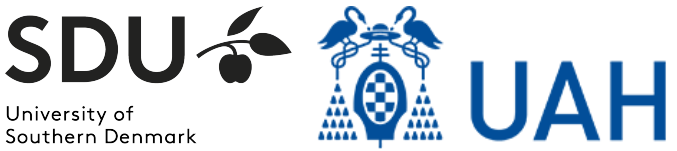
Putting the Periphery on the Map: Tourism Activity measured with Big Data

Karol J. Borowiecki Maja U. Pedersen Marco Palomeque

University of Southern Denmark & Universidad de Alcalá

April 11, 2024



Big Data measurement

Figure: Differences between a case study and a Big Data study

Campina de Faro (Portugal)

AI collage on European periphery

Outline

- Motivation and contributions
- Literature review
- Data description
- TripAdvisor data validation
- Study of peripheral tourism
- Conclusion

Motivation: sustainability and redistribution

Tourism challenges:

- **Sustainability:** tourism accounted for 8% of global green house emissions (Lenzen et al., 2018) and increasing (World Tourism Organization, 2019);
- **Redistribution:** overtourism has negative external effects on the environment and the local population (Bobic and Akhavan, 2022).

Motivation: sustainability and redistribution

Tourism challenges:

- **Sustainability:** tourism accounted for 8% of global green house emissions (Lenzen et al., 2018) and increasing (World Tourism Organization, 2019);
- **Redistribution:** overtourism has negative external effects on the environment and the local population (Bobic and Akhavan, 2022).

Possible solution, enhance peripheral tourism:

- Promotion of more **sustainable tourism** practices;
- Preservation of **cultural diversity**;
- **Economic development** of disadvantaged areas (Vu and Turner, 2009).

Motivation: sustainability and redistribution

Tourism challenges:

- **Sustainability:** tourism accounted for 8% of global green house emissions (Lenzen et al., 2018) and increasing (World Tourism Organization, 2019);
- **Redistribution:** overtourism has negative external effects on the environment and the local population (Bobic and Akhavan, 2022).

Possible solution, enhance peripheral tourism:

- Promotion of more **sustainable tourism** practices;
- Preservation of **cultural diversity**;
- **Economic development** of disadvantaged areas (Vu and Turner, 2009).

Difficulties:

- Poor infrastructure and aging population constitute a **challenge for policy makers**;
- Quantity measures are largely **unavailable**.

Motivation: sustainability and redistribution

Tourism challenges:

- **Sustainability:** tourism accounted for 8% of global green house emissions (Lenzen et al., 2018) and increasing (World Tourism Organization, 2019);
- **Redistribution:** overtourism has negative external effects on the environment and the local population (Bobic and Akhavan, 2022).

Possible solution, enhance peripheral tourism:

- Promotion of more **sustainable tourism** practices;
- Preservation of **cultural diversity**;
- **Economic development** of disadvantaged areas (Vu and Turner, 2009).

Difficulties:

- Poor infrastructure and aging population constitute a **challenge for policy makers**;
- Quantity measures are largely **unavailable**.
  - $\Rightarrow$ We propose a new approach to create a big database with detailed information about both the tourists and the attractions in the periphery.

Motivation: methodological approach

International tourism statistics are notorious for being:

1) Over-aggregated, usually at the country level

2) No information about the tourist

3) Available with a lag of at least several months

4) Available only at best monthly

5) Not comparable peripheral data

Motivation: methodological approach

International tourism statistics are notorious for being:

1) Over-aggregated, usually at the country level
   Our data is at the attraction-level!
2) No information about the tourist
   Our data contains a treasure of information about the tourist!
3) Available with a lag of at least several months
   Our data is close to real-time!
4) Available only at best monthly
   Our data is daily!
5) Not comparable peripheral data
   Our data is objective and allows international comparisons!

Motivation: methodological approach

International tourism statistics are notorious for being:

1) Over-aggregated, usually at the country level
   Our data is at the attraction-level!
2) No information about the tourist
   Our data contains a treasure of information about the tourist!
3) Available with a lag of at least several months
   Our data is close to real-time!
4) Available only at best monthly
   Our data is daily!
5) Not comparable peripheral data
   Our data is objective and allows international comparisons!

We validate the data and show one application using the peripheral regions from the INCULTUM project.

Contributions

This research provides two sets of contributions:

- **Methodology contributions:**
  - $\Rightarrow$ We demonstrate and validate the possibility to construct a large database on tourism activity based on alternative sources;
  - $\Rightarrow$ We obtain comparable international tourism data of peripheral areas, mostly unavailable before;
  - $\Rightarrow$ We illustrate tourism mobility with unprecedented depth and precision.

Contributions

This research provides two sets of contributions:

- **Methodology contributions:**
  - ⇒ We demonstrate and validate the possibility to construct a large database on tourism activity based on alternative sources;
  - ⇒ We obtain comparable international tourism data of peripheral areas, mostly unavailable before;
  - ⇒ We illustrate tourism mobility with unprecedented depth and precision.
- **Novel insights on peripheral tourism:**
  - ⇒ Increasing interest in these regions, characterized by unique cultural and natural attractions;
  - ⇒ Notable shift towards domestic and European tourism.

Literature review

We contribute to different strands of the literature.

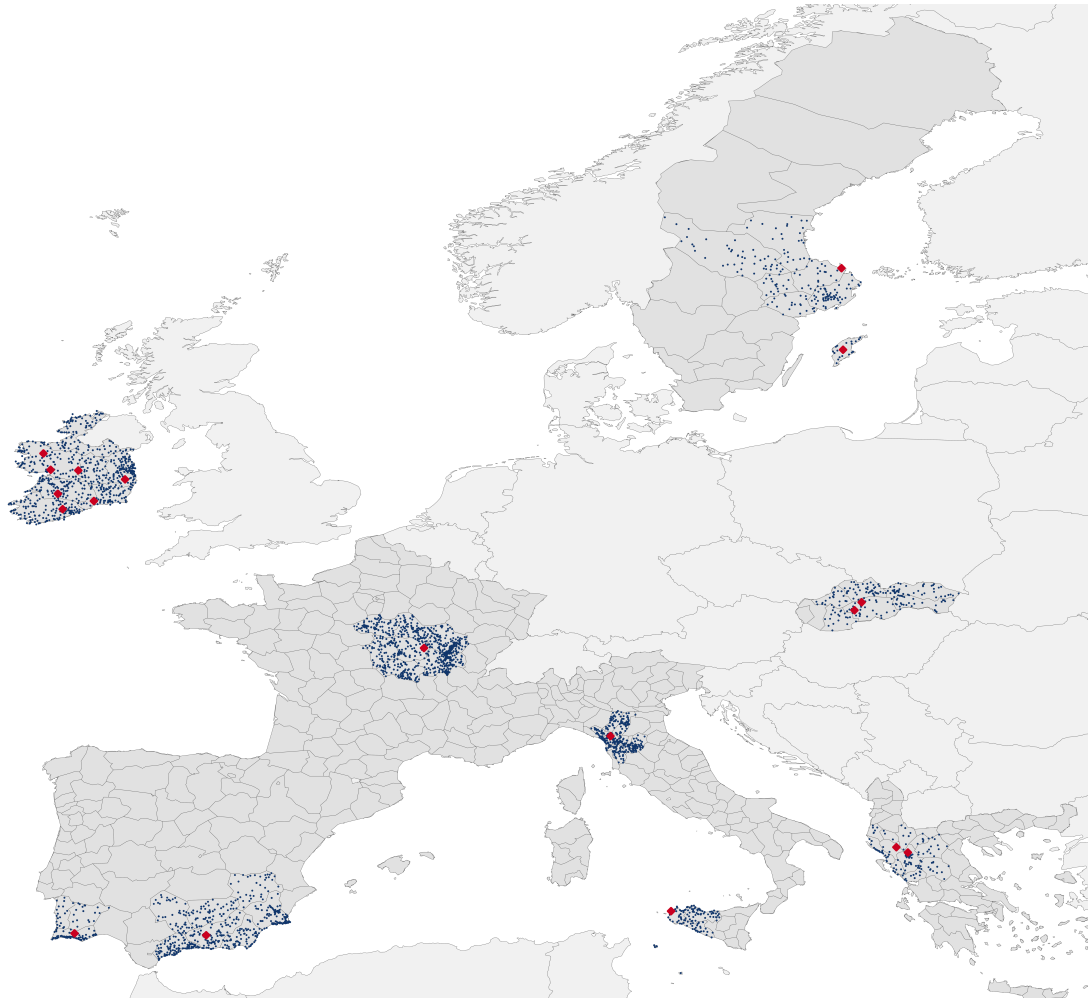
- Rural and nature tourism:
  - ⇒ Leick et al. (2024); Ye et al. (2019); Komppula (2005); Molera and Pilar Albaladejo (2007), Bel et al. (2015); Frochot (2005); Calero and Turner (2020).
- Cultural tourism:
  - ⇒ Borowiecki and Castiglione (2014); Bertacchini et al. (2021); Cui et al. (2024); Brandano and Meleddu (2021); Brandano and Crociata (2023); Panzer-Krause (2019).
- Big data tourism measurement techniques:
  - ⇒ Borowiecki et al. (2024); Leick et al. (2024); Mahat and Hanafiah (2020).
- Social media influence on tourism:
  - ⇒ Chen and Chang (2018); Cheng and Jin (2019); Falk et al. (2019); Sainaghi (2020); Martin-Fuentes et al. (2020); Jean et al. (2019); Leick et al. (2022); Kim et al. (2017).
- Impact of Covid-19 on rural tourism:
  - ⇒ Gil-Alana and Poza (2022); Plzáková and Smeral (2022); Wang et al. (2022); Vaishar and Šťastná (2022); Seraphin and Dosquet (2020).

