FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Data-Driven Personalization of Digital Interactions Using Communication and Behavioural Metadata to Enhance Efficiency in Debt Collection

João António Teixeira Coelho

Mestrado em Engenharia Informática e Computação

Supervisor: Prof. Luís Paulo Reis

Industrial Supervisor: Eng. Pedro Mendes

July 26, 2026

Data-Driven Personalization of Digital Interactions Using Communication and Behavioural Metadata to Enhance Efficiency in Debt Collection

João António Teixeira Coelho

Mestrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

President: Prof. Mariana Malta
Examiner: Prof. Paulo Novais
Supervisor: Prof. Luís Paulo Reis

July 26, 2026

Resumo

O atraso no pagamento é um desafio persistente para as organizações que dependem da regularização atempada de valores em dívida, e as práticas de comunicação usadas na cobrança de dívidas em atraso são frequentemente uniformes entre devedores e ao longo do ano, aplicando o mesmo calendário de contacto independentemente do momento em que os destinatários estão mais aptos a agir. O trabalho existente tende a abordar partes isoladas do problema, sendo limitada a evidência de sistemas integrados que reúnam análise temporal, lógica de decisão, agendamento de comunicações e conformidade regulatória num único fluxo.

Esta dissertação apresenta um sistema de apoio à decisão que personaliza o momento e a cadência das comunicações de cobrança através do alinhamento temporal, e não da otimização causal. O sistema segue uma arquitetura em quatro camadas: ingestão e pré-processamento de dados de pagamento heterogéneos; extração de características temporais através de decomposição sazonal-tendência e deteção de picos; lógica de decisão filtrada por restrições, que propõe janelas e cadências de contacto dentro de limites regulatórios e comportamentais; e uma camada de saída que fornece janelas, níveis de confiança, explicabilidade e um registo de auditoria. A abordagem é avaliada sobre duas fontes complementares: 87 meses de atividade de pagamento agregada do SIBS MULTIBANCO, que capturam a sazonalidade ao nível da população, e registos operacionais de mensagens e pagamentos do Invisible Collector.

As janelas propostas alcançam um Temporal Alignment Score (TAS) médio de 0,485, face a 0,21 de uma baseline heurística pericial e 0,42 de uma baseline uniforme, satisfazendo todas as restrições codificadas por construção. A análise revela dois regimes sazonais de pagamento distintos e indica que alinhar o contacto com períodos de maior intensidade de pagamento observada está associado a diferenças mensuráveis na resposta observada. Um estudo de utilidade percebida junto de devedores e gestores complementa a avaliação quantitativa, avaliando equidade e intrusividade.

Palavras-chave: cobrança de dívidas • alinhamento temporal • decomposição sazonal • sistema de apoio à decisão • modelos de decisão heurísticos • cadência de comunicação • metadados comportamentais • observabilidade parcial

Abstract

Late payment is a persistent challenge for organisations that rely on being paid on time, and the communication practices used to recover overdue debts are frequently uniform across debtors and across the year, applying the same contact schedule regardless of when recipients are most able to act. Existing work tends to address isolated parts of the problem, and the literature provides limited evidence of deployable end-to-end systems that integrate temporal analysis, decision logic, communication scheduling, and regulatory compliance within a single pipeline.

This dissertation presents a decision-support system that personalises the timing and cadence of debt-collection communications through temporal alignment rather than causal optimisation. The system follows a four-layer architecture: ingestion and preprocessing of heterogeneous payment data; extraction of temporal features through seasonal-trend decomposition and peak detection; constraint-filtered decision logic that proposes contact windows and cadences within regulatory and behavioural limits; and an output layer providing windows, confidence levels, explainability, and an audit trail. The approach is evaluated on two complementary sources: 87 months of aggregated SIBS MULTIBANCO payment activity, capturing population-level seasonality, and operational message and payment records from Invisible Collector.

The proposed windows achieve a mean Temporal Alignment Score (TAS) of 0.485, against 0.21 for an expert heuristic baseline and 0.42 for a uniform baseline, while satisfying all encoded constraints by construction. The analysis reveals two distinct seasonal payment regimes and indicates that aligning contact with periods of higher observed payment intensity is associated with measurable differences in observed response. A perceived-utility study with debtors and managers complements the quantitative evaluation by assessing fairness and intrusiveness.

Keywords: debt collection • temporal alignment • seasonal decomposition • decision support system • heuristic decision models • communication cadence • behavioural metadata • partial observability

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework for achieving a better and more sustainable future for all. It includes 17 goals[1] to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice. This dissertation contributes most directly to the goals concerned with innovation, equitable treatment, and accountable institutions. By developing a data-driven decision-support system for debt-collection communication that operates within regulatory and behavioural constraints, the work advances technological innovation (SDG 9) while promoting fairer, less intrusive treatment of debtors (SDG 10) and transparent, accountable collection practices (SDG 16). It also supports economic resilience by improving the efficiency of receivables recovery, with particular relevance to small and medium enterprises that are most exposed to late payment (SDG 8). The specific Sustainable Development Goals mentioned have the following names:

SDG 8 Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

SDG 9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

SDG 10 Reduce inequality within and among countries

SDG 16 Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels

[1] https://sdgs.un.org/goals

SDG	Target	Contribution	Performance Indicators and Metrics
8	8.3	More efficient receivables recovery aims to support creditor cash flow, disproportionately benefiting micro, small and medium enterprises most exposed to late payment.	Potential reduction in days-to-resolution, to be evaluated prospectively; observed response-rate differences relative to uniform and expert baselines under retrospective attribution.
	8.10	Aligning collection actions with periods of higher observed payment responsiveness aims to strengthen the operational capacity of financial-services providers managing overdue receivables.	Observed temporal response-rate differences under retrospective attribution.
9	9.5	The dissertation delivers a deployable decision-support system integrating temporal analysis, decision logic and compliance constraints, addressing an identified research gap.	Mean Temporal Alignment Score (TAS) of 0.485 against 0.21 for the expert baseline; reproducible evaluation metrics (precision, coverage, Constraint Compliance Rate (CCR)).
10	10.2	Prioritising proportional, non-intrusive contact aims to reduce over-contacting of financially vulnerable debtors and promote more inclusive treatment.	Perceived intrusiveness and fairness scores; contact-frequency caps enforced under the encoded `ConstraintPolicy`.
	10.3	Fairness and respect for debtor rights are explicitly evaluated, aiming to support more equitable collection outcomes across debtor profiles.	Potential reduction in unnecessary contacts under configured policy constraints; perceived-fairness scores across compared strategies.
16	16.6	Compliance-by-design, with an audit trail and traceable outputs, supports accountable and transparent collection practices.	CCR of 1.00 with respect to encoded constraints; audit-log completeness; traceability coverage of issued recommendations.

Acknowledgements

This dissertation was developed in close collaboration with Invisible Cloud, and I owe a great deal to Miguel Rangel and Pedro Mendes. Their advice and guidance were a constant source of insight throughout the work, and the access they gave me to a real operational setting shaped this thesis in ways that would not otherwise have been possible.

I am also deeply grateful to Professor Luís Paulo Reis for his expertise and guidance. His direction helped me frame the problem rigorously and kept the work grounded whenever it risked drifting.

To my great friends, who were always there for me through every stage of this journey, thank you for the support, the patience, and the welcome distractions.

Finally, to my parents, Isabel and António, my brother Eduardo and my family, thank you for always putting me first and for making me the person I am today. None of this would have been possible without you.

João Coelho

Declaration of honour

I declare that the present dissertation is my own work and has not been previously submitted or assessed at this or any other institution. References to other authors (statements, ideas, thoughts), as well as the use of published works, scrupulously respect the rules of attribution, and are duly indicated in the text and in the bibliographical references, in accordance with the referencing standards. I am aware that the practice of plagiarism or self-plagiarism constitutes an academic offence, as established in the Code of Ethics of the University of Porto.
I further declare that I used Artificial Intelligence tools in the preparation of part(s) of the present dissertation, and that such use and its results are duly indicated and were subject to careful verification for the detection of errors, inaccuracies, unfounded attributions, or other forms of bias in content and arguments, as well as the determination of authorship.

João António Teixeira Coelho

“If everything seems under control, you’re not going fast enough”

Mario Andretti

Contents

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence

NLP Natural Language Processing

LLM Large Language Model

ML Machine Learning

CX Customer Experience

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RQ Research Question

FinTech Financial Technology

SMS Short Message Service

NPL Non-Performing Loan

FD Financial Data

CCD Contact Center Data

HMM Hidden Markov Model

VaDE-Seq2Seq Variational Deep Embedding with Sequence-to-Sequence

IVR Interactive Voice Response

GDPR General Data Protection Regulation

CCD2 Consumer Credit Directive II

MDP Markov Decision Process

RL Reinforcement Learning

FRS-DRL Flexible Recommendation System using Deep Reinforcement Learning

FDCPA Fair Debt Collection Practices Act

EBA European Banking Authority

PSD2 Payment Services Directive 2

DSR Design Science Research

TAS Temporal Alignment Score

ACAQL Adversarial Conservative Alternating Q-Learning

NACE Nomenclature des Activités Économiques dans la Communauté Européenne

SIBS SIBS MULTIBANCO

IC Invisible Collector

CCR Constraint Compliance Rate

STL Seasonal and Trend Decomposition using Loess

ETL Extract, Transform, Load

RR Response Rate

CTA Call to Action

CSV Comma-Separated Values

VMS Voice Message Service

SotA State of the Art

P50 Median

UI User Interface

Chapter 1

Introduction

Debt arises when goods, services, or money are provided but not paid for on time, whether immediately or by an agreed later date. The origins of credit and debt management trace back to ancient civilizations, with evidence of monetary credit practices emerging around 2500 BC among the Babylonians and in Ancient Egypt. Debt collection mechanisms were already present in Sumer around 3000 BC, where failure to repay often resulted in forms of "debt slavery" involving not only debtors but also their families or servants [18]. Over the centuries, these early practices evolved into more structured financial systems. By the late 15th century, with the rise of the Bourgeois class, credit became institutionalized and widely recorded through accounting methods [34]. In recent decades, the collection sector has not seen the same level of development and sophistication as lead generation, customer acquisition, and after-sales service [30]. Existing market solutions in the technological context are still characterized by significant manual processes and ad hoc decision-making, where collectors rely on fixed schedules and personal experience rather than behavioral analytics. Even large financial institutions continue to manage collection workflows in ways "they did in the past, with paper and pencil" [2]. Early studies already explored the potential of applying machine learning and reinforcement learning to debt collection [1, 52], yet the practical integration of these models into commercial workflows remained limited. More recent works continue to highlight the limited use of predictive modeling and behavioral features in credit management, despite notable advances in data science, automation, and content generation [39, 42, 43].

1.1 Motivation

Debt collection is not only a financial process but also a relational one, as it directly involves human behavior, trust, and communication [44]. As societies and markets become increasingly digital, organizations face the challenge of preserving empathy and fairness while also achieving operational efficiency [22]. The motivation for this research stems from the need to better reconcile automation with personalization, understood here not as linguistic customization but as the careful management of communication timing, frequency, and sequencing. The goal is to ensure that technological innovations **improve the quality of human interaction in financial recovery rather than make it more impersonal or stressful for the people involved**.

The Invisible Collector platform (`https://www.invisiblecollector.com`), owned by the proposing company, provides the technological context for this research. The platform enables automated communication with end users to support collection processes through configurable contact policies, temporal rules, and behavioral segmentation, while seeking to maintain balanced relationships between creditors and debtors. Within this project, the platform will serve as both an operational and experimental environment, supporting the analysis of debtor behavior, the exploration of adaptive *timing and cadence* strategies, and the evaluation of data-driven decision-support mechanisms.

From a scientific standpoint, this research combines ideas from artificial intelligence, behavioral science, and ethics to design debt collection strategies that are both effective and human-centered. Here, empathy guides the design as an ethical principle rather than as a matter of message wording, shaping contact policies that are non-intrusive, fair, and compliant. By aligning data-driven methods with behavioral and regulatory constraints, this work aims to support a more responsible approach to debt settlement.

1.2 Problem Statement

Despite significant advances in automation and digitalization, debt collection systems still lack integrated models that can dynamically adapt *when* and *how often* to contact debtors based on behavioral, temporal, and contextual information. In practice, collection workflows are still largely driven by static schedules, manual heuristics, or broad segmentation rules. They give little weight to income cycles, liquidity constraints, past interaction patterns, or regulatory limits on how often debtors can be contacted.

Although predictive models for default risk and repayment propensity have been widely studied, their integration into operational decision-making processes, particularly those governing contact timing, cadence, and sequencing, remains fragmented. As a result, organizations often fail to align collection actions with periods of higher debtor responsiveness, leading to inefficient recovery efforts, unnecessary friction, and an increased risk of non-compliance.

This work addresses this gap by proposing the development of an intelligent debt collection strategy engine focused on aligning communication timing and cadence with periods of higher observed payment responsiveness. The proposed approach uses temporal and behavioral metadata to support decision-making under operational, ethical, and legal constraints, with the aim of improving recovery efficiency while minimizing intrusiveness and preserving balanced creditor–debtor relationships, avoiding the slow loss of trust that poorly timed or excessive contact can cause [5, 41].

1.2.1 Scope of the Proposed Engine

This thesis develops a *deployable decision-support engine* for debt collection, focused on aligning communication timing and contact cadence with periods of higher observed payment responsiveness under conditions of partial observability.

The proposed engine is designed to:

- Consume aggregated payment transaction data structured along temporal, sectoral, geographic, and channel dimensions;
- Apply statistical and inferential methods to extract temporal structure, including seasonality, liquidity cycles, and payment concentration patterns;
- Generate actionable recommendations for communication timing, contact windows, and cadence constraints aligned with inferred periods of higher repayment responsiveness;
- Operate under explicit regulatory and ethical constraints, including contact frequency limits and data protection requirements;
- Produce auditable and explainable outputs grounded in interpretable decision logic rather than opaque learned policies.

The engine does *not*, in its current scope:

- Learn policies from individual action–outcome feedback loops, which would require micro-level interaction logs unavailable for this research;
- Claim formal optimality guarantees in a Markov Decision Process sense, due to partial state observability and the absence of attributed rewards;
- Perform individualized message content generation or linguistic personalization, as message content is treated as an exogenous, expert-curated input assumed to be legally compliant, ethically appropriate, and sufficiently informative to prompt action, rather than a decision variable optimized by the engine.
- Infer intra-day timing windows (e.g., hour-of-day contact optimization), as the primary data sources operate at monthly aggregation granularity; day-offset cadence within a dunning sequence is supported, but hour-level recommendations require timestamped interaction logs beyond the current baseline scenario.

Despite these constraints, the engine is *deployable in practice*: it produces concrete timing and cadence recommendations that can be integrated into existing collection workflows as a decision-support layer. The architecture is explicitly designed to be extensible, allowing future integration of micro-level interaction data and policy learning mechanisms when such data becomes available.

1.3 Hypothesis

The following hypothesis is formulated from the problem statement:

The personalization of communication timing and contact cadence, based on temporal and behavioral metadata, supports faster resolution of debt-related disputes by aligning contact actions with periods of higher observed payment responsiveness without increasing perceived intrusiveness or deteriorating the debtor's relationship with the creditor.

This hypothesis assumes that aligning collection actions with periods of higher debtor responsiveness, such as liquidity cycles, historical interaction patterns, and engagement states, can positively influence recovery outcomes.

1.4 Research Questions

Based on the hypothesis, the following research questions are formulated. They aim to investigate how temporal and behavioral metadata can be used to design adaptive, efficient, and ethically compliant debt collection strategies, with a particular focus on communication timing, cadence, and regulatory constraints.

RQ1: Which temporal, sectoral, and geographic features derived from aggregated payment dynamics are most informative for inferring high-responsiveness contact windows in debt collection processes?

RQ2: How can data-driven models infer effective contact timing windows that align with periods of higher observed payment intensity and historical collection responsiveness, under conditions of partial observability?

RQ3: How can contact cadence and sequencing be adapted dynamically to balance recovery effectiveness with intrusiveness and regulatory compliance?

RQ4: How can heuristic-based decision models support adaptive communication strategies under ethical, operational, and legal constraints?

RQ5: How can the impact of a deployable, timing-based debt collection strategy engine be evaluated under partial observability, using temporal proxies, simulation, and counterfactual reasoning in the absence of fully attributed action–outcome data?

RQ6: How do regulatory and compliance constraints shape the feasible policy space and performance trade-offs of a deployable automated debt collection strategy engine?

1.5 Document Structure

The remainder of this dissertation is organised as follows. Chapter 2 describes the research methodology of the literature review, including the search strategy, data sources, and study-selection process. Chapter 3 surveys the state of the art, framing debt collection as a digital communication problem and reviewing predictive modelling, debtor segmentation, communication-strategy optimisation, and decision models for adaptive collection, before consolidating the research gaps that motivate this work. Chapter 4 presents the methodology and implementation, detailing the system architecture, the macro- and micro-level analytical layers, the decision layer, and the evaluation framework. Chapter 5 reports the results and evaluation across the three analytical layers, together with the quantitative and perceived-utility validation. Chapter 6 summarises the main findings, revisits the research questions and hypothesis, and outlines directions for future work.

Chapter 2

Literature Review

This chapter presents the Systematic Literature Review conducted to identify, select, and synthesize peer-reviewed research relevant to debt collection processes, communication personalization, and AI-driven decision support in financial contexts. The review follows the PRISMA 2020 guidelines and produced a corpus of 23 studies that inform the State of the Art presented in Chapter 3.

2.1 Research Methodology

The Literature Review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 [37] guidelines for systematic reviews, which support transparency and reproducibility. The objective was to identify, select, and synthesize peer-reviewed work relevant to (i) debt collection and credit recovery processes, (ii) personalization and communication strategies, and (iii) artificial intelligence, machine learning, and communication metadata applied to financial interactions.

2.1.1 Search Strategy and Data Sources

A structured search was performed across two major scientific databases widely recognized for coverage in computer science, information systems, and applied economics:

- ***Scopus*** [46]
- ***IEEE Xplore*** [23]

Scopus is a comprehensive abstract and citation database covering a wide range of fields including engineering, business, decision sciences, and artificial intelligence.

IEEE Xplore hosts high-impact publications in electrical engineering, computer science, machine learning, and information technologies. It is particularly strong for Artificial Intelligence (AI) and predictive modeling research.

Each database required customized queries adapted to its search features and indexing structure. The search strings combined concepts related to debt collection, credit recovery, machine learning, AI-driven personalization and digital communication optimization. Boolean operators, field filters (title, abstract, keywords), and subject restrictions were applied where supported.

All database searches were performed using institutional access provided by the Universidade do Porto between 14/12/2025 and 05/01/2026. Database versions and access dates were recorded to ensure reproducibility and to document the exact state of the indexing systems at the time of the review.

A detailed description of the queries used for each database is presented in Subsection 2.1.2.

2.1.2 Search Queries

The search strategy was designed to balance precision and recall while capturing the multidisciplinary nature of AI-driven personalization in debt-related financial contexts. Rather than relying on a single literal formulation such as "debt collection", the queries deliberately incorporated a broader set of financial, behavioral, and communication-related terms commonly used in the literature to describe repayment processes, delinquency, and customer interaction.

This approach reflects established practice in financial analytics and behavioral research, where functionally equivalent problems are framed using diverse terminology, including *credit risk*, *loan default*, *payment behavior*, *repayment probability*, and *financial delinquency*. Similarly, personalization and communication optimization are often discussed through constructs such as *engagement prediction*, *customer interaction modeling*, *digital nudging*, *reminder strategies*, or *persuasive communication*, without explicit reference to personalization as a standalone concept.

Two complementary query families were defined to cover both the predictive and communicative dimensions. The first focuses on behavioral prediction and repayment modeling using machine learning techniques, while the second targets communication and interaction strategies in debt- and credit-related contexts. Together, these queries capture both the analytical foundations and the applied personalization mechanisms relevant to modern debt collection systems.

Query A: Behavioral Prediction and Repayment Modeling

Query A targets literature addressing payment behavior, delinquency, and repayment outcomes through machine learning and behavioral modeling approaches.

Scopus

```
TITLE-ABS-KEY (
 ( "payment behavior" OR "repayment behavior"
   OR "repayment probability" OR "default prediction"
   OR "financial delinquency" OR arrears
   OR "debt collection" OR "collection strategy" )
 AND
 ( "machine learning" OR "deep learning"
   OR "predictive modeling" OR "behavioral scoring"
   OR "behavioral modeling" OR classification OR clustering )
)
```

```
AND PUBYEAR > 2014 AND PUBYEAR < 2026
```

IEEE Xplore

```
("payment behavior" OR "repayment behavior"
 OR "repayment probability" OR "default prediction"
 OR "financial delinquency")
AND
("machine learning" OR "deep learning"
 OR "predictive modeling" OR "behavioral scoring"
 OR "behavioral modeling" OR classification OR clustering)
```

Query B: Communication and Interaction Strategy Optimization

Query B focuses on studies that explicitly examine customer interaction, engagement, and communication strategies in financial delinquency or credit-risk settings. This query emphasizes the application of predictive and optimization techniques to contact strategies, reminders, and persuasive communication.

