

Data Warehousing

Module Code - ACCA6026

Date: 14th May 2026

Tyler Williamson

Student No. 2237644

Table of Contents

1	Introduction.....	2
2	Task 1.....	2
2.1	Online Transaction Processing (OLTP) vs Data Warehouse (DW)	2
2.2	Data Warehouse Architecture and ETL	3
2.3	Data Marts (Targeted Insights)	4
2.4	Normalised OLTP vs Dimensional Data Warehouse	5
3	Task 2.....	7
3.1	Internal and External Source Systems.....	7
3.2	Star Schema	7
3.3	Data Marts and the Schema	9
3.4	Simplified ETL Plan	9
3.5	Data Quality and Information Governance	10
4	Conclusion	10
5	References	11
Figure 1 - Multi-Tier Data Architecture		3
Figure 2 - Multi-Tier Star Schema showing Aggregate and Metadata tables		6

1 Introduction

This report presents a strategic blueprint for the development and implementation of a modern data warehouse solution for Rags'n'Riches Ltd. Currently, the organisation's decision-making is compromised by fragmented data silos distributed across six regional warehouses, leading to significant logistics inefficiencies and an annual transport expenditure exceeding £500,000.

The objective is to establish a technical framework that transitions the company from a transactional focus to a high-maturity, data-driven enterprise. By leveraging Dimensional Modelling and a multi-tier architecture, the proposed system will synthesise internal metrics, such as fuel card data and inventory status with external film industry trends. This single version of the truth will empower the board to rationalise operations, optimise the current £2,100,000 profit margin, and provide the empirical foundation required to capitalise on emerging UK film industry tax incentives[1].

2 Task 1

2.1 Online Transaction Processing (OLTP) vs Data Warehouse (DW)

In modernising the IT infrastructure for Rags'n'Riches Ltd, it is essential to distinguish between the functional requirements of operational databases and analytical repositories. The company's current infrastructure relies on OLTP systems. These databases are highly normalised to support concurrent write operations, such as the real-time tracking of rental inventory across the six UK warehouses[2]. The primary focus of an OLTP environment is the atomic transaction: recording precisely when a costume is dispatched or when a delivery is signed for on-site.

Conversely, a DW is a subject-oriented, integrated, and non-volatile collection of data designed to support read-heavy analytical queries and management's decision-making[3]. While the OLTP system is vital for daily operations, it is unsuitable for complex analysis. Executing multi-year queries on an operational database can lead to resource contention, degrading the performance of the order entry system at the Raglan administrative office [2].

Strategic advantages for the board:

- **Historical Repository:** Unlike OLTP systems, which often overwrite data to reflect current status, a DW preserves a permanent historical record[4]. This allows for time-variant analysis, enabling the board to identify shifts in the popularity of "Room in a Box" sets versus individual hires.

- **Data Integration:** The DW acts as a unified platform that integrates disparate internal silos, such as transport fuel records and inventory data, with external variables, like UK film tax incentives.
- **Rationalisation Capability:** Leveraging a DW allows senior management to transition to prescriptive analytics[3]. This provides the empirical evidence needed to rationalise operations and optimise the £2,100,000 profit margin.

2.2 Data Warehouse Architecture and ETL

To implement a robust analytical environment, a Three-Tier architecture is required[5]. This ensures strategic data analysis remains physically and logically separate from daily rental transactions.

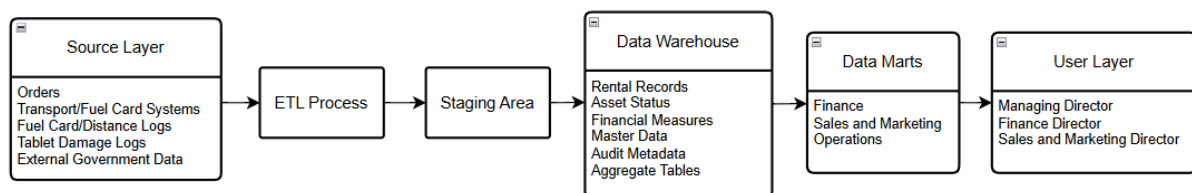


Figure 1 - Multi-Tier Data Architecture

The Multi-Tier Architecture

The architecture typically comprises three distinct layers: the Source Layer, the Data Staging Area, and the Data Warehouse Layer.