Data description: Peripheral regions

	Country	Location as described by pilot	NUTS3 region name	NUTS3 region code	NUTS 2 region name	NUTS 2 region code
1	Spain	The Altiplano	Granada	ES614	Andalusia	ES61
2	Portugal	Campina de Faro	Algarve	PT150	Algarve	PT15
3	Slovakia	Banská Bystrica, Banská Štiavnica	Banskobystrický kraj	SK032	Central Slovakia	SK03
4	Italy	Monti di Trapani, Calatafimi-Segesta, Custonaci, Buseto Palizzolo	Trapani	ITG11	Sicilia	ITG1
5	Italy	San Pellegrino, Alpe, Tuscan-Emilian Appennines	Modena Lucca	ITH54 ITI12	Emilia-Romagna Toscana	ITH5 ITI1
6	France	Regional Natural Park	Nièvre	FRC12	Bourgogne	FRC1
7	Greece	Aaos Valley, Konitsa	Ionnina	EL543	Epirus	EL54
8	Albania	Upper Vjosa Valley, Përmet	Gjirokastër	AL033	Southern Albania	AL03
9	Ireland	County Mayo County Galway County Limerick County Cork County Waterford County Wicklow	West Region West Region Mid-West Region South-West Region South-East Region Mid-East Region	IE042 IE042 IE051 IE053 IE052 IE062	Northern and Western Region Northern and Western Region Southern Region Southern Region Southern Region Eastern and Midland Region	IE04 IE04 IE05 IE05 IE05 IE06
10	Sweden	Gotland Öregrund	Gotlands län Uppsala län	SE214 SE121	Småland and the Islands East Middle Sweden	SE21 SE12

Data description: Peripheral regions

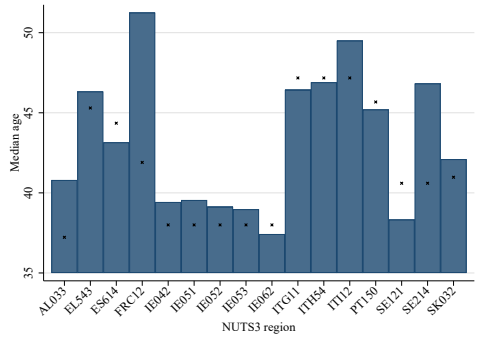
Figure: Location of attractions in peripheral and neighboring regions



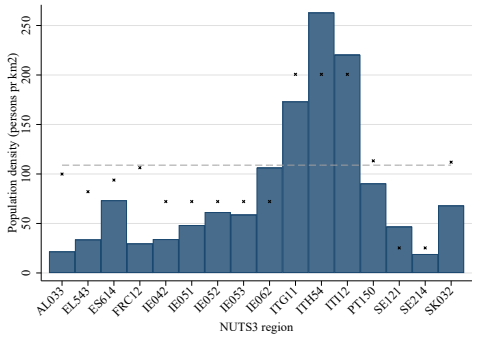
Notes: This Figure shows the attractions (blue) and pilot site (red) of peripheral and neighboring regions. *Source:* Own data collected from Tripadvisor.

Data description: Peripheral regions

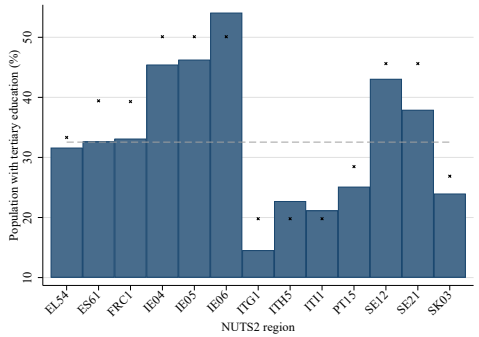
Figure: Indicators in peripheral regions



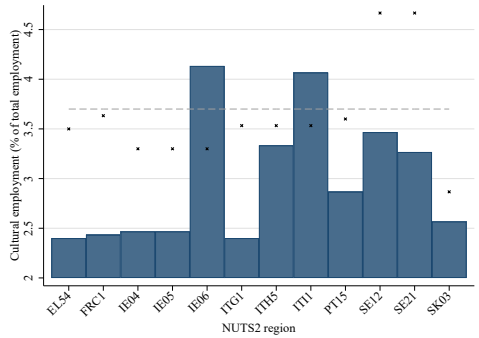
Median age



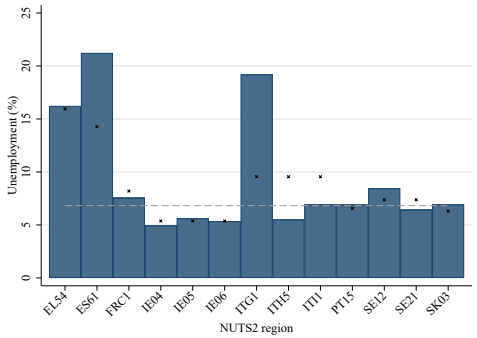
Population density



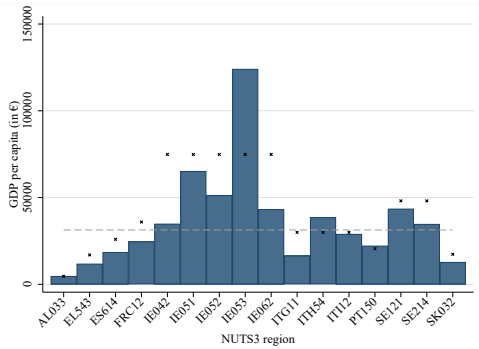
Tertiary education



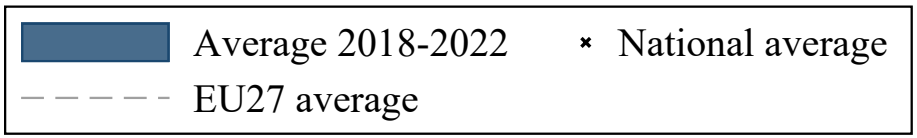
Cultural employment



Unemployment rates



GDP per capita

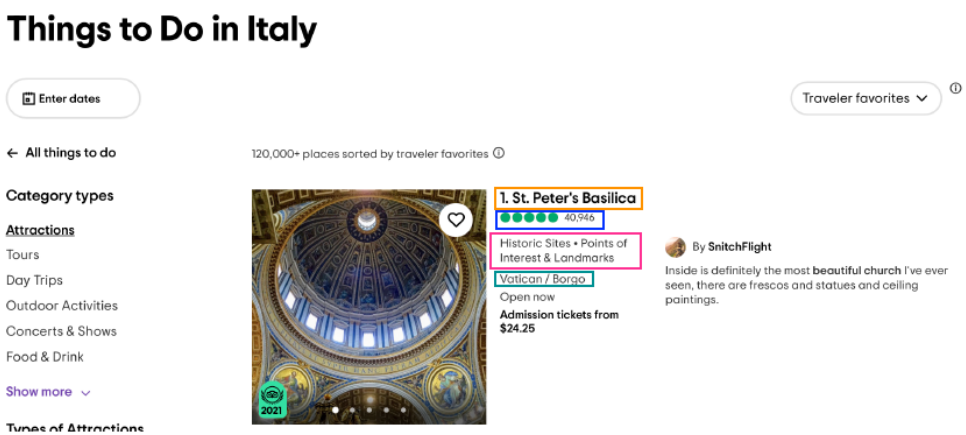


Data description: Tripadvisor data

Figure: Data collected from Tripadvisor

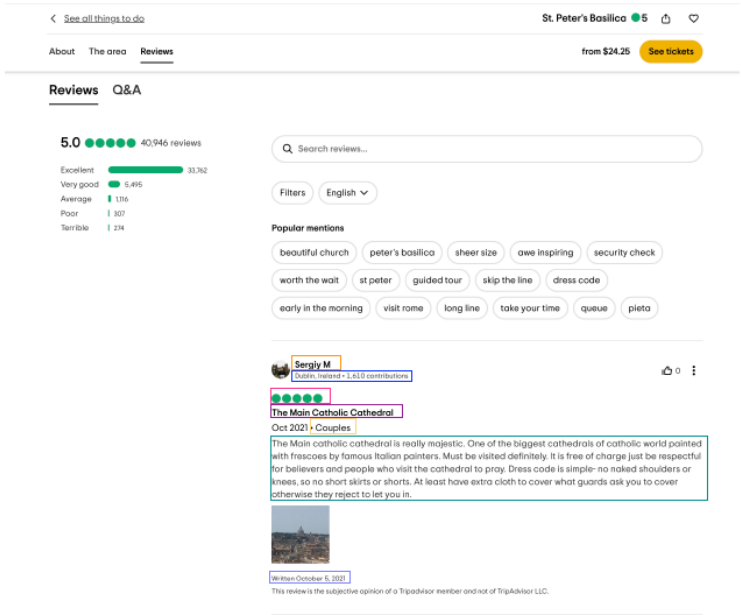
Data to collect — How it appears in UI

- Name of attraction
- Rank number
- Hyperlink to attraction reviews page
- Rating
- Number of reviews
- Type of attraction
- Location of attraction



