EUROPEAN DATAWAREHOUSE

Explanatory Report

Standardised geographic information for European Loan Level Data

Analyst Contacts

Ludovic Thebault, PhD
Vice President
+49 (0) 69 8088 4302
ludovic.thebault@eurodw.eu

Bruno Pereira
Senior Project Manager
+49 (0) 69 8088 4312
bruno.pereira@eurodw.eu

Special thanks to Eimear Leahy, Gisela Herkner, Jiamin Lou and Matthew Aitken for their contribution to the paper

European DataWarehouse GmbH
Walther-von-Cronberg-Platz 2
60594 Frankfurt am Main
www.eurodw.eu

How and why European DataWarehouse (ED) produces standardised geographic indicators

Standardised region names are now available for almost all loans in ED as fields Geo_1 and Geo_2. This report concludes the three previous ED publications about standardised geographic information for French, Italian and Spanish loan level data by addressing the other countries represented in ED. ED intends to improve and update its processes to determine region names so as to reflect the actual reporting practices of data providers. Please refer to Appendix 1 for detailed instructions regarding the generation of fields Geo_1 and Geo_2.

The geographic origin of the loans is reported in Edwin as a postcode or NUTS[1] identifier (see Appendix 2). For SME and RMBS loans, postcode information is mandatory while NUTS is optional, whereas for leasing deals, postcode information is optional and NUTS is mandatory. Because reporting practices vary considerably, particularly for the postcode, identifying the exact regional origin of loans can be work intensive. ED's calculated fields Geo_1 and Geo_2 make this task fast and user friendly as the necessary preparation work is done at the ED level.

In order to determine the Geo_1 and Geo_2 fields, ED uses the information reported by data providers as efficiently as possible.[2] ED's queries typically use the mandatory information in priority and use optional information when mandatory information is insufficient. Country-specific mappings[3] and procedures identify the most precise geographic description possible given the available data. Exhibit 1 shows the regional unit used for each country as a basis for the identification of the Geo fields. The regional unit

---

[1] See http://ec.europa.eu/eurostat/web/nuts/overview and http://ec.europa.eu/eurostat/web/nuts/history
[2] ED endeavours to make the best use out of the data provided, even when it is not reported according to taxonomy. If, for instance, a NUTS identifier is reported in a postcode field, ED's query may still use it to find the appropriate Geo field, while this entry would also be recognised as a data quality problem and addressed as such with the data provider. ED's mappings typically account for updates in NUTS so that a region can still be identified following NUTS revisions. and
[3] See in the appendices of ED's reports for France, Italy and Spain.

used as a basis for Geo_2 is often NUTS 3.[4] When the first digits of the postcode sufficed to determine NUTS 2, NUTS 2 was selected for Geo_2.

Exhibit 1: Regional unit used to define Geo_1 and Geo_2

	Geo_1 based on	Geo_2 based on
Austria	9 * NUTS 2 regions (Bundesland)	35 * NUTS 3 regions (Gruppen von Bezirken)
Belgium	3 * NUTS 1 regions	11 * NUTS 2 regions (provinces + Brussels)
Finland	5 * NUTS 1 regions	20 * NUTS 2 regions (maakunta)
France	22* NUTS 2 regions (régions) + overseas	96 * NUTS 3 regions (départements)
Germany	16 * NUTS 1 regions (Bundesland)	39 * NUTS 2 regions (Regierungsbezirke)
Ireland	8 * NUTS 3 regions	26 * counties
Italy	21 * NUTS 2 (regions)	95 * NUTS 3 (provinces)
Netherlands	12 * NUTS 2 regions (province)	40 * NUTS 3 regions (COROP regions)
Norway	7 * NUTS 2 regions	19 * NUTS 3 regions (counties)
Portugal	18 * distritos + Açores + Madeira	Postcode areas
Spain	19 * NUTS 2 regions (autonomous community)	52 * NUTS 3 regions (province)
Sweden	8 * NUTS 2 regions (Riksområden)	21 * NUTS 3 regions (counties/Län)
United Kingdom	12 * NUTS 1 regions + 13 Crown dependencies and overseas territories (identified by postcode only)	37 * NUTS 2 (counties/districts/council areas (unitary authorities) when NUTS is the mandatory information; 119 * Postcode areas

Source: European DataWarehouse;