Scopus

```
TITLE-ABS-KEY (
 ( "credit risk" OR "loan default"
   OR "debt collection" OR arrears
   OR "financial delinquency" )
 AND
 ( "customer interaction" OR "engagement prediction"
   OR "contact strategy" OR "digital nudging"
   OR "persuasive communication" OR "reminder"
   OR "notification" OR "communication optimization" )
 AND
 ( "machine learning" OR "predictive"
   OR "optimization" OR "personalization" )
)
AND PUBYEAR > 2014 AND PUBYEAR < 2026
```

IEEE Xplore

```
(("credit risk" OR "loan default"
  OR "debt collection" OR "payment delinquency")
 AND
 ("customer interaction" OR "engagement prediction"
  OR "contact strategy" OR "digital nudging"
  OR "persuasive communication" OR "reminder strategy")
 AND
```

```
("machine learning" OR "predictive model"
 OR "behavioral model"))
```

Structuring the search around both predictive modeling and communication-oriented optimization covers the AI and machine learning approaches most relevant to personalized, ethical, and effective debt collection.

2.1.3 Inclusion and Exclusion Criteria

A set of inclusion and exclusion criteria was applied during full-text screening to keep the selected studies relevant and consistent. These criteria were designed to identify empirical research addressing AI-driven personalization of digital communication strategies in debt-related financial contexts.

Inclusion Criteria (IC):

1. **IC1 – Domain and context relevance:** The study addresses digital communication or interaction systems in contexts involving creditor–debtor relationships, payment recovery, financial obligation management, or analogous scenarios requiring persuasive or restorative communication under regulatory or institutional constraints.

2. **IC2 – Personalization objective:** The study proposes or evaluates personalized or adaptive communication or collection strategies in which one or more decision dimensions vary according to user or contextual characteristics, such as contact timing, communication cadence, sequencing of actions, contact channel, or action intensity. Studies focused primarily on linguistic content, message wording, tone, or framing are excluded unless these are secondary to a temporal or sequential decision strategy.

3. **IC3 – Use of communication or transactional metadata:** The study leverages communication metadata (e.g., interaction history, temporal patterns, engagement signals) and/or transactional or behavioral data (e.g., payment history, settlement behavior, financial consistency) to model debtor behavior or optimize communication and contact strategies.

4. **IC4 – AI-driven decision-making:** The study employs artificial intelligence or computational intelligence techniques, such as machine learning, predictive modeling, heuristics, optimization methods, or sequential decision models, to support or automate decisions related to communication timing, cadence, or strategy selection.

5. **IC5 – Empirical or applied validation:** The work includes empirical evaluation, experimental results, simulations, case studies, or evidence of real-world deployment, enabling assessment of the proposed approach.

Exclusion Criteria (EC):

1. **EC1 – Language exclusion:** The publication is not written in English or Portuguese.

2. **EC2 – No full-text access:** The full text of the publication is unavailable after reasonable access attempts.

3. **EC3 – Absence of personalization:** The study focuses on generic automation, bulk communication, or uniform treatment strategies, without adaptive or user-specific personalization.

4. **EC4 – Non-communication-focused AI:** The AI application is unrelated to digital communication optimization, instead addressing areas such as user interface customization, general recommendation systems, or unrelated decision-support tasks.

5. **EC5 – Conceptual or low-evidence studies:** The work is purely theoretical, opinion-based, or lacks sufficient methodological detail, validation, or evaluation to support its conclusions.

Publications were limited to works published from 2015 onward. Restricting the corpus to this period improves relevance and makes the studies more comparable, reflecting the maturation of machine learning–based personalization, data-driven decision-making, and behavioral modeling in financial contexts.

2.1.4 Alignment Between Inclusion Criteria and Research Questions

The inclusion criteria were defined to align the selected studies clearly with the research questions guiding this work (Section 1.4). Each criterion reflects one or more analytical dimensions required to examine how temporal, behavioral, and regulatory factors influence the design and effectiveness of debt collection communication strategies.

- **IC1 – Domain and context relevance** keeps the selected studies situated within creditor–debtor relationships or comparable financial obligation scenarios. This criterion primarily supports **RQ5** (measurable impact on repayment outcomes and relationship preservation) and **RQ6** (regulatory influence on process efficiency), by anchoring the review in realistic, regulated financial communication environments.

- **IC2 – Personalization objective** captures personalization as a dynamic, context-dependent process focused on adapting communication timing, cadence, sequencing, or channel selection. It is closely aligned with **RQ1** (features indicative of debtor responsiveness), **RQ2** (inference of effective contact timing windows), and **RQ3** (dynamic adaptation of contact cadence and sequencing).

- **IC3 – Use of communication or transactional metadata** focuses on studies that leverage behavioral, temporal, or transactional signals derived from interaction histories or payment behavior. This directly informs **RQ1** by identifying predictive features, and **RQ2** and **RQ3** by supporting data-driven timing and cadence decisions.

- **IC4 – AI-driven decision-making** limits the review to studies that employ computational intelligence techniques to support adaptive decision-making related to communication strategies. This criterion contributes to **RQ2** and **RQ3** through data-driven inference and optimization, and to **RQ4** by examining heuristic-based decision models operating under ethical, operational, and legal constraints.
- **IC5 – Empirical or applied validation** requires that the selected studies include empirical evaluation, experimental results, simulations, or real-world evidence. This supports **RQ5** by enabling assessment of the measurable impact of timing and cadence personalization, and reinforces the overall empirical grounding of the review.

Overall, this alignment means that each included study contributes to one or more research questions, strengthening the coherence, relevance, and analytical depth of the systematic literature review while remaining consistent with the defined scope of this work.

2.1.5 Study Selection Process

The study selection followed the PRISMA 2020 workflow and consisted of four stages: identification, screening, eligibility assessment, and inclusion. During the screening stage, titles and abstracts were reviewed to assess preliminary relevance. During the eligibility assessment stage, full-text articles were retrieved and evaluated against the complete set of inclusion and exclusion criteria (Section 2.1.3). Records retrieved from all databases were exported, merged, and deduplicated before screening. No protocol was preregistered for this review.

Snowballing Approach

Backward snowballing was conducted using the final set of studies that met all inclusion and exclusion criteria after full-text screening. The reference lists of these included studies were examined to identify additional relevant work. Any newly identified papers were retrieved and evaluated using the same eligibility criteria defined in Section 2.1.3.

This process was applied iteratively until no further eligible studies were identified. Snowballing was deliberately limited to the full-text–included studies in order to ensure conceptual relevance and methodological consistency, while minimizing the inclusion of tangential or weakly related work. The additional studies identified through this process contributed to consolidating the final dataset and reflect the interdisciplinary nature of research on AI-driven personalization in debt-related financial communication.

Scopus

The Scopus search returned a total of **758 records**, comprising **753** results from Query A and **5** results from Query B. All records satisfied the publication year criterion ($\geq$2015) and were subjected to title and abstract screening to assess relevance to the scope of this review.

IEEE Xplore

The IEEE Xplore search returned a total of **163 records**, consisting of **159** results from Query A and **4** results from Query B. All records satisfied the publication year criterion ($\geq$2015) and were subjected to title and abstract screening to evaluate their relevance to the scope of this review.

Duplicate Removal

All records retrieved from the selected databases were imported into Mendeley for reference management and deduplication. Automatic duplicate detection was applied, followed by a manual verification to ensure accuracy. Out of an initial set of **921 records**, **87 duplicates** were identified and removed. This process resulted in **834 unique records**, which were carried forward to the screening phase.

Screening Stage

During the screening stage, titles and abstracts of the 834 unique records were reviewed to assess their preliminary relevance to the objectives of this review. Screening decisions were guided by the inclusion and exclusion criteria defined in Section 2.1.3, with particular emphasis on identifying studies that went beyond static risk modeling and contributed to actionable, behavior-aware decision support in debt-related communication contexts.

A study was retained at this stage only if the title or abstract clearly indicated at least one of the following characteristics:

- **A. Behavioral or temporal modeling beyond static scoring**, such as payment behavior over time, repayment trajectories, overdue timing windows, survival analysis, sequence or time-series models (e.g., LSTM, GRU), or behavior change detection. Studies limited to one-shot or static classifiers were excluded.

- **B. Decision support for collection, prioritization, or action**, including invoice payment prediction, debt collection prioritization, arrears management, assignment or resource optimization, or risk-based intervention. Papers focused solely on model accuracy comparison without operational or decision-support implications were excluded.

- **C. Explicit interaction, communication, or intervention context**, such as reminders, nudging, engagement modeling, customer interaction, payment advice, personalized services, or timing of actions. Studies addressing credit risk without clear actionability or communication relevance were excluded.

- **D. Use of non-traditional or behavioral data**, including transaction sequences, consumption behavior, usage patterns, soft information, or customer activity data. Studies relying exclusively on financial ratios or balance-sheet variables were excluded.

During this phase, papers whose primary contribution consisted of benchmarking machine learning models, feature selection exercises, or comparative accuracy studies (e.g., "loan default prediction using XGBoost", "ensemble models improving AUC", or generic "credit risk assessment frameworks") were systematically excluded, even when advanced Machine Learning (ML) techniques were employed. This ensured alignment with the thesis focus on AI-driven personalization, behavioral modeling, and interaction optimization.

Following title and abstract screening, **65 studies** were retained for full-text retrieval and eligibility assessment.

Eligibility Assessment Stage

Full-text retrieval was attempted for the **65 studies** retained after the screening stage. Each retrieved full-text article was assessed against the complete set of inclusion and exclusion criteria defined in Section 2.1.3.

During this phase, the following exclusions were applied:

- **2 studies** were excluded due to language constraints, as they were not written in English or Portuguese (EC1).
- **8 studies** were excluded because full-text access could not be obtained after reasonable retrieval attempts (EC2).
- **25 studies** were excluded for failing to meet both the personalization and communication focus requirements (EC3 and EC4), typically due to reliance on static prediction models or lack of interaction-oriented decision support.
- **5 studies** were excluded for failing EC4, as they applied AI techniques unrelated to digital communication or contact optimization.
- **1 study** was excluded for failing EC5, due to insufficient methodological detail or lack of empirical validation.

After applying these criteria, **23 studies** remained eligible and were included in the final qualitative synthesis. A substantial number of additional papers identified during the search and screening phases, while excluded from the systematic review due to the strict personalization and communication-focused criteria, were retained for use as supporting material in the related work and background sections, where they provide complementary theoretical, methodological, and domain context.

Final Dataset

After the eligibility assessment, **18 studies** from the database searches were included. Additionally, **5 studies** were identified through backward snowballing. These 5 external sources were evaluated using the same eligibility criteria and met all requirements for inclusion.

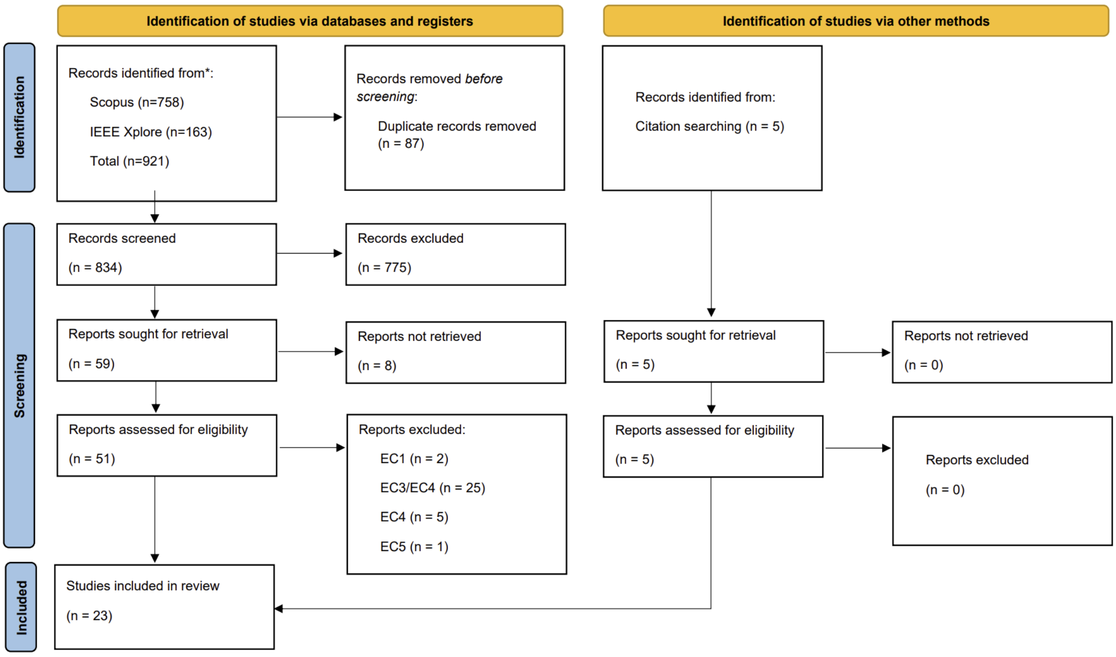

Figure 2.1: PRISMA 2020 flow diagram illustrating the systematic literature review process.

The final dataset thus consisted of **23 studies** (**18** from database searches + **5** from external sources) that were selected for qualitative synthesis and will be analyzed in detail in the State of the Art chapter.

The complete selection and filtering process is summarized in the PRISMA 2020 flow diagram (Figure 2.1), which documents the number of records at each stage of identification, screening, eligibility assessment, and final inclusion.

Beyond the systematic literature review, this thesis benefits from access to empirical operational data from the Invisible Collector platform, comprising historical communication logs and payment records. This dataset provides direct evidence of collection outcomes that complements the academic literature and addresses the gap identified in prior work regarding the limited availability of real-world interaction data in debt collection research.

Chapter 3

State of the Art

3.1 Scope and Organization of the State of the Art

This chapter presents a critical and thematic synthesis of the studies selected through the Systematic Literature Review described in Chapter 2. Rather than listing individual contributions, the goal of this State of the Art is to integrate and analyze the existing literature from a conceptual and analytical perspective, with a focus on identifying key foundations, limitations, and research gaps relevant to data-driven debt collection.

The chapter frames debt collection not simply as a financial optimization task, but as a problem of digitally mediated communication and sequential decision-making under behavioral, ethical, and regulatory constraints. Within this perspective, personalization is understood as the adaptive orchestration of communication timing, cadence, and interaction strategies based on temporal and behavioral metadata, rather than as linguistic or content-level customization.

The organization of the chapter reflects both the historical evolution of the sector and a logical decomposition of decision-making processes in modern debt collection systems. Section 3.2 positions debt collection as a digital communication problem, highlighting the growing gap between traditional collection practices and contemporary data-driven interaction paradigms. Sections 3.3 and 3.4 then review predictive modeling and debtor segmentation approaches, examining how temporal, behavioral, and socio-demographic features are used to infer debtor responsiveness and payment dynamics.

Building on this foundation, Section 3.5 examines communication strategy optimization, with a focus on contact timing, cadence, and sequencing. This includes a discussion of digital nudging and interaction design techniques used to influence repayment behavior. These interaction-level strategies are analyzed under constraints such as intrusiveness, compliance, and behavioral fatigue, without addressing linguistic content generation.

Section 3.6 then reviews decision-oriented models for adaptive collection strategies, including heuristic approaches, optimization frameworks, and sequential decision-making methods. These approaches are evaluated in terms of their ability to combine predictive signals with operational and regulatory constraints, enabling dynamic strategy selection over time.

The structure of this chapter is aligned with the research questions formulated in Chapter 1. Sections 3.3 and 3.4 primarily support Research Question (RQ)1 by reviewing how temporal and behavioral patterns are used in the literature to infer responsiveness, providing the conceptual basis for applying similar principles to aggregated payment data. Section 3.5 directly informs RQ2 and RQ3 by reviewing evidence and models for identifying effective contact timing windows, and by analyzing how cadence and sequencing can be adapted dynamically under behavioral limits (e.g., annoyance costs), operational constraints, and compliance requirements. Section 3.6 supports RQ4 by examining decision-oriented approaches, ranging from heuristic workflows to sequential optimization methods, that enable adaptive communication strategies under ethical, operational, and legal constraints. Finally, the critical discussion in Section 3.7 consolidates cross-cutting gaps that motivate RQ5 and RQ6, clarifying how the impact of timing-based personalization can be assessed and how regulatory constraints shape both efficiency and design choices in automated debt collection processes.

In this way, the State of the Art establishes a coherent analytical foundation for the chapters that follow, defining the conceptual space in which a data-driven, ethically grounded, and operationally viable approach to intelligent debt collection can be developed.

3.2 Debt Collection as a Digital Communication Problem

Over the last decade, debt collection has undergone a significant transformation driven by the digitization of financial services and the expansion of online lending platforms and Financial Technology (FinTech) ecosystems. Since approximately 2015, the growth of digitally mediated consumer credit has substantially increased both the volume and heterogeneity of delinquent accounts, creating new operational and strategic challenges for creditors and collection agencies, as documented in predictive and data-driven collection studies [2, 42, 43]. In parallel, regulatory clarification across several jurisdictions has reduced uncertainty regarding the use of digital communication channels, enabling large-scale adoption of Short Message Service (SMS), e-mail, and automated calling systems in collection processes, while simultaneously imposing constraints on contact frequency and volume that incentivize optimization [21].

As a result, debt collection has progressively shifted from a predominantly physical and call-center–centric activity toward a digitally mediated interaction between creditors and debtors. This transition has amplified the role of communication timing, channel selection, and interaction sequencing, positioning debt collection not just as a financial recovery task, but as a problem of managing repeated digital interactions under legal, ethical, and behavioral constraints.

Debt Collection as a Relational and Behavioral Process

A growing body of literature suggests that debt collection should be understood as a relational and behavioral process rather than a purely transactional one. Decisions made by debtors in financial distress are shaped by contextual factors related to the interaction itself, including the perceived intrusiveness of contact attempts, prior interaction history, and sensitivity to repeated or cumulative

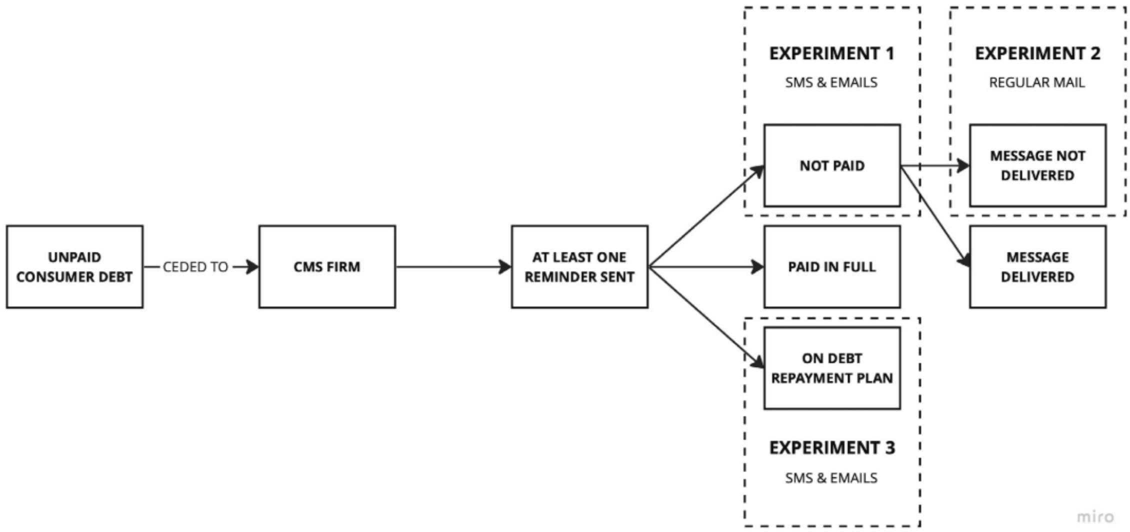

Figure 3.1: Simplified representation of the debt recovery process highlighting repeated communication interventions and their cumulative effects over time. Adapted from field experiment settings in [45].

interventions. These factors influence debtor responsiveness and engagement over time, even when the underlying financial obligation remains unchanged.

