DEUTSCHE BUNDESBANK
EUROSYSTEM

Profiling banks

Clustering payment profiles of TARGET2 participants

Marc Glowka, Payments and Settlement Systems, Deutsche Bundesbank

16th Payment and Settlement System Simulation Seminar
Helsinki, 30 August 2018

Disclaimer

The author of this paper is a member of one of the user groups with access to TARGET2 data in accordance with Article 1(2) of Decision ECB/2010/9 of 29 July 2010 on access to and use of certain TARGET2 data. The Deutsche Bundesbank and the MIPC have checked the paper against the rules for guaranteeing the confidentiality of transaction-level data imposed by the PSSC pursuant to Article 1(4) of the above mentioned issue. The views expressed in the paper are solely those of the author and do not necessarily represent the views of the Eurosystem.

Introduction
About payment behaviour, profiles and clustering

Payment behaviour

- Describe how participants introduce their transactions intraday
- Relevant for a number of FMI analyses (Liquidity distribution, risk identification. fraud detection etc.)
- Deviations from regular payment behaviour could indicate a risk event

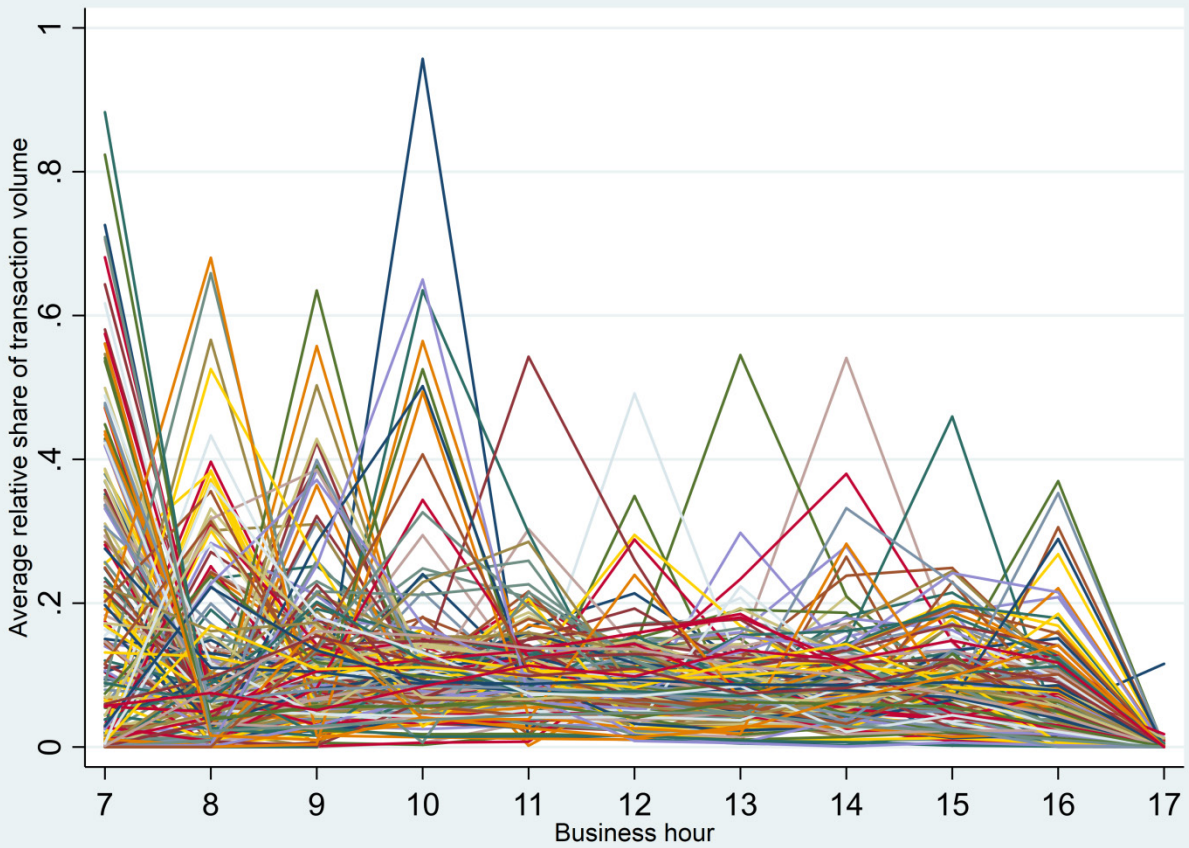
Payment Profile

- Based on regular payment behaviour
- Several participants with an equal payment behaviour establish a profile
- Profiles provide insights into the liquidity needs, intraday distribution of transactions and relations between participants

Cluster Analysis

- Groups objects based on their similarity of underlying structures and characteristics
- Explorative data technique (unsupervised learning algorithm)
- Wide variety of different clustering techniques

Introduction
About payment behaviour, profiles and clustering



Source: TARGET2; Own calculations

Data Set

Transformation from multiple to three dimensional clustering data set

Data Basis: TARGET2 transaction data

Data Selection:

- Large participants (at least 0.05 % transaction volume in TARGET2)
- Participant-entered transactions
- Day-time settlement cycle transactions
- Transactions settled in 2017

Data Preparation:

- Aggregation to one-hour intervals
- Relative shares of each one-hour interval on daily transaction volume
- Daily average of relative shares for each one-hour interval