The following sections of this report explain how the standardised geographic indicators were determined for the main markets that were not addressed in previous reports. For Belgium, data is generally sufficient and the mapping was rather straightforward. For the Netherlands, data is generally sufficient except for about 10% of the sample. Indeed, when only 2 postcode digits are available as sole source of information, ED's mapping becomes less accurate. For Portugal and the UK, NUTS and postcode regions do not perfectly match, ED therefore used a mapping based on the postcode regions and districts, rather than a postcode primarily following NUTS regions. In the case of the UK, ED mapped the postal districts to the NUTS 1 regions. In the postcode field, Irish data providers tend to report the actual county name instead of the postcode. Irish counties were used as Geo_2 and were also mapped to NUTS 3 regions (used as Geo_1). For the other countries, ED's mapping reflects the NUTS information provided, either

---

[4] It is implicitly understood that the first digits of a country should identify the region and that sufficient details should be provided for this purpose. In France and Spain, two digits are generally sufficient. In the Netherlands, four digits are needed. In some countries such as Germany however, identification of the region from the postcode can be tricky even when the full postcode is provided. For some deals, available information may allow more precision than what Geo_2 provides. For example, in some cases Geo_2 is mapped to NUTS 2 even though NUTS 3 is provided.

because the postcode information is not sufficient to determine the region (and thus because a region name can only be provided when NUTS information is provided), or because that country is represented by deals for which NUTS is the main information available (for auto deals for instance).

Belgium

In Belgium, the NUTS 2 region (province) is identified with the first two digits of a postcode. Almost all loans report at least one precise geographic identifier. 81% of RMBS loans and all SME loans for Belgium, report both NUTS and the postcode, which almost always match (Exhibit 2). The distribution of the loan exposures per region broadly matches the contribution of these regions to the Belgian GDP (Appendix 3). The capital region (Brussels), however, is underrepresented compared to its weight in the Belgian economy.

Exhibit 2: Regional postcodes and NUTS generally match for Belgian RMBS and SME loans

	RMBS	SME
A) Total Number of Loans*	955,156	213,002
B) Of which loans with no geographical info available**	428	-
C) Proportion match possible (=1-B/A)	100.0%	100.0%
D) Loans reporting both postcodes and NUTS	773,381	213,002
E) loans reporting both postcodes and NUTS as % of total loans (=D/A)	81.0%	100.0%
F) Cases where the province indicated by postcodes NUTS match	771,274	213,002
G) Cases where postcodes and NUTS do not match (=D-F)	2,107	-
H) Matching provinces as % of cases where both postcodes and NUTS are available (=F/D)	99.7%	100.0%
I) Incoherences as proportion of total number of loans (=G/A)	0.2%	0.0%

*Counting only the number of active loans in active deals as of last reporting date available
**Counting "other region" NUTS as not available

Source: European DataWarehouse, data providers;

Netherlands

The Netherlands are divided into 12 administrative provinces (corresponding to NUTS 2 and ED's Geo_1), which, in turn, are divided into municipalities and regrouped into COROP[5] regions (corresponding to NUTS 3 and ED's Geo_2). As shown in Exhibit 3a, 99.9% of Dutch RMBS loans report only 2 postcode digits of which 89.6% also report NUTS 3. In cases where four-digit postcodes are provided, these can be mapped to the NUTS 3 regions using specific matching tables provided by Eurostat.[6] ED can determine Geo_1 and Geo_2 based on the four-digit postcode when available, or based on the NUTS code if no four-digit postcode is provided. When only two digits of the postcode are provided without

---

[5] COROP stands for "Coördinatiecommissie Regionaal Onderzoeksprogramma", i.e. Coordination Commission Regional Research Programme.
[6] See http://ec.europa.eu/eurostat/tercet/flatfiles.do

NUTS (as is the case in Exhibit 3a for 293,561 loans) Geo_1 can still be identified, but the mapping is then less precise and does not allow the identification of Geo_2.[7] Two-digit postcodes are thus used as a last resort.

Exhibit 3a: Reporting of geographic information for Dutch RMBS LLD

# Postcode Digits	NUTS provided*	Number of loan parts	% of Total
2	Yes	2,562,402	89.6%
2	No	293,561	10.3%
3	Yes	1,772	0.1%
3	No	-	0.0%
4	Yes	-	0.0%
4	No	288	0.0%
5	Yes	-	0.0%
5	No	719	0.0%
6	Yes	-	0.0%
6	No	-	0.0%
		2,858,742	100%

*Source: European DataWarehouse, data providers; *treated as "No" where NUTS is reported as "NLZZZ"*

2.8 million loans (or loan parts) are available for Dutch RMBS, and almost all provide a usable geographic indicator. While postcodes and NUTS are available for most RMBS loans (2,564,174 out of 2,858,742 i.e. 89.7%), information on both NUTS and postcodes is only available for 22% of SME loans (Exhibit 3a, Exhibit 3b). In such cases, the province (Geo_1) matches in 93.6% and 92% of cases for RMBS and SME loans respectively (Exhibit 4). This represents a substantially worse fit than is the case in other markets. In fact, it seems that NUTS is often also provided, precisely because the 2-digit postcode is not sufficient to determine the province with certainty. For 4,009 SME loans out of 31,592, no usable NUTS code is provided and only the first two or three postcode digits are available. In this sample, our mapping using two digits should determine Geo_1 accurately about 90% of the time.