From this perspective, effective debt recovery depends not only on the content of communications, but also on how interactions are sequenced, timed, and accumulated throughout the collection lifecycle. Empirical and data-driven studies indicate that excessive or poorly timed contact strategies are associated with lower responsiveness and reduced engagement, whereas appropriately spaced and context-aware interventions are more likely to sustain cooperation over time [47].

The relational dimension of debt collection is particularly relevant in the context of non-performing loans (NPLs). Controlled field experiments reported by Saulītis [45] show that defaulted debtors may perceive repeated collection efforts as intrusive or stigmatizing, which can increase the risk of behavioral reactance and strategic disengagement. These findings indicate that the effectiveness of reminders and follow-up actions is highly sensitive to their timing and cumulative frequency, with adverse effects emerging when interaction intensity exceeds perceived tolerance thresholds.

Consumer Trust as a Structural Constraint in Debt Collection

Beyond immediate recovery outcomes, debt collection operates within a relational context in which *trust* is what keeps engagement going over time. Trust influences whether debtors interpret creditor outreach as legitimate, proportionate, and procedurally fair, shaping willingness to cooperate, to disclose constraints, and to maintain communication channels over time. Across disciplines, trust is consistently characterized as a mechanism that reduces perceived risk and uncertainty in social and economic exchange, enabling cooperation in repeated interactions [32, 41].

In digitally mediated debt collection, trust is not just an ethical nicety; it directly affects how well the system performs. Repeated outreach sequences create a dynamic interaction history in which perceptions of intrusiveness, fairness, and proportionality accumulate. When contact attempts are perceived as excessive or poorly timed, debtors may experience a threat to autonomy and respond with psychological reactance, increasing resistance and disengagement [5]. Such responses can manifest as channel avoidance (e.g., ignoring calls or messages), reduced responsiveness, or delayed repayment even when liquidity constraints are temporary. Conversely, pacing communication in a way that respects temporal availability and tolerance thresholds can help preserve perceived legitimacy and sustain cooperation.

This perspective is particularly relevant under the constraints of partial observability. When individual intent, situational frictions, and dispute status are not directly observable, aggressive escalation based on incomplete signals can amplify distrust and impose long-term costs that are difficult to attribute to any single action. As a result, trust preservation can be viewed as a *structural constraint* on timing and cadence optimization: the objective is not just to maximize short-term repayment probability, but to do so while limiting cumulative interaction intensity in ways that may erode trust.

Accordingly, this thesis treats the preservation of balanced creditor–debtor relationships as a design requirement grounded in trust maintenance, operationalized through timing and cadence-level decisions rather than through linguistic or content-level personalization [5, 41].

Limitations of Rule-Based Collection Workflows

Despite the increasing complexity of debtor behavior and the proliferation of digital communication channels, many debt collection systems continue to rely on rule-based workflows and predefined escalation sequences. These approaches typically encode expert-defined heuristics, such as fixed dunning schedules or severity ladders, which are applied uniformly across broad debtor segments [24].

Several studies have identified structural limitations in this paradigm. From a decision-theoretic perspective, Liu et al. [31] show that rule-based collection strategies tend to be myopic, as they prioritize immediate outcomes while failing to account for delayed feedback and longer-term effects of sequential decision-making. In addition, operational collection data often combines heterogeneous temporal granularities, including static customer attributes, periodic repayment histories, and high-frequency interaction signals. Treating these features uniformly can limit the expressiveness of state representations and constrain downstream decision-making.

From an operational standpoint, workflow-driven systems based on finite-state machines are costly to maintain and scale. Wang et al. [51] demonstrate that such architectures require extensive manual configuration and frequent updates to accommodate evolving interaction patterns, making them poorly suited to capture the diversity of real-world debtor behavior. Similarly, Nascimento et al. [35] report that applying uniform collection rules to heterogeneous debtor populations leads to inefficient resource allocation and suboptimal recovery performance.

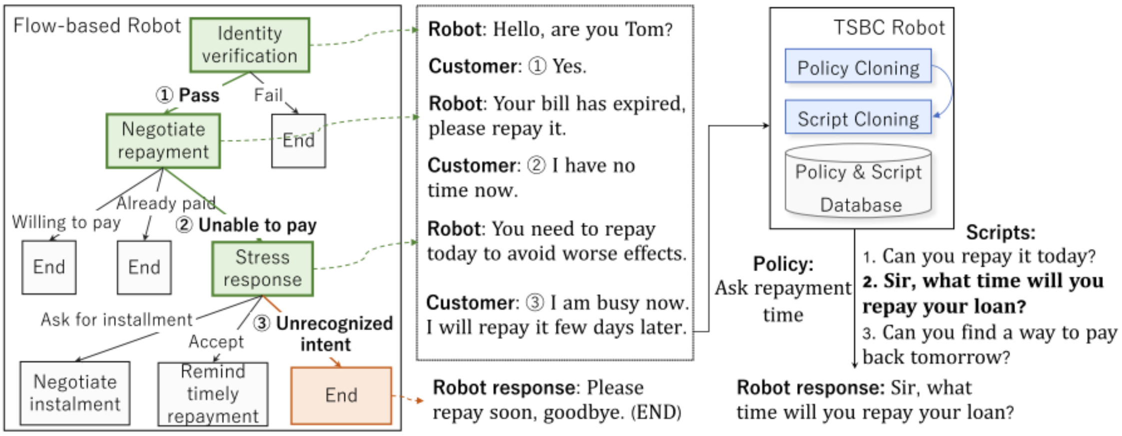

Figure 3.2: Conceptual comparison between traditional flow-based debt collection workflows and adaptive dialogue-driven systems. Flow-based approaches rely on predefined finite-state transitions, while adaptive systems incorporate conversational context and delayed feedback [51].

As illustrated in Figure 3.2, traditional flow-based collection systems rely on rigid predefined workflows, whereas adaptive dialogue systems are designed to respond dynamically to conversational context and delayed feedback, improving flexibility and reducing long-term maintenance overhead.

Escalation strategies commonly used in practice are also often grounded in expert judgment rather than empirical optimization. Duman et al. [16] describe how fixed escalation sequences that ignore individual behavioral history or sensitivity to contact intensity are associated with higher risks of disengagement and reduced collection effectiveness.

Lack of Sophistication Compared to Marketing and Customer Experience (CX)

A recurring observation across the literature is the contrast between the sophistication of customer acquisition and marketing systems and the relative immaturity of debt collection technologies. While modern marketing platforms employ advanced segmentation, attribution modeling, and temporal optimization, debt collection processes often remain manual, fragmented, and weakly data-driven [22, 30].

This technological gap is particularly evident in the treatment of communication channels. Halsey et al. [21] draw an analogy between debt collection and digital advertising, noting that attributing a payment to a specific communication action presents challenges similar to attributing a conversion to a marketing touchpoint. In both domains, outcomes arise from sequences of interactions rather than isolated events, requiring models capable of capturing temporal dependencies and cumulative effects.

Research on personalized communication strategies reinforces this analogy. Ghaffari et al. [17] show that failing to account for debtor heterogeneity leads to one-size-fits-all approaches that underperform compared to behaviorally informed segmentation. At the same time, Sánchez et al. [43] observe that, despite extensive focus on credit risk and scoring, debt collection remains comparatively underexplored with respect to adaptive communication strategies and sequential decision-making.

These findings position debt collection at the intersection of digital communication, behavioral modeling, and decision optimization, three areas that remain poorly integrated. This gap motivates the need for approaches that treat debt recovery as a sequential communication process supported by data-driven decision models, rather than as a static sequence of predefined actions.

3.3 Predictive Modeling of Debtor Behavior

Predictive modeling represents one of the most extensively explored areas within the debt collection literature. Most existing approaches focus on estimating the likelihood that a debtor will repay an outstanding balance within a predefined time horizon, using historical financial, demographic, and interaction data. These models are primarily intended to support account prioritization, workload allocation, and early intervention by identifying cases with higher or lower repayment propensity.

Early and contemporary work in this area typically formulates debt collection as a supervised learning problem. Nascimento et al. [35] evaluate multiple machine learning algorithms, including Random Forests, Gradient Boosting, and Multilayer Perceptrons, to predict repayment within 30-, 60-, and 90-day windows. Their results indicate that ensemble-based methods can achieve strong predictive performance, with area-under-the-curve (AUC) values exceeding 0.90 in several configurations. Similar findings are reported by Appel et al. [2], who apply machine learning techniques to invoice-to-cash processes, and by Jankowski and Paliński [24], who emphasize interpretable decision-tree models to derive explicit business rules for repayment classification.

Beyond binary repayment prediction, some studies consider richer predictive targets. Bellotti et al. [4] extend this predictive framing by modeling expected recovery rates on non-performing loans (NPLs) as a continuous outcome rather than merely estimating the probability of recovery. While this approach provides a more granular assessment of portfolio value and loss severity, it remains primarily descriptive in nature, supporting valuation and prioritization tasks rather than informing which collection actions should be taken to influence recovery outcomes.

Temporal and Sequential Perspectives on Repayment

While many predictive models focus on estimating whether repayment will occur, several studies argue that the timing of repayment is equally relevant for operational decision-making. Duman et al. [16] identify payment timing after collection actions as a central modelling issue in collection scorecard development. Rather than treating repayment as a static binary outcome, they analyse the number of days between the last collection action and payment, finding that among customers who eventually paid, 94% did so within three days. They therefore label actions according to whether payment occurs within this short post-action window and develop a logistic-regression collection scorecard to estimate action-specific payment probability. They discuss survival analysis as a possible alternative formulation but do not adopt it for the implemented scorecard.

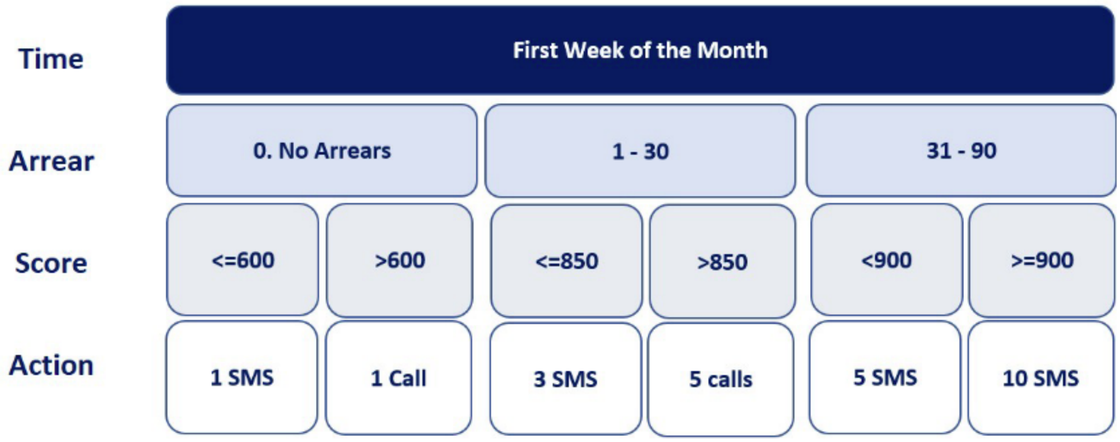

Figure 3.3: Illustrative segmentation of the debt collection process by arrears stage and corresponding predictive performance degradation, adapted from Carrera and Benalcázar [8]. The figure highlights how repayment likelihood and model discrimination vary across delinquency stages, underscoring the limitations of monolithic predictive scores in dynamic collection environments.

However, even when temporal aspects are considered, most predictive approaches rely on simplified representations of time. Liu et al. [31] point out that operational debt collection data often spans multiple temporal granularities, including static descriptors, monthly repayment histories, and high-frequency interaction signals. Many models treat these heterogeneous temporal features uniformly, which can degrade state representation and limit the ability to capture delayed or cumulative effects across repeated interactions.

Recent work by Carrera and Benalcázar [8] further illustrates this limitation by showing that predictive performance varies substantially across different stages of arrears. Their results indicate that models trained on early-stage delinquency data perform markedly worse when applied to later stages, motivating the use of separate scoring models for distinct arrears segments. This stage-dependent degradation of predictive power reflects the inherently non-stationary nature of debtor behavior and reinforces the view that static predictive scores inadequately represent the evolving debtor state over time.

Behavioral Versus Financial Features

A recurring finding across the debt collection literature is that behavioral and interaction-based variables often provide stronger predictive signals than purely financial features. Sánchez et al. [43] explicitly distinguish between financial descriptors (FD) and contact and communication data (CCD), showing that variables such as call outcomes, promises to pay, response delays, and prior engagement patterns lead to significant improvements in predictive accuracy. Their results indicate that past interaction behavior is frequently more informative than outstanding balance or income-related proxies when considered in isolation.

This observation is reinforced by studies incorporating richer behavioral representations. Sivamayilvelan et al. [47] organize debtor features into basic information, payment history, and income-related attributes, mapping these inputs to a latent representation of debtor state within a reinforcement learning framework. Similarly, Liu et al. [31] report the use of large feature sets spanning demographics, historical repayment behavior, and high-frequency interaction signals, underscoring the growing importance of behavioral metadata in predictive and sequential modeling.

From a modeling perspective, Murteira and Augusto [33] decompose repayment behavior into two distinct stages: the decision to repay and the conditional amount repaid. This hurdle-based formulation exposes the structural heterogeneity of debtor behavior and suggests that repayment dynamics cannot be adequately captured by a single predictive outcome variable.

Limitations of the Predictive Framing: The *What* Versus the *How*

Despite their demonstrated accuracy, predictive models exhibit fundamental limitations when applied to debt collection decision-making. Most approaches answer the question of *what* is likely to happen (e.g., whether or when a debtor will repay), but provide limited guidance on *how* a creditor should act to influence that outcome. As emphasized by Liu et al. [31], debt collection is inherently a sequential decision problem in which actions taken at one point affect future states, observations, and opportunities.

In practice, predictive scores are often operationalized for internal efficiency rather than behavioral intervention. For example, Kim and Kang [27] use late payment prediction to ensure fair allocation of customer contact lists among call center agents, illustrating that even accurate models are frequently decoupled from personalized strategy selection.

Attributing outcomes to specific actions also remains challenging in environments characterized by repeated interactions. Duman et al. [16] note that in multi-contact processes it is difficult to determine which communication triggered a payment, reducing the practical usefulness of static predictive scores for strategy design. This attribution challenge is analogous to the carry-over effects modeled in digital advertising using Adstock formulations, where the influence of past exposures decays over time and cannot be attributed to a single interaction [21]. As a result, while predictive models are effective for ranking and prioritization, they are insufficient for defining adaptive and personalized collection strategies on their own.

Early attempts to move beyond purely predictive framing can be found in decision-oriented valuation models. Chehrazi and Weber [10] model delinquent credit-card accounts as dynamically evolving assets whose value depends on future recovery trajectories, illustrating a shift toward action-aware valuation, albeit without explicitly modeling communication or interaction strategies.

In summary, the literature shows that predictive models can reliably estimate repayment likelihood, timing, and expected recovery using both financial and behavioral data. However, these models largely operate independently of decision mechanisms, motivating the need for approaches that move beyond prediction toward integrated, action-aware strategy formulation. Additional empirical results supporting these findings are reported by Nascimento et al. [35] and summarized in Table A.1 in Appendix A.1.

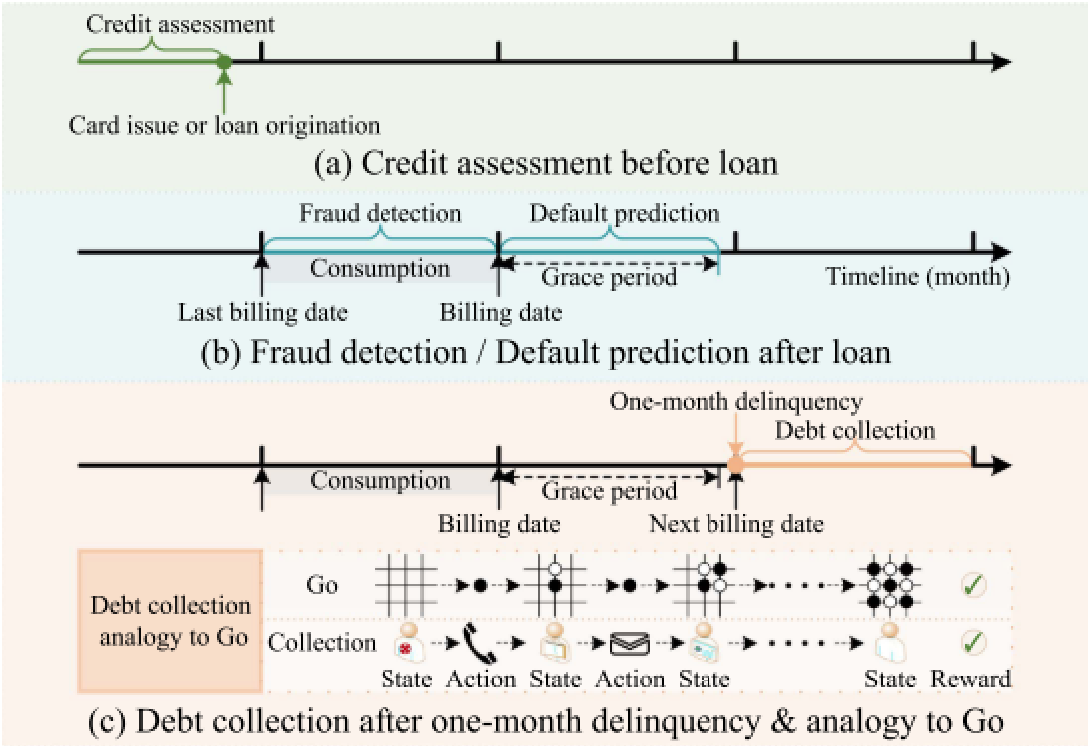

Figure 3.4: Conceptual distinction between credit risk assessment, delinquency forecasting, and debt collection as a sequential decision-making problem. While the first two focus on predictive estimation, delinquency forecasting remains fundamentally myopic, optimizing short-term repayment likelihood rather than the long-term cumulative value of the creditor–debtor relationship, whereas debt collection requires action-dependent reasoning under delayed feedback [31].

Table 3.1: Debtor typology framework based on willingness to pay, payment ability, financial organization, and behavioral rationality, adapted from [17].

Willingness to Pay	Payment Ability	Financial Organization	Rationality	Typology
Willing	Able	Organized	Rational	WAOR
Willing	Able	Organized	Emotional	WAOE
Willing	Able	Chaotic	Emotional	WACE
Defiant	Insolvent	Chaotic	Emotional	DICE
Defiant	Insolvent	Organized	Rational	DICR

3.4 Segmentation and Debtor Typologies

Traditional debt collection systems have historically relied on static risk segmentation, grouping debtors primarily according to their estimated probability of default or repayment. While such approaches are useful for high-level prioritization, recent literature argues that they fail to capture the behavioral heterogeneity underlying non-payment. As a result, identical collection strategies are often applied to debtors with fundamentally different motivations, constraints, and responsiveness profiles.

Behavioral Personas and Willingness versus Ability to Pay

A significant shift toward behaviorally grounded segmentation is proposed by Ghaffari et al. [17], who introduce a multidimensional framework for debtor typology construction. Their approach is based on four key dimensions: willingness to pay, ability to pay, financial organization, and rational behavior. By combining these axes, the authors derive 16 distinct debtor personas, such as *WAOR* (Willing–Able–Organized–Rational) and *DICE* (Defiant–Insolvent–Chaotic–Emotional), each associated with markedly different response patterns to collection interventions.

Empirically, Ghaffari et al. [17] show that five dominant typologies, including WAOR, WAOE, WACE, DICE, and DICR, account for approximately 63% of the debtors in their large-scale dataset, indicating that these personas are not marginal but structurally representative. Reaction rates also vary sharply across profiles, ranging from nearly 98% for WAOR debtors to approximately 20% for DICE profiles, quantitatively demonstrating the behavioral heterogeneity masked by traditional risk segmentation.

Behavioral Metadata as a Segmentation Substrate

The construction of such typologies relies heavily on behavioral and interaction-based metadata. Ghaffari et al. [17] employ over 70 features, including email validity, presence of the debtor's name in the email address, response latency, and prior interaction outcomes. These signals reflect engagement and attentiveness rather than financial capacity alone.