- **Source Layer:** This contains the disparate operational systems mentioned in the case study, including the Orders database, the transport department's databases (handling stock movements and fuel cards), and the on-site tablet data uploaded via 4G/5G links.
- **Data Staging Area:** This is a crucial intermediate buffer zone. It is here where the ETL (Extract, Transform, Load) process occurs[6]. By performing data cleaning and integration in the isolated area, the company ensure that computationally intensive tasks do not impact the performance of the administrative office in Raglan.
- **Data Warehouse Layer:** The final destination where integrated, high-quality data is stored in a dimensional format, ready for management to query.

The ETL Process

The ETL process is the plumbing of the data warehouse, responsible for converting raw operational data into consistent business information[7].

- 1. Extraction:** Data is pulled from internal sources (e.g. historical costume records from Durham) and external sources (e.g. government tax incentives or film industry production calendars). To ensure no disruption to daily operations, extraction is typically scheduled during low-traffic periods, such as overnight[8].
- 2. Transformation:** This involves standardising warehouse names or client IDs to ensure a single version, mapping rental scheme codes into a uniform format and calculating the 10% or 15% financial logic. This phase ensures a single version of the truth across the enterprise.
- 3. Loading:** The cleaned and formatted data is physically loaded into the DW. Because the warehouse is non-volatile, this data is not overwritten, allowing the board to track the lifecycle of props from acquisition to write-off[6].

Integration of External Data

A primary justification for this architecture is the ability to ingest external data. By integrating UK government film tax break data with internal rental volume, the board can perform correlation analysis. This allows the company to predict which warehouses (such as Evesham for militaria) will require increased staffing or transport capacity in response to projected shifts in the UK film production landscape.

2.3 Data Marts (Targeted Insights)

While the enterprise DW provides a centralised, holistic repository of all Rags'n'Riches Ltd data, its scale can be overwhelming for individual departments seeking specific insights. To address the unique requirements of the Managing Director, Finance Director, and the Sales and Marketing Director, the architecture should employ Data Marts[9].

A data mart is a functional subset of the data warehouse, focused on a specific business unit or subject area. Following the Kimball “Bottom-Up” philosophy, these marts are populated from the central warehouse but are structured to facilitate rapid querying for particular departmental Key Performance Indicators (KPIs)[4][10].

Departmental Applications and Extraction Logic:

- Finance Data Mart:** This mart integrates incomings and outgoings, fuel card expenditures, and write-off costs. For the Finance Director, this allows for precise profitability analysis by comparing delivery charges against actual transport costs for specific warehouses, such as Freuchie or Blackburn[11]. The extraction process utilises the Fact_Rentals table filtered by Fuel_Cost_Per_Rental and Daily_Rental_Rate, providing the Finance Director with a high-speed view of overheads versus revenue.

- **Sales and Marketing Data Mart:** Focused on popularity and profitability, this mart enables the analysis of rental scheme effectiveness. By extracting data linked to Dim_Rental_Scheme and Dim_Client, the Sales and Marketing Director can evaluate whether the 10% discount on additional costumes in complete sets is generating higher volume compared to individual hires. It further identifies if the 15% emergency surcharges is deterring high-value film studio clients or if they remain price-insensitive due to tight production deadlines.
- **Operations (Managing Director) Data Mart:** This mart prioritises rationalisation data. By tracking Internal Stock Movements through the Dim_Warehouse_Location and Dim_Item dimensions, the Managing Director can identify if specific items, like pantomime costumes from Abergavenny, are frequently being shipped to other locations[4]. This empirical evidence supports strategic decisions on stock relocation or warehouse specialisation adjustments to improve profit margins and reduce the logistics burden.
- **The Extraction Paradigm:** Each mart is created via a process of dimensional filtering. Instead of providing the entire warehouse, the ETL (Extract, Transform, Load) logic slices the star schema to present only the relevant attributes. This ensures that the Managing Director is not bogged down by granular fuel card data, while the Finance Director has immediate access to the exact fiscal metrics required for the board's monthly review.

2.4 Normalised OLTP vs Dimensional Data Warehouse

The IT Director must transition from Normalisation to Dimensional Modelling. OLTP systems use highly normalised structures (3rd Normal Form) to eliminate data redundancy[11]. However, this results in a snowflake of complex, joined tables that are computationally expensive and slow to query for analytical purposes[2].