---

[7] The region can be identified with two digits in about 90% of cases. An analysis of the mismatched cases (when the two-digit postcode points to a different region than the NUTS), shows that they almost systematically affect regions sharing a border. For RMBS, of the 164,974 mismatches observed (using the first two digits of the postcode), 85,640 involve the region Gelderland, which shares borders with 6 other regions (Overijssel, Flevoland, Utrecht, Zuid-Holland, Noord-Brabant and Limburg). In 99.99% of these cases involving Gelderland, the other region involved in the mismatch is one of these 6 regions it shares a border with.

Exhibit 3b: Reporting of geographic information for Dutch SME deals

# Postcode Digits	NUTS provided*	Number of loan parts	% of Total
2	Yes	-	0.0%
2	No	4,009	12.7%
3	Yes	6,936	22.0%
3	No	-	0.0%
4	Yes	-	0.0%
4	No	1,762	5.6%
5	Yes	-	0.0%
5	No	-	0.0%
6	Yes	-	0.0%
6	No	18,885	59.8%
		31,592	100%

*Source: European DataWarehouse, data providers; *treated as "No" where NUTS is reported as "NLZZZ"*

Exhibit 4: Regional postcodes and NUTS generally match for Dutch RMBS and SME deals

	RMBS	SME
A) Total Number of Loans*	2,858,742	31,592
B) Of which loans with no geographical info available**	2,088	2
C) Proportion match possible (=1-B/A)	99.9%	100.0%
D) Loans reporting both postcodes and NUTS	2,564,174	6,936
E) loans reporting both postcodes and NUTS as % of total loans (=D/A)	89.7%	22.0%
F) Cases where the province indicated by postcodes NUTS match	2,399,200	6,379
G) Cases where postcodes and region do not match (=D-F)	164,974	557
H) Matching provinces as % of cases where both postcodes and NUTS are available (=F/D)	93.6%	92.0%
I) Incoherences as proportion of total number of loans (=G/A)	5.8%	1.8%

*Counting only the number of active loans in active deals as of last reporting date available
**For which dummy or incorrect postcode info was provided

Source: European DataWarehouse, data providers;

Overall, the sample of loans in ED's database broadly reflects the geographical distribution of the Dutch economy (see Appendix 3). Nevertheless, the Noord-Holland region, which includes Amsterdam, is underrepresented compared to its weight in the Dutch economy.

Portugal

For most of the Portuguese RMBS and SME deals, detailed postcode information is provided, while NUTS typically is not. The NUTS regions for Portugal, often do not match the postcode areas (postcode areas and NUTS tend to overlap). Thus, a mapping between NUTS codes and postcodes would be impractical. Nevertheless, all postcodes can be mapped to each one of the Portuguese "distritos" (used as Geo_1)

and to more granular administrative units (Geo_2), such as a municipality or town district, identifiable with the postcode data provided. Therefore, all Geo_2, can be mapped to a Geo_1 region. In cases where postcode information is not provided but NUTS 3 is available, and where NUTS and postcodes do not overlap, (as is the case for Algarve), it is possible to identify Geo_1 (district level).

United Kingdom

In the United Kingdom, the first letters of a postcode identify a postcode area centred around a town, but the boundaries of these postcode areas do not perfectly match the boundaries of the UK's regions (thus, NUTS 1 codes are used to generate Geo_1). For the Geo_2 fields, we use two sets of names, depending on data availability. For deals providing postcodes (RMBS, SME), Geo_2 corresponds to the postcode area names. For Deals for which NUTS is provided instead, Geo_2 names correspond to NUTS 2 names.[8] ED maps the postcode areas to the UK regions (Government Office Regions, used for NUTS 1), based on the name of the town around which the postcode area is centred. Thus, the area of Peterborough (PE, used as Geo_2) is split between the regions East of England and East Midlands, but we map it to East of England, because the city of Peterborough is in that region. In about a third of cases for UK loans, geographic information reported is insufficient to even identify Geo_1. For the loans for which sufficient data is available, ED finds that the distribution of the loan exposures per region broadly matches the contribution of these regions to British GDP (see Appendix 3).