Data to collect — How it appears in UI

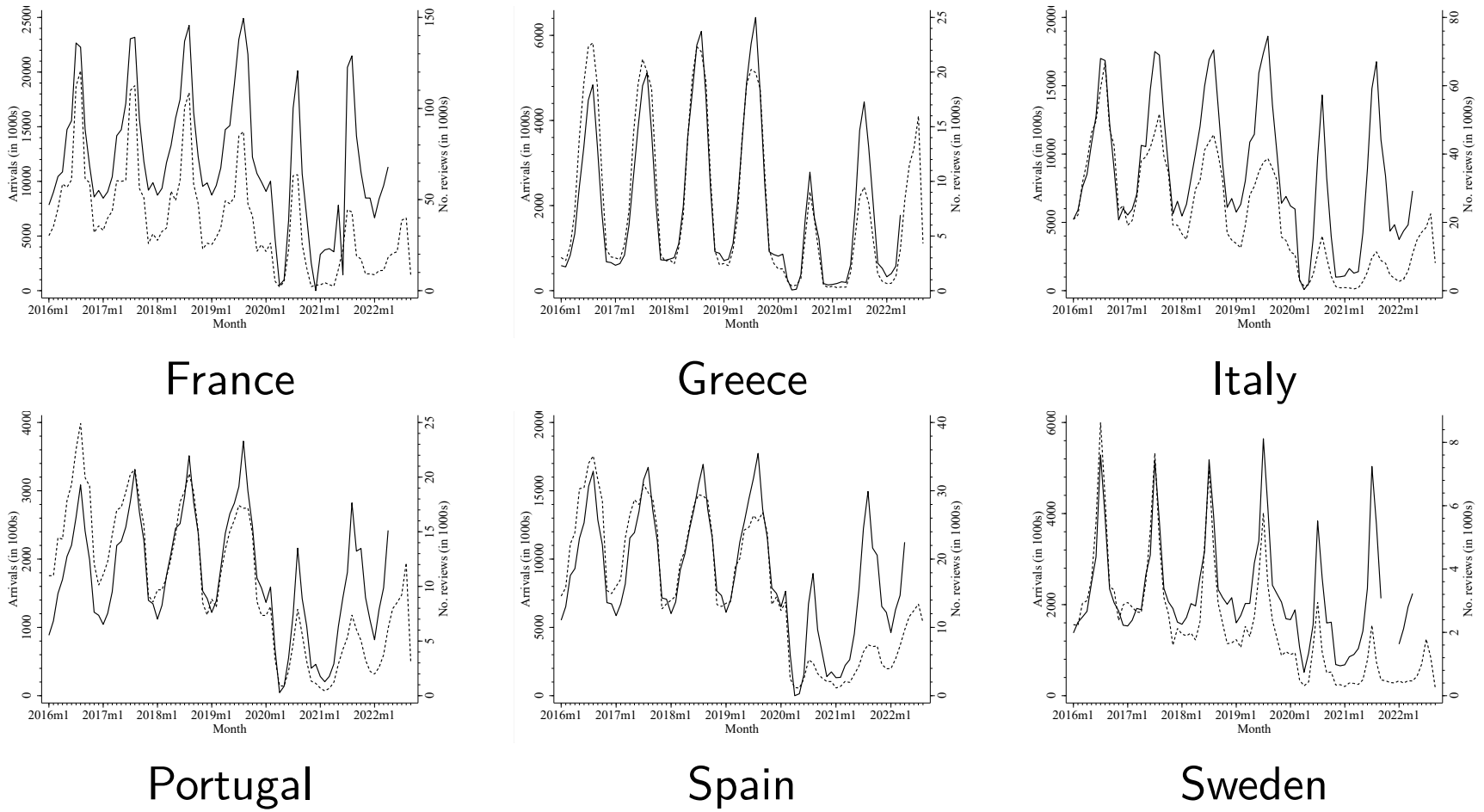
- User name
- Hyperlink to user profile page
- User location
- Number of contributions
- Rating
- Title of review
- Review text
- Type of tourist
- Date review was written
- Date of visit / date of experience (if available)
  - Note: this is available on some reviews, I couldn't find an example on this particular page but you can see an example in the user profile section of this document



Notes: Example of the information collected from Tripadvisor regarding the attraction and the reviews. *Source:* Tripadvisor

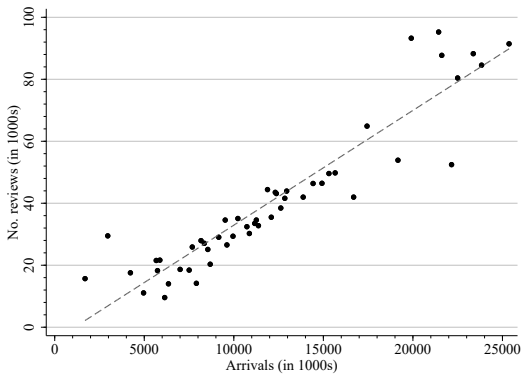
Validating the Tripadvisor data

Figure: Total number of arrivals and reviews over time

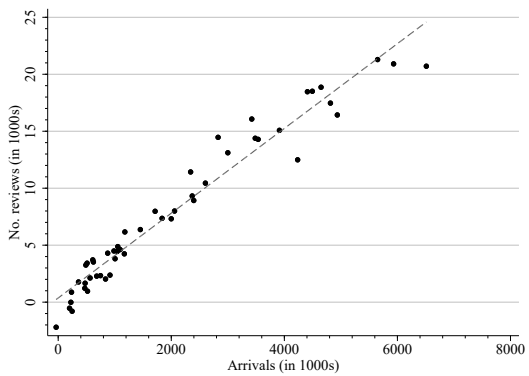


Validating the Tripadvisor data

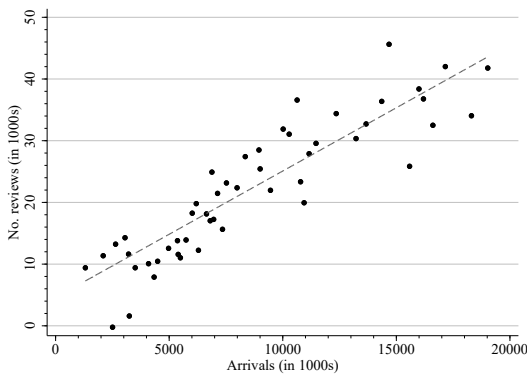
Figure: Monthly correlation between tourist arrivals and number of reviews



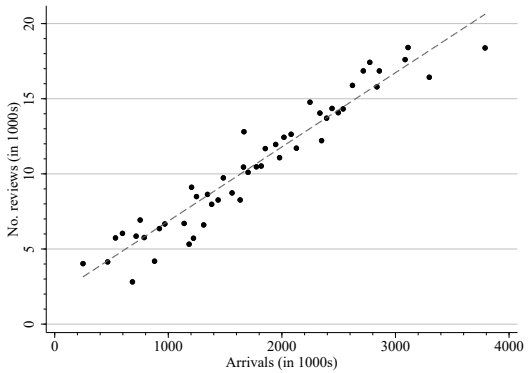
France



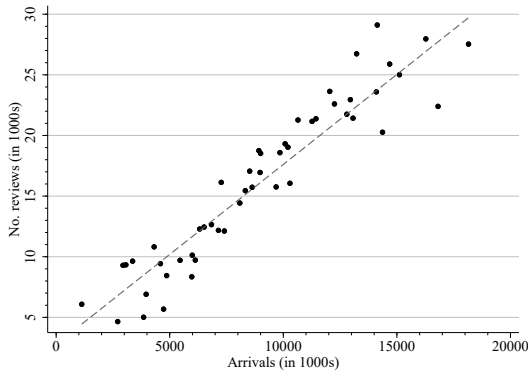
Greece



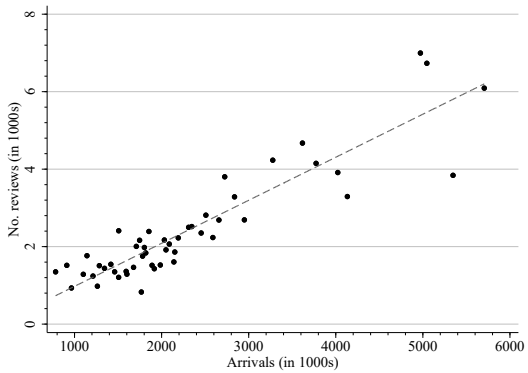
Italy



Portugal



Spain



Sweden

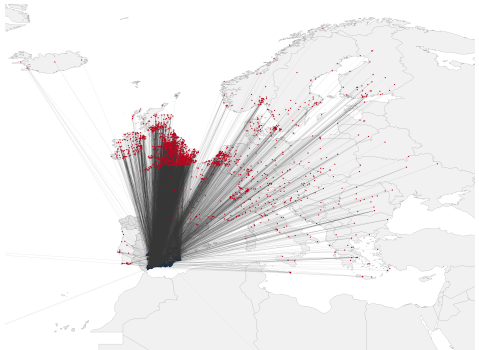
Outline

- Motivation and contributions
- Literature review
- Data description
- TripAdvisor data validation
- **Study of peripheral tourism**
- Conclusion

Travel patterns

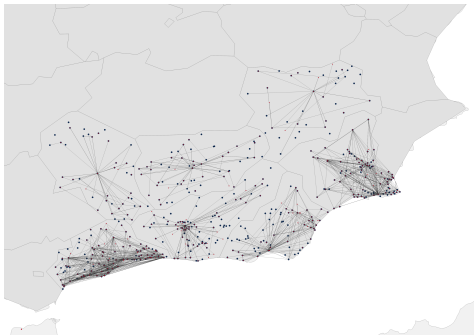
Figure: Travel patterns of visitors to peripheral and neighboring regions - Spain