The Dimensional Model (Star Schema)

In a DW, the Star Schema is the standard for performance and usability consisting of a central Fact Table integrated with Dimension and Aggregate Tables[4].

- **The Grain:** The fundamental 'Grain' is a single rental transaction per item per day, enabling a granular level of analysis[4].
- **Fact Table:** Stores quantitative metrics of a business process; rental events are the core facts.
- **Dimension Tables:** Contain descriptive attributes that provide context, allowing users to filter the data by who, what, where and when.

Illustrative Example (Figure 2):

- **Fact_Rentals:** Contains keys to dimensions and metrics such as Daily_Rental_Rate, Discount_Amount, and Surcharge_Amount.
- **Agg_Monthly_Warehouse_Revenue:** Pre-calculated summary table for rapid executive reporting.
- **Dim_Item:** Attributes include Item_ID, Item_Name, Category and Status (defining item type as Specialist or Common Core).
- **Dim_Warehouse_Location:** Attributes include Location_Name (e.g. Colchester, Durham) and Warehouse_Specialisation.
- **Dim_Time:** Enables analysis by Day, Month, Financial Quarter, or specific Film Seasons to track industry-specific peaks.

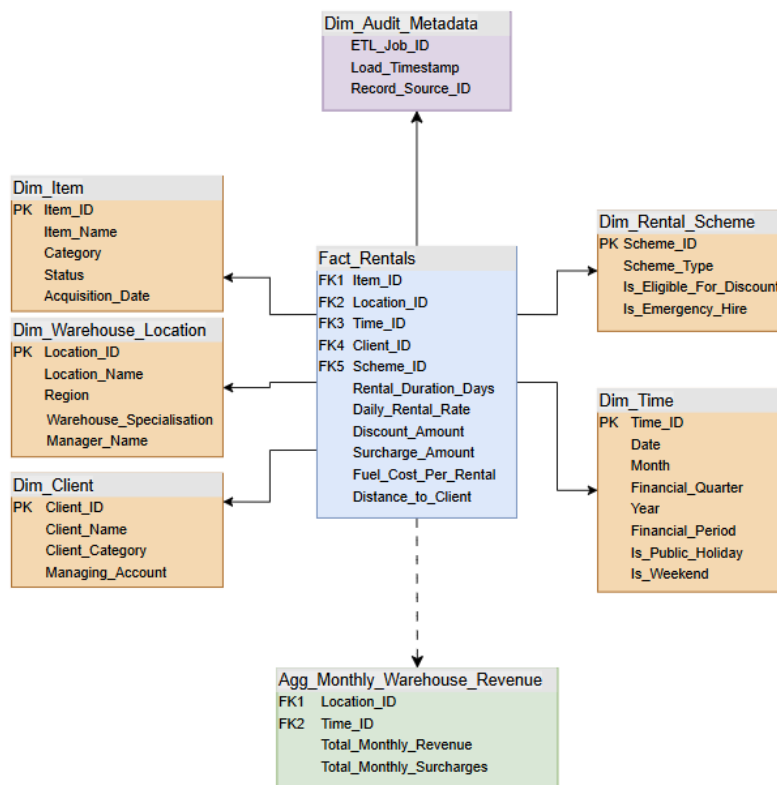


Figure 2 - Multi-Tier Star Schema

This structure is inherently intuitive for management; it removes the need for complex SQL knowledge and allows a non-technical user to easily ask: "What was the total revenue (Fact) for Historical Costumes (Dim_Item) in Durham (Dim_Location) during December (Dim_Time)?" By reducing the number of joins the warehouse provides the near-instantaneous response times required for executive decision-making.

3 Task 2

3.1 Internal and External Source Systems

Rags'n'Riches currently operates with fragmented data across multiple locations. To populate the warehouse, data must be ingested from the following:

- **Internal Sources:**
 - **Rental Orders Database:** The primary source for client transactions, item IDs, and rental durations from the Raglan office.
 - **Transport and Logistics System:** Records internal stock movements between the six warehouses and manages fuel card data to track delivery expenditure[12].
 - **Maintenance and Inventory (Tablet Logs):** Real-time data on item condition, damage assessments, and write-offs uploaded via 4G/5G links.
- **Externals Sources:**
 - **Geospatial/Mapping APIs:** Used to calculate the Distance_to_Client metric for audit and surcharge verification[13].
 - **UK Film Industry Data:** Government tax break information and production calendars to provide context for demand forecasting.