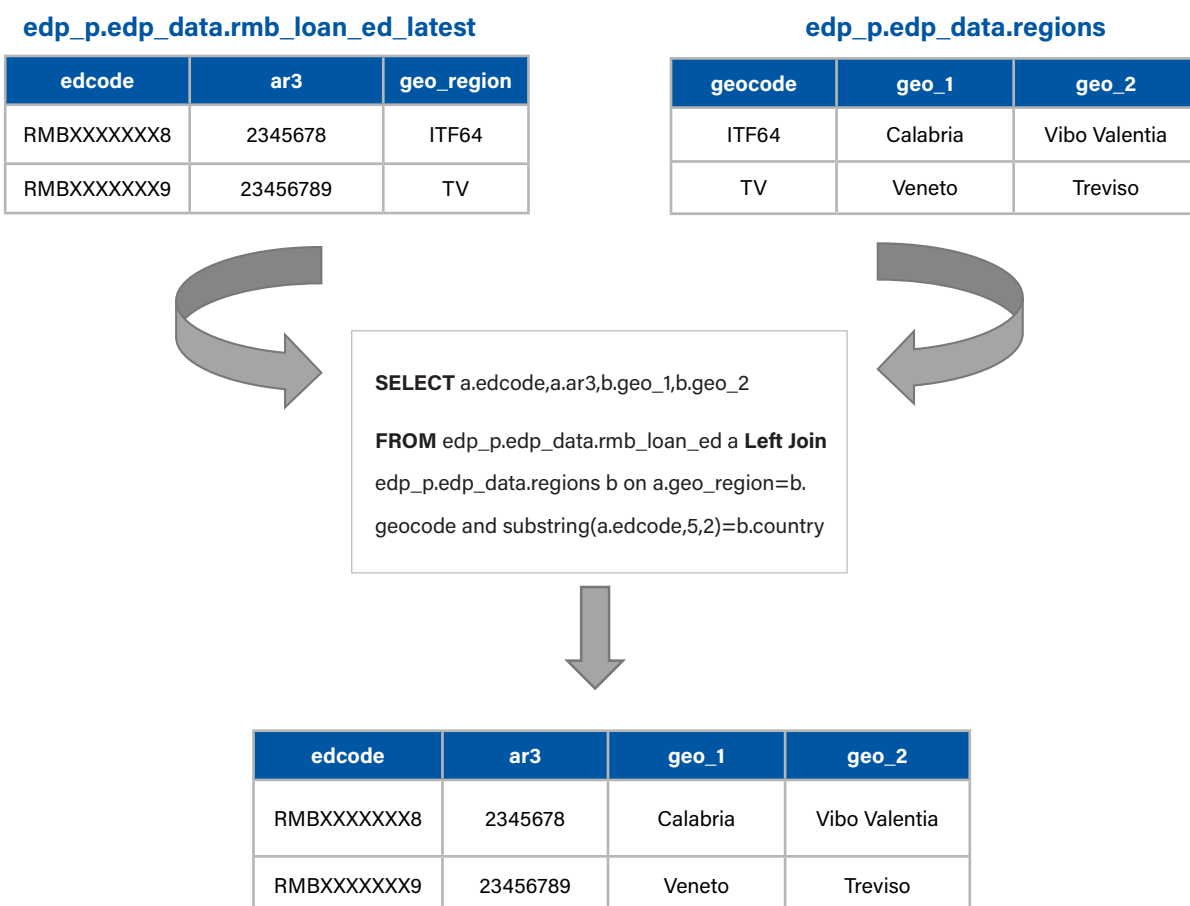## Austria, Finland, Ireland, Germany, Norway, Sweden

For these countries, the mapping remains based on the NUTS information. The deals for these countries are mostly auto deals, in which case the data template only asks for the NUTS geographic region. Regarding Irish RMBS deals, the field AR129 (postcode) is usually filled in with the actual county name (Geo_2), which ED maps to the NUTS 1 names (Geo_1). In the case of Germany, it is often not possible to identify with certainty the "Land" for a loan, based only on the first digits of its postcode, as these often point to several possible "Länder". Typically, either two or three digits are provided for German RMBS deals, but even four digits would not suffice to identify the Land with certainty. Thus, Geo fields are not generated based on postcode information for Germany and NUTS, when available, is used instead. In the case of Sweden, there is only one terminated RMBS deal, for which NUTS is available and used. For these countries, only the German auto sector and the Irish RMBS sector are shown in Appendix 3, given that for the other markets, too few deals are available to be really representative.