This emphasis on interaction metadata is supported by Sánchez et al. [43], who formally distinguish between Financial Data (FD) and Contact Center Data (CCD). Their results show that incorporating CCD, such as call attempts, promises to pay, and interaction outcomes, leads to

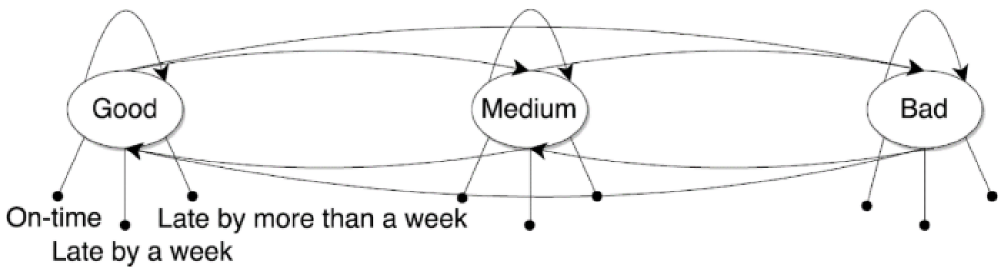

Figure 3.5: Illustrative Hidden Markov Model of latent debtor financial states and observed repayment behavior, adapted from Philip et al. [38]. The model represents debtor typologies as evolving latent states with probabilistic transitions, rather than static categories.

substantially higher classification accuracy than using financial variables alone. These findings reinforce the notion that debtor behavior within communication channels constitutes a critical substrate for meaningful segmentation.

Complementary perspectives are offered by Jankowski and Paliński [24], who demonstrate that debtor characteristics such as legal form, geographic region, and contact history significantly affect repayment patterns in mass receivables. Together, these studies suggest that segmentation grounded in behavioral metadata provides a more actionable representation of debtor state than static financial descriptors.

Static Versus Dynamic Segmentation

Despite increasing granularity in feature engineering, many segmentation approaches remain static in practice, assigning debtors to fixed categories that do not evolve in response to new interactions. Nascimento et al. [35] illustrate limitations of uniform collection strategies, showing that applying identical treatment rules to heterogeneous debtor populations leads to inefficient effort allocation and reduced recovery performance.

More recent work argues for dynamic segmentation based on evolving state representations. Liu et al. [31] emphasize that operational debt collection data spans multiple temporal granularities, including static demographics, periodic financial aggregates, and high-frequency interaction signals. Segmentation approaches that fail to integrate these temporal layers risk misrepresenting the debtor's evolving state and may result in poorly aligned interventions.

A formal treatment of dynamic debtor typologies is proposed by Philip et al. [38], who model repayment behavior using partially constrained hidden Markov models (HMMs). In this framework, debtors are characterized by latent financial and behavioral states that evolve over time, with observed repayment patterns treated as emissions from these hidden states. The authors show that state transition dynamics are closely related to underlying financial health, while emission probabilities capture observable repayment behavior, providing a dynamic and data-driven analogue to the distinction between latent capacity and observed repayment behavior discussed in earlier typology-based approaches [17].

Non-Delinquency	Delinquency
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4]
[0, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 2, 3, 3, 2, 2, 2, 2, 2, 2, 1]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 3, 3, 3, 3, 2, 3, 3, 3, 3, 2]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 3, 3, 3, 3, 2, 1, 1, 1, 1, 1]
[0, 2, 3, 3, 3, 3, 3, 3, 3, 3, 2, 3]	[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[4, 0, 2, 3, 3, 1, 1, 1, 1, 1, 1, 1]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 1, 2, 2, 3, 2, 1, 1, 2, 2, 2, 1]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 2, 3]	[0, 3, 2, 1, 4, 4, 4, 4, 4, 4, 4, 4]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1]
[0, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 2, 3, 3, 3, 3, 3, 2, 3, 3, 3, 2]
[0, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 2, 1]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 2, 2, 0, 1, 1, 1, 1, 0, 1, 1]
[0, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	[0, 3, 2, 2, 2, 2, 2, 1, 1, 1, 4, 4]
...	...

Figure 3.6: Latent cluster assignment trajectories over time for delinquent and non-delinquent users, adapted from Kim et al. [28]. Stable cluster membership characterizes non-delinquent behavior, while delinquency is associated with frequent transitions between latent states.

Recent advances in deep generative modeling further extend this dynamic perspective. Kim et al. [28] propose a Variational Deep Embedding with Sequence-to-Sequence (VaDE-Seq2Seq) architecture that jointly performs sequential delinquency prediction and latent clustering. Their results show that non-delinquent users tend to remain in relatively stable latent clusters over time, whereas delinquent users exhibit more frequent transitions between clusters. These findings reinforce the view that debtor typologies are better understood as temporal trajectories rather than fixed profiles.

Limits of Segmentation Without Adaptive Strategy

Despite the increasing granularity of debtor profiles, the literature identifies a critical limitation: segmentation alone does not resolve the myopia of collection workflows. As noted by Liu et al. [31], even refined segmentation schemes remain insufficient when associated decision rules are static and fail to integrate multiple temporal granularities of data, such as high-frequency interaction signals alongside periodic financial snapshots. In such settings, segmentation improves descriptive resolution but does not, by itself, enable effective sequential decision-making.

Related evidence from field experiments further illustrates the risks of strategy design that ignores behavioral feedback. Saulītis [45] shows that repeated or poorly calibrated interventions can reduce repayment likelihood, particularly when debtor sensitivity to contact intensity is not accounted for. These findings suggest that segmentation without adaptive control of intervention timing and frequency may lead to disengagement even among debtors who are otherwise willing to repay.

Both Hidden Markov Model (HMM)–based and deep generative segmentation approaches remain fundamentally descriptive. While they uncover latent states and behavioral trajectories, they do not specify which actions should be taken to influence state transitions. Consequently, although debtor typologies provide a necessary behavioral substrate, they must be coupled with adaptive, sequential decision models to move from descriptive segmentation toward prescriptive strategy optimization.

3.5 Communication Strategy Optimization and Digital Nudging

While predictive modeling and segmentation provide insight into *who* is likely to pay and *under which conditions*, the effectiveness of debt collection ultimately depends on *how*, *when*, and *through which channel* communication is executed. Recent literature increasingly frames debt collection as an optimization problem over communication strategies, where messages are no longer isolated events but components of a sequential, behavior-sensitive intervention process.

This section reviews empirical and analytical work on communication strategy optimization, with a focus on timing, frequency, message framing, and multichannel interaction effects. Together, these studies establish the technical foundation for treating communication as an adaptive control variable rather than a static workflow. Complementing this line of work, Sancarlos et al. [42] explicitly frame *collection strategy* as an operational layer driven by propensity-to-pay estimates (e.g., adjusting contact intensity or escalation), highlighting how predictive outputs are already used as implicit strategic inputs even when communication actions are not modeled explicitly.

Behavioral and Economic Foundations of Repayment Timing

A substantial body of research in behavioral economics and household finance shows that debt repayment behavior is inherently temporal and shaped by liquidity constraints, cognitive budgeting processes, and social norms. These foundations provide essential context for understanding why repayment outcomes vary systematically over time and why communication timing can play a critical role in payment effectiveness, independently of debtor intent or moral disposition.

Household budget allocation and mental accounting. Households do not treat income as a fully fungible resource. Instead, income is often implicitly allocated across mental accounts corresponding to categories such as housing, utilities, food, transportation, healthcare, and discretionary consumption. Debt obligations compete with these categories for both financial liquidity and cognitive attention. Seminal work on mental accounting shows that individuals evaluate financial decisions within category-specific budgets rather than through global optimization, leading to

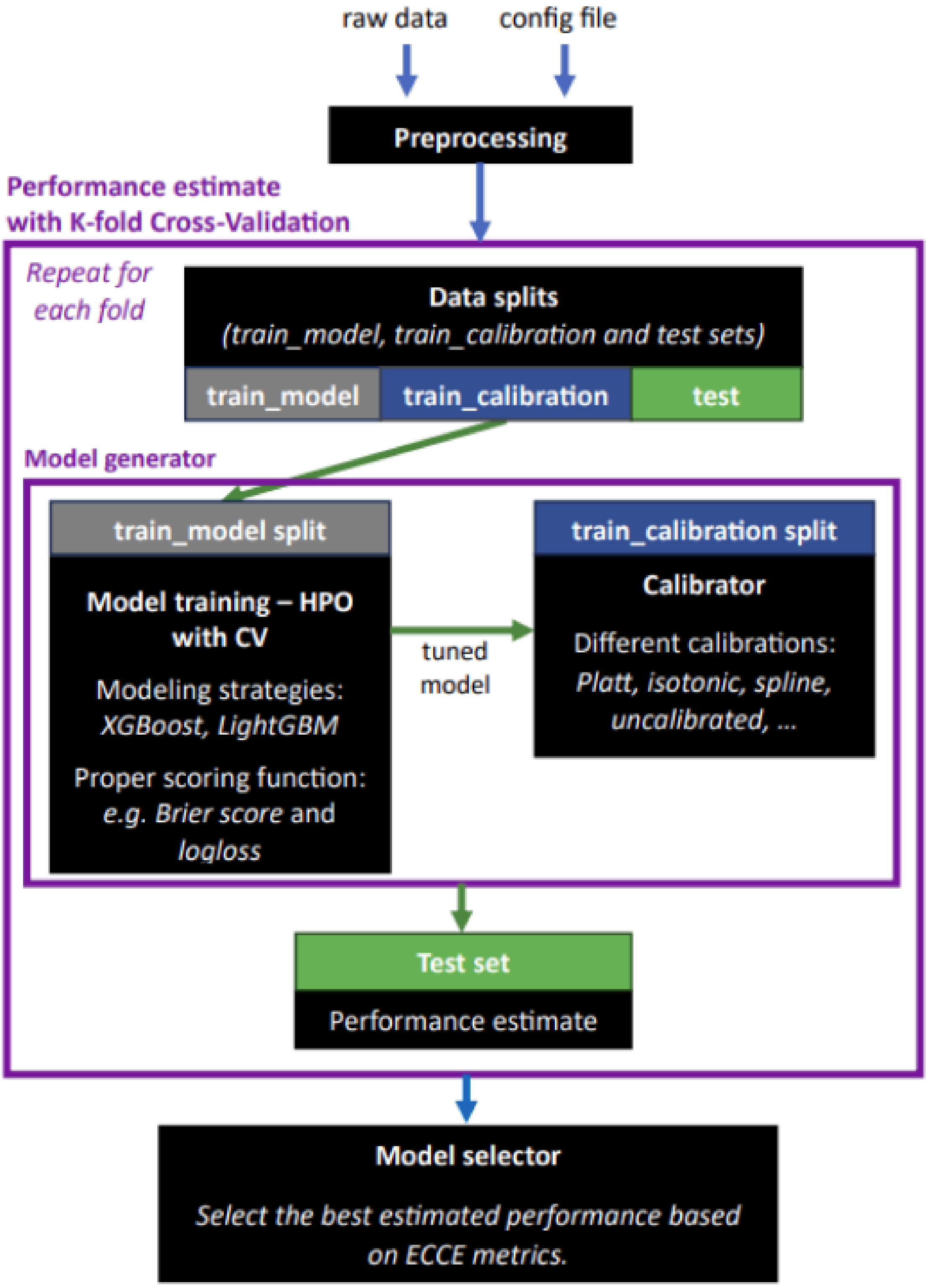

Figure 3.7: Diagram of an advanced machine-learning pipeline designed to produce calibrated propensity-to-pay estimates intended to guide operational collection strategy decisions (e.g., intensity adjustment and escalation), adapted from Sancarlos et al. [42]. The pipeline illustrates how strategy is often driven by calibrated scoring, while remaining decoupled from explicit communication policy modeling.

context-dependent payment behavior [49, 50]. As a result, repayment decisions are influenced not only by willingness to comply but also by whether debt obligations are cognitively prioritized within the current budgeting cycle, making repayment behavior sensitive to timing relative to budget availability.

Income cycles, liquidity constraints, and temporal availability. Regular income cycles, such as salaries, pensions, or social benefits, create predictable fluctuations in household liquidity. Empirical evidence shows that consumption and payment behavior are synchronized with these cycles, with expenditures and repayments increasing shortly after income receipt and declining toward the end of the cycle [48]. Liquidity constraints further amplify this effect, as households with limited access to credit or savings are less able to smooth consumption or repayment over time [19]. Consequently, the ability to repay varies systematically within the income cycle, even when total income and debt levels remain unchanged.

Payment bunching, valleys, and peaks as behavioral proxies. Observed payment data frequently exhibit periods of low activity followed by concentrated repayment peaks. Rather than reflecting deliberate non-compliance or abrupt changes in willingness to pay, such patterns are consistent with liquidity-driven delays and prioritization dynamics. Studies of credit card repayment behavior document payment bunching around statement dates, income arrival, or minimum payment thresholds, driven by present bias and financial constraints [26, 29]. These temporal patterns can therefore be interpreted as behavioral proxies for delayed repayment under constraint, providing informative, though imperfect, signals of repayment dynamics without requiring assumptions about debtor intent.

Willingness versus ability to pay as orthogonal dimensions. A central distinction in household finance is between willingness to pay and ability to pay [7, 53]. Non-payment may arise from liquidity shortages, competing financial obligations, or timing mismatches, even when the debtor is willing to settle the obligation. Timing-based interventions may thus influence repayment outcomes by aligning contact with periods of higher liquidity rather than by changing underlying preferences or intent.

Comfort, moral perception, and social norms. Debtors' responses to collection efforts are also shaped by perceptions of fairness, proportionality, and social norms. Behavioral evidence suggests that overly frequent or poorly timed contact can reduce responsiveness or increase avoidance, while appropriately timed reminders may be perceived as supportive prompts rather than coercive pressure [6, 25]. These effects operate independently of formal enforcement mechanisms and highlight the importance of psychological and social constraints in communication design.

Debt as a symptom of broader issues. Outstanding debt doesn't always mean someone is refusing to pay. It can also stem from administrative issues, billing disputes, or claims that are still being sorted out [12]. In these situations, jumping to aggressive collection tactics or contacting people at the wrong time can backfire, creating more problems than it solves [45].

Implications for communication strategy design. Together, these behavioral and economic insights suggest that timing and cadence are fundamental behavioral constraints, not secondary tuning parameters. Static communication schedules ignore income cycles, liquidity constraints, cognitive budgeting, and resolution delays. This motivates the use of timing-aware and state-dependent communication policies that align contact attempts with periods of higher repayment capacity and lower psychological resistance, forming the conceptual foundation for data-driven and AI-supported communication optimization examined in subsequent sections.

Timing Optimization

A growing body of literature identifies contact timing as a critical determinant of repayment outcomes in debt collection. Rather than being uniformly distributed over time, debtor responsiveness exhibits strong temporal structure, shaped by income cycles, recent interactions, and the evolving state of delinquency. As a result, the effectiveness of a communication attempt depends not only on whether contact occurs, but on *when* it occurs relative to these temporal constraints.

At the calendar level, several studies document systematic liquidity-driven effects. Halsey et al. [21] show that repayment probability increases around salary receipt periods, with pronounced peaks toward the end of the month and on specific weekdays. These patterns suggest that aligning contact attempts with periods of higher liquidity availability materially increases the likelihood of repayment, independently of debtor risk classification.

Beyond coarse calendar effects, finer-grained temporal analyses reveal substantial heterogeneity in optimal contact windows across debtor profiles. Ghaffari et al. [17] analyze reaction rates across discrete time windows during the day and show that responsiveness varies systematically across debtor typologies. While these results are often discussed in relation to message framing, they also highlight that temporal availability and attentional capacity are not uniform, reinforcing the need for time-aware contact policies rather than fixed schedules.

A complementary perspective is offered by studies examining repayment behavior relative to prior collection actions. Duman et al. [16] report that in their banking dataset, among customers who eventually paid after a collection action, 94% did so within three days. This concentration of repayment events defines a narrow post-contact observation window, after which the marginal effectiveness of additional reminders declines sharply. Such findings motivate short-horizon monitoring strategies and delayed re-contact policies that respect the temporal decay of responsiveness.

The feasibility and effectiveness of timing decisions also depend on the stage of delinquency. Carrera and Benalcázar [8] show that repayment dynamics and predictive performance vary substantially across arrears stages, implying that optimal contact timing cannot be defined independently of delinquency duration. Early-stage cases may benefit from prompt, minimally intrusive contact aligned with liquidity windows, whereas later-stage cases often exhibit weaker temporal sensitivity and require more cautious pacing to avoid disengagement.

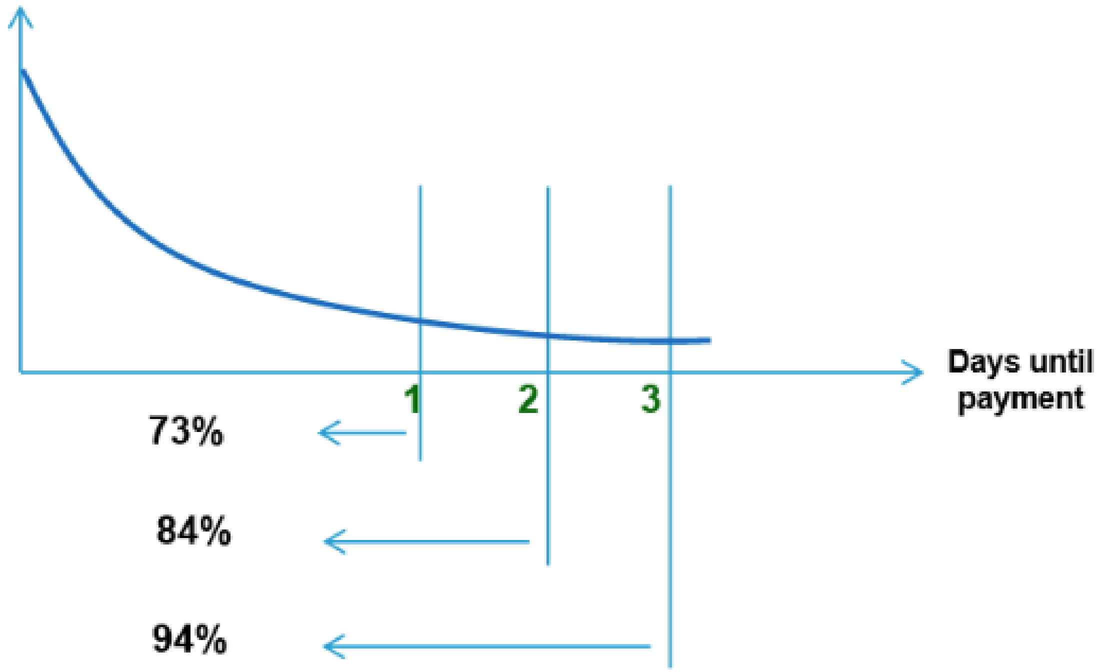

Figure 3.8: Distribution of payments after the last collection action, showing that 73%, 84%, and 94% of paying customers paid within one, two, and three days, respectively, in the banking dataset analysed by Duman et al. [16]. This temporal concentration motivates short post-contact attribution windows and adaptive re-contact strategies.

Together, these studies establish timing as a constrained variable shaped by liquidity cycles, behavioral fatigue, and the structure of prior interactions. Instead of a universal schedule, effective debt collection needs timing policies that adapt to the evolving state of the case, balancing responsiveness against the risk of premature or excessive contact.

Frequency Versus Irritation: Annoyance Costs

While increasing contact frequency can raise the probability of repayment in the short term, a growing body of literature points to the behavioral costs associated with excessive communication. Repeated contact attempts impose cognitive and emotional burden on debtors and may be perceived as intrusive or unfair, particularly when poorly timed or misaligned with the debtor's ability to respond.

Saulītis [45] provides empirical evidence of such behavioral costs in the context of debt collection, showing that repeated reminders can reduce repayment likelihood when perceived as excessive. In these cases, debtors may disengage or delay repayment, even when liquidity constraints alone do not fully account for non-payment. These findings indicate that non-response should not be interpreted solely as unwillingness to pay, but may also reflect behavioral overload or resistance to perceived pressure.

From a decision-theoretic perspective, Liu et al. [31] argue that indiscriminate increases in contact frequency tend to optimize short-term recovery metrics while failing to account for longer-term effects on process efficiency and future engagement. Applying uniform escalation policies to debtors who may have missed a payment due to inattention or temporary constraints can therefore increase irritation without materially improving recovery outcomes.

As a result, contact frequency emerges as a constrained optimization variable subject to diminishing returns and negative behavioral externalities. Beyond a certain point, additional contact attempts reduce marginal effectiveness and increase the risk of disengagement. Effective collection strategies must therefore balance the potential benefits of increased frequency against behavioral tolerance limits, taking into account recent interaction history, delinquency stage, and regulatory constraints on intrusiveness.

Contextual Sensitivity and Behavioral Limits of Communication Interventions

A complementary line of work highlights that the effectiveness of communication interventions in debt collection is sensitive to contextual and behavioral factors, even when timing and frequency are held constant. Drawing on insights from behavioral economics, several studies show that debtors' responses are shaped not only by the presence of a reminder, but also by how it is perceived in terms of fairness, salience, and proportionality.