3.2 Star Schema

The proposed star schema (as illustrated in Figure 2) is designed to handle the complexity of staggered rentals, diverse inventory categories, and multi-site logistics[4]. By decoupling descriptive attributes from quantitative metrics, it provides a high-performance environment for executive decision-making[11].

The Architectural Foundation

Transitioning to this dimensional model is facilitated by the staging area. This is a critical architectural layer where the transform phase of ETL occurs[6]. It ensures that data from disparate sources, such as the tablet damage logs or the transport system, is standardised before it enters the dimension tables. This prevents dirty data (like inconsistent naming conventions for items) from compromising the integrity of the board's reports. By processing these complex calculations in staging, we protect the performance of the central warehouse[14].

Master Data Management

Master data represents the consistent, non-volatile entities of the business. In this schema, Dim_Item and Dim_Client serve as the single version of the truth. To ensure long-term stability, these dimensions utilise surrogate keys[4]. This isolates the warehouse from changes in the operational source systems (Natural Keys), ensuring that if a warehouse renames an item ID locally, the historical record in the warehouse remains intact.

- **Dim_Item:** Maintains a unique global identity for costumes. By employing SCD Type 2, the system allows the Managing Director to track an item's performance history as it moves between sites (e.g. Abergavenny and Durham) rather than just its current state.
- **Dim_Client:** By standardising client records, the system can distinguish high-frequency studios from one-off amateur dramatics groups. This master data allows the Sales and Marketing Director to perform customer lifetime value analysis, which is impossible in the current fragmented OLTP system.

Metadata Integration

Metadata, data about the data, is the glue that ensures technical and operational transparency. Within the schema, attributes like Acquisition_Date in Dim_Item and Managing_Account in Dim_Client provide immediate business context regarding recording lineage and ownership.

- **Auditability:** The ETL process will append hidden technical metadata, such as Load_Timestamp, Record_Source_ID, and ETL_Job_ID, to every row in the Fact_Rentals table. This allows the IT Director to perform precise root cause analysis if data discrepancies arise between the mobile logs and the central repository.
- **Temporal Context:** Metadata in Dim_Time, such as Is_Public_Holiday, allows the board to see if data anomalies (like spikes in emergency hires) correlate with specific external events or maintenance windows.

Aggregate Data and Performance

To support the Finance and the Sales and Marketing Directors, the system will utilise dedicated aggregate tables. In a standard warehouse, querying millions of rows in Fact_Rentals for a simple monthly report can be resource intensive.

- **Agg_Monthly_Warehouse_Revenue:** This table pre-calculates the sum of Daily_Rental_Rate minus Discount_Amount (the 10% set logic) and adds the Surcharge_Amount (the 15% emergency fee) per Warehouse_Specialisation.

- **Strategic Speed:** By storing these pre-summed values, the board can view profitability trends across different facets (e.g. Militaria vs Small Props) in near real-time. This eliminates the performance lag often seen in the Raglan administrative office's current reporting, allowing for rapid, empirical responses to market shifts.

3.3 Data Marts and the Schema

To ensure department-specific efficiency and data security, three distinct Data Marts will be extracted from the central warehouse[9]. These marts utilise a dimension view approach, where the underlying Star Schema is filtered to show only attributes relevant to a specific director's KPIs.

1. **Finance Data Mart:** This mart focuses exclusively on Incomings and Outgoings. By extracting metrics from Fact_Rentals alongside Dim_Warehouse_Location, the Finance Director can perform a direct audit of 15% emergency surcharge revenue against rising actual fuel expenditure. This allows for the identification of warehouses underperforming due to high logistics overheads.
2. **Sales and Marketing Data Mart:** Designed for market segmentation, this mart extracts Dim_Client, Dim_Rental_Scheme, and rental volume facts. It serves as a testing ground for the 10% complete set discount; the Sales and Marketing Director can evaluate if this incentive successfully shifts the customer mix from amateur theatre rentals to high-value, long-term film studios contracts.
3. **Operations Data Mart:** This mart prioritises inventory lifecycle and rationalisation. It extracts Dim_Item and Dim_Warehouse_Location to track internal stock movement orders. This visibility helps the Managing Director identify which common core items are frequently in transit, supporting the business case for relocating stock to reduce the £500,000 annual transport bill.
4. **The Extraction Paradigm:** Each mart is generated through denormalised extraction, formatting data for speed. This ensures directors access specific dashboards without needing to understand complex join logic, accelerating the better strategic decisions requested in the project brief.