World



Europe

Domestic

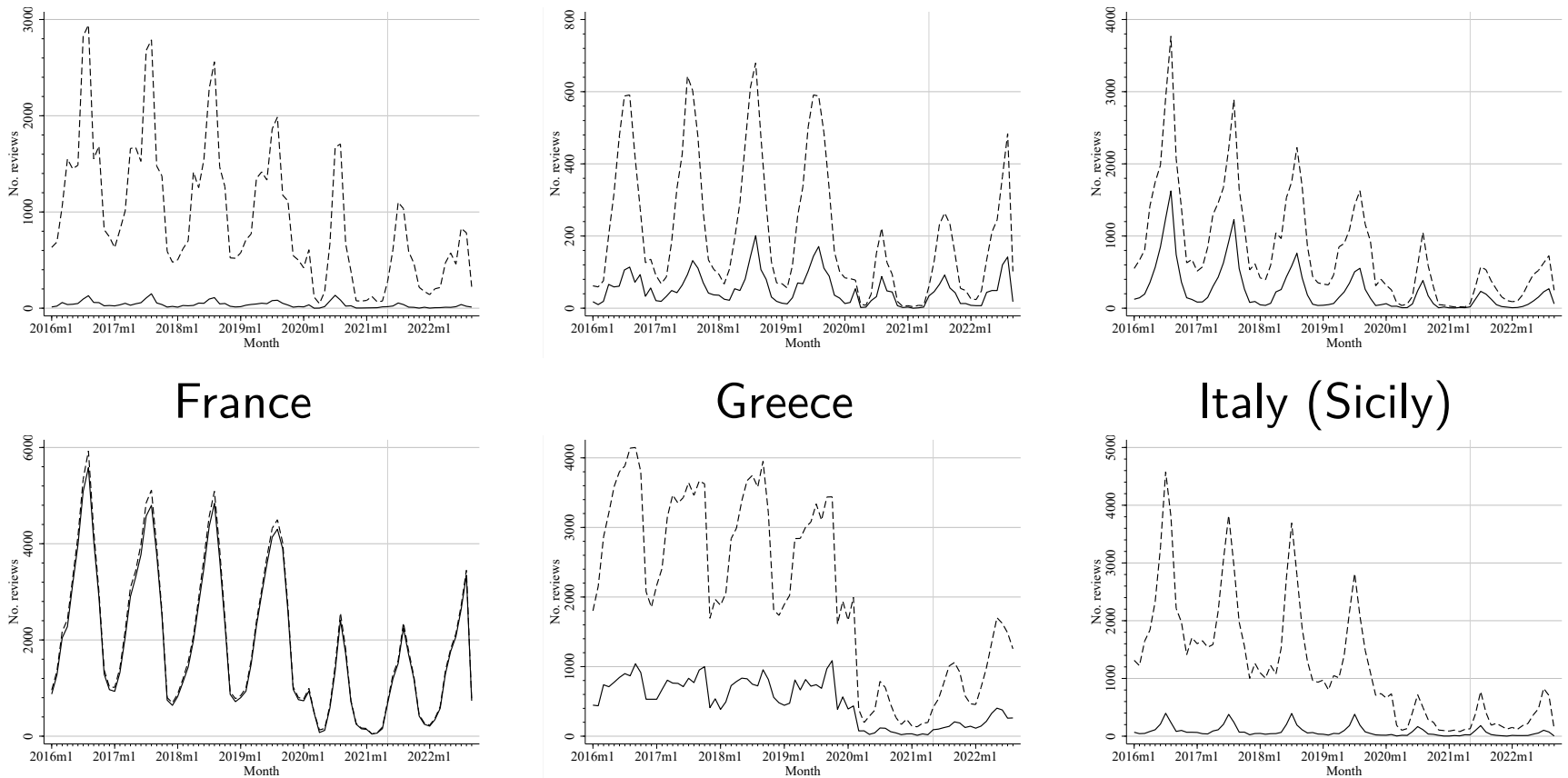


Local

Notes: The red dots represent the location of users and the blue dots the location of attractions *Source:* Own data collected from Tripadvisor.

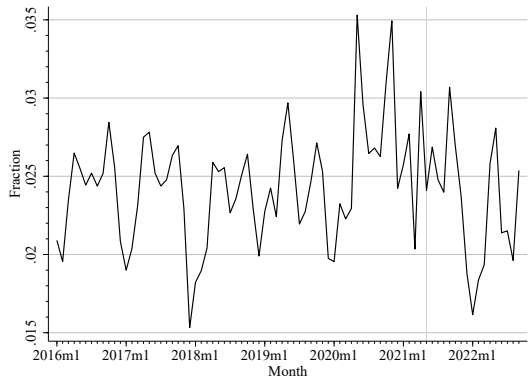
Time trends of tourism activity

Figure: Number of reviews over time for peripheral and neighboring regions

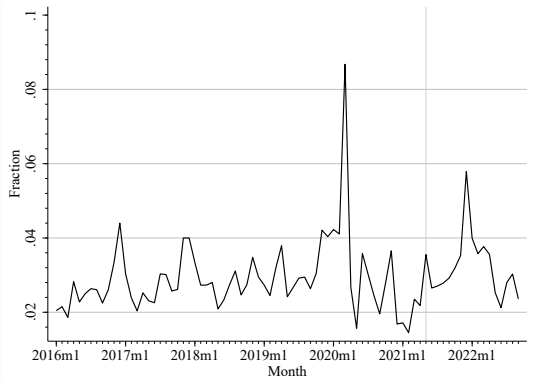


Time trends of tourism activity

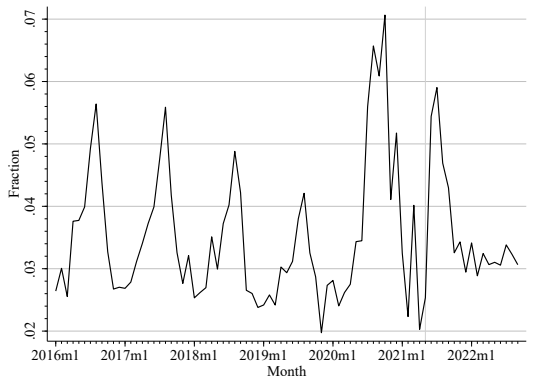
Figure: Fraction of the number of reviews in peripheral and neighboring regions out of all reviews in country over time



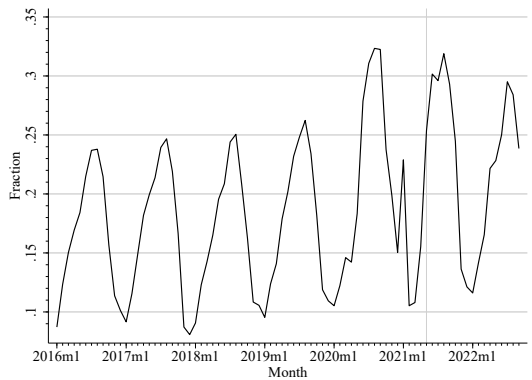
France



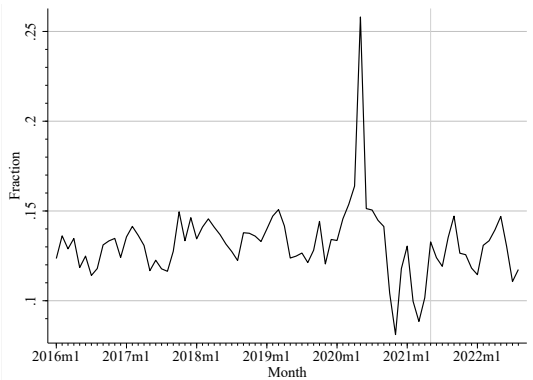
Greece



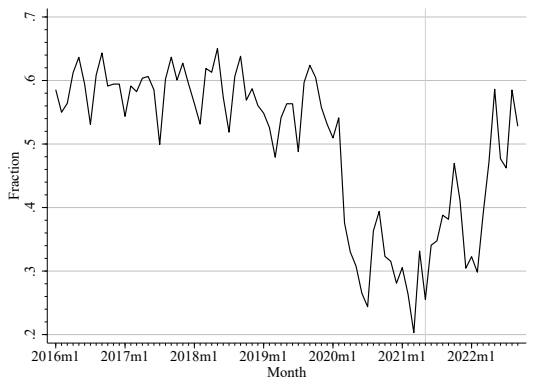
Italy (Sicily)



Portugal



Spain

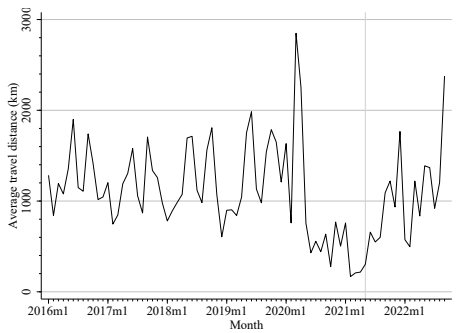


Sweden

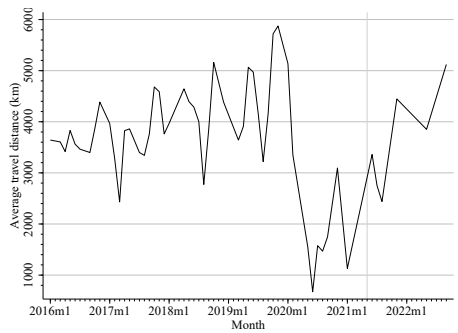
Note: The vertical line indicates when the INCULTUM pilot started. *Source:* Own data collected from Tripadvisor.

Time trends of tourism activity

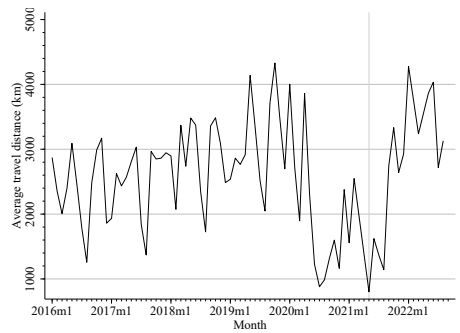
Figure: Distance traveled over time in peripheral and neighboring regions



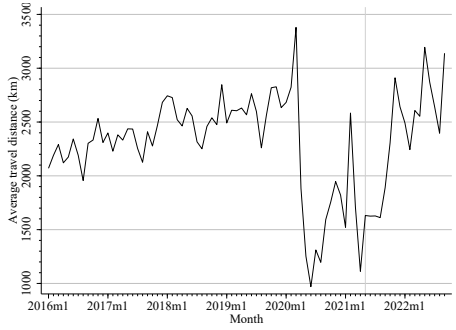
France



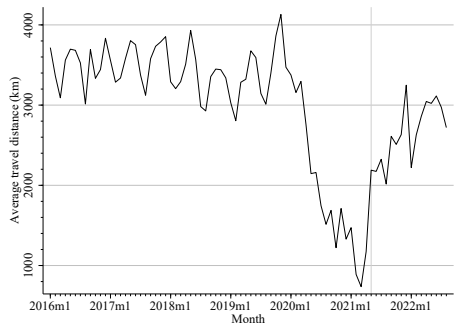
Greece



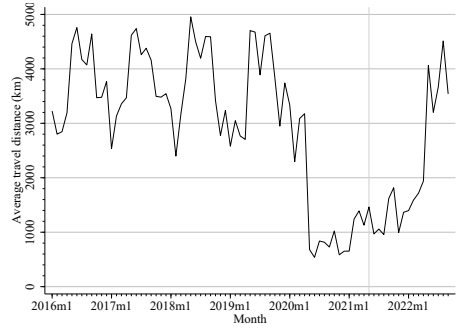
Italy (Sicily)