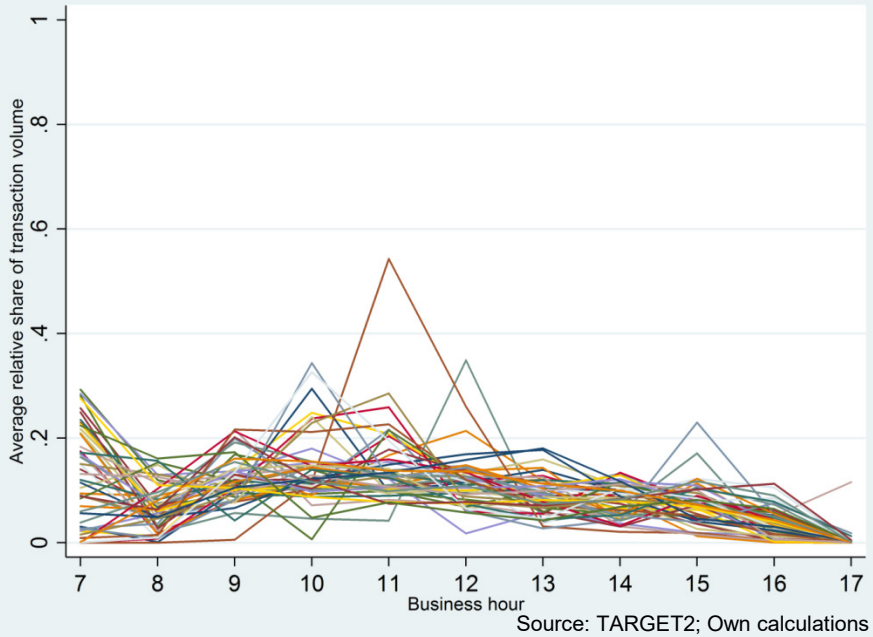
Multiple k-means cluster analysis

Need for an individual cluster technique

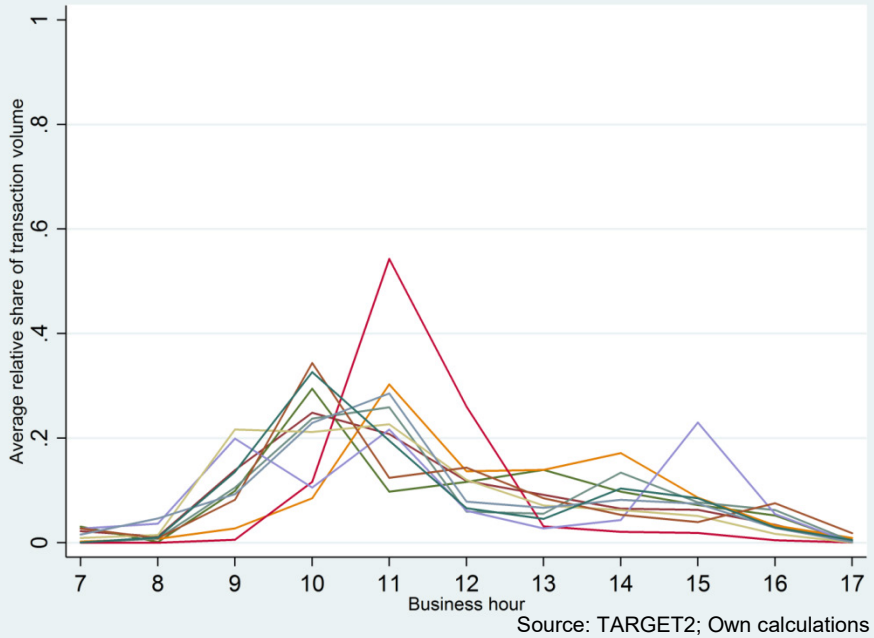
- Results are depending on seed setting
- Pre-defining the number of clusters to be grouped
- Non-combination of different similarity measures
- Advanced machine learning algorithms do not lead to meaningful results

Multiple k-means cluster analysis
Need for an individual cluster technique

Results are depending on seed setting



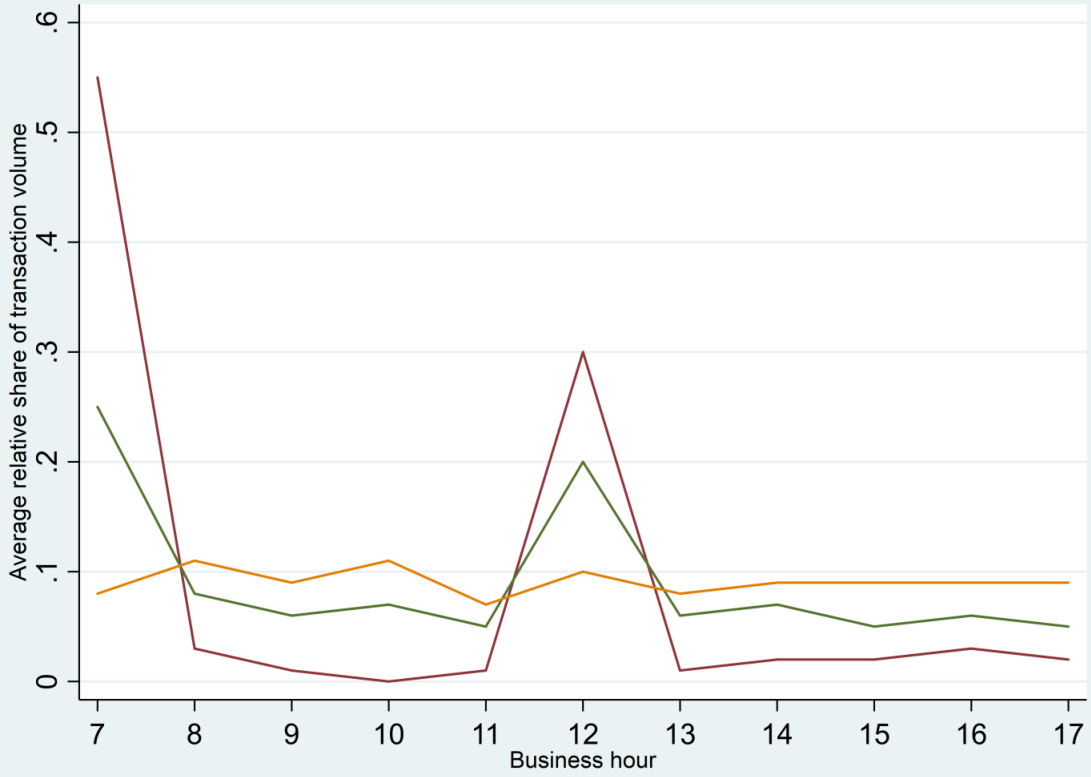
Seed: 896204172; Cluster: 8; Start option: Random; Measure: Euclidean Distance



Seed: 39541; Cluster: 8; Start option: Random; Measure: Euclidean Distance

Multiple k-means cluster analysis
Need for an individual cluster technique

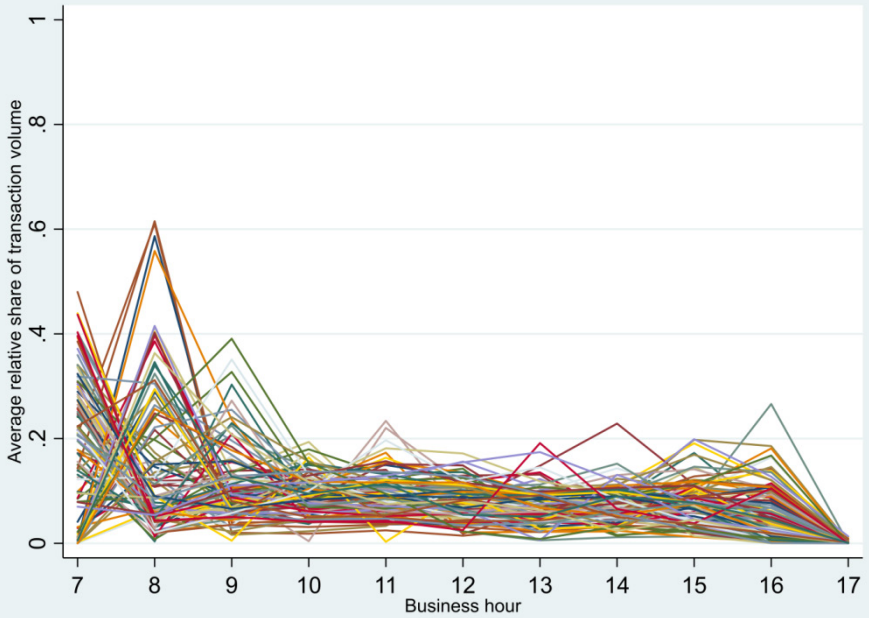
Non-combination of different similarity measures



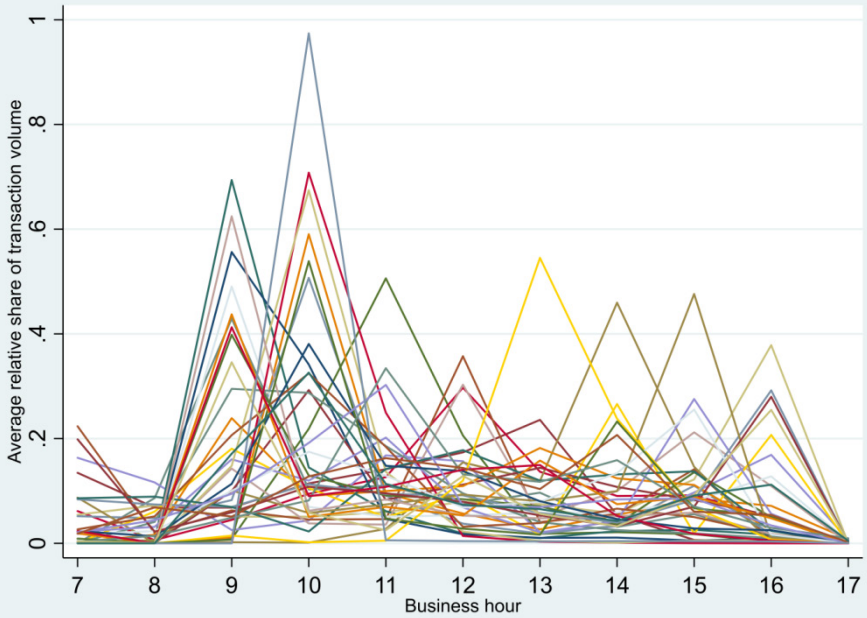
Note: Self-created data set for illustration of differences