3.4 Simplified ETL Plan

The ETL (Extract, Transform, Load) process occurs in a dedicated staging area to prevent disrupting the daily operations of the Raglan administrative office[6].

- **Source:** Raw data is extracted nightly from warehouse databases, fuel card logs, and mobile tablet reports[7].
- **Transformations:**

- **Logic Application:** If `Is_Eligible_For_Discount` is `TRUE`, the system automatically calculates a 10% reduction on additional items.
- **Surcharge Calculation:** If the interval between the order and rental start dates is less than 14 days, a 15% `Surcharge_Amount` is generated.
- **Data Cleaning:** This phase standardises inconsistent entries, such as correcting location codes (e.g. “ABER” becomes “Abergavenny”) and uniforms date formats.
- **Load:** Validated data is loaded into the Star Schema. This includes generating surrogate keys to maintain referential integrity and ensure the board receives a single version of truth for strategic reporting[6].

3.5 Data Quality and Information Governance

During the ETL process, data quality is maintained through NULL checks and referential integrity, ensuring every `Item_ID` exists in `Dim_Item`[15][16][17]. Information governance is critical; access to `Dim_Client` must be restricted via role-based access control for ensuring GDPR compliance[18]. Sensitive contact details and the `Managing_Account` field are secured, remaining visible only to authorised personnel[19][20]. To ensure exceptional oversight, Raglan will appoint Data Stewards, formalising operational accountability and ensuring technical security measures align with business-wide data integrity.

4 Conclusion

The implementation of a centralised data warehouse and the associated Star Schema architecture represents a strategic pivot in the operational strategy of Rags’n’Riches Ltd. By transitioning from the limitations of fragmented OLTP silos to a robust dimensional model, the organisation moves beyond basic inventory tracking toward sophisticated, prescriptive analytics.

This framework directly addresses the board’s core concerns regarding logistics overheads and warehouse rationalisation. By auditing transport costs against specific rental revenues through a single version of the truth, senior management is now equipped with the empirical evidence needed to protect and grow the £2,100,000 annual profit margin. Ultimately, this data-driven foundation ensures the company remains agile, scalable, and capable of maintaining its market-leading position within the evolving global film and theatre production industry.