Portugal



Spain

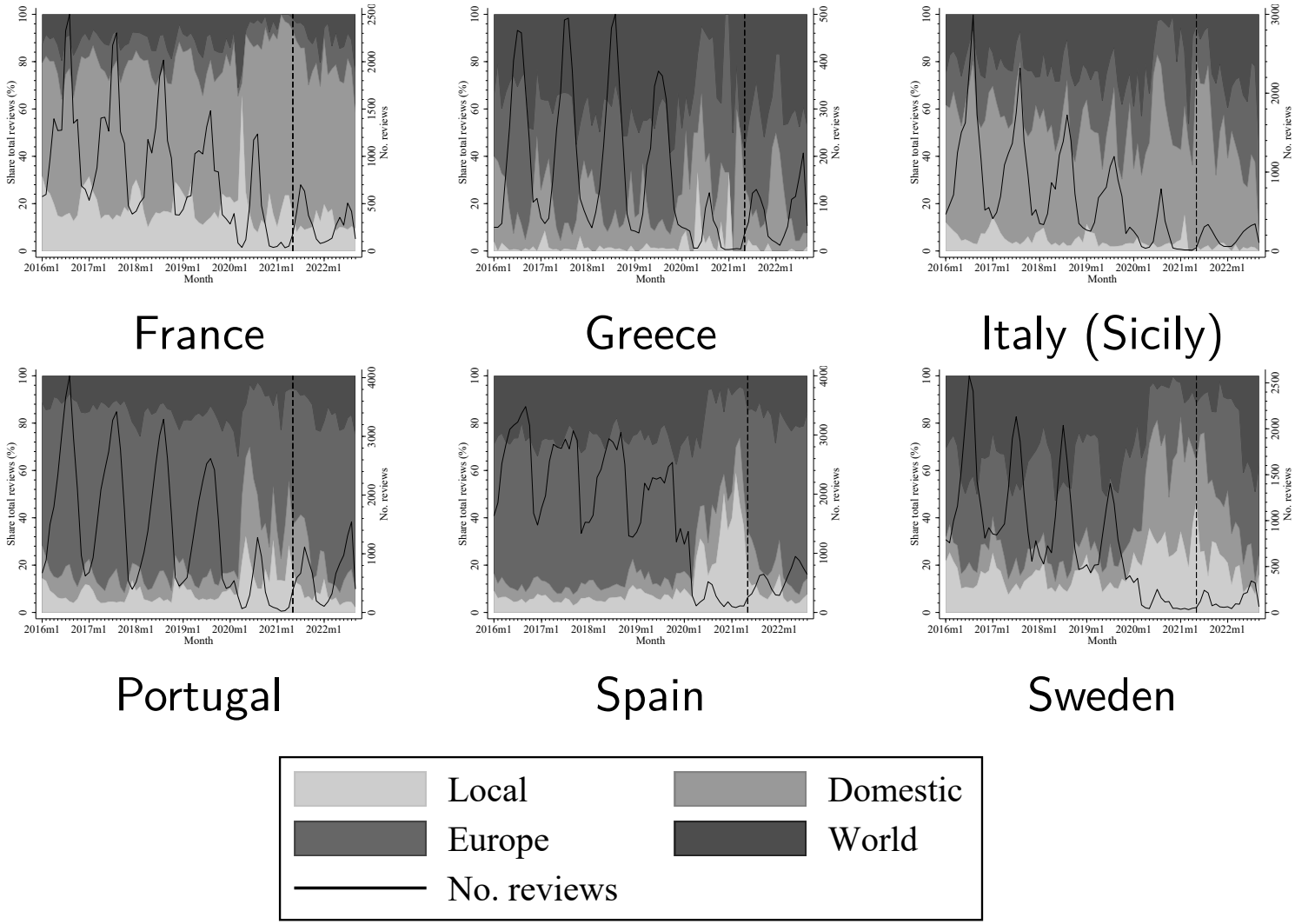


Sweden

Note: The vertical line indicates when the INCULTUM pilot started. *Source:* Own data collected from Tripadvisor.

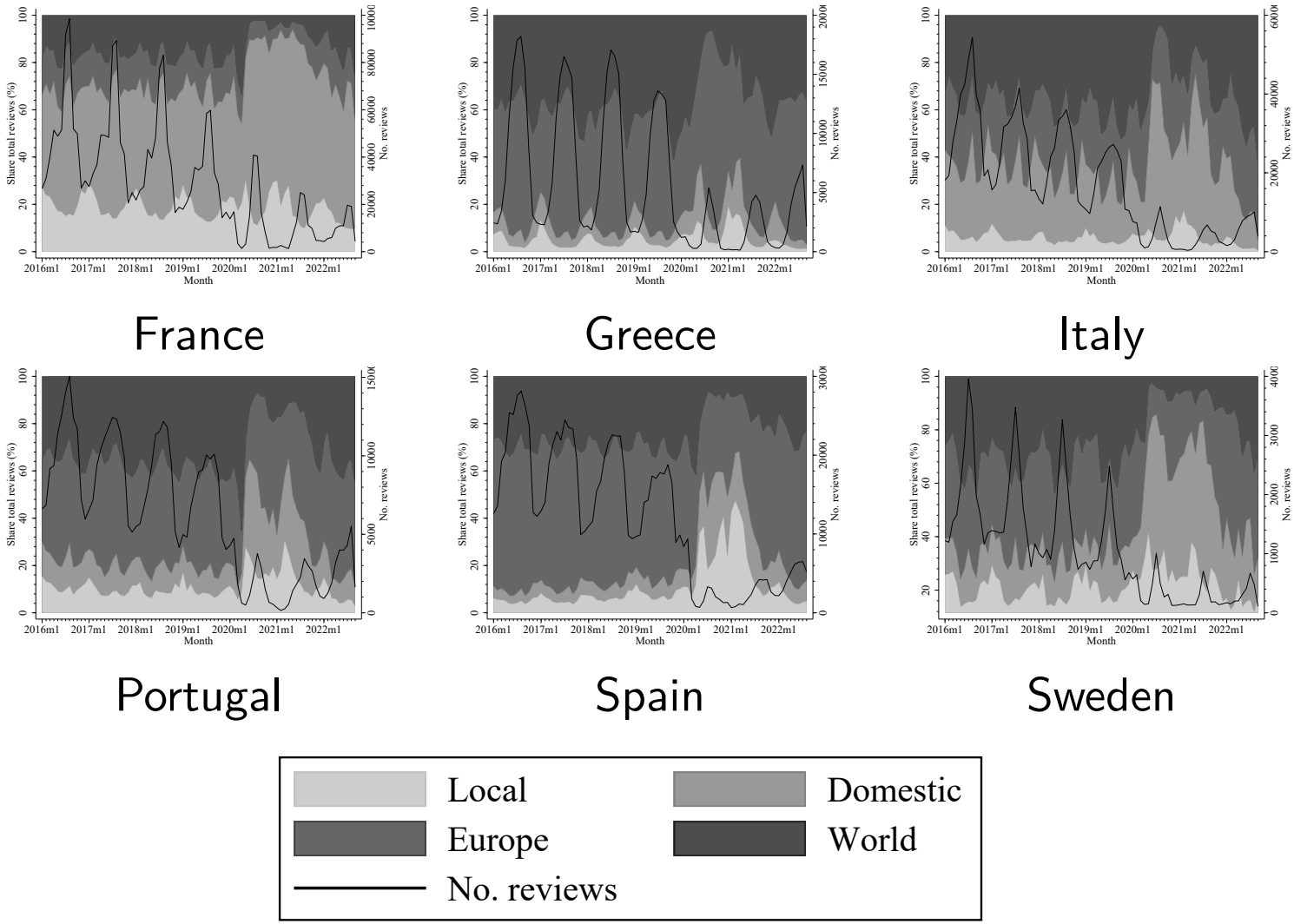
Time trends of tourism activity

Figure: Share of reviews for different travel categories over time for peripheral and neighboring regions



Time trends of tourism activity

Figure: Share of reviews for different travel categories over time for national level



Conclusion

- **We have demonstrated and validated the use of Tripadvisor. Advantages: daily availability, granularity, user data, and comparable and objective data.**

Conclusion

- We have demonstrated and validated the use of Tripadvisor. Advantages: daily availability, granularity, user data, and comparable and objective data.
- **We showed a resilient interest in the peripheral regions despite the COVID-19 pandemic and increasing visits from both their respective countries and across Europe.**

Conclusion

- We have demonstrated and validated the use of Tripadvisor. Advantages: daily availability, granularity, user data, and comparable and objective data.
- We showed a resilient interest in the peripheral regions despite the COVID-19 pandemic and increasing visits from both their respective countries and across Europe.
- **Our methodology allowed us to measure peripheral tourism in a wide and diverse set of countries across Europe with consistent results.**