Empirical evidence indicates that uniform communication approaches can generate heterogeneous outcomes across debtor profiles. Ghaffari et al. [17] show that debtor typologies exhibit systematically different responsiveness patterns to collection interventions, suggesting that identical actions may be perceived as supportive by some debtors and as intrusive or excessive by others. While these differences are often discussed in relation to message framing, they more broadly reflect variation in behavioral sensitivity and tolerance to intervention.

Field experiments further illustrate the risks associated with poorly calibrated communication strategies. Saulītis [45] evaluates the impact of salience-enhancing and norm-based reminders in the context of non-performing loans and finds that such mechanisms can be effective in early stages of delinquency, but may lose effectiveness or even reduce repayment as arrears accumulate. In particular, interventions that increase perceived pressure or unfairness were associated with lower repayment rates, suggesting that increased salience without alignment to debtor state can amplify avoidance rather than compliance.

These findings point to an important limitation for communication strategy optimization: even well-intentioned interventions may become counterproductive when applied without regard for debtor state, prior interaction history, or accumulated interaction intensity. Behavioral sensitivity therefore acts as a constraint on communication design, reinforcing the need to prioritize timing, pacing, and sequencing decisions that limit perceived intrusiveness and help preserve cooperation over time, rather than relying on increasingly salient or aggressive interventions.

Multichannel Strategies and Cross-Channel Effects

Debt collection processes increasingly operate across multiple communication channels, including SMS, e-mail, automated calls, and human agents. These channels do not act independently. Instead, debtor responsiveness to a given interaction is shaped by prior exposures across channels, introducing temporal dependencies and carry-over effects that complicate isolated channel optimization.

Halsey et al. [21] adopt an Adstock-based modeling approach to capture such cross-channel dynamics, demonstrating that exposure through one channel can increase responsiveness in subsequent interactions via another channel. For example, prior contact through e-mail or SMS is associated with a higher likelihood that a debtor will answer a later phone call. These results show that communication effectiveness accumulates over time and across channels, rather than being attributable to single interactions.

Building on this perspective, Liu et al. [31] propose a hierarchical distinction between *light* channels (e.g., SMS, Interactive Voice Response (IVR)) and *heavy* channels (e.g., human agents). While heavy channels may accelerate repayment when used appropriately, their excessive or premature use increases operational costs and the risk of debtor disengagement. As a result, channel selection cannot be treated as a static choice, but must be coordinated over time as part of a broader communication sequence.

This hierarchy reflects both operational considerations and behavioral constraints. Light channels typically impose lower cognitive and emotional load, whereas heavy channels introduce higher perceived intrusiveness and pressure. Escalating too quickly to more intrusive channels may improve short-term outcomes but can undermine long-term cooperation and increase annoyance, reinforcing the need for carefully paced and state-aware channel sequencing rather than isolated channel-level optimization.

Empirical Outcomes and Limitations

Empirical evaluations indicate that communication strategy optimization can yield measurable performance gains. Liu et al. [31] report improvements in realized recovery rates following the deployment of offline reinforcement learning to optimize communication policies in a large-scale banking setting. Similarly, Halsey et al. [21] observe statistically significant increases in repayment likelihood associated with e-mail engagement and cross-channel exposure effects.

However, these gains are accompanied by clear limitations. Message effectiveness is not monotonic: it decays over time, varies sharply across debtor profiles, and may reverse under excessive pressure. As demonstrated by Saulītis [45], poorly calibrated nudges can reduce repayment rates, underscoring the ethical and behavioral constraints inherent in communication-based interventions. In parallel, optimization efforts in practice often target adjacent operational subproblems rather than end-to-end communication policy design. For example, Çevik Onar et al. [36] propose a decision-support approach for optimizing debt collection assignments, illustrating how operational optimization frequently remains fragmented from communication policy orchestration.

Overall, the literature establishes that timing, frequency, framing, and channel selection are powerful but interdependent levers in debt collection. Optimizing these dimensions requires models capable of balancing short-term recovery objectives against longer-term engagement and process efficiency under uncertainty and delayed feedback. These requirements motivate the need for integrated decision models, discussed in Section 3.6, that can dynamically orchestrate communication strategies rather than relying on static or locally optimized rules.

Attribution, Carry-Over Effects, and the Limits of Causal Inference in Current Work

Although multiple studies provide empirical evidence that contact timing, cadence, or channel selection influence repayment behaviour, the majority of findings in the literature remain fundamentally correlational. Field experiments such as those by Saulītis [45] and observational analyses such as Halsey et al. [21] estimate associations between communication characteristics and repayment outcomes, but they do not identify causal effects of individual interventions within a sequence of contacts.

The core methodological challenge arises from attribution. Debt collection outcomes are the result of sequences of heterogeneous interactions, often exhibiting temporal carry-over effects and shaped by unobserved debtor characteristics. As in digital advertising, multiple contact attempts jointly contribute to a repayment event, making it difficult to isolate the incremental impact of any specific action. Models such as Adstock [21] capture temporal decay and accumulation of influence across interactions, but they do not establish counterfactual outcomes, namely, what would have occurred had a particular intervention not taken place.

Most operational datasets also lack the randomization or quasi-experimental variation required for robust causal inference. Factors such as arrears stage, collector workload, debtor liquidity cycles, and prior engagement history act as confounders that bias naive comparisons of "treated" versus "untreated" interactions. As a consequence, observed performance improvements cannot be reliably attributed to the timing or sequencing of communication actions themselves, rather than to underlying temporal or behavioral differences across debtor populations.

This methodological gap directly motivates RQ5. Measuring the impact of personalized timing and cadence requires evaluation approaches that explicitly account for sequential decisions, delayed rewards, and cross-channel dependencies under partial observability. Establishing credible attribution, whether through randomized interventions, uplift modeling, or causal structural approaches, is essential for assessing the true effectiveness of automated collection strategies and for avoiding spurious correlations that may lead to inefficient or ethically problematic interventions.

3.6 Decision Models for Adaptive Collection Strategies

While predictive models, segmentation frameworks, and communication optimization techniques provide increasingly accurate descriptions of debtor behavior, they do not, by themselves, determine how collection actions should be orchestrated over time. The central limitation identified across the literature is that debt collection is inherently a sequential decision-making problem, where each action influences future states, observations, and available options.

This section reviews decision models proposed for adaptive debt collection strategies, contrasting traditional heuristic-based systems with recent advances in Reinforcement Learning (RL) and sequential optimization. These approaches address the transition from isolated predictors to integrated strategy engines capable of balancing recovery performance, operational constraints, behavioral risk, and execution context.

Heuristic-Based Decision Systems

Most operational debt collection systems rely on heuristic decision rules, commonly referred to as *dunning rules*. These rules define fixed escalation paths specifying which action to take (e.g., SMS, call, legal notice) and when to execute it. Nascimento et al. [35] describe such systems as deterministic workflows that apply uniformly to broad debtor categories, often ignoring individual behavioral responses.

Similarly, Duman et al. [16] note that banking practice frequently encodes expert judgment into predefined severity ladders, such as progressing from soft reminders to more intrusive contact after a fixed delay. While these heuristics are transparent and easy to implement, they are inherently static and fail to adapt to feedback generated by debtor interactions.

Jankowski and Paliński [24] propose extracting stable decision rules from decision trees to automate mass debt collection, enabling early elimination of low-potential cases. Although such approaches improve scalability and explainability, they remain limited to myopic, one-step decision logic. As observed by Sivamayilvelan et al. [47], deterministic policies are only effective when feature values are complete and debtor behavior is stable, conditions rarely met in real-world collection environments.

Debt Collection as a Sequential Decision Problem

Recent literature increasingly frames debt collection as a sequential decision process rather than a static classification task. Liu et al. [31] formalize this perspective by modeling collection as a constrained Markov Decision Process (MDP), where the objective is to maximize cumulative recovery under delayed feedback, behavioral uncertainty, and regulatory constraints.

A key contribution of the ACAQL framework is the recognition that collection decisions are hierarchical. Rather than selecting a single action directly, the system first chooses a channel (e.g., SMS, call center, human agent) and subsequently determines the specific interaction strategy within that channel. This hierarchical structure departs from standard MDP formulations and better reflects operational realities.

This formulation highlights why purely predictive approaches are insufficient: knowing the probability of repayment does not determine the optimal sequence of actions required to achieve it. Empirical studies further suggest that optimal policies differ systematically across delinquency

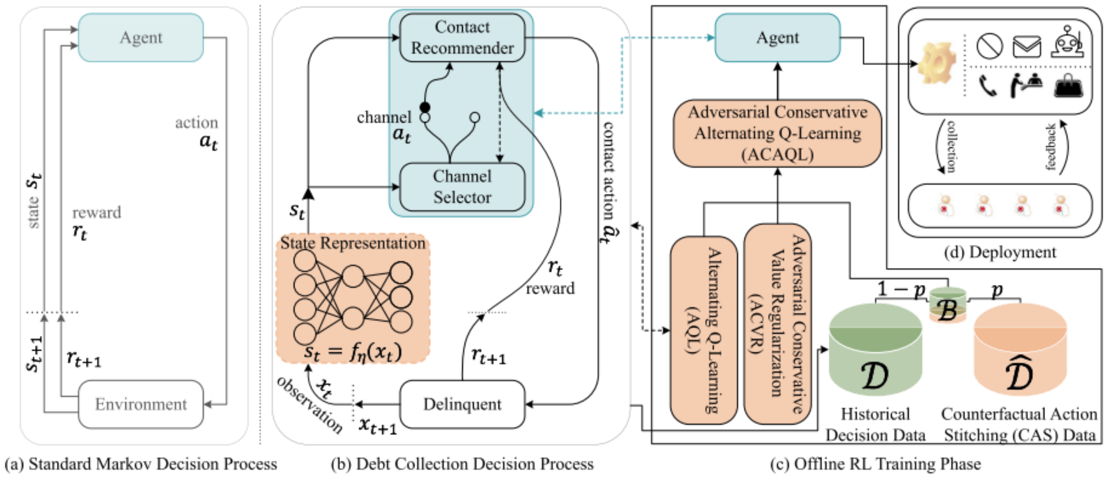

Figure 3.9: Comparison between a standard MDP and the hierarchical decision structure of debt collection, where channel selection and action execution occur at different decision layers [31].

stages, reinforcing the interpretation of arrears stage as a latent state variable within sequential decision models rather than a static segmentation attribute [8].

RL for Adaptive Strategy Optimization

RL has emerged as a promising paradigm for learning adaptive collection policies from historical interaction data. Liu et al. [31] introduce Adversarial Conservative Alternating Q-Learning (Adversarial Conservative Alternating Q-Learning (ACAQL)), an offline RL approach designed to mitigate the risks associated with exploration in high-stakes financial environments. By learning exclusively from logged interaction data rather than real-time trial-and-error, the framework seeks to improve long-term recovery while limiting the deployment of unsafe or untested actions.

Sivamayilvelan et al. [47] propose a deep reinforcement learning–based recommendation framework (Flexible Recommendation System using Deep Reinforcement Learning (FRS-DRL)) that employs a hybrid actor–critic architecture to suggest flexible actions such as visit scheduling and payment plan adjustments. Their approach reduces reliance on manually defined policy rules and adapts recommendations in response to observed debtor behavior. The authors also note that the realized effectiveness of recommended actions depends on collector-specific factors, such as experience and behavioral consistency, highlighting that the execution context influences observed outcomes.

Related work by Halsey et al. [21] suggests framing collection decisions as a *Next Best Action* problem, drawing analogies with marketing optimization. In this formulation, sequential decision models are used to coordinate channel choice, timing, and interaction intensity in order to maximize cumulative recovery, rather than immediate response to individual contacts.

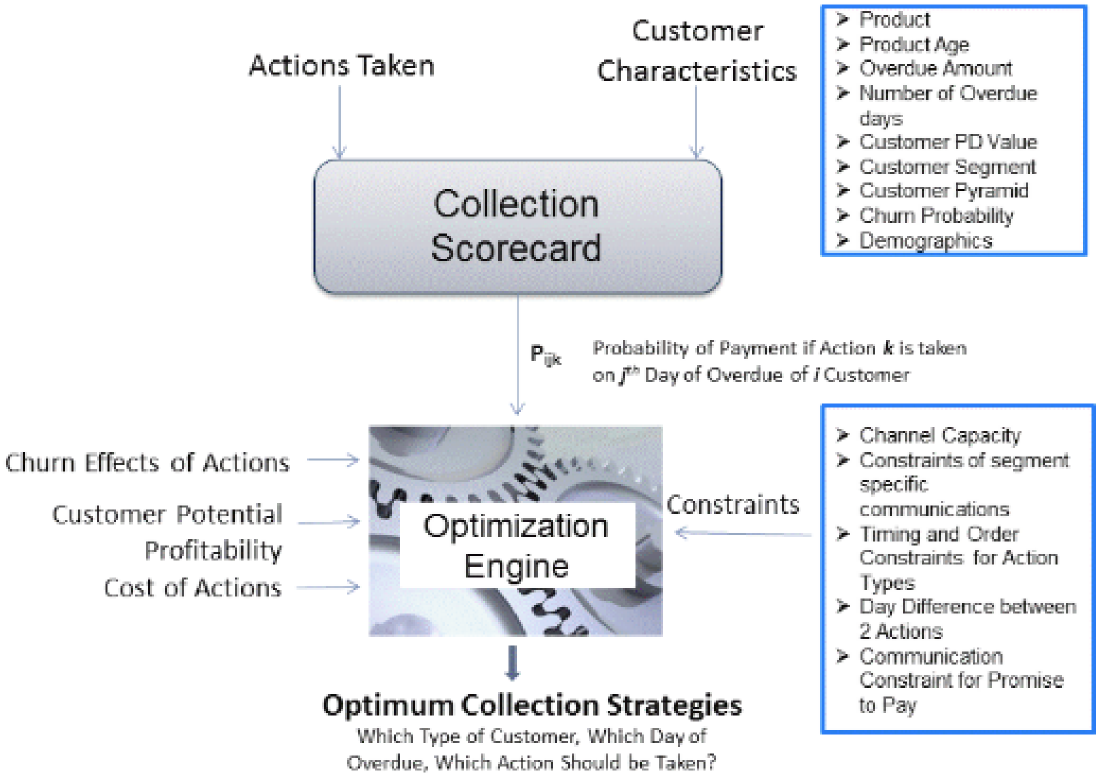

Figure 3.10: Separation between predictive scorecards and optimization engines in modern debt collection systems, illustrating the transition from probability estimation to sequential decision-making, adapted from Duman et al. [16].

Strategy Engines versus Isolated Predictors

A recurring theme across these studies is the distinction between predictive scorecards and decision engines. Duman et al. [16] explicitly separate the *Collection Scorecard*, which estimates repayment likelihood, from the *Optimization Engine*, which determines the optimal sequence of actions.

This separation clarifies why high predictive accuracy does not automatically translate into effective strategy. As Liu et al. [31] argue, predictive models are inherently myopic: they optimize short-term forecasts rather than long-term cumulative outcomes shaped by repeated interactions.

Behavioral Constraints and Strategic Responses

Decision models must also account for behavioral responses to collection pressure. Saulītis [45] provides empirical evidence that excessive or poorly timed communication can generate behavioral costs, leading to reactance and reduced repayment. In such cases, debtors may disengage from the collection process or delay payment in response to perceived pressure or unfairness, rather than as a consequence of financial inability alone.

Practical Limitations and Integration Challenges

Despite their promise, adaptive decision models face significant barriers to deployment. Duman et al. [16] identify channel capacity constraints, timing constraints, and promise-to-pay constraints as critical obstacles to unconstrained optimization. Liu et al. [31] caution that offline RL may produce over-optimistic policies when historical data lacks sufficient coverage of alternative actions.

In addition, real-world decision-making frequently involves operational constraints that lie outside the scope of most RL formulations. Çevik Onar et al. [36] demonstrate that many production systems focus on optimizing assignment and workload distribution across collectors, subject to service-level and capacity constraints, while remaining agnostic to message-level personalization. Similarly, Kim and Kang [27] show that fair allocation of debtor contact lists must balance predicted repayment likelihood against agent workload equity and performance heterogeneity, introducing fairness constraints that complicate purely reward-driven optimization.

Jankowski and Paliński [24] also emphasize the need for explainability and human validation in operational environments, which limits the complexity of deployable models.

In short, the literature converges on a central conclusion: effective debt collection requires decision models that operate over sequences of actions, integrate behavioral feedback, and respect operational, ethical, and execution-level constraints. RL provides a powerful theoretical foundation for this transition, but practical implementations remain constrained by data availability, compliance requirements, capacity limits, and the need for human oversight. These challenges motivate constrained, sequential decision frameworks that combine data-driven optimization with explicit policy and compliance constraints.

3.7 Critical Discussion and Research Gaps

The literature reviewed in this chapter demonstrates substantial progress in the application of data-driven methods to debt collection, particularly in predictive modeling, behavioral segmentation, communication optimization, and interaction-level behavioral analysis. Despite these advances, several structural gaps remain unresolved. Collectively, these gaps limit both the practical effectiveness and theoretical maturity of current debt collection systems, motivating the need for a more integrated research approach.

Lack of End-to-End Integrated Models

A central limitation of the current state of the art is the absence of end-to-end systems that are both deployable in regulated environments and capable of integrating prediction, sequential decision-making, communication execution, and compliance-aware policy constraints within a single operational engine. Most existing approaches address these components in isolation: predictive models estimate repayment likelihood, segmentation frameworks classify debtors into behavioral groups, and communication studies optimize individual attributes such as timing or framing. These components are rarely combined into adaptive systems that close the feedback loop between action, observation, and outcome.

As a result, information flows remain fragmented. Predictive scores are often consumed by rule-based workflows, segmentation outputs are manually mapped to fixed strategies, and communication responses are not systematically incorporated into subsequent decision-making. This architectural separation limits adaptability and prevents systems from learning from their own interventions over time. Fragmentation also manifests along multiple dimensions: temporal (static predictions versus sequential processes), behavioral (debtor-centric models versus multi-actor systems), operational (message design versus capacity and assignment constraints), and feedback-related (myopic optimization versus cumulative outcome learning). The lack of models that jointly address these dimensions constrains the development of coherent and adaptive collection strategies.

Prediction, Decision, and Communication as Misaligned Optimization Layers

Closely related to this fragmentation is not only the separation between prediction, decision, and communication, but the fact that these layers are typically optimized toward incompatible objectives. Predictive models prioritize statistical accuracy or calibration, operational systems emphasize workload balance and cost efficiency, communication strategies optimize local response rates, and sequential decision models aim to maximize long-term cumulative outcomes. These objectives are rarely aligned or jointly optimized.

The literature consistently shows that strong predictive performance does not automatically translate into effective action selection. Knowing *what* is likely to happen, for example, whether a debtor will repay, does not specify *how* a creditor should intervene to influence that outcome. Similarly, communication strategies are often optimized locally, such as for timing or framing, without being embedded in a broader sequential decision context. This leads to myopic optimization, where actions are chosen based on immediate expected gains rather than long-term system dynamics. The absence of a unifying optimization objective across prediction, decision, and communication therefore remains a fundamental gap in existing approaches.

Cross-Cutting Limitations in Causal Inference

As discussed in Section 3.5, much of the empirical evidence in the literature remains correlational. Debt collection outcomes emerge from sequences of heterogeneous interactions and are influenced by behavioral confounding, delayed effects, and operational constraints that bias naive comparisons. This attribution challenge is not limited to communication strategy evaluation; it also affects segmentation validity, predictive model calibration, and the offline evaluation of sequential decision policies. The absence of robust causal identification strategies is thus a domain-wide limitation, and it is what RQ5 sets out to address: measuring the effects of personalized communication under sequential decisions, temporal dependencies, and heterogeneous debtor trajectories.

Execution Context Ignored in Decision Models

A further limitation is that many decision models implicitly treat actions as atomic and context-independent, assuming uniform execution regardless of operational conditions. In practice, however, the outcome of a collection action depends not only on the selected message, channel, or timing, but also on *who* executes the action, under what workload conditions, and with what level of experience or consistency.

Empirical evidence indicates that collector performance, behavioral traits, and execution quality materially influence realized outcomes, while workload distribution and capacity constraints shape the feasible action space. Despite this, execution context is rarely incorporated into state representations or policy parameters. As a result, many approaches optimize abstract actions without accounting for execution variability, limiting their applicability in real-world collection environments where human agents and operational constraints play a central role.