---

[8]ED noticed that in some cases, data providers report NUTS instead of the postcode in field AR129. Our mapping will still be able to determine Geo_1 in such cases, although this will separately be treated as a data quality issue. Cases where no geographic info is provided are also treated as data quality issues.

Appendix 1: Using ED's calculated fields Geo_1 and Geo_2

"The extra field geo_region is available in the stratification tables in the ED Cloud database,[9] combined with the mapping available in the table regions it is possible to identify geo_1 and geo_2 regions. The example below shows how to extract these two fields for Italian RMBS deals.[10]

In the tables "edp_p.edp_data.<asset_class>_loan_ed" and "edp_p.edp_data .<asset_class>_loan_ed_latest", the field geo_region is now available containing a key that can be used to join both tables with the "edp_p.edp_data.regions" table. By joining the tables, (using geocode and country fields) users are able to identify the region names.

edp_p.edp_data.rmb_loan_ed_latest

edcode	ar3	geo_region
RMBXXXXXXX8	2345678	ITF64
RMBXXXXXXX9	23456789	TV

edp_p.edp_data.regions

geocode	geo_1	geo_2
ITF64	Calabria	Vibo Valentia
TV	Veneto	Treviso

SELECT a.edcode,a.ar3,b.geo_1,b.geo_2

FROM edp_p.edp_data.rmb_loan_ed a **Left Join**
edp_p.edp_data.regions b on a.geo_region=b.
geocode and substring(a.edcode,5,2)=b.country

edcode	ar3	geo_1	geo_2
RMBXXXXXXX8	2345678	Calabria	Vibo Valentia
RMBXXXXXXX9	23456789	Veneto	Treviso

---

[9]See https://eurodw.eu/ed-cloud-pro
[10]The following query will provide ge3o_1 and Geo_2 for deal RMBSITXXX:
  SELECT a.edcode,a.ar3,b.geo_1,b.geo_2 **FROM** edp_p.edp_data.rmb_loan_ed_latest a Left Join edp_p.edp_data.regions b on a.geo_region=b.
  geocode and substring(a.edcode,5,2)=b.country where edcode = 'RMBSITXXX'

Appendix 2: ECB template geographic disclosure requirements[11]

ECB Loan Level Data - Reporting Template for RMBS - June 2013

AR128	Optional	Static	Geographic Region List	List	The region description of where the property is located. See Taxonomy for relevant choices.
AR129	Mandatory	Static	Property Postcode	Text/ Numeric	First 2 or 3 characters must be provided at a minimum. See Taxonomy for relevant choices.

ECB Loan Level Data - Reporting Template for SME - December 2012

AS16	Mandatory	Static	Postcode	Text	First 2 or 3 characters must be provided at a minimum. Do not supply the full postcode. See Taxonomy for relevant choices.
AS17	Optional	Static	Geographic Region	List	The region description of where the obligor is located based on the Nomenclature of Territorial Units for Statistics (NUTS) using regional coding format (NUTS2). Refer to taxonomy for guidance.

ECB Loan Level Data - Reporting Template for Leasing ABS - September 2013

AL16	Optional	Static	Postcode	Text/ Numeric	Lessee's postcode. First 2 - 4 characters should be provided only. Do not supply the full postcode. All 'No Data' options may be used in this optional field.
AL17	Mandatory	Static	Geographic Region	List	The region where the obligor is located as at underwriting. NUTS3 2006 region coding to be used. See "Geographic Region List" in taxonomy for relevant choices. All 'No Data' options may be used in this field.

ECB Loan Level Data - Reporting Template for Auto ABS - September 2013

AA21	Mandatory	Static	Geographic Region	List	The region where the borrower is located as at underwriting. NUTS3 2006 region coding to be used. See "Geographic Region List" in taxonomy for relevant choices. All 'No Data' options may be used in this field.

Loan Level Data - Reporting Template for Consumer Finance ABS - September 2013

AN20	Mandatory	Static	Geographic Region	List	The region where the borrower is located. NUTS3 2006 classification to be used. See "Geographic Region List" in taxonomy for relevant choices. All 'No Data' options may be used in this field.