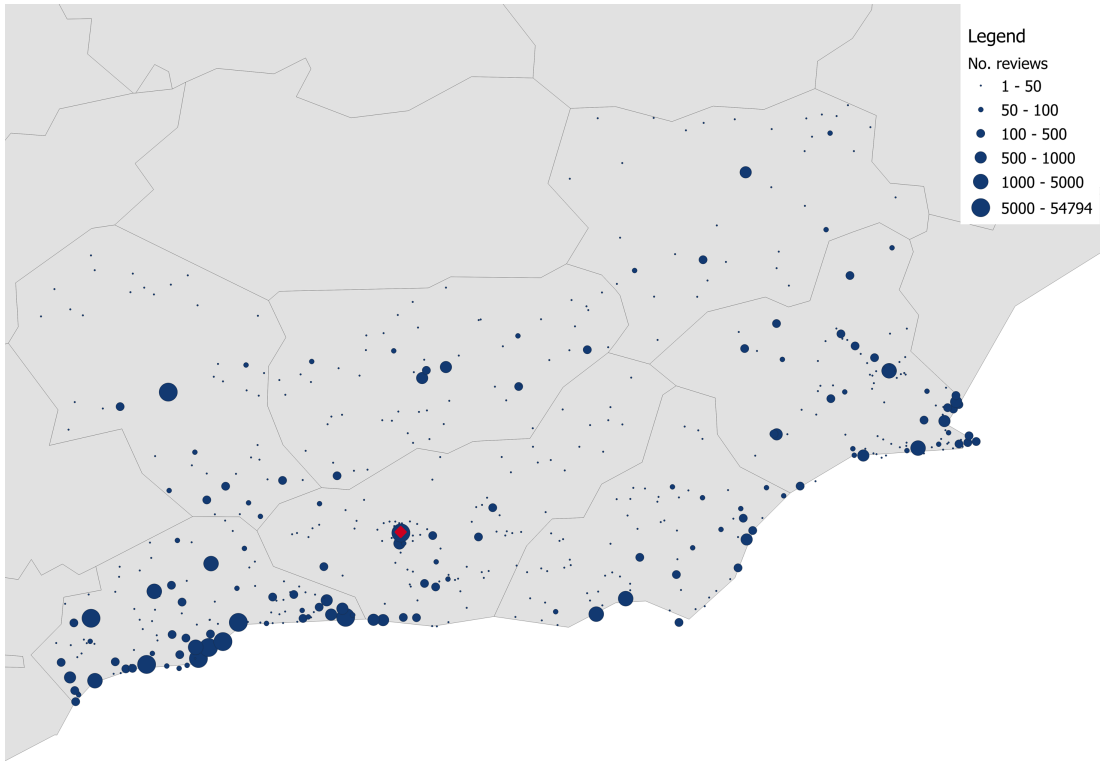
Conclusion

- We have demonstrated and validated the use of Tripadvisor. Advantages: daily availability, granularity, user data, and comparable and objective data.
- We showed a resilient interest in the peripheral regions despite the COVID-19 pandemic and increasing visits from both their respective countries and across Europe.
- Our methodology allowed us to measure peripheral tourism in a wide and diverse set of countries across Europe with consistent results.
- **We highlight the potential of nature and cultural attractions for both regional development and sustainable practices in managing tourism growth.**

Thank You!

Appendix: Data description: Peripheral regions

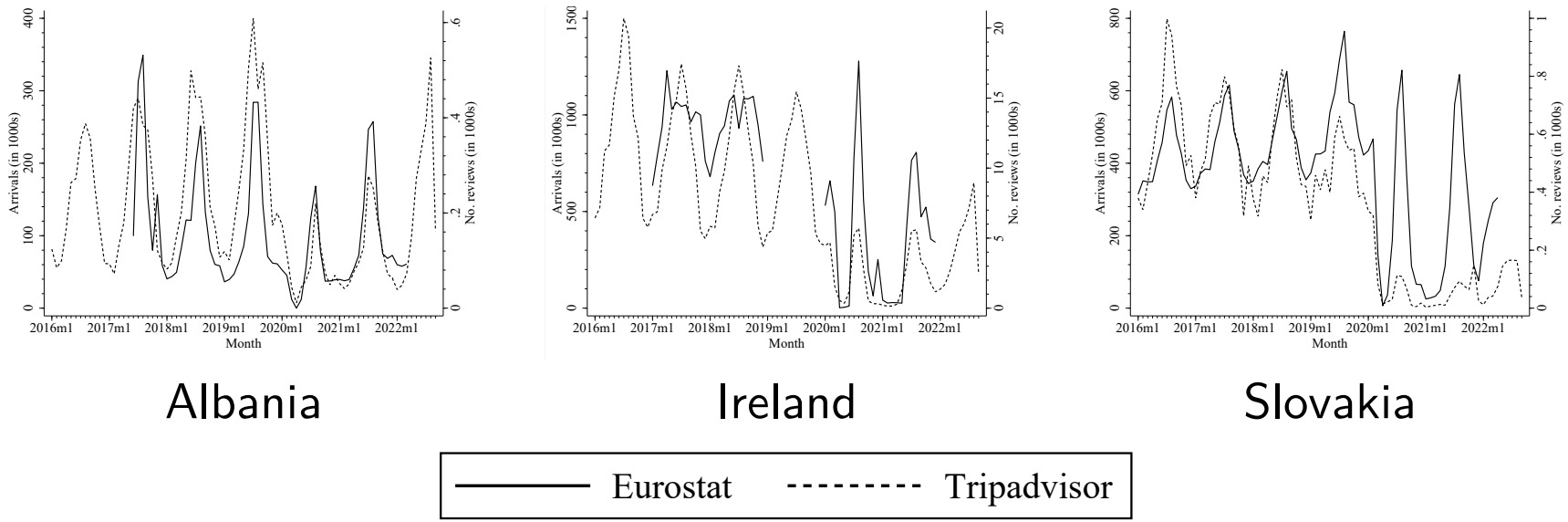
Figure: Location of attractions in peripheral and neighboring regions



Notes: This Figure shows the attractions (blue) and pilot site (red) of peripheral and neighboring regions. *Source:* Own data collected from Tripadvisor.

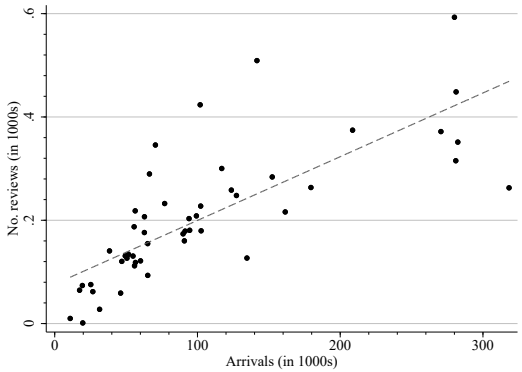
Appendix: Validating the Tripadvisor data

Figure: Total number of arrivals and reviews over time

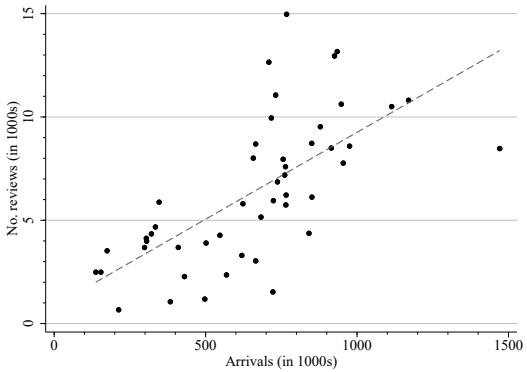


Appendix: Validating the Tripadvisor data

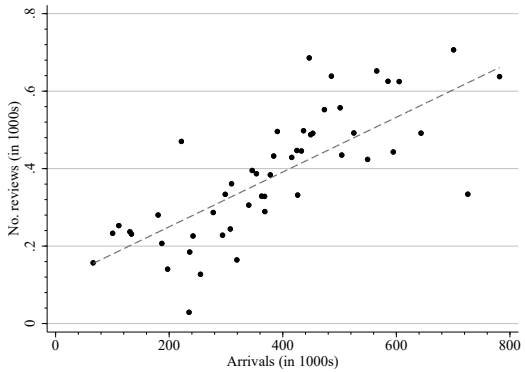
Figure: Monthly correlation between tourist arrivals and number of reviews



Albania



Ireland



Slovakia

Appendix: Validating the Tripadvisor data

Table: Validity Test: Regression results for full sample

	(1) ln(Reviews)	(2) ln(Domestic reviews)	(3) ln(Foreign reviews)	(4) ln(Reviews)
ln(Arrivals)	0.673***			
	(0.025)			
ln(Domestic arrivals)		0.610***		
		(0.036)		
ln(Foreign arrivals)			0.688***	
			(0.022)	
Occupancy rate				0.036***
				(0.002)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
N	632	623	630	635
R^2	0.964	0.963	0.960	0.946

Notes: Robust standard errors in parentheses. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$.
Source: Official tourism statistics from Eurostat and own data collected from Tripadvisor.

Appendix: Validating the Tripadvisor data

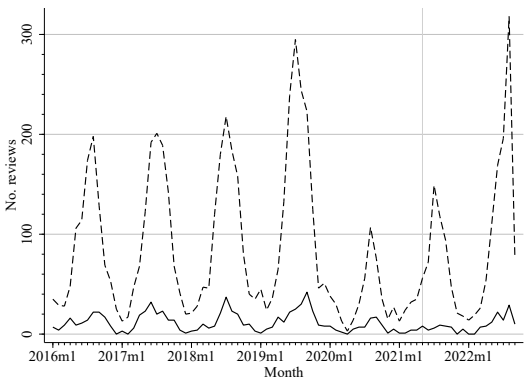
Table: Validity Test: Correlation coefficients between Tripadvisor reviews and Eurostat tourism measures

	All	Albania	France	Greece	Ireland	Italy	Portugal	Slovakia	Spain	Sweden
					(a) No. arrivals					
No. reviews	0.862	0.726	0.875	0.847	0.843	0.790	0.8149	0.702	0.832	0.800
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					(b) No. domestic arrivals					
No. domestic reviews	0.796	0.295	0.786	0.862	0.789	0.640	0.632	0.283	0.688	0.771
	(0.000)	(0.025)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.013)	(0.000)	(0.000)
					(c) No. foreign arrivals					
No. foreign reviews	0.892	0.707	0.854	0.908	0.935	0.843	0.873	0.831	0.882	0.826
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					(d) Occupancy rate					
No. reviews	0.536	0.762	0.827	0.938	0.927	0.695	0.887	0.685	0.867	0.835
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

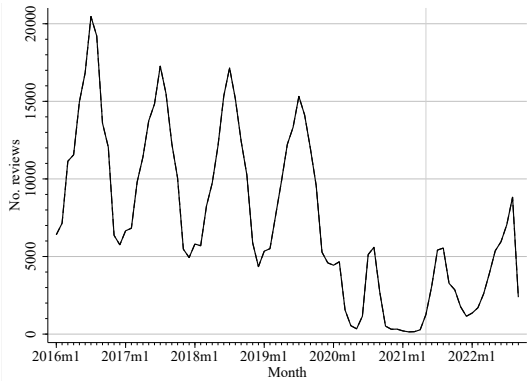
Source: Official tourism statistics from Eurostat and own data collected from Tripadvisor.

Appendix: Time trends of tourism activity

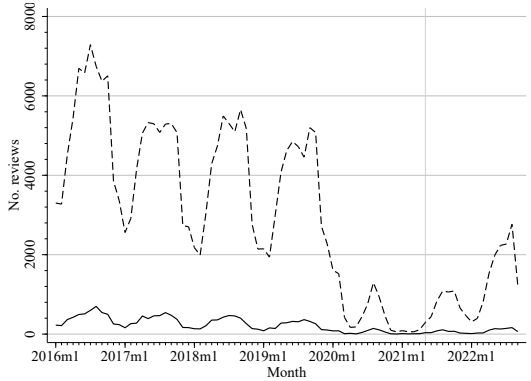
Figure: Number of reviews over time for peripheral and neighboring regions



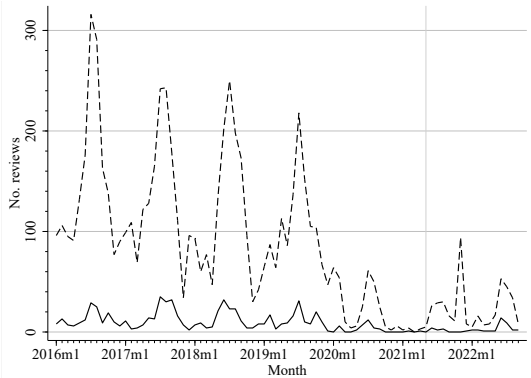
Albania



Ireland



Italy (Tuscany)

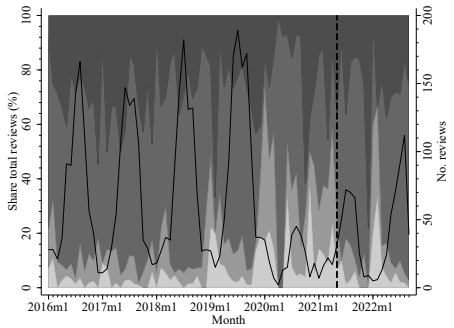


Slovakia

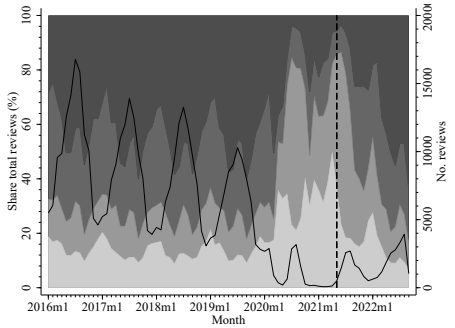
INCULTUM region — — — — — INCULTUM + bordering regions

Appendix: Time trends of tourism activity

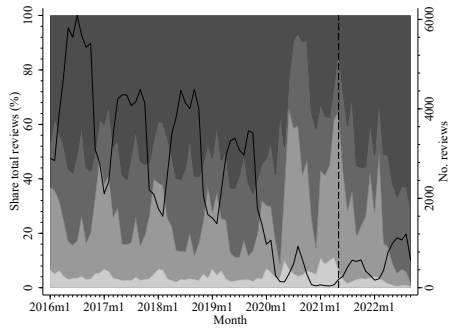
Figure: Share of reviews for different travel categories over time for peripheral and neighboring regions



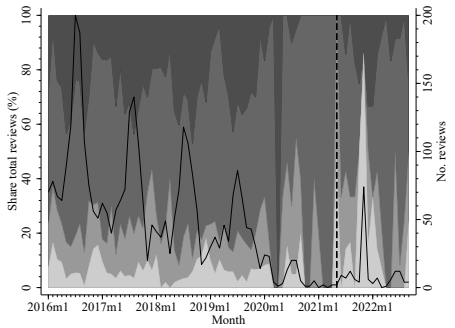
Albania



Ireland



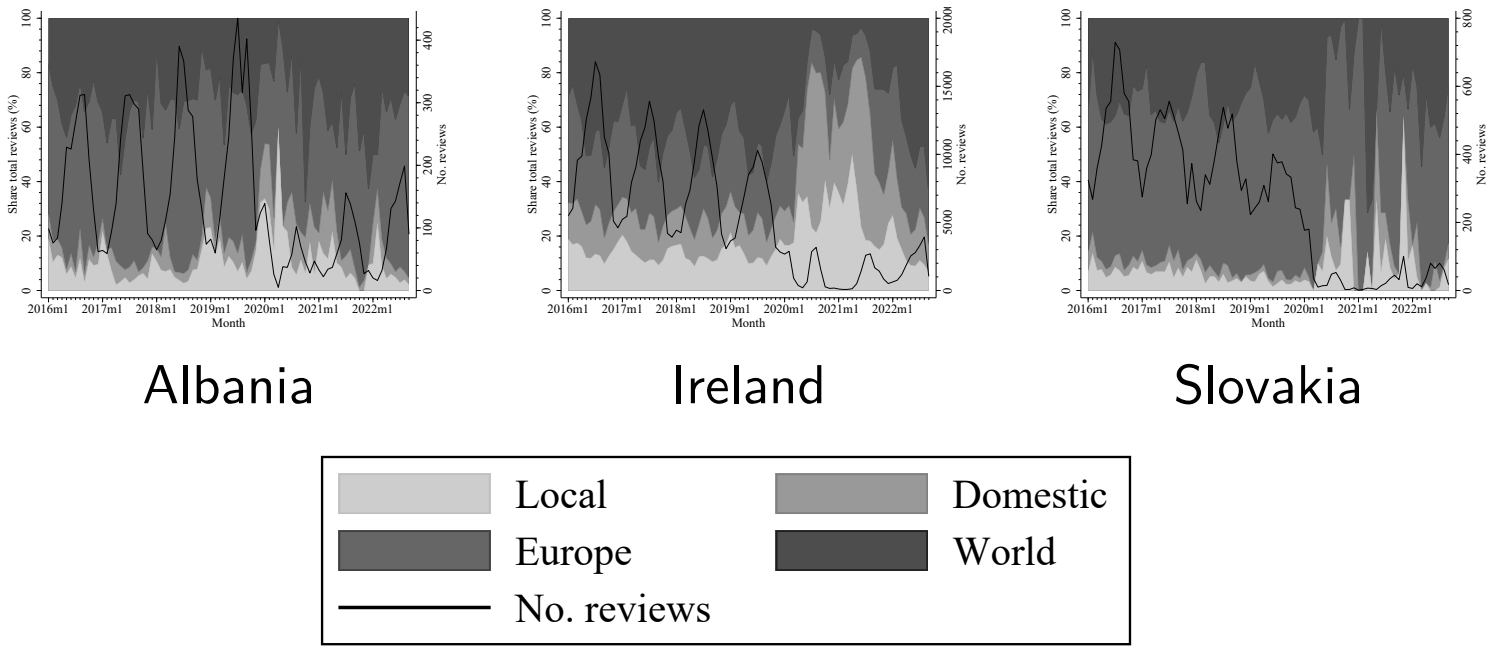
Italy (Tuscany)



Slovakia

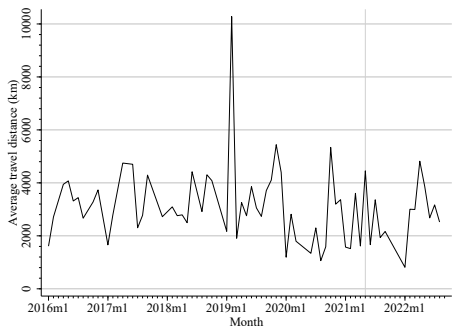
Appendix: Time trends of tourism activity

Figure: Share of reviews for different travel categories over time for national level

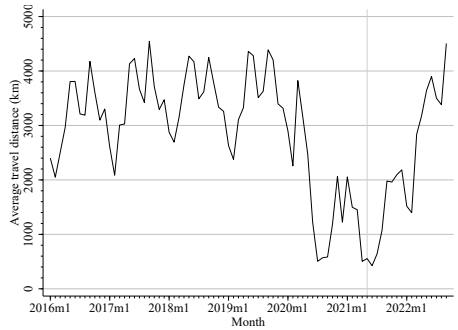


Appendix: Time trends of tourism activity

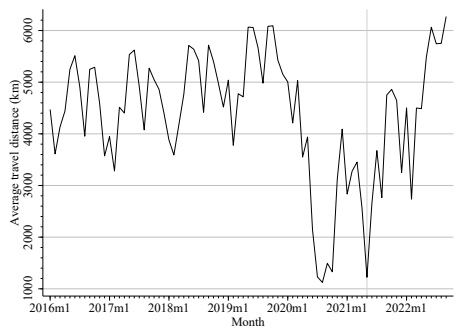
Figure: Distance travelled over time in peripheral and neighboring regions



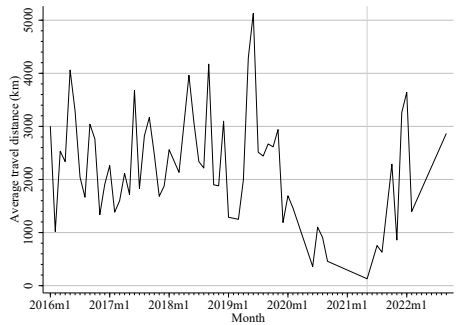
Albania



Ireland



Italy (Tuscany)