Multiple k-means cluster analysis
Need for an individual cluster technique

Advanced machine learning algorithms do not lead to meaningful results



Source: TARGET2; Own calculations



Source: TARGET2; Own calculations

Optimal Cluster = 4 VEV-Model

Multiple k-means cluster analysis
Methodology

Multiple runs of k-means algorithm:
- Randomly chosen seeds
- Distance (Euclidean Distance) and similarity (Correlation) measures
- Randomly chosen number of pre-defined cluster solutions (between 8 and 16 clusters)
- Randomly chosen start centrum from a uniform distribution over the data range

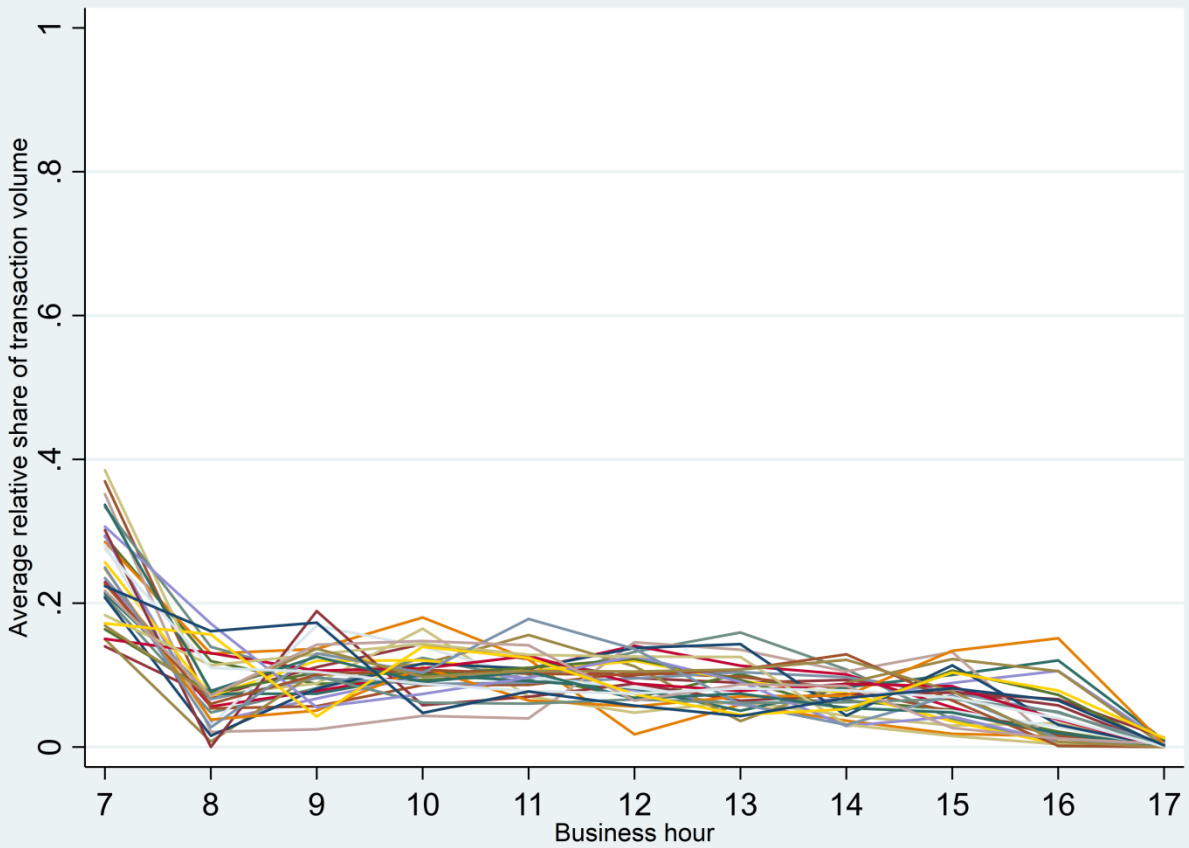
Pooling of similar cluster solutions
- Calculation of normalised cluster solution statistics (average, median and standard deviation for each one-hour interval and cluster solution)
- Application of k-means clustering algorithm on the normalised cluster solution statistics

Determination of the dominant cluster solution
- Calculating the number of assignments from multiple to pooling clustering results per participant
- Choosing the pooling cluster solution with the maximum of assignments

Merging with primary data set

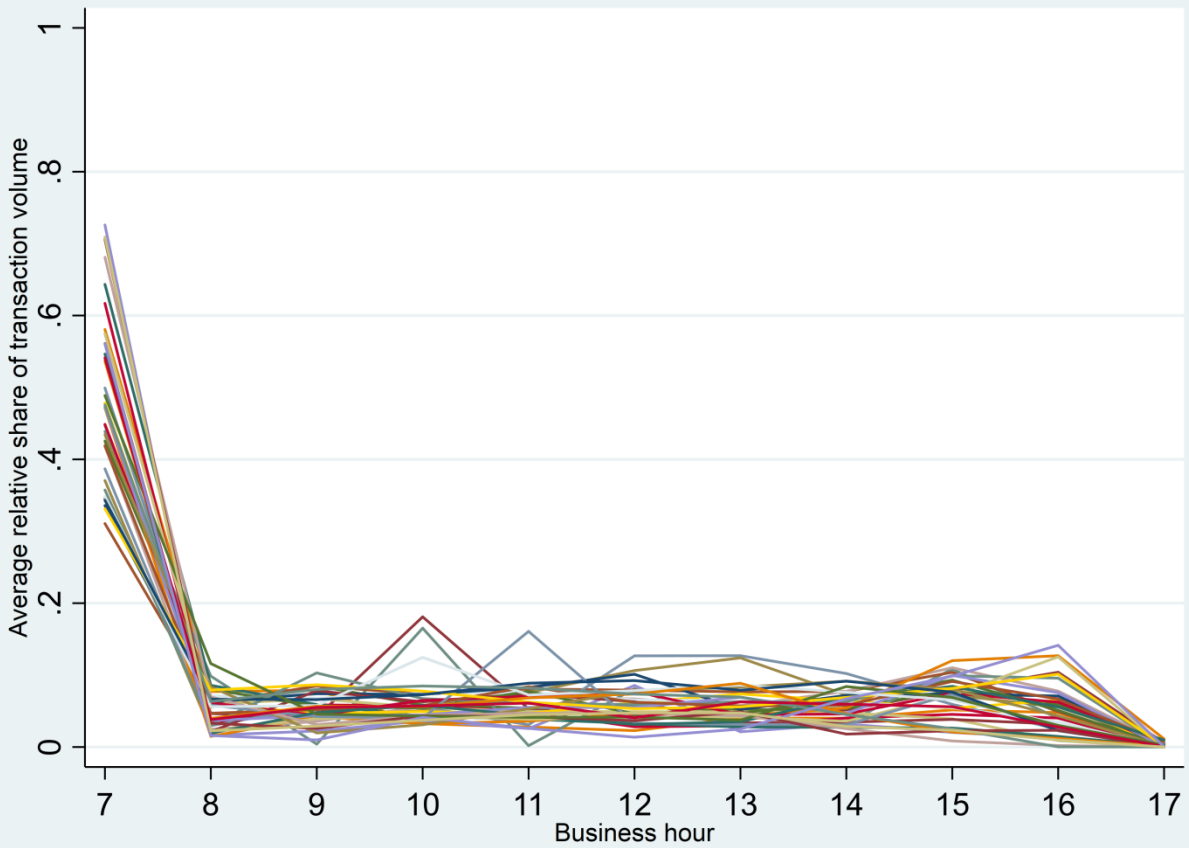
Outlier Analysis
- Identify outliers (Outside 75 % quantile + 1.5 IQR and above interval share of 20% on total transaction volume or less than three participant in cluster solution)
- Clustering outliers again

Payment Profiles
Early birds



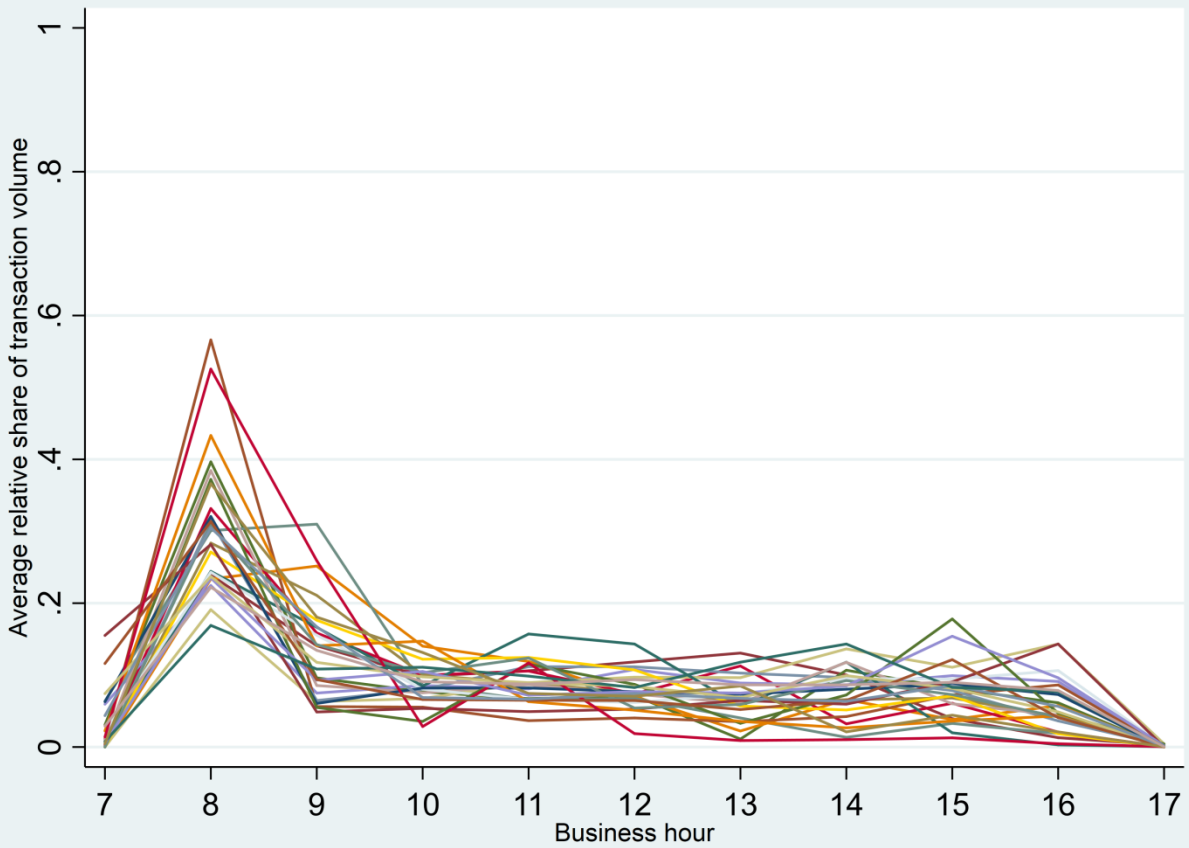
Source: TARGET2; Own calculations

Payment Profiles
Extreme early birds



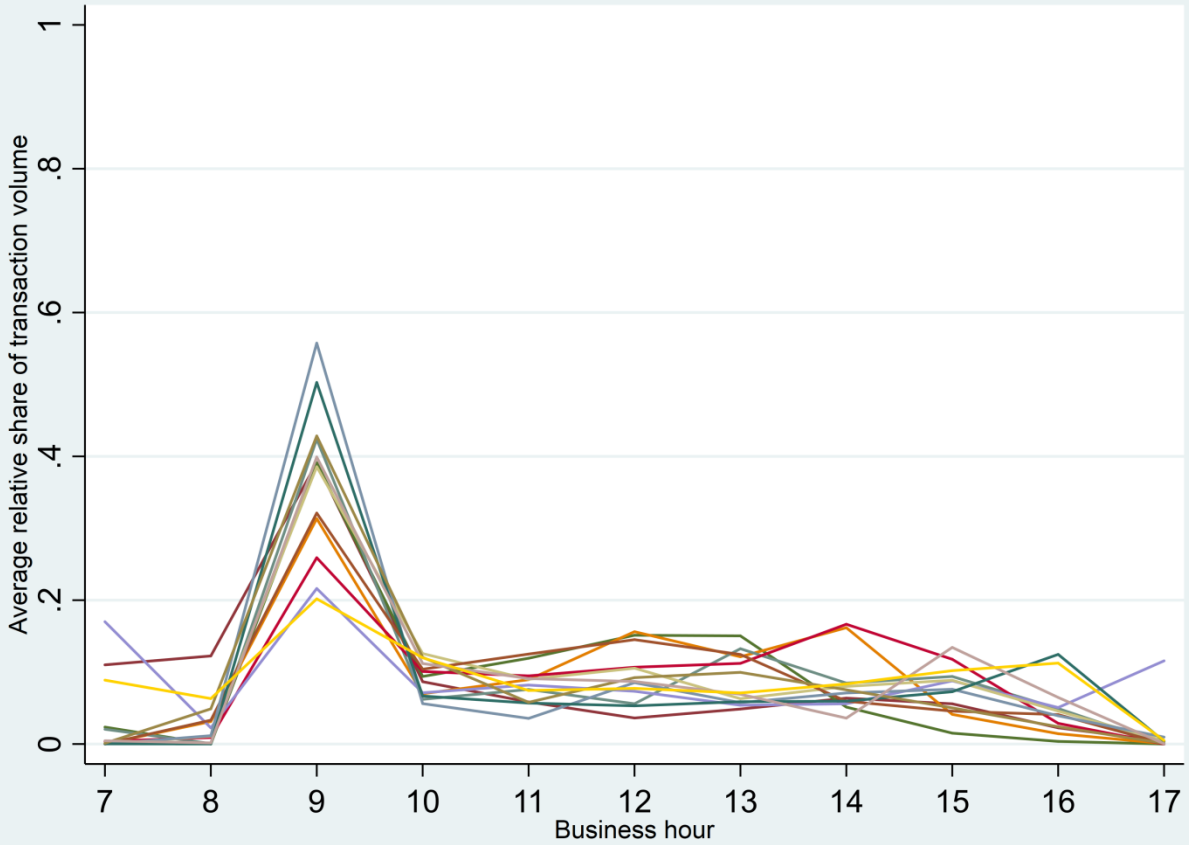
Source: TARGET2; Own calculations

Payment Profiles
Second wave



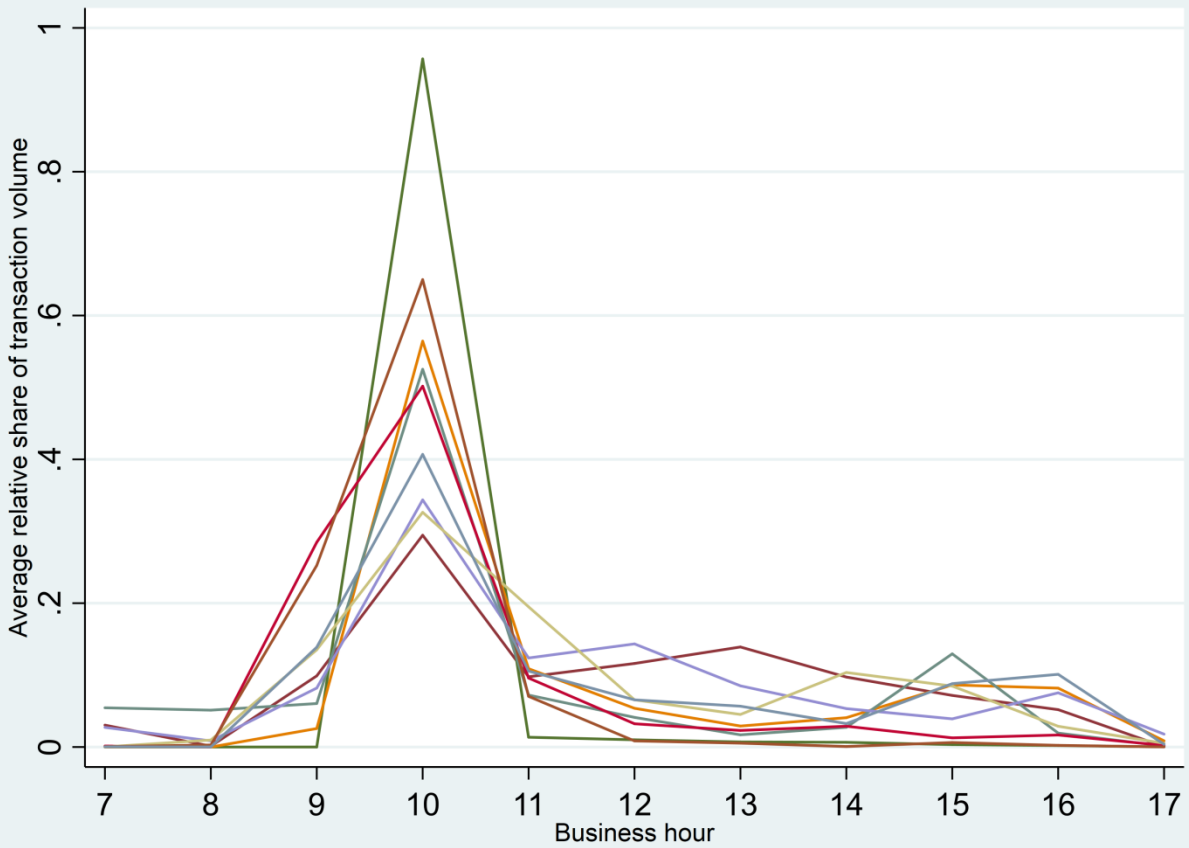
Source: TARGET2; Own calculations

Payment Profiles
Third wave



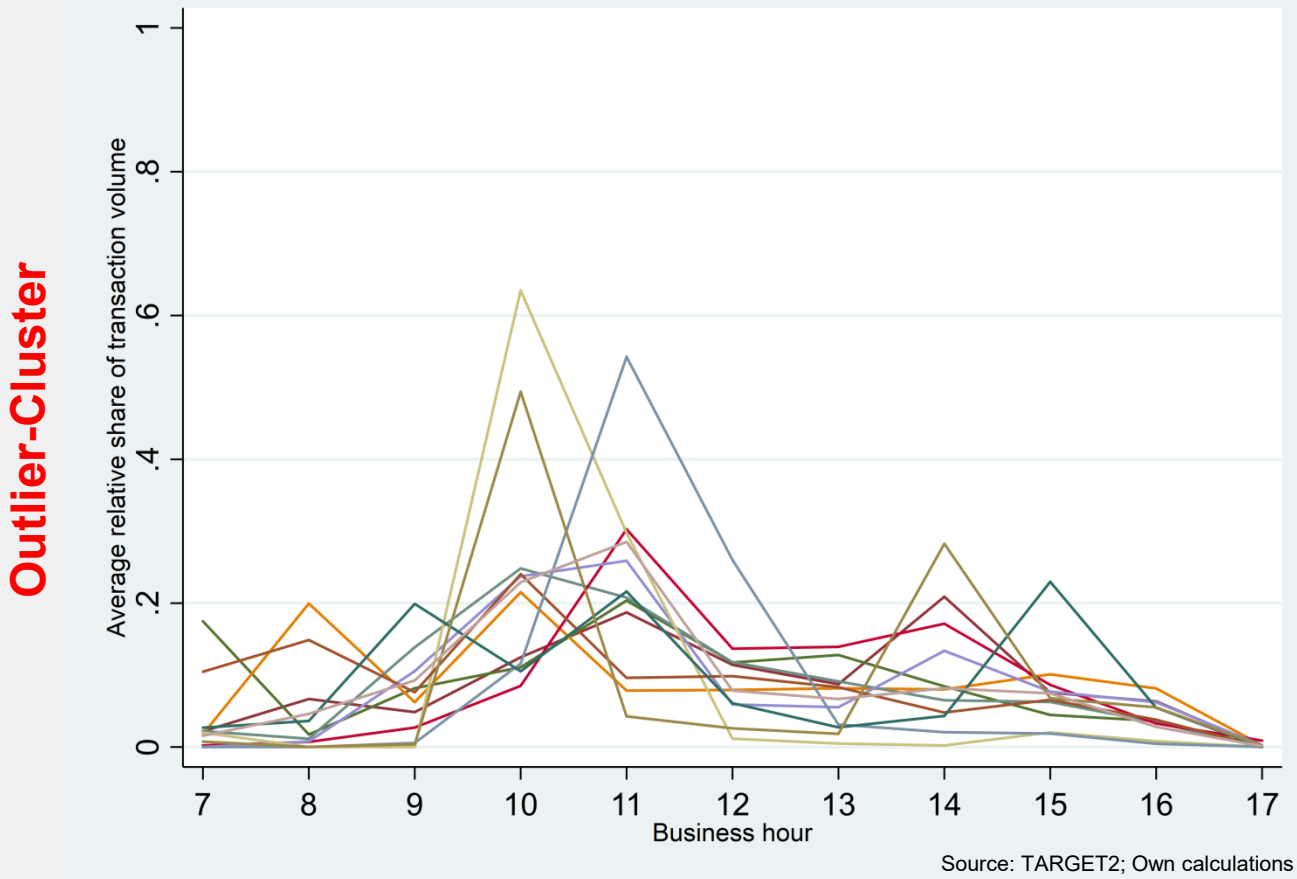
Source: TARGET2; Own calculations

Payment Profiles
Long sleepers

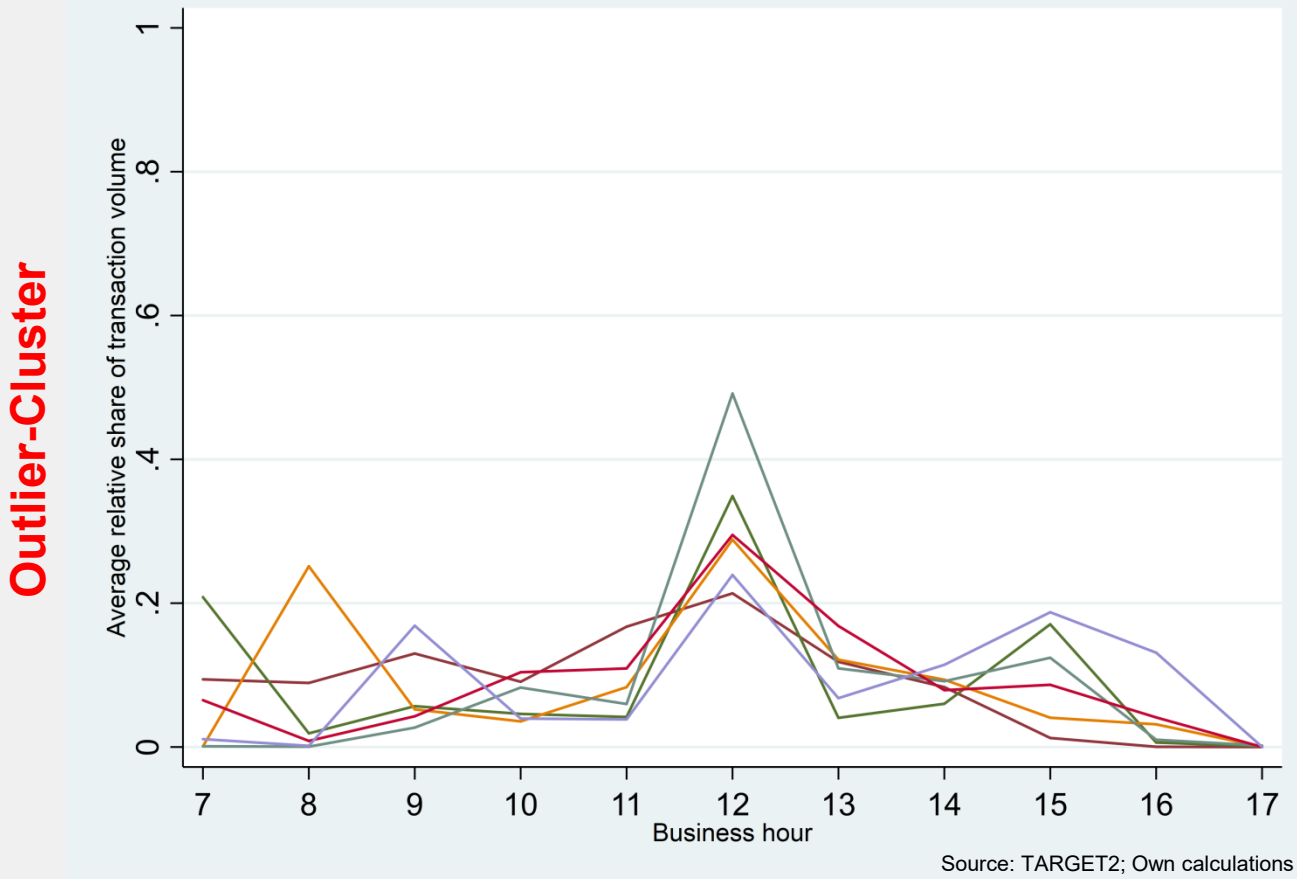


Source: TARGET2; Own calculations

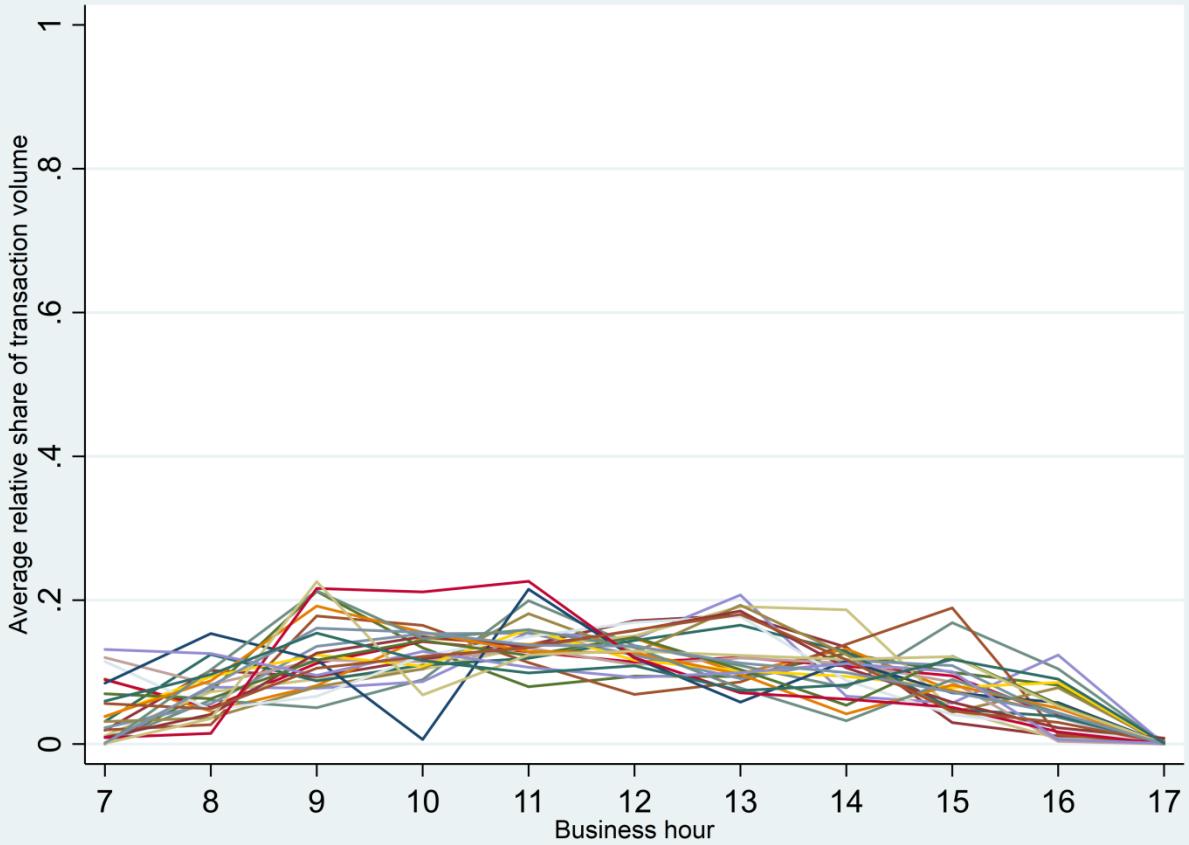
Payment Profiles
Late morning payers



Payment Profiles
Noon payers

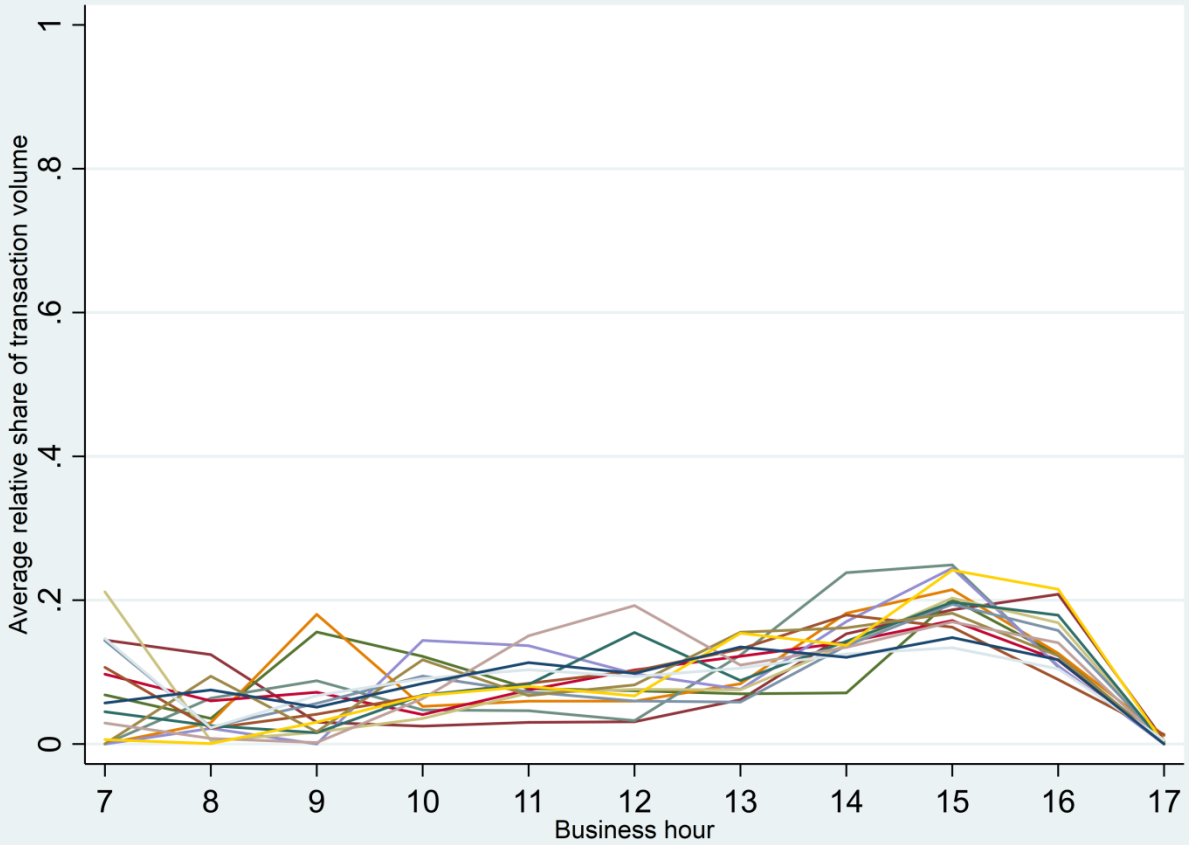


Payment Profiles
Time-independent payers



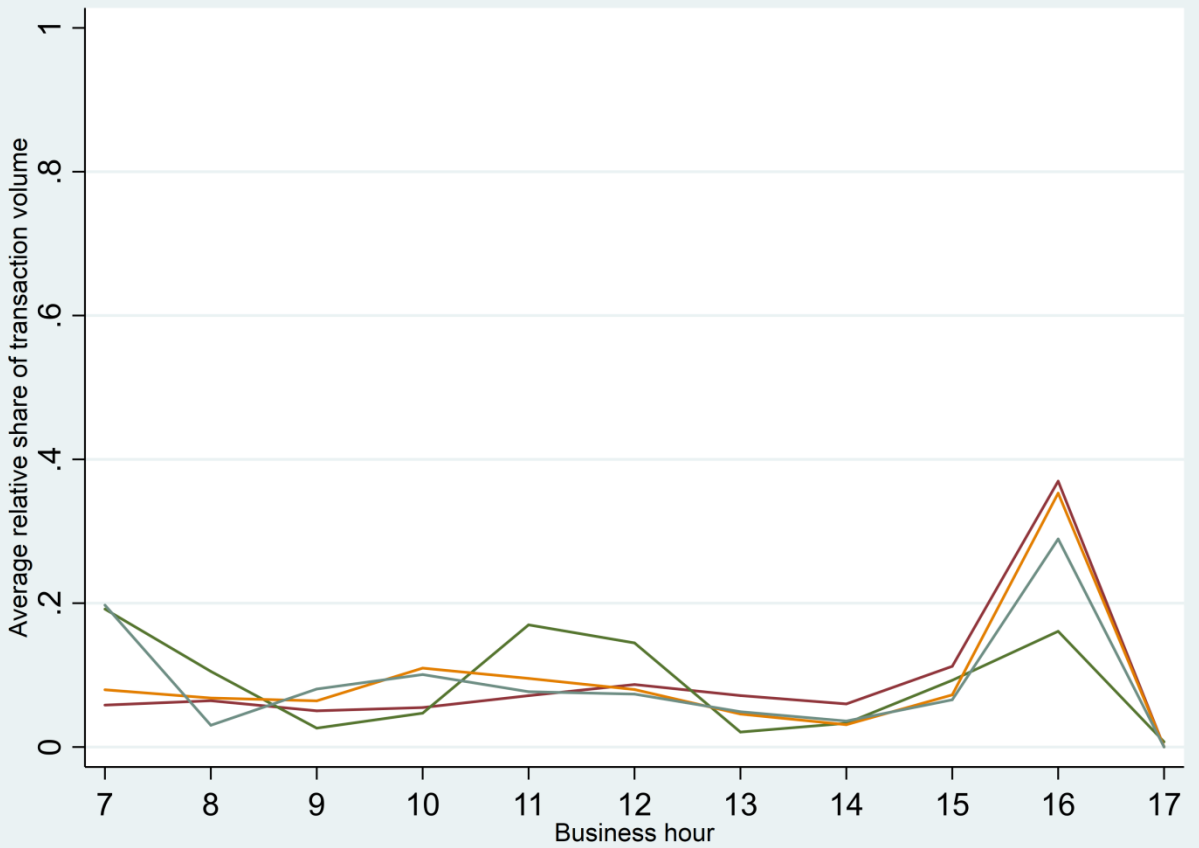
Source: TARGET2; Own calculations

Payment Profiles
Tea-time payers



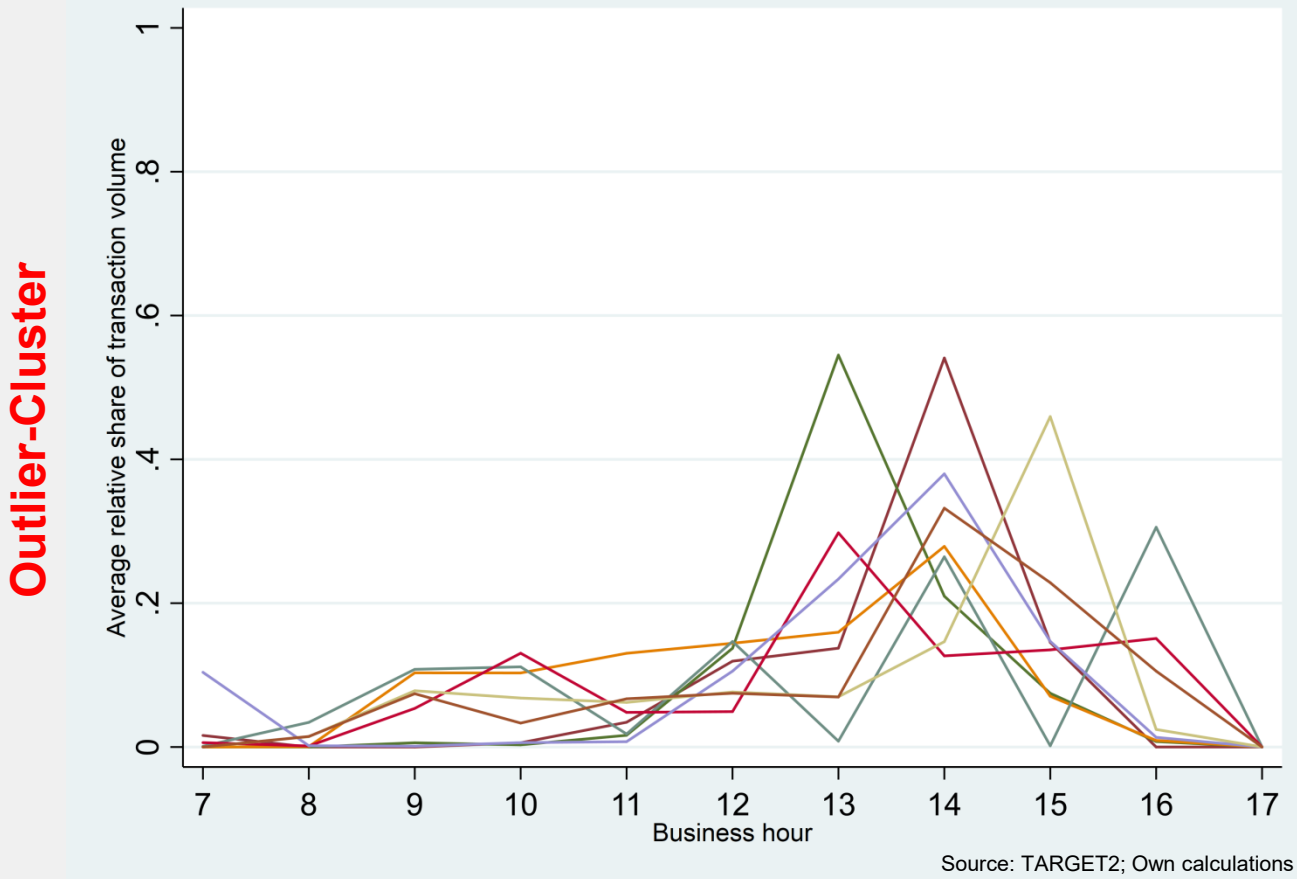
Source: TARGET2; Own calculations

Payment Profiles
Late payers (1)



Source: TARGET2; Own calculations

Payment Profiles
Late payers (2)



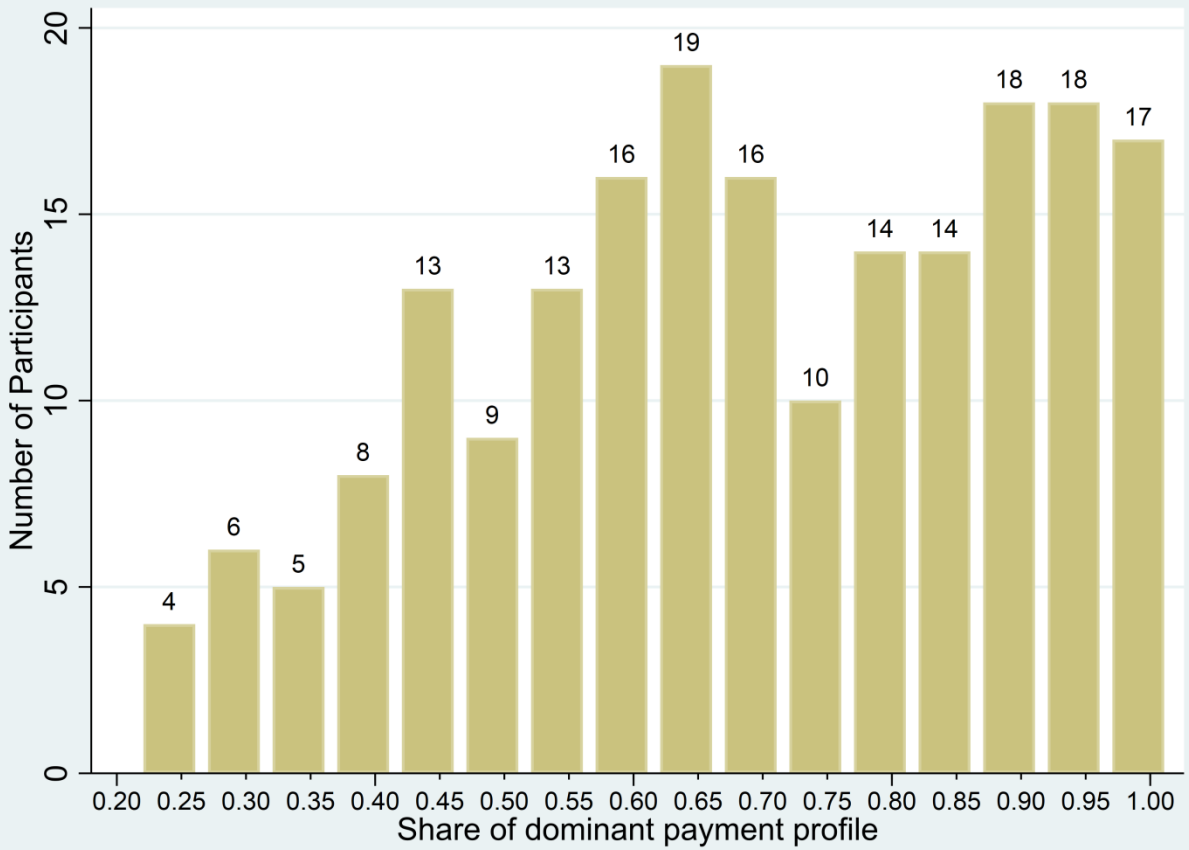
Payment profiles
Characteristics (1)

Payment Profiles	Dominant characteristics
Early birds	**Between 20 and 40 %** of the daily transactions are introduced in the **first business hour** of TARGET2. In addition this is also the **maximum** of the day.
Extreme early birds	**The maximum of the day and more than 40 %** of the daily transactions are introduced in the **first business hour** of TARGET2.
Second wave	**More than 20 %** of the daily transactions are introduced in the **between 8 and 9 o'clock** and in addition this is also the **maximum** of the day.
Third wave	**More than 20 %** of the daily transactions are introduced in the **third business hour** of TARGET2 and in addition this is also the **maximum** of the day.
Long sleepers	**The maximum of the day and more than 20 %** of the daily transactions are introduced between 1**0 and 11 o'clock.**
Late morning payers	**More than 20 %** of the daily transactions are introduced in **the fifth business hour of TARGET2**. In addition this is **also the maximum of the day.**

Payment profiles
Characteristics (2)

Payment Profiles	Dominant characteristics
Noon payers	**More than 20 %** of the daily transactions are introduced **between 12 and 13 o'clock** and in addition this is also the **maximum of the day**.
Time-independent payers	The participants with these profiles **equally distributed their payment activity over the day** with l**ess transactions in the morning or evening**. No **one-hour interval exceeds 20 %** of the transaction share.
Tea-time payers	The **transaction volume share increases over the day** and **reaches a maximum between 15 and 17 o‘clock.** In addition the transaction volume share **remains usually below 20 %** over the day.
Late payers	**The maximum and more than 20 %** of the daily transactions are introduced in the afternoon **between 13 and 17 o'clock**.

Payment profiles
Stability?



Source: TARGET2; Own calculations

Note: Days without payment activity are excluded

Conclusion and way forward

A promising start but not the end of the way

- Results independent from seed setting
- Combining different similarity measures
- Different meaningful payment profiles are identified

- Aggregated data set
- Focus only on one year
- Payment profiles are not stable
- Additional characteristics were not considered

Way Forward:

- Usage of different data sets (different year, varying time dimension, include more participants)
- Relationships between participants with different payment profiles
- Payment profiles depending on transaction classes
- Investigate the deviation on payment profiles
- Refinement of payment profiles

Thank you for your attention!

Marc Glowka
Deutsche Bundesbank
Payments and Settlement Systems
Marc.Glowka@bundesbank.de