---

[11] See www.ecb.europa.eu/paym/coll/loanlevel/transmission/html/index.en.html for individual taxonomies.
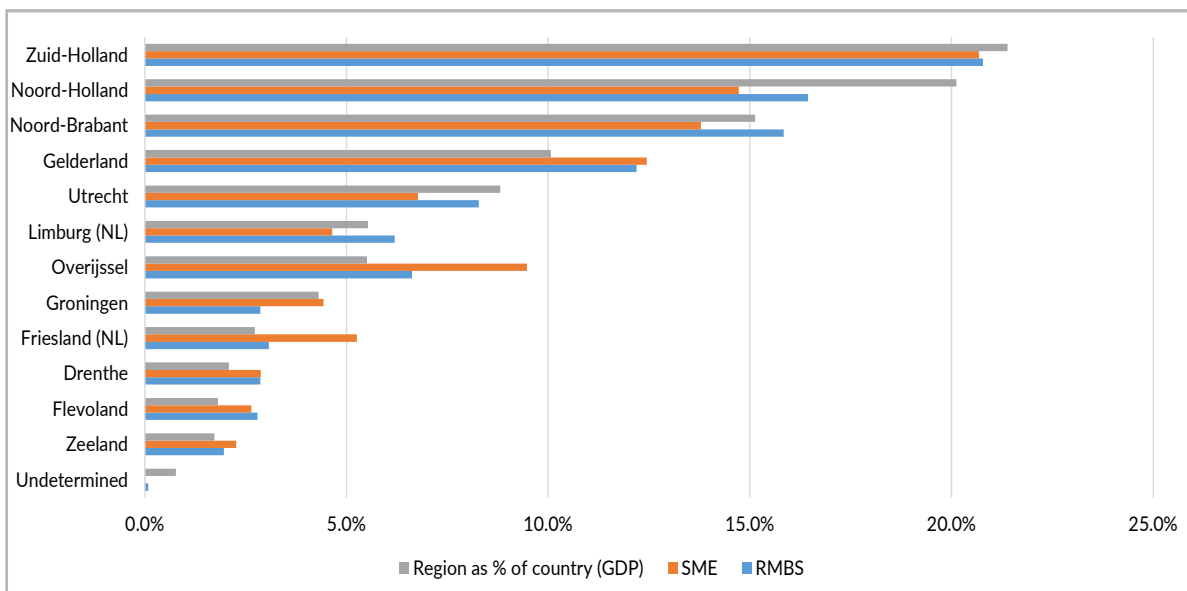
Appendix 3: Distribution of loans per country

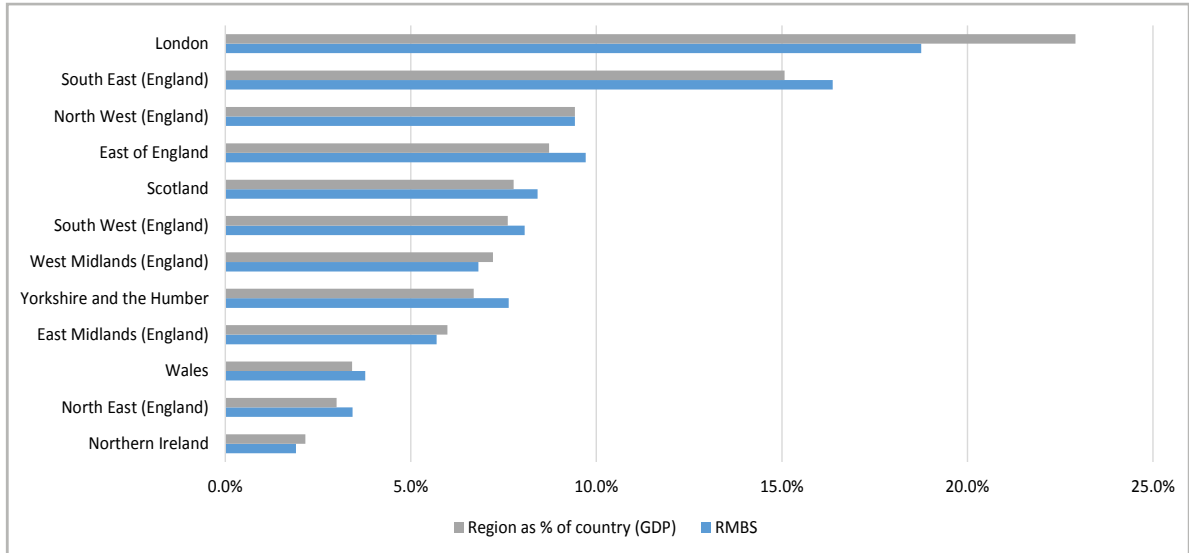### Exhibit 5a: Loan exposure compared with GDP per region for Belgium
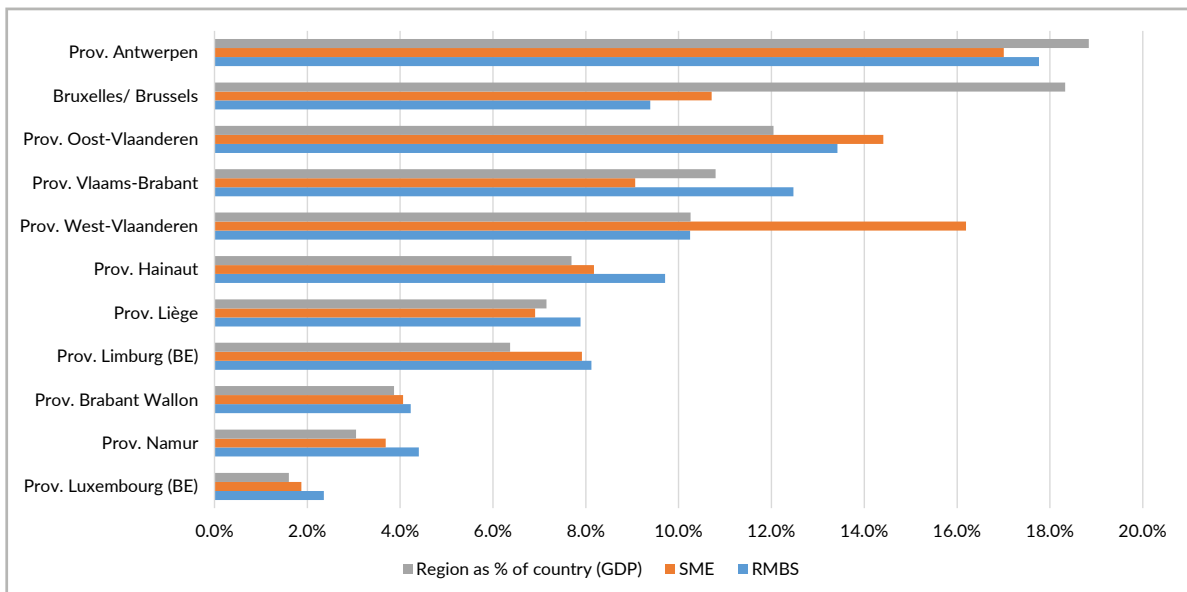
Source: European DataWarehouse, data providers; Eurostat

Exhibit 5b: Loan exposure compared with GDP per region for the Netherlands
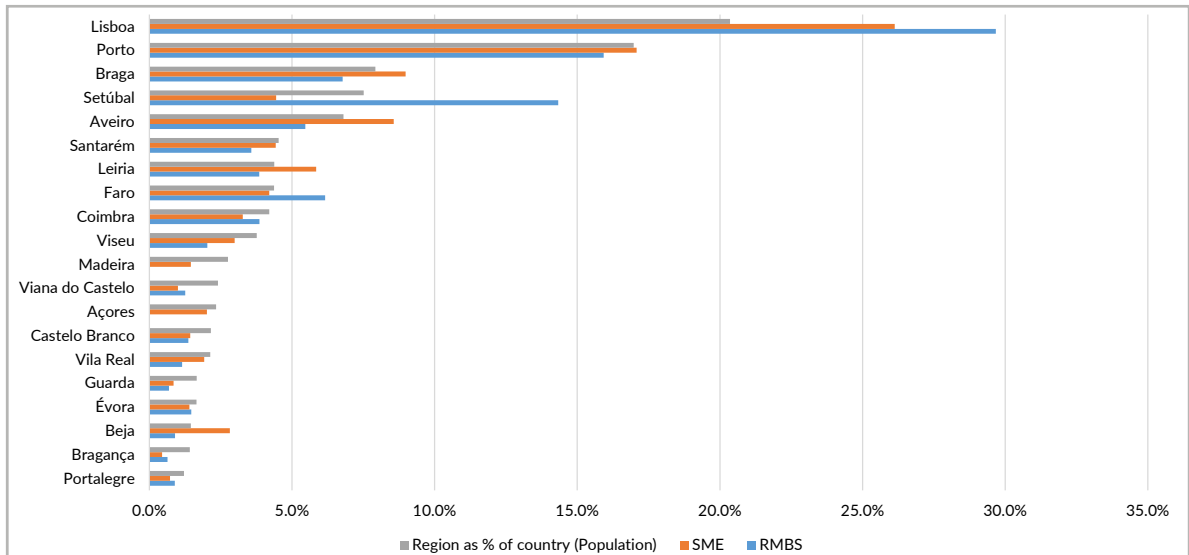

Source: European DataWarehouse, data providers; Eurostat

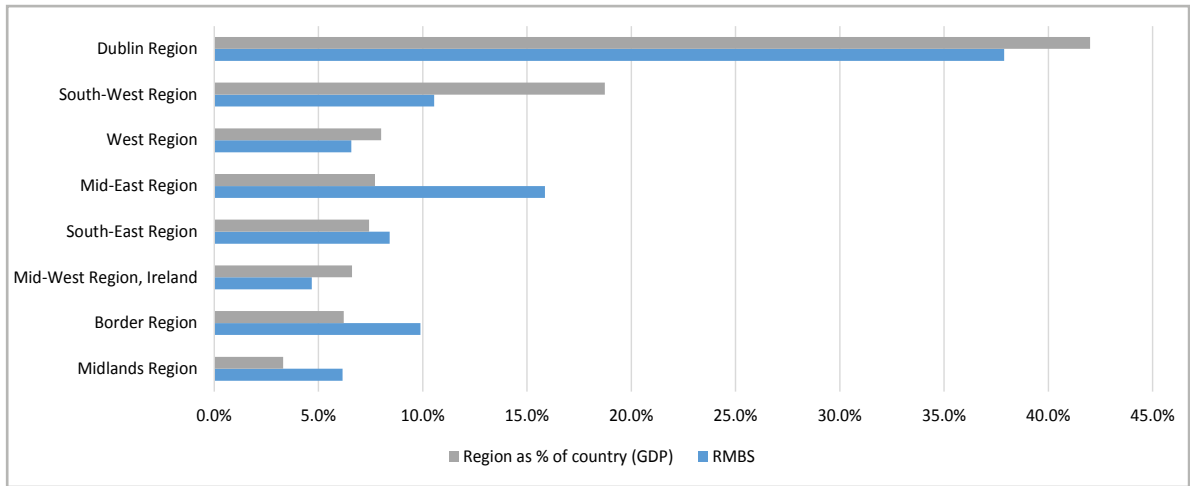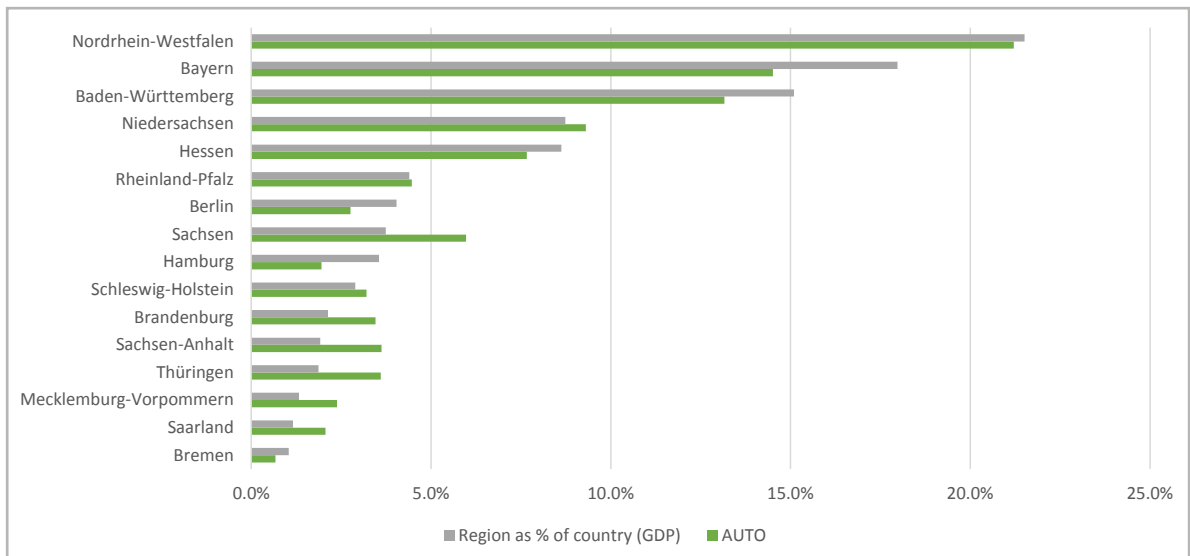## Exhibit 5c: Loan exposure compared with GDP per region for the United Kingdom

Source:European DataWarehouse, data providers; Eurostat

Exhibit 5d: Loan exposure compared with GDP per region for Portugal

Source:European DataWarehouse, data providers; Eurostat

Exhibit 5e: Loan exposure compared with GDP per region for Ireland (RMBS only)

Source: European DataWarehouse, data providers; Central Statistics Office (http://www.cso.ie)

Exhibit 5f: Loan exposure compared with GDP per region for Germany (Auto only)

Source: European DataWarehouse, data providers; Eurostat