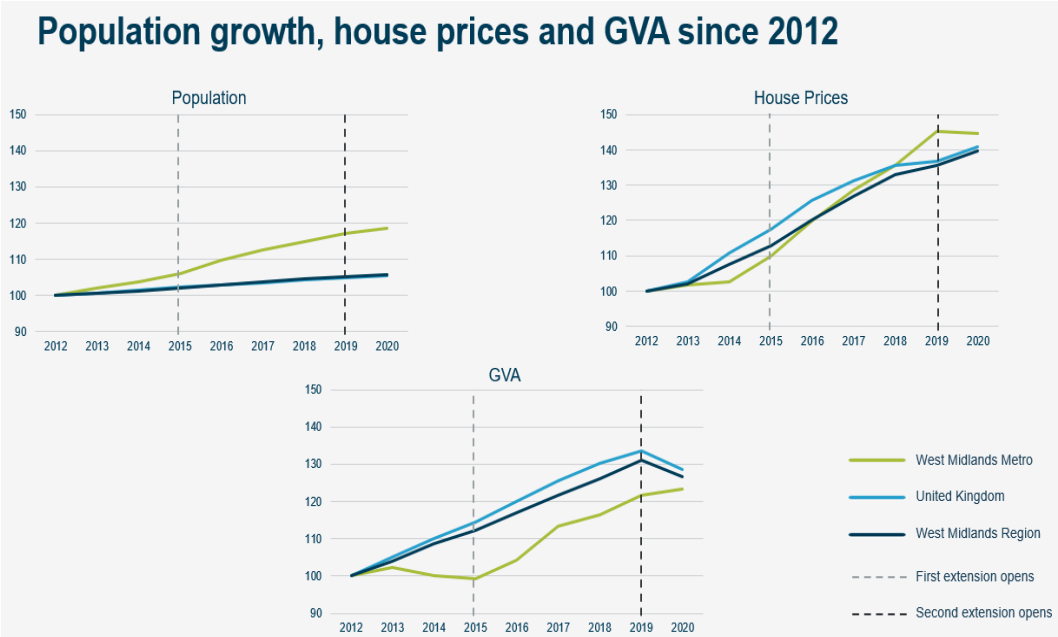
Figure 2
GVA increased at a faster pace in the WMMR from 2015 onwards compared with the UK average



Source: *GVA data – ONS statistics in development*

52. The construction of the West Midlands Metro line extension was first announced in 2012. If we zoom into the period 2012 to 2020, plotting population and GVA growth, we see that population grew at a faster rate in the Metro region compared with the UK and the West Midlands region more generally. Adding house prices to the picture shows that house prices in the Metro region swung up from 2014 onwards.

Figure 3
Population and house prices grew at faster rates in the West Midlands Metro region than the UK average from 2015 onwards



Sources: Population – Census Output Area population estimates – West Midlands, England (supporting information); House prices – UK house price statistics and West Midlands Metro Region house prices;GVA – GVA data – ONS statistics in development

53. The extension opened in 2015 was the first since the line first opened in 1999. As a result, there was high interest in this development. House prices in the West Midlands Metro Region appear to have had the strongest growth rate following the opening of the extension in 2015, with another uptick between 2018 and 2019 when the second extension opened. Both events coincided with the increased rate of growth in GVA seen starting around in the same period to bring the overall growth rate in line with the rest of the UK. It is likely that the extension of the line increased the desirability of living along the metro line. However, it is important to highlight that we cannot imply causal inference from this presentation of statistical indicators.

V. Conclusion

54. The availability of economic indicators at subnational level continues to gain momentum. We have highlighted that there are basic conditions that must be in place before attempts to produce subnational statistics. Whether a top-down or bottom-up approach is adopted for producing subnational statistics, it is important to ensure that any subnational versions sum to same national total figures. Both approaches have advantages and disadvantages. Which one to adopt depends on the availability of sufficiently decent quality data and reasonable assumptions to produce subnational data of excellent quality. Also, the assumptions must be communicated clearly to users.

55. We highlighted the importance of securing regular access to administrative and other data sources that provide sufficient coverage of smaller businesses to facilitate the development of subnational statistics of sufficient quality. There is also need for an up-to-date business register, which facilitates the allocation of economic activity to locations around the country. Lastly, we have highlighted that there is need for sufficiently linked geographical hierarchies with a common foundation.

56. This paper has shown how these conditions have been applied in the UK's ambitious programme to produce and disseminate subnational statistics to inform local research and policy analysis. We produced GVA statistics for small areas and explained how we ensured that there is built-in flexibility to construct larger geographical areas for analysis. Lastly, we have shown how the GVA building blocks data can be combined with other economic indicators to generate a detailed profile of an area.

References

Bean, C. (2016) Independent Review of UK Economic Statistics. https://assets.publishing.service.gov.uk/media/5a7f603440f0b62305b86c45/2904936_Bean_Review_Web_Accessible.pdf

Northern Ireland Statistics and Research Agency (NISRA) Geography (2019) NISRA Geography Fact Sheet. https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/NISRA%20Geography%20Fact%20Sheet%20-%20March%202019.pdf

Office for National Statistics (2021) Census Geography: An overview of the various geographies used in the production of statistics collected via the UK census. https://webarchive.nationalarchives.gov.uk/ukgwa/20220401212146/https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

Scottish Government (2021) Intermediate Zone Boundaries 2011. https://www.data.gov.uk/dataset/133d4983-c57d-4ded-bc59-390c962ea280/intermediate-zone-boundaries-2011

Smits, J. and Permanyer, I. (2019) The subnational human development database. https://www.nature.com/articles/sdata201938

United Nations (2023), The Sustainable Development Goals Report Special edition. https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023.pdf

United Nations (2008) International Standard Industrial Classification of All Economic Activities Revision 4. Statistical papers Series M No. 4/Rev.4.