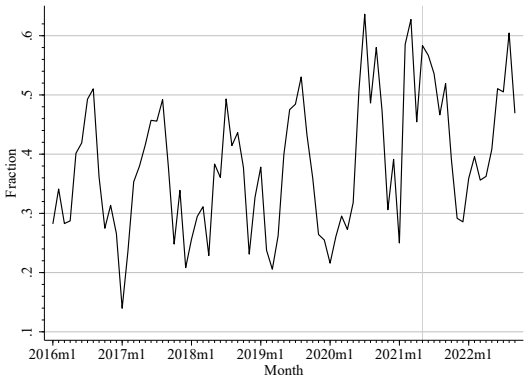


Slovakia

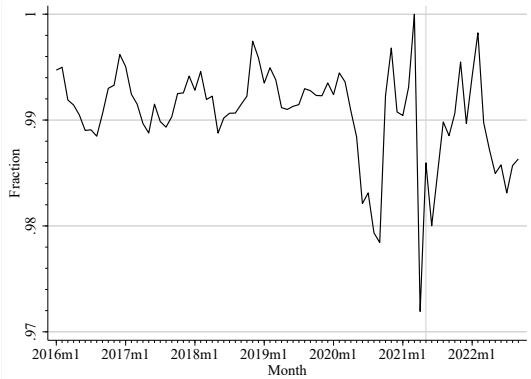
Note: The vertical line indicates when the INCULTUM pilot started. *Source:* Own data collected from Tripadvisor.

Appendix: Comparison of peripheral and neighboring regions

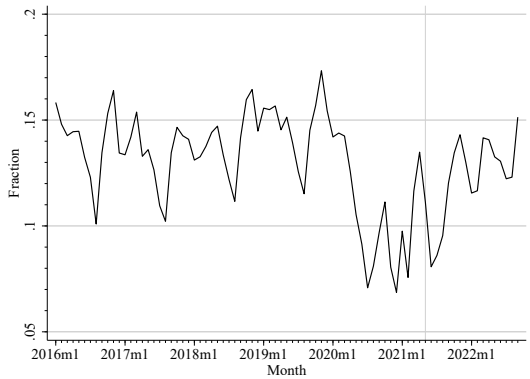
Figure: Fraction of the number of reviews in peripheral and neighboring regions out of all reviews in country over time



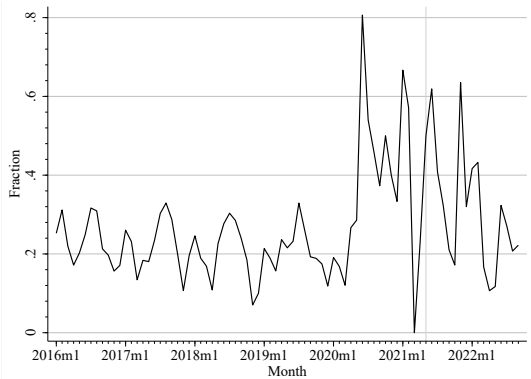
Albania



Ireland



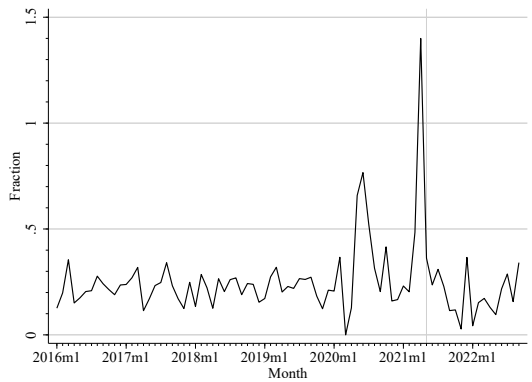
Italy (Tuscany)



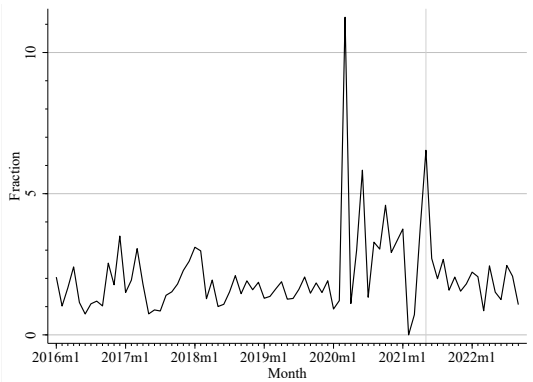
Slovakia

Appendix: Comparison of peripheral and neighboring regions

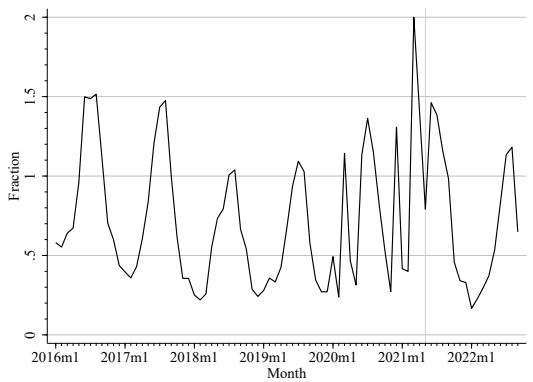
Figure: Fraction of number of reviews in peripheral regions out of average number in neighboring regions over time



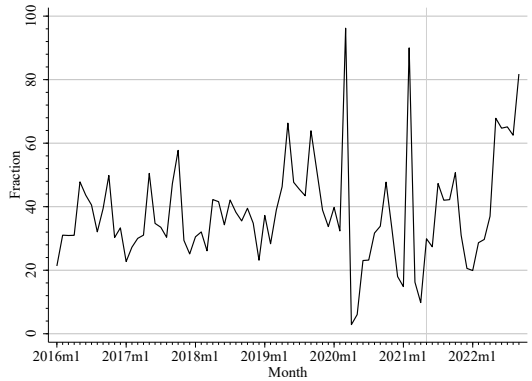
France



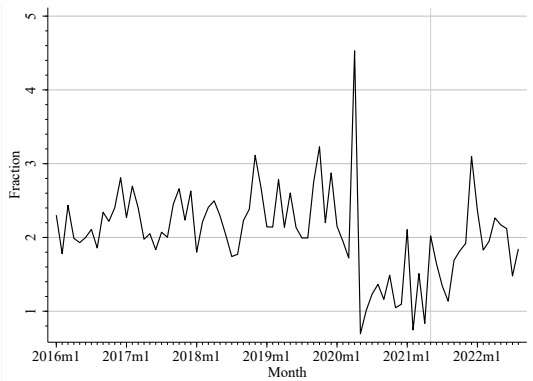
Greece



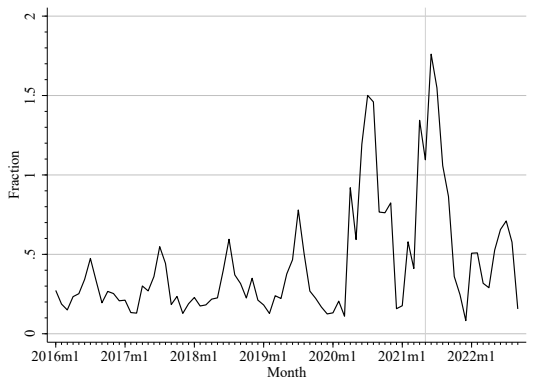
Italy (Sicily)



Portugal



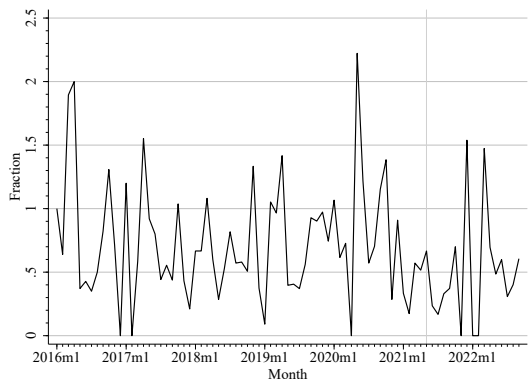
Spain



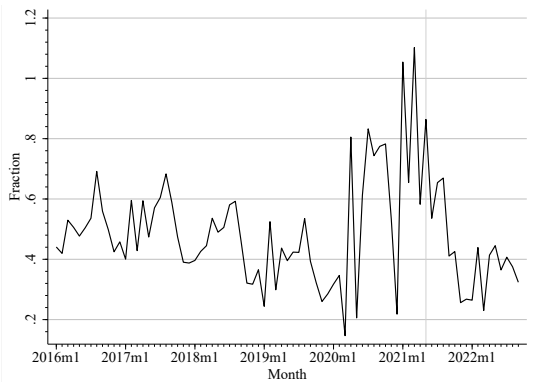
Sweden

Appendix: Comparison of peripheral and neighboring regions

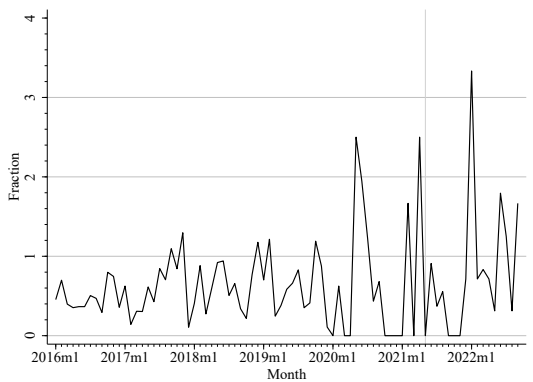
Figure: Fraction of number of reviews in peripheral regions out of average number in neighboring regions over time



Albania



Italy (Tuscany)



Slovakia

Note: The vertical line indicates when the INCULTUM pilot started. *Source:* Own data collected from Tripadvisor.