Limitations of Explicit Delinquency Labels and the Use of Temporal Proxies. An additional cross-cutting limitation concerns the availability of explicit delinquency or due-date information in real-world datasets. In many operational contexts, particularly when working with aggregated or sector-level payment data, direct observation of payment delay is not feasible. Consequently, both prior work and the present thesis rely on indirect temporal indicators, such as payment clustering, post-valley repayment peaks, and changepoints in activity, to approximate delayed repayment behavior. While these proxies do not provide causal identification of delinquency, they enable the analysis of relative timing effects and the evaluation of communication strategies under realistic data constraints. This methodological choice reflects a trade-off between ecological validity and label precision, and motivates the use of complementary data sources and robustness checks when available.

Ethics and Regulation Treated as Constraints Rather Than Variables

Debt collection operates within a tightly regulated environment, in which legal and ethical requirements directly shape the space of admissible actions. Regulatory frameworks such as the Fair Debt Collection Practices Act (FDCPA) in the United States and EU-level consumer credit and data protection legislation impose explicit limits on contact frequency, permissible communication channels, consent requirements, data usage, and the timing and tone of interactions. These rules are not peripheral constraints but structural determinants of the strategy space: they define which actions are legally permissible, when they may be executed, and how personal data may be processed or profiled. As observed in operational studies such as Duman et al. [16], compliance requirements frequently override purely data-driven recommendations, particularly when suggested actions conflict with channel capacity, consent limitations, or harassment thresholds.

In the European context, these restrictions are grounded in several sector-defining regulatory instruments. The EU General Data Protection Regulation (General Data Protection Regulation (GDPR)) [40] establishes principles of consent, data minimization, purpose limitation, and explainability that directly constrain the design of personalization systems. The Credit Servicers and Credit Purchasers Directive (Directive (EU) 2021/2167) [14] regulates how credit servicers and purchasers of non-performing loans interact with borrowers, including transparency requirements, complaint-handling obligations, and the principle that arrears management must avoid undue harassment, coercion, or misleading practices. The revised Consumer Credit Directive (Consumer Credit Directive II (CCD2), Directive (EU) 2023/2225) [15] updates the consumer-credit framework that applies upstream of arrears, including pre-contractual information and creditworthiness assessments. For financial institutions, the European Banking Authority (European Banking Authority (EBA)) Guidelines on the management of Non-Performing Loans (NPLs) and forborne exposures [20] further impose requirements on engagement processes, affordability assessments, and the fair treatment of distressed borrowers. Payment Services Directive 2 (PSD2) and related open banking rules [13] additionally shape the permissible use of payment and account-access data for inferring liquidity signals or repayment capacity. Across these instruments, contact-frequency limits are framed as proportionality and non-harassment principles rather than as specific numeric caps; numeric defaults adopted in this work therefore reflect organisational-policy ceilings inspired by these principles, not literal regulatory limits.

Together, these instruments define the feasible action space of any automated communication or decision model and show that regulation is not just a boundary condition but a force that shapes strategic behaviour, trust formation, and long-term system dynamics.

At the same time, regulatory constraints interact with debtor behavior in ways that influence optimal policy design. Restrictions on contact intensity limit aggressive escalation strategies and shift the balance toward cooperative or informational interventions. Privacy requirements under GDPR also constrain the degree of personalization that can be applied and necessitate mechanisms for explainability, contestability, and the right to information. These dynamics suggest that regulation should not be treated as a fixed exogenous limit, but as an endogenous component of the decision environment that modulates both action effectiveness and debtor response.

A further limitation of the existing literature is that ethical and regulatory considerations are predominantly modeled as hard constraints applied *after* strategy design. In most studies, compliance requirements, such as contact limits, consent rules, or prohibitions on harassment, are enforced as external filters on otherwise unconstrained optimization. While this approach ensures legal adherence, it fails to incorporate ethical and regulatory effects into the optimization process itself. Behavioral evidence indicates that debtor responses are sensitive to perceived fairness, intrusiveness, and trust; poorly calibrated interventions can trigger annoyance, reactance, or disengagement, thereby altering future engagement and repayment dynamics. These effects, however, are rarely treated as explicit variables within decision models.

As a result, systems may satisfy formal constraints while still producing outcomes that are ethically questionable, reputationally damaging, or psychologically harmful. Critically, ethical violations do not just breach external rules; they change how debtor behaviour evolves and affect long-term system dynamics. This motivates treating ethics and regulation as part of the decision logic itself rather than as external boundary conditions. This misalignment is what RQ6 examines: how regulatory and compliance constraints shape the feasible policy space and performance trade-offs of a deployable automated debt collection strategy engine.

Data Availability and Observability Constraints

The literature reviewed in Sections 3.2–3.6 consistently demonstrates that interaction-level and behavioral metadata often outperform purely financial features in explaining repayment outcomes. Studies such as Ghaffari et al. [17] rely on individual behavioral signals (e.g., contact responsiveness and communication history), while work on sequential optimization and reinforcement learning assumes access to action–outcome tuples that enable attribution and policy evaluation [16, 31]. These approaches implicitly operate under conditions of near-complete observability of debtor state, actions, and outcomes.

In practice, however, such data availability assumptions are difficult to satisfy outside controlled institutional environments. Micro-level interaction datasets are typically proprietary, commercially sensitive, and restricted to single organizations. Access is further constrained by legal, contractual, and competitive considerations, limiting reproducibility and cross-organizational comparability. Regulatory frameworks such as the GDPR also impose strict limitations on the automated processing, storage, and sharing of individual-level debtor data, particularly when used for profiling or decision-making.

As a consequence, many real-world debt collection contexts operate under conditions of *partial observability*. While aggregate payment behavior, sectoral dynamics, and temporal patterns may be observable at portfolio or population level, individual action histories, explicit delinquency labels, and direct action–response attribution are frequently unavailable. This disconnect creates a methodological gap between models that assume full observability and deployment environments in which only indirect or aggregated signals can be exploited.

Despite the central role of timing, cadence, and sequencing decisions in debt collection practice, relatively little work has examined how such decisions can be informed under partial observability. Existing approaches predominantly focus on optimizing *which* action to take given a fully specified individual state, while treating the broader temporal and economic context as exogenous or fixed. The problem of inferring *when* debtors are collectively more likely to be responsive, in the absence of individual-level action–outcome data, therefore remains underexplored.

These data availability and observability constraints shape the methodological approach developed in Chapter 4, which addresses timing and cadence optimization using indirect temporal signals and population-level patterns rather than fully observable interaction histories.

Positioning of This Thesis

This thesis addresses the identified gaps by proposing a unified view of debt collection as a sequential, communication-driven decision problem. Rather than treating prediction, segmentation, communication, and compliance as loosely coupled modules optimised in isolation, the proposed approach integrates these elements into a single decision-support framework in which temporal signals, behavioural constraints, and regulatory requirements jointly shape the recommendation space.

Instead of relying on individual interaction histories or causal action–outcome attribution, the proposed framework operates under conditions of partial observability, emphasizing contextual inference derived from aggregated temporal and sectoral payment dynamics to inform the timing and cadence of communication strategies. Ethical and legal considerations are not imposed solely as external filters applied after strategy design, but are embedded directly into the decision logic as structural constraints that define the feasible recommendation space.

This framework is empirically grounded through access to historical communication metadata from the Invisible Collector platform, enabling validation of timing recommendations against observed collection response patterns under operational conditions.

By treating debt collection as a unified decision-support problem rather than a pipeline of isolated optimisers, this thesis moves beyond static workflows toward adaptive, interpretable, and compliance-by-design collection strategies. The approach prioritises deployability and auditability under realistic data constraints, and provides the conceptual foundation for the methodological framework developed in Chapter 4.

3.8 Chapter Summary

This chapter presented a structured and critical review of the state of the art in data-driven debt collection, framing the problem as one of digital communication, behavioural modelling, and sequential decision-making. Rather than treating debt recovery as a purely financial or operational task, the literature was examined through the lens of interaction design, personalization, and adaptive control.

Section 3.2 reframed debt collection as a digitally mediated communication process, emphasising the relational, emotional, and behavioural dynamics that shape debtor responses. Sections 3.3 and 3.4 reviewed predictive modelling and debtor segmentation approaches, showing that while modern ML techniques can accurately estimate repayment likelihood and identify meaningful debtor typologies, these models primarily address the *what* rather than the *how* of debt collection.

Section 3.5 examined communication strategy optimization and digital nudging, demonstrating that timing, frequency, framing, and channel selection are powerful but highly interdependent levers. Empirical evidence showed that poorly calibrated communication strategies may backfire, increasing annoyance costs and, in some contexts, triggering disengagement or strategic non-payment. These findings expose the limitations of static rules and one-size-fits-all workflows.

Section 3.6 then reviewed decision models for adaptive collection strategies, positioning debt collection as a sequential and hierarchical decision problem. The literature indicates that heuristic workflows and isolated predictive scores are inherently myopic, whereas RL-based and sequential optimization approaches provide a principled framework for learning adaptive policies under delayed feedback, operational constraints, and behavioural trade-offs. These approaches remain constrained in practice by data availability, compliance requirements, and the need for human oversight.

Overall, this chapter identified a central research gap: the absence of integrated, end-to-end systems that jointly address prediction, segmentation, communication, and decision-making under regulatory constraints, while treating ethical and compliance considerations as structural components of the decision environment rather than as external, post hoc filters.

The next chapter addresses this gap by presenting the proposed methodological approach. Building on the insights synthesised in this review, the methodology proposes a dual-layer, timing-aware decision-support framework that combines aggregated payment signals and empirical communication-response patterns to generate interpretable, compliance-by-design recommendations under conditions of partial observability.

Chapter 4

Methodology and Implementation

4.1 Introduction

This chapter presents the design and implementation of a deployable decision-support system for aligning **communication timing and cadence** in debt collection with periods of higher observed payment responsiveness, using **temporal and behavioral metadata**, under conditions of **partial observability**.

Building on the gap analysis in Chapter 3, the proposed approach addresses the lack of timing-aware, compliance-by-design systems that remain interpretable and auditable when direct action–outcome attribution is limited. Rather than attempting to infer causal effects of individual communication actions, the system focuses on identifying and exploiting *temporal alignment signals* derived from both population-level payment behavior and historical collection interactions.

It is structured around three complementary components. First, a **macro-level layer** analyzes aggregated payment activity using *SIBS MULTIBANCO* data to identify seasonal patterns and calendar windows associated with higher observed payment intensity. Second, a **micro-level layer** processes anonymized communication and payment data from the *Invisible Collector* platform to estimate empirical response patterns across channels, steps, and calendar periods. These two layers are developed and analyzed independently, providing distinct but complementary signals. Third, a **decision layer** combines these signals into actionable portfolio-level recommendations, translating temporal insights into compliant contact sequences and cadence strategies.

This dual-layer design lets the system operate under partial observability while drawing on *independent data sources that point in the same direction*. The objective is not to establish causal relationships between actions and outcomes, but to support decision-making by aligning communication strategies with historically favorable temporal conditions.

The chapter is structured as follows. Section 4.2 presents the system architecture, detailing the macro (SIBS) and micro (IC) analytical layers and their integration into a unified decision framework. Sections 4.3 and 4.4 describe the methodological approach and evaluation criteria, including metrics designed for temporal alignment rather than causal inference.

4.2 Conceptual Model of the Solution

4.2.1 Design Goals and Principles

The proposed **Intelligent Debt Collection Strategy Engine** generates **calendar-level timing windows** and **cadence constraints** for contact strategies at the portfolio level. It is designed for environments where the system cannot reliably observe or causally attribute outcomes to specific contact actions (**partial observability**).

Accordingly, the engine adopts a *heuristic-informed, data-driven* approach that prioritizes:

- **Deployability**: integration-ready outputs aligned with existing debt collection workflows;
- **Compliance-by-design**: explicit constraints define the feasible recommendation space;
- **Interpretability and auditability**: recommendations are traceable to underlying data signals, assumptions, and constraint activations, with structured metadata and persistent audit logs.

Rather than optimizing individual actions, the system focuses on *temporal alignment*: identifying periods in which communication is more likely to coincide with favorable payment conditions, based on independent macro- and micro-level signals.

4.2.2 High-Level Architecture

The system is structured around three complementary components:

1. **Macro layer (SIBS)**: analyzes aggregated payment activity to identify seasonal patterns and calendar windows associated with higher observed payment intensity;
2. **Micro layer (IC)**: processes anonymized communication and payment data from the Invisible Collector platform to estimate empirical response patterns across channels, steps, and calendar periods;
3. **Decision layer**: combines macro- and micro-level signals into actionable portfolio-level recommendations, translating temporal insights into compliant contact sequences and cadence strategies.

These components are supported by an integration layer responsible for structured outputs, explainability metadata, and persistent audit logging.

A key design characteristic is that the macro and micro layers are developed and analyzed independently, using distinct data sources and processing pipelines. This separation lets the system combine evidence from two independent sources without assuming causal relationships between communication actions and payment outcomes.

Table 4.1: Input categories consumed by the engine.

Category	Examples of fields	Notes
Aggregated payment data	timestamp (monthly), transaction volume/-value, payment channel (online/physical), sector/Nomenclature des Activités Économiques dans la Communauté Européenne (NACE), geography (region/municipality)	Anonymized and aggregated; no debtor identity required
Communication metadata (IC)	message timestamps, channel (email/SMS/VMS), delivery status, step index, payment timestamps	Derived from anonymized platform data; used for empirical response estimation
Constraint policies	prohibited time windows, max frequency, channel restrictions (opt-in), arrears-stage escalation rules, consent flags	Hard limits define feasible recommendation space
Optional descriptors	portfolio-level indicators (e.g., median arrears age, median debt value), prior response summaries	Used in the decision layer; system remains functional without them

4.2.3 Inputs, Processing, and Outputs (Specifications)

To reduce narrative repetition, inputs and outputs are summarized in Tables 4.1 and 4.2.

4.2.3.1 Core processing pipeline (temporal inference)

The engine extracts temporal structure from both aggregated payment signals and historical communication data to build a *responsiveness map*.

At the macro level, monthly temporal resolution provides a practical lower bound for identifying recurring liquidity-driven patterns (e.g., salary cycles, seasonal expenditure), particularly under partial observability. The primary processing pipeline is:

- **Decomposition**: identify trend and seasonal components (e.g., STL) from aggregated payment series;
- **Peak/valley identification**: detect periods of higher and lower payment activity;
- **Stability checks**: evaluate persistence of seasonal patterns across time and segments (e.g., sector, geography);
- **Window proposal**: define calendar-level contact windows aligned with observed payment intensity;
- **Constraint filtering**: remove suggestions that violate regulatory or organizational policies;
- **Scoring**: assign confidence scores based on signal strength and stability.

Table 4.2: Outputs generated by the engine.

Output	Description
Recommended contact windows	Calendar-level seasonal windows (e.g., "May–July", "March–July") with associated confidence scores
Cadence schedules	Day-offset contact sequences (e.g., D+0 email, D+2 SMS, D+5 VMS) subject to constraint policies
Channel guidance (contextual)	Channel usage informed by empirical response patterns (when available) and policy constraints; not derived from SIBS data
Constraint violations/alerts	Flags when proposed actions conflict with policy constraints or missing consent
Explainability metadata	Traceable metadata linking recommendations to temporal signals, heuristic rules, and constraints
Audit log entries	Timestamped, versioned outputs supporting reproducibility and review

At the micro level, the system estimates empirical response patterns by associating delivered communication events with subsequent payments within defined temporal windows. This process yields response rate estimates by channel, step, and calendar period, without assuming causal attribution.

4.2.3.2 Cadence recommendation (constraints-first)

Cadence is produced as *constraint-compliant schedules* rather than unconstrained optima. The decision layer translates temporal signals into executable contact sequences by enforcing:

- **frequency caps** (regulatory and organizational),
- **minimum spacing** (to mitigate behavioral fatigue and psychological reactance [5, 45]),
- **stage-dependent escalation pacing** (early vs. late arrears).

This layer operates as a rule-based translation mechanism that converts timing signals into operational strategies while preserving compliance and interpretability.

The decision layer classifies each portfolio along two dimensions derived from the IC portfolio profile: *urgency* (based on median days overdue) and *value tier* (based on median pending balance relative to a company-specific or global threshold). This classification determines the contact sequence, channel usage, and escalation timing.

The SIBS seasonal signal acts as an intensity modifier: when the current period falls within a recommended calendar window, the contact sequence is intensified (e.g., additional contacts and reduced spacing); otherwise, the base cadence is applied. The campaign begins immediately upon portfolio initialization (D+0), regardless of the SIBS window status.

4.2.4 Scope Boundaries and Non-Goals

Within scope: timing window inference from aggregated temporal patterns; empirical estimation of response patterns from communication metadata; cadence definition under regulatory and organizational constraints; operation under partial observability; interpretable and auditable portfolio-level recommendations.

Out of scope: content personalization (Natural Language Processing (NLP)/Large Language Model (LLM)s); individual-level action–outcome modeling; formal MDP optimality claims; real-time adaptive control; causal attribution of payments to specific communication actions.

4.3 Research Design and Methodology

4.3.1 Research Design and Rationale

The dissertation follows a **Design Science Research (DSR)** orientation in which a decision-support artifact is designed, implemented, and evaluated using empirical evidence derived from real-world data. The approach is **quantitative and applied**, combining statistical analysis of temporal signals with rule-based decision logic under explicit constraints.

The methodology is structured around three complementary components:

- **Macro-level analysis**: extraction of seasonal payment patterns from aggregated transaction data;
- **Micro-level analysis**: empirical estimation of communication response patterns from historical interaction data;
- **Decision synthesis**: translation of these signals into operational recommendations under compliance constraints.

This design addresses both:

- the *analytical question* of whether temporal structure in payment and interaction data can support timing decisions under partial observability; and
- the *engineering question* of whether such decisions can be produced in a deployable, interpretable, and auditable form.

4.3.2 Research Context

The study targets digitally mediated debt collection processes characterized by:

- **Partial observability**: limited ability to attribute payments to specific communication actions;
- **Compliance constraints**: restrictions on contact timing, frequency, and channel usage (EU-/Portuguese regulatory context);

- **Heterogeneity**: sectoral and geographic variation in payment behavior;
- **Operational constraints**: requirement for integration into existing collection workflows;
- **Ethical considerations**: need for transparency and minimization of unnecessary intrusiveness.

The Invisible Collector platform provides the operational environment and micro-level data, while the macro-level analysis is based on aggregated SIBS MULTIBANCO payment data. The system design remains platform-agnostic.

4.3.3 Population and Sampling

Two complementary data populations are considered:

Macro population (SIBS): aggregated payment transactions, grouped by temporal, sectoral, geographic, and channel dimensions.

Micro population (IC): communication events (messages) and associated payments recorded in the Invisible Collector platform, anonymized and processed across historical backups.

Sampling unit:

- macro level: temporal clusters (e.g., sector $\times$ region $\times$ month);
- micro level: delivered communication events and subsequent payment occurrences within defined temporal windows.

Inclusion criteria:

- macro: sufficient temporal span to capture seasonality (target ≥ 12 months);
- micro: delivered communication events with valid timestamps and associated payment data.

Exclusion criteria:

- incomplete or inconsistent records;
- payments without attributable context (e.g., missing linkage to documents);
- extreme outliers indicative of bulk or non-standard settlement behavior.

4.3.4 Data Collection and Preprocessing

Macro data (SIBS): aggregated monthly transaction data (CSV format) covering multiple years, segmented by sector, geography, and payment channel.

The SIBS MULTIBANCO (SIBS) dataset comprises 294 CSV files covering 87 months of aggregated MULTIBANCO payment activity (January 2019 to March 2026), segmented by sector, payment channel (online/physical), and geography.

Table 4.3: Variables: features, outcomes, and controls.

Type	Variables
Macro features	Month of year, sector/NACE, geography, payment channel (online/physical), transaction volume/value
Micro features	Message timestamp, channel (email/SMS/VMS), delivery status, step index, relative position in sequence
Evaluation outcomes	Temporal Alignment Score, Coverage, Precision, Constraint Compliance Rate
Contextual indicators	Portfolio-level aggregates (e.g., arrears age, debt value) used in decision layer

Micro data (Invisible Collector (IC)): historical platform data extracted from daily database backups, anonymized and transformed into structured analytical datasets (Parquet format). Multiple backups are processed to reconstruct historical records that may no longer exist due to data retention policies.

The IC dataset comprises 537 daily database backups spanning November 2024 to April 2026, from which 9 anonymized Parquet tables were constructed. These contain approximately 991,000 raw message records, 396,000 documents, 259,000 payments, and 1.68 million document–message linkage records. After filtering to known message types, 937,756 messages are retained for analysis; exact filtered counts and attribution figures are reported in Section 5.3.1. Multiple backups are processed to reconstruct historical records that may no longer exist due to GDPR-compliant data retention policies.