5 References

- [1] K. Spanaki, D. Dennehy, T. Papadopoulos and R. Dubey, "Data-driven digital transformation in operations and supply chain management," vol. 284, pp.109599, Jun 1 2025, doi: 10.1016/j.ijpe.2025.109599. Available at: <https://dx.doi.org/10.1016/j.ijpe.2025.109599>.
- [2] D. Joiner, et al., "DW vs OLTP Performance Optimization in the Cloud on PostgreSQL (A Case Study)", pp. 1–8. [Online].doi: 10.1109/RASSE54974.2022.9989603.Available: <https://ieeexplore.ieee.org/document/9989603>
- [3] D. Loshin, *Business Intelligence*, 2nd ed. Ed. 2013. doi: 10.1016/C2010-0-67240-3[Online]. Available: <http://www.books24x7.com/marc.asp?bookid=51062>
- [4] R. Kimball and M. Ross, *The data warehouse toolkit : the definitive guide to dimensional modeling*, Indianapolis, Ind: Wiley, 2013.
- [5] S. Lyu, S. Craig, G. O'Reilly and D. Taniar, "The development and use of data warehousing in clinical settings: a scoping review," vol. 7, pp.1599514, 2025, doi: 10.3389/fdgth.2025.1599514. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/40538572>.
- [6] R. Kimball and J. Caserta, *The data warehouse ETL toolkit*, 1st Ed. Indianapolis: Wiley, 2004. [Online]. Available: <http://www.loc.gov/catdir/enhancements/fy0616/2004016909-b.html>
- [7] B. Khan, W. Khan, S. Jan and M. I. Chughtai, "An Overview of ETL Techniques, Tools, Processes and Evaluations in Data Warehousing," vol. 6, no. 1, pp.1–20, 2024, doi: 10.32604/jbd.2023.046223. Available at: <https://www.proquest.com/docview/3200128360>.
- [8] K. Maharaj and K. Kumar, "Enhancing Data Warehouse Efficiency by Optimizing ETL Processing in Near Real Time Data Integration Environment," vol. 13864, pp.289–304, 2023, doi: 10.1007/978-981-99-2233-8_21. Available at: http://ebookcentral.proquest.com/lib/SITE_ID/reader.action?docID=7243091&ppg=304&c=UERG.
- [9] L. A. Lenert, et al., "Automated Production of Research Data Marts from a Canonical Fast Healthcare Interoperability Resource (FHIR) Data Repository: Applications to COVID-19 Research," Mar 12 2021, doi: 10.1101/2021.03.11.21253384. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/33758877>.
- [10] H. Zhang, S. Ren, X. Li, H. Baharin, A. Alghamdi and O. A. Alghamdi, "Developing scalable management information system with big financial data using data mart and mining architecture," vol. 60, no. 3, pp.103326, May 2023, doi: 10.1016/j.ipm.2023.103326. Available at: <https://dx.doi.org/10.1016/j.ipm.2023.103326>.
- [11] M. Golfarelli and S. Rizzi, *Data warehouse design : modern principles and methodologies*, New York: McGraw-Hill, 2009. [Online]. Available: <https://cir.nii.ac.jp/crid/1970023484980686004>
- [12] S. Jasti and D. Ravi, "Integrating ETL with other data management approaches," vol. 6, pp.122–126, 2025, doi: 10.63282/3050-922X.IJERET-V6I2P114. Available at: <https://ijeret.org/index.php/ijeret/article/download/474/453>.

- [13] P. Koutroumpis, A. Leiponen and L. D. W. Thomas, "Markets for data," vol. 29, no. 3, pp.645–660, Jun 1 2020, doi: 10.1093/icc/dtaa002. Available at: <http://www.econis.eu/PPNSET?PPN=1701106167>.
- [14] A. Lamer, C. Saint-Dizier, N. Paris and E. Chazard, "Data Lake, Data Warehouse, Datamart, and Feature Store: Their Contributions to the Complete Data Reuse Pipeline," vol. 12, pp.e54590, Jul 17 2024, doi: 10.2196/54590. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/39037339>.
- [15] C. Batini and M. Scannapieco, *Data and Information Quality: Concepts, Instrumentation and Techniques*. 2016. doi: 10.1007/978-3-319-24106-7[Online]. Available: https://bvbr.bib-bvb.de:443/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=028800653&sequence=000002&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA
- [16] T. Nguyen, H. Nguyen and T. Nguyen-Hoang, "Data quality management in big data: Strategies, tools, and educational implications," vol. 200, pp.105067, Jun 2025, doi: 10.1016/j.jpdc.2025.105067. Available at: <https://dx.doi.org/10.1016/j.jpdc.2025.105067>.
- [17] H. Hassani, X. Huang and S. MacFeely, "Mapping the evolution of data governance scientific research," vol. 7, Jan 1 2025, doi: 10.1017/dap.2025.10014. Available at: <https://www.proquest.com/docview/3231771706>.
- [18] Data Management Association International, *DAMA-DMBOK : data management body of knowledge : print ed*, 2nd edition, revised Ed. Technics Publications, 2024.
- [19] M. J. Walsh, J. McAvoy and D. Sammon, "The Data Governance Journey in Practice: Insights from Case Study Research," vol. 42, no. 3, pp.471–488, Jun 1 2025, doi: 10.1080/10580530.2025.2477459. Available at: <https://www.tandfonline.com/doi/abs/10.1080/10580530.2025.2477459>.
- [20] B. M. V. Bernardo, H. P. S. Mamede, J. M. P. Barroso and V. Santor, "Data governance & quality management: Innovation and breakthroughs across different fields," vol. 9, no. 4, pp.1–35, Oct 1 2024, doi: 10.1016/j.jik.2024.100598. Available at: <https://www.econstor.eu/handle/10419/327500>.