Preprocessing pipeline (reproducible scripts):

- parsing and normalization of heterogeneous data formats;
- temporal standardization and ordering;
- consistency checks (missing values, duplicates, referential integrity);
- transformation into analysis-ready structures (time series for SIBS, event-based datasets for IC);
- anonymization and removal of direct identifiers.

4.3.5 Variables and Measurements

Key variables are summarized in Table 4.3.

4.3.6 Analysis Methods

To ensure coherence, the methodology follows a **primary analysis pipeline** aligned with the system architecture.

Macro analysis (SIBS):

- temporal decomposition (e.g., STL) to extract seasonal structure;
- identification of peak and trough periods in payment activity;
- stability analysis across time and segments;
- generation of candidate calendar windows with associated confidence scores.

Micro analysis (IC):

- attribution of payments to prior communication events within defined temporal windows;
- computation of empirical response rates by channel, step, and calendar period;
- comparison of response patterns across temporal segments.

The attribution window for each delivered message is defined as the interval between its timestamp and the next *alarm_action* within the same alarm, using a fallback window of 7 days when no subsequent action exists. Payments occurring after the next action are treated as unattributed. This heuristic does not imply causal attribution, but enables a consistent and reproducible association between payment events and preceding communication steps.

Decision synthesis:

- combination of macro- and micro-level signals;
- application of constraint-compliant cadence rules;
- generation of portfolio-level recommendations.

Predictive modeling: not central to the approach. When used, it is limited to exploratory validation (e.g., assessing feature relevance) and does not replace the interpretable decision logic.

Software/tooling: Python for data processing and analysis; statistical libraries for decomposition and signal detection; structured storage formats (e.g., Parquet) for efficient data handling; relational databases for audit logging and reproducibility.

4.4 Evaluation Framework

4.4.1 Evaluation Objectives

The evaluation addresses the challenge of assessing a timing-based decision-support system under conditions of **partial observability**. The objective is not to establish causal impact, but to assess whether the proposed recommendations exhibit meaningful *temporal alignment* with observed payment and response patterns.

Specifically, the evaluation tests whether the system:

- aligns with historically observed payment intensity (macro-level validation);

Table 4.4: Primary evaluation metrics.

Metric	Definition / intuition	Goal
Temporal Alignment Score (TAS)	Share of total payment volume occurring within recommended temporal windows; measures alignment with observed payment intensity	Higher
Coverage	Fraction of the calendar horizon covered by recommended windows; controls window breadth	Moderate
Precision	Concentration of payments within the recommended windows relative to their duration (TAS / Coverage)	Higher
Constraint Compliance Rate	Proportion of generated recommendations satisfying the active encoded `ConstraintPolicy` (frequency caps, spacing rules, prohibited windows). As the engine filters non-compliant recommendations prior to output, CCR is expected to be 1.00 with respect to the active encoded policy, irrespective of how that policy is configured.	Near 100%

- aligns with empirical communication response patterns (micro-level validation);
- outperforms baseline timing strategies in retrospective comparisons;
- satisfies compliance constraints by construction;
- remains stable across temporal, sectoral, and geographic variations.
- is perceived as fair, non-intrusive, and operationally useful by affected stakeholders (perceived-utility validation).

4.4.2 Primary Metrics

Metrics are designed to be computable under aggregation constraints and to reflect alignment rather than causal effectiveness (Table 4.4).
Example definition (TAS):

$$\text{TAS} = \frac{\sum_{t \in W_{\text{rec}}} P(t)}{\sum_{t \in T} P(t)}.$$

4.4.3 Micro-Level Evaluation (IC Data)

In addition to macro-level alignment, the evaluation incorporates empirical analysis of communication response patterns derived from the Invisible Collector dataset.

The following indicators are computed:

- **Response rates**: proportion of delivered messages followed by payment within a defined temporal window;
- **Channel performance**: response rates segmented by communication channel (e.g., email, SMS, VMS);

- **Step-level effectiveness**: response variation across positions within a contact sequence;
- **Temporal lift**: relative differences in response rates across calendar periods.

These metrics provide an independent empirical signal that complements the macro-level analysis, without implying causal attribution between individual messages and payments.

4.4.4 Baselines and Comparative Benchmarks

Recommendations are compared against the following baselines:

- **Uniform timing**: evenly distributed contact windows across the calendar;
- **Random timing**: randomly selected windows of equivalent duration;
- **Heuristic rules**: simple fixed schedules (e.g., early-month or end-of-month contact).

These baselines provide reference points for assessing whether the proposed approach captures meaningful temporal structure. Cadence-only baselines (varying frequency without temporal alignment) are deferred to future work, as their evaluation requires a counterfactual on contact intensity that the retrospective IC data do not support without an experimental arm.

4.4.5 Validation Strategy

Generalization is assessed primarily through temporal validation: fitting on earlier periods and evaluating on future periods. Sectoral and geographic hold-out validation (holding out specific sector groups or regions and municipalities) are supported by the engine architecture and the SIBS data structure, but their systematic evaluation is deferred to future work; the present results are reported for the financial-sector aggregate.

4.4.6 Perceived Utility Evaluation

The retrospective metrics above assess temporal alignment but cannot capture how the resulting contact strategies are perceived by the people they affect. To complement them, a perceived-utility evaluation is conducted through two structured questionnaires, one eliciting the debtor perspective and one the perspective of collection managers, assessing whether the personalised strategies are perceived as fairer, less intrusive, and more operationally useful than non-personalised baselines. Respondents evaluate three strategies forming an ablation gradient (uniform, segmented, and the full engine) on five-point Likert items, in a within-subject design analysed with non-parametric tests.

This strand is methodologically distinct from the retrospective analysis: it involves primary data collection from human respondents rather than the reuse of historical records, and it measures perception and stated intent rather than observed payment behaviour. It is therefore subject to the limitations of self-report and does not constitute causal evidence. The full instrument design, within-subject analysis plan, and statistical power analysis are detailed in Section 5.5.2, and the instruments themselves are reproduced in Appendix B.

4.4.7 Interpretation and Limitations

Given the absence of controlled experiments and full action–outcome observability, all evaluation results are interpreted as **associational** rather than causal.

Observed improvements in alignment or response patterns indicate that the system successfully identifies temporally favorable conditions, but do not imply that the recommended actions directly cause increased repayment.

Where applicable, comparative analyses (e.g., differences in response rates across periods) are presented as descriptive indicators of temporal variation rather than estimates of causal uplift.

4.5 Ethical Considerations

The system is designed to minimize privacy risks and prevent harmful or overly intrusive automation in debt collection processes.

- **Privacy and GDPR compliance**: all data used in the analysis is anonymized or aggregated prior to processing. SIBS data is inherently aggregated, while Invisible Collector data is anonymized during the ETL process, with direct identifiers (e.g., names, emails, phone numbers, tax identifiers) removed or replaced with synthetic values. Historical backups are processed to reconstruct temporal patterns without retaining personally identifiable information. Data storage and access follow the principle of data minimization and appropriate security controls.
- **Human-in-the-loop**: the system operates as a decision-support tool and does not automate actions. All recommendations are advisory and intended for human review. Outputs include structured metadata and audit logs to support traceability and accountability.
- **Fairness and non-discrimination**: the system does not rely on sensitive personal attributes. Inputs are restricted to temporal, contextual, and operational features. Where segmentation is applied (e.g., by sector or geography), it reflects aggregate behavioral patterns rather than individual characteristics. Potential disparate effects should be monitored in deployment contexts.
- **Responsible use and proportionality**: recommendations are designed to reduce unnecessary or poorly timed contact attempts by avoiding periods of low payment activity and respecting regulatory and organizational constraints. The system prioritizes proportional communication strategies and should be deployed with oversight, pilot testing, and periodic audits.

- **Questionnaire participation**: the perceived-utility evaluation involves human respondents who participate voluntarily and anonymously. No personal or identifying data is collected, respondents are informed of the academic purpose of the study, and responses are used solely for research. The hypothetical scenarios presented involve no real debt or real debtor data.

4.6 Contributions and Significance

This dissertation contributes to both the scientific understanding and practical implementation of timing-aware decision-support systems in debt collection under partial observability.

Scientific contributions:

- a dual-layer analytical framework combining macro-level temporal patterns (aggregated payment data) and micro-level empirical response patterns (communication metadata), enabling timing-aware decision-making without relying on causal attribution;
- an evaluation methodology based on temporal alignment metrics and retrospective benchmarking, suitable for environments with limited action–outcome observability;
- an empirical characterization of seasonal payment behavior and communication response patterns across time, sectors, and channels, based on independent data sources.

Practical contributions:

- a deployable and auditable decision-support system that generates timing windows and cadence recommendations for portfolio-level debt collection strategies;
- a compliance-by-design approach integrating regulatory constraints directly into the decision logic;
- integration-ready outputs and audit mechanisms supporting real-world deployment within existing collection workflows.

4.7 Chapter Summary

This chapter presented the methodology and implementation of a decision-support system for timing and cadence recommendations in debt collection under partial observability. It defined the system architecture, including macro- and micro-level analytical components, and described the research design, evaluation framework, and implementation process. The following chapter presents the empirical results obtained from the application of this approach.

Chapter 5

Results and Evaluation

5.1 Introduction

This chapter presents the empirical results obtained from the design, implementation, and evaluation of a decision-support engine for communication timing and cadence in debt collection under partial observability. The work builds on the methodology described in Chapter 4 and is organised around three complementary analytical layers.

The first layer, presented in Section 5.2, reports the macro-level temporal analysis conducted over aggregated payment data from SIBS. This layer identifies seasonal patterns and calendar windows associated with higher observed payment intensity across consumer segments and transaction metrics.

The second layer, presented in Section 5.3, reports the micro-level attribution analysis derived from anonymised communication and payment records provided by the IC platform. This layer estimates empirical response rates and temporal lift across contact channels, policy steps, and calendar periods.

The third layer, presented in Section 5.4, describes the portfolio strategy engine that integrates the signals from both preceding layers into actionable, constraint-compliant cadence recommendations. The section details the design rationale, the urgency-value cadence matrix, the SIBS modifier, and illustrative engine output.

Finally, Section 5.5 presents the validation approach, combining retrospective quantitative metrics with a perceived utility evaluation based on structured questionnaires administered to debtors and collection managers.

Throughout this chapter, all results are interpreted as associational rather than causal. The absence of controlled experiments and the conditions of partial observability under which the engine operates preclude claims of causal attribution between recommended actions and observed payment outcomes. Observed improvements in temporal alignment and response patterns indicate that the engine successfully identifies historically favorable conditions for contact, but do not imply that the recommended actions directly cause increased repayment.

5.2 Macro-Level Temporal Analysis: SIBS Data

5.2.1 Dataset Overview and Preprocessing

The macro-level analysis is based on aggregated payment data provided by SIBS MULTIBANCO, covering 87 months of transaction activity from January 2019 to March 2026. The dataset comprises 294 Comma-Separated Values (CSV) files organised across eight structural variants, each encoding a different aggregation of the underlying payment activity: `WIDE_TS`, `WIDE_TS_CANAL`, `WIDE_TS_FOREIGN`, `WIDE_FREQ`, `SCALAR`, `MULTI_SCALAR`, `DIST_TABLE`, and `ORIGIN_TABLE`. These variants capture complementary views of the data, ranging from time series of transaction volume and value per sector to distribution profiles of customer origin and loyalty indicators.

Three analytical dimensions are considered throughout the macro-level analysis. The first is the payment modality, with three levels: the full population (All), physical point-of-sale transactions (Physical), and online transactions (Online). It is important to note that these modalities refer to the payment channel recorded in the SIBS data, and are distinct from the contact channels used by the collection engine (email, SMS, and Voice Message Service (VMS)). The second dimension is the transaction metric, with three variants: number of purchases (`N.Compras`), total transaction value (`Val.Compras`), and average transaction value (`Val.Médio`). The third is the calendar month, used as the primary temporal unit for seasonality inference. This yields nine modality–metric combinations that are analysed independently.

Preprocessing followed a structured pipeline implemented in Python. Each file was parsed according to its structural variant and normalised to a common monthly time series format. Missing values, encoded as `-` in the raw CSV files, were handled via linear interpolation prior to decomposition.

5.2.2 Seasonal Decomposition

Seasonal structure was extracted from each of the nine modality–metric time series using Seasonal and Trend Decomposition using Loess (STL) decomposition [11], which separates each series into trend, seasonal, and residual components in a robust and iterative manner. STL was selected for its tolerance of outliers and its ability to handle series where the seasonal component varies gradually over time. Both properties are relevant here, given that the dataset spans the 2020–2021 period of disrupted economic activity, during which payment patterns deviated substantially from their long-run seasonal structure.

Figure 5.1 shows the STL decomposition for the All–`N.Compras` series, which exhibits the dominant mid-year seasonal regime identified across the dataset.

Two distinct seasonal regimes were identified across the analysed combinations. The first is a mid-year peak, concentrated approximately between May and October, and most pronounced in physical transactions and in `N.Compras`-based metrics. The second is a year-end peak, observed approximately between October and March, and more strongly associated with `Val.Médio`-based

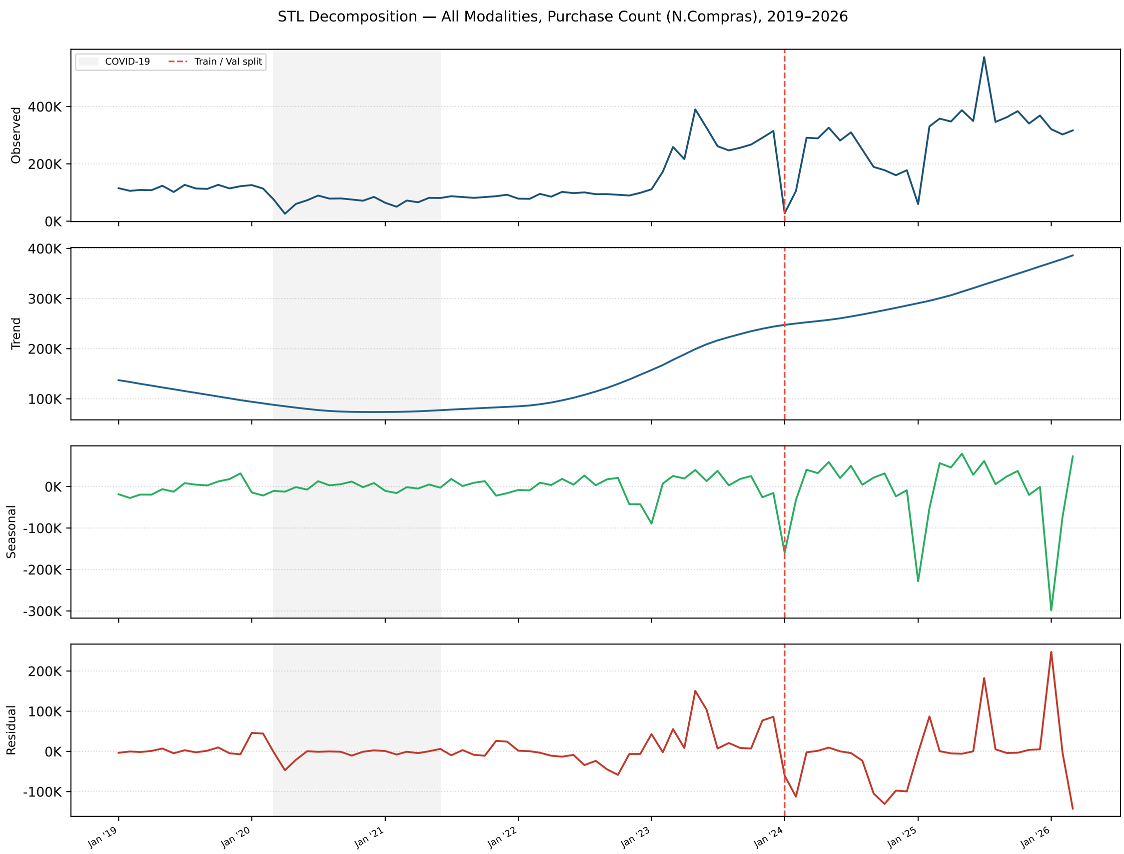

Figure 5.1: STL decomposition of the All–`N.Compras` time series (purchase count, all payment modalities combined), covering January 2019 to March 2026. The four panels show the observed series, the extracted trend, the seasonal component, and the residual. The grey shaded region marks the COVID-19 disruption period (2020–2021), and the red dashed line indicates the train/-validation split at January 2024. Boundary artefacts are visible in the seasonal panel near the end of the series (2025–2026), a known property of STL at series boundaries, and should be interpreted with caution.

metrics. The two regimes share October as a transition month, which partly explains the divergence in collection effectiveness observed in that month, discussed further in Section 5.3.4. Within each regime, however, the productive months for collection action are a sub-period of the seasonal window: as the micro-level lift analysis in Section 5.3.3 shows, June–July anchors the mid-year regime and January–February anchors the year-end regime, with the remaining months of each window contributing diminishing or negative returns.

5.2.3 Peak and Valley Detection

Peak and valley detection was applied to the extracted seasonal components to identify calendar windows of elevated and reduced payment intensity. The detection algorithm was configured with a half-width of 2 months and a maximum window length of 9 months (`half_width=2`, `max_window_months=9`), balancing sensitivity to genuine seasonal structure against the risk of producing windows that are too broad to be operationally useful.

Table 5.1 summarises the detected windows for each modality–metric combination, together with the corresponding TAS, precision, and coverage scores. Entries with a confidence score below 0.5 were excluded from this summary, as they correspond to combinations where the seasonal signal was insufficiently stable to support reliable window inference.

Table 5.1: Detected contact windows per modality–metric combination, evaluated on the held-out validation period (January 2024–March 2026). Only entries with confidence > 0.5 are shown. Each row reflects the highest-confidence entry per combination, drawn from one of two SIBS file sources: the `WIDE_TS` aggregate series (60-month training slice, January 2019–December 2023) for All/Val.Médio and All/N.Compras, where the longer history resolves a higher-confidence peak; and the `WIDE_TS_CANAL` per-channel series (24-month training slice, January 2022–December 2023) for the Physical and Online rows and for All/Val.Compras. The engine's live recommendation path runs STL on the full `WIDE_TS_CANAL` series and may resolve a shifted peak (see Section 5.4.3). TAS is the share of total payment volume within the recommended window. Precision is TAS divided by coverage.

Modality	**Metric**	**Window**	**TAS**	**Precision**	**Confidence**
All	Val.Médio	Oct–Feb	0.636	1.527	1.00
All	N.Compras	Jun–Oct	0.574	1.378	1.00
All	Val.Compras	Oct–Feb	0.477	1.144	0.56
Physical	Val.Médio	Nov–Mar	0.448	1.075	1.00
Physical	Val.Compras	May–Sep	0.440	1.055	0.77
Physical	N.Compras	Jun–Oct	0.387	0.928	0.90
Online	N.Compras	Oct–Feb	0.461	1.106	0.69
Online	Val.Compras	Oct–Feb	0.457	1.096	0.63
Mean (confidence $>$ 0.5)			0.485	1.164	–

Figure 5.2 provides a calendar view of the same results, with colour intensity encoding the confidence score of each window.

Three combinations reach the maximum confidence of 1.00. The two aggregate-modality results carry the highest TAS: All/`Val.Médio` (Oct–Feb) is the densest of all detected windows, concentrating 63.6% of total `Val.Médio` payment volume in five months at 52.7% above uniform expectation, and anchors the year-end regime; All/`N.Compras` (Jun–Oct) achieves TAS 0.574 and precision 1.38, anchors the mid-year regime used as the running worked example in Section 5.4.4, and concentrates 57.4% of the total annual payment volume in a five-month window at 38% above uniform expectation. The third confident window, Physical/`Val.Médio` (Nov–Mar), also reaches confidence 1.00 but with a more modest TAS of 0.448; it corroborates the year-end regime from the channel-disaggregated file and is not headlined here precisely because its concentration is closer to uniform.

The Online modality produced windows for two out of three metrics above the confidence threshold: `N.Compras` and `Val.Compras` both yielded an October–February window with confidence scores of 0.69 and 0.63 respectively. The `Val.Médio` combination fell below the threshold (confidence below 0.5) and was excluded. The Online window is the inverse of the dominant physical pattern: rather than concentrating in the mid-year months, Online payment activity peaks in the October–February period, aligning more closely with the year-end regime. This distinction is discussed further in Section 5.2.5.

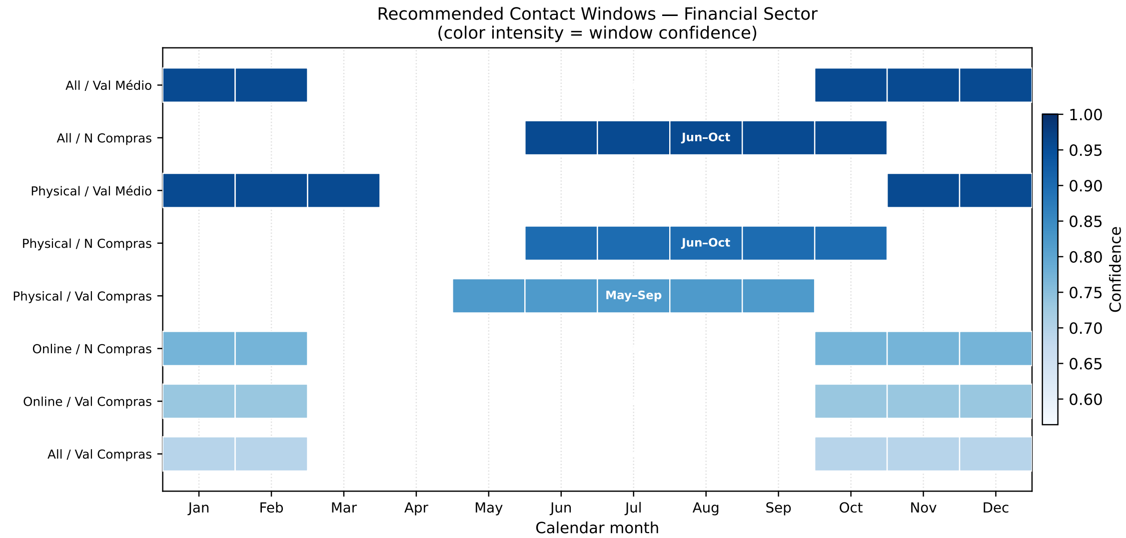

Figure 5.2: Recommended contact windows for the financial sector across all modality–metric combinations with confidence above 0.5, displayed as a calendar Gantt chart. Colour intensity encodes the window confidence score. Two clusters are visible: a mid-year group (approximately May–October), dominant across physical and aggregate series; and a year-end group (approximately October–March), covering the Online modality, Physical/Val.Médio, and the All/-Val.Médio and All/Val.Compras series. Within the year-end group, most series concentrate in October–February, with Physical/Val.Médio extending one month later (November–March).

5.2.4 Comparison with Baselines

The engine's recommended windows were compared against three baselines: uniform timing (windows distributed evenly across the calendar), random timing (randomly selected windows of equivalent total duration), and an expert heuristic baseline corresponding to a January–February contact window.

Across all modality–metric combinations with confidence above 0.5, the engine achieved a mean TAS of 0.485 and a mean precision of 1.164, compared to a TAS of 0.42 under the uniform baseline. The expert heuristic, which targets January and February, is a deliberately weak reference point: this window coincides with a period of reduced consumer liquidity following end-of-year expenditure, and serves as a lower bound rather than a competitive alternative (mean TAS ≈ 0.21). Figure 5.3 shows the per-combination TAS scores across all evaluated strategies.

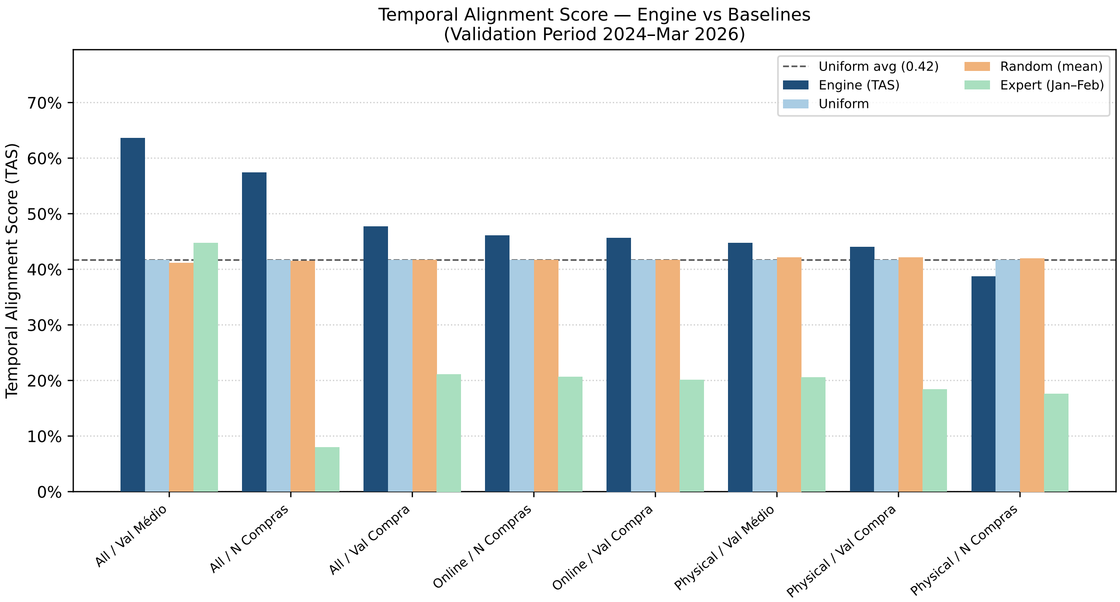

Figure 5.3: TAS scores per modality–metric combination for the engine's recommended windows and all baselines, evaluated on the validation period (2024–March 2026). Only combinations with confidence above 0.5 are shown. The dashed line marks the uniform baseline average ($\overline{\text{TAS}} = 0.42$). The expert heuristic (January–February) serves as a lower-bound reference.

5.2.5 Discussion

The two seasonal regimes identified in Section 5.2.2 reflect distinct payment dynamics, consistent with the well-documented seasonality of aggregate consumption and economic activity [3]. The mid-year peak, concentrated in physical transactions and driven by `N.Compras`, is consistent with increased consumer transaction volume during spring and summer months, a pattern linked to discretionary spending and broader economic activity cycles. The year-end peak, more visible in `Val.Médio`, likely reflects larger individual transactions associated with the December salary supplement (*subsídio de Natal*) and related year-end financial settlements, rather than a general increase in the number of transactions.

The divergence between the Physical and Online modalities is one of the more substantive findings of the macro-level analysis. While physical transactions concentrate in the mid-year window (May–October), online transactions exhibit a year-end concentration (October–February), consistent with the `Val.Médio`-driven regime observed in the aggregated data. This suggests that the two modalities are driven by different behavioural patterns: physical transactions appear more sensitive to general consumer activity cycles, while online transactions may be more closely tied to end-of-year financial events such as salary supplements and year-end settlements. From a practical standpoint, this means that a single population-level window is insufficient to characterise all payment behaviour, and that modality-level segmentation should be considered when mapping SIBS signals to collection portfolio contexts.

The boundary artefacts observed in the STL seasonal component near the end of the series

(2025–2026) are a known property of the method and do not affect the core results, which are anchored to the stable central portion of the series. Since TAS is computed as the fraction of observed payment volume falling within windows inferred from the training period, rather than from the STL seasonal component itself, the boundary artefacts do not affect the reported alignment scores. Caution is nonetheless warranted when interpreting seasonal estimates for the most recent months.

5.3 Micro-Level Attribution Analysis: IC Data

5.3.1 Dataset Overview and Extract, Transform, Load (ETL) Pipeline

The micro-level analysis is based on anonymised communication and payment data extracted from the IC platform. The dataset was reconstructed from 537 daily database backups spanning November 2024 to April 2026, processed through a custom ETL pipeline. Each backup was restored to a temporary MySQL instance, from which records were extracted, anonymised, and appended to a set of structured Parquet files. Direct identifiers, including customer names, contact details, tax identifiers, and document references, were replaced with synthetic values during extraction. Deduplication was applied by primary key. Where records carry an `updated_at` field, rows are sorted by that field and the latest version is retained, so that mutable status fields (such as delivery confirmation or payment linkage) reflect the most recent observed state. For tables without an update timestamp, the last occurrence encountered during the chronological backup pass is retained, which yields the same behaviour. Records present in earlier backups but removed from later ones are nonetheless preserved, allowing reconstruction of historical states that may have been deleted from the live database under GDPR-compliant retention policies.

The resulting dataset comprises nine anonymised Parquet tables. The tables relevant to this analysis contain approximately 991,000 raw message records, 396,000 documents, 259,000 payments, and 1.68 million document–message linkage records, covering 30 companies operating under 7 distinct collection policies. After filtering to known message types (email, SMS, and VMS), 937,756 messages were retained for analysis, of which 502,818 were classified as delivered. Of the 258,953 payment records, 25,447 could not be linked to any document and were excluded from attribution; the remaining 233,506 payments form the basis of all response rate and lift calculations reported in this section.

The key database tables used in attribution are: `alarms` (debt recovery cases), `alarm_actions` (individual contact steps, encoding per-company policy sequences), `messages` (outbound communications with delivery status), and `payment_lines` (linking payments to documents).

5.3.2 Response Rates, Step Effectiveness, and Payment Timing

Overall response rates were computed as the proportion of delivered messages followed by a payment within the attribution window, defined as the interval between consecutive steps in the policy sequence (typically 1–7 days, median 3 days), with a 7-day fallback applied to the final step in each sequence. For email, messages with open or click events were counted as delivered, in addition to those with an explicit delivery confirmation.

Response rates varied across contact channels. SMS achieved the highest rate at 1.5%, followed by email at 1.3%, and VMS at 0.4%. These figures reflect the aggregate rate across all companies, policies, and calendar periods, and should be interpreted as population-level indicators rather than causal estimates of channel effectiveness. The relatively low VMS rate is consistent with the channel's role in the collection sequence: VMS is typically deployed at later stages and for lower-value debts, where response rates are structurally lower regardless of channel.

Response rates also varied substantially by position within the contact sequence. The first step achieved the highest rate at 4.7%, followed by the second at 3.3%. Rates generally declined for later steps, with a partial recovery at step 4 (1.4%), consistent with the onset of escalation to a new channel. This pattern supports the principle of initiating the most impactful contact attempts early in the sequence, particularly at the opening of temporally favourable periods.

Among payments that occurred within the attribution window, the median time between message delivery and payment was 2.6 days. This rapid response pattern is consistent with findings in the State of the Art (SotA) literature, which report that among customers who pay after a collection action, the large majority do so within 72 hours of contact [16], and supports the use of short attribution windows in the analysis.

5.3.3 Temporal Lift Analysis

Temporal lift was computed for each calendar month as the relative deviation of that month's response rate from the grand mean response rate across all delivered messages (1.28%). A positive lift indicates that contacts delivered in that month were followed by payment at a higher rate than average; a negative lift indicates the opposite. This measure is purely descriptive and does not control for confounding factors such as portfolio composition or contact volume.

The results reveal pronounced seasonal variation in collection effectiveness. The five months with the strongest positive lift are July (+44.6%), June (+23.2%), January (+15.4%), May (+13.8%), and February (+12.5%). The months with the strongest negative lift are October (−35.1%), August (−19.2%), November (−16.4%), and September (−9.6%). All lift estimates are based on sample sizes of at least 20,000 delivered messages per month and are therefore considered stable.

5.3.4 Discussion

The temporal lift results provide an independent empirical signal that partially converges with, and partially refines, the macro-level windows identified in Section 5.2. The relationship between the two layers is not one of simple agreement but of complementary constraint.

The five highest-lift months all fall within at least one SIBS window. July and June, the two strongest months (+44.6% and +23.2%), sit inside the mid-year windows identified for All/`N.Compras` (June–October), Physical/`N.Compras` (June–October), and Physical/`Val.Compras` (May–September). January and February (+15.4% and +12.5%) fall inside the year-end windows

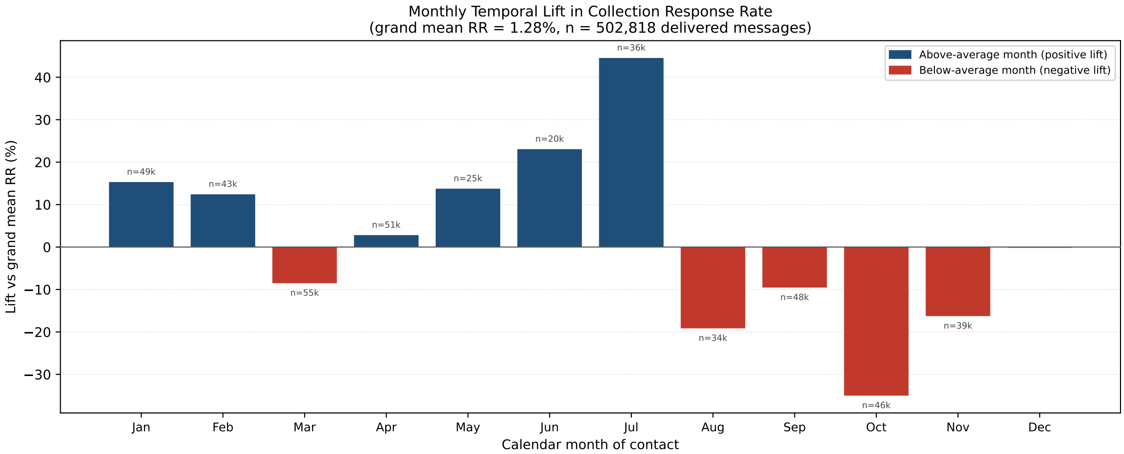

Figure 5.4: Monthly temporal lift in collection response rate relative to the grand mean (1.28%), computed across all 502,818 delivered messages from 30 companies. Sample sizes per month are shown adjacent to each bar. Positive lift (blue) indicates months where contacts were followed by payment at an above-average rate; negative lift (red) indicates the opposite. December recorded near-zero lift (−0.3%) and is shown without a sample size annotation.

covering All/`Val.Médio`, All/`Val.Compras`, and the Online combinations (October–February). This alignment supports the core assumption of the macro-level analysis: that seasonal patterns in aggregate payment activity are reflected, at least partially, in debtor payment behaviour.

However, the same SIBS windows also contain months with strongly negative lift. October (−35.1%) opens six SIBS windows yet produces the worst collection outcomes in the dataset. August (−19.2%) and September (−9.6%) sit deep inside the mid-year windows. November (−16.4%) opens the Physical/Val.Médio window (November–March) and sits as the second month of the October–February windows, despite being one of the worst-performing months for collection. This divergence is not a contradiction but a clarification: SIBS measures when consumers transact, while IC lift measures when contacts are most likely to trigger payment. These are related but distinct signals.

The pattern that emerges from reading both layers together is that each seasonal regime has a high-lift sub-period and a low- or negative-lift remainder. For the mid-year regime, the productive months are June and July; surrounding months show declining or negative returns despite remaining inside the SIBS window. For the year-end regime, the productive months are January and February; October and November, which also sit inside year-end SIBS windows, produce negative lift. This finding has a direct implication for the portfolio strategy engine: the SIBS layer determines *whether* the modifier is engaged (window membership), while operators are expected to initiate during the high-lift sub-period of each window (June–July for the mid-year regime and January–February for the year-end regime) so that the intensified sequence lands on the productive months rather than on the low-lift remainder. The SIBS layer identifies the season; the IC layer identifies where within that season to act.

5.3.5 Limitations and Scope

Several limitations apply to the micro-level analysis and should be taken into account when interpreting the results.

First, all response rate and lift estimates are associational. The attribution methodology links payments to preceding messages within a defined temporal window, but cannot establish that the message caused the payment. Debtors who were already intending to pay may have done so regardless of contact, and the observed rates therefore represent an upper bound on the causal effect of communication.

Second, orphaned payments, defined as payment records with no linkage to any document in the dataset, were excluded from all attribution calculations. These 25,447 records (9.8% of total payments) may represent legitimate payments that were processed through channels not captured in the data, or records affected by data retention deletions. Their exclusion may bias reported response-rate estimates downward if they correspond to genuine post-contact payments, but the magnitude and even the direction of the bias cannot be fully established without recovering the missing document linkages.

Third, the attribution window is policy-defined and variable. For steps with a defined successor, the window closes when the next step fires (median 3 days). For the final step in a sequence, a fixed 7-day fallback window is applied. This means that response rates for later steps may be understated if some debtors pay after the fallback window closes.

Fourth, the lift analysis does not control for portfolio composition. Months with higher proportions of recently assigned, lower-value, or more cooperative debtors may show higher response rates for reasons unrelated to calendar timing. The observed lift should be interpreted as descriptive evidence of temporal variation rather than as an estimate of causal seasonal effect.

5.4 Portfolio Strategy Engine

5.4.1 Design Rationale

Debt collection shares fundamental mechanics with other Call to Action (CTA)-driven communication domains such as sales and marketing: a defined set of contact stimuli, a measurable response (payment), and a trackable interaction history. This structural similarity means that data-driven approaches developed in adjacent domains are applicable here, provided the stimuli and tracking infrastructure are well defined. The IC platform provides exactly this infrastructure, making portfolio-level strategy recommendation tractable.

It is worth stating the granularity boundary up front. The SIBS layer operates at monthly resolution and contributes only the seasonal/calendar-window component (which months are favourable). The day-offset cadence within a contact sequence is produced by the heuristic cadence matrix described in Section 5.4.2, calibrated against the operational response patterns recovered from the IC data; it is not, and could not be, inferred from the monthly SIBS signal.

A key design constraint is that collection difficulty is not uniform across a portfolio. Carrera and Benalcázar [8] show that repayment dynamics vary substantially across arrears stages, with later-stage cases exhibiting weaker temporal sensitivity and requiring more cautious pacing. This motivates a strategy that adapts contact intensity and escalation timing to the urgency profile of the portfolio rather than applying a uniform schedule.

A second constraint concerns contact frequency. Saulītis [45] provides empirical evidence that repeated reminders can reduce repayment likelihood when perceived as excessive, triggering behavioral reactance and strategic disengagement. Liu et al. [31] further show that indiscriminate increases in contact frequency tend to optimise short-term metrics while degrading longer-term engagement. These findings motivate an explicit frequency cap and minimum spacing policy. The engine exposes a configurable `ConstraintPolicy` object that defines upper-bound caps on contacts per month and per week and a two-mode minimum inter-contact spacing: a base floor of 3 days for sequences generated without the SIBS modifier, and a reduced floor of 1 day for sequences produced inside an active SIBS window, where the modifier's 20% compression and inserted seasonally-motivated contacts can land adjacent to existing steps. Deployment contexts can tighten either floor to match organisational or regulatory ceilings. The per-cell cadences described in Section 5.4.2 are calibrated to the cell's urgency and value tier within those configurable bounds, scaling from 5 contacts over 14 days for the least urgent cases to 12 contacts over 30 days for the most critical. Constraint compliance is treated as a design requirement rather than a post-hoc filter: sequences violating the active policy (base or SIBS-modifier floor, depending on whether the modifier is engaged) are excluded prior to output.

The engine is therefore designed around three principles: adapt contact intensity to portfolio urgency and value tier; engage the seasonal modifier when the initialisation date falls inside a SIBS window, under the operational expectation that campaigns are initiated during the window's high-lift sub-period rather than its low-lift remainder; and enforce compliance constraints by construction.

5.4.2 Cadence Matrix

The core decision structure is a 3×2 cadence matrix, shown in Figure 5.5, that maps each portfolio to a contact sequence based on two dimensions derived from the portfolio profile.

The first dimension is urgency, defined by the median days overdue across the portfolio: Recent (< 30 days), Moderate (30–60 days), and Critical (> 60 days). The second dimension is value tier, defined by the median pending balance relative to the company's historical Median (P50) (global fallback: €58, the IC dataset median). Portfolios at or below the threshold are classified as Low Value; those above as High Value.

The six cells differ in contact count, horizon, channel mix, and escalation threshold. Contact counts range from 5 contacts over 14 days for Recent+Low to 12 contacts over 30 days for Critical+High. Escalation to legal or external action is triggered at D+90 for Recent, D+60 for Moderate, and D+45 for Critical, reflecting the declining recovery probability with delinquency duration [8].

VMS is absent from Low Value cells in the Recent and Moderate rows, where the cost of the channel is not recoverable given the pending balance. For Critical+Low, VMS is introduced from D+14 onward, where escalation pressure justifies the additional cost despite the lower value tier.

Within each cell, the role assignment of the three channels reflects their operational asymmetries. Email opens each sequence and recurs at formal-notice steps: it is asynchronous, carries no per-length cost, and is the appropriate medium for content that does not fit a 160-character window (full payment references, regulatory disclosures, pre-escalation notices). SMS dominates the follow-up positions, where the highest empirical response rate (1.5%) and low cognitive load make it the most efficient reminder channel. VMS is reserved for the late steps of high-urgency or high-value sequences, where its higher cost is amortised by the proximity of escalation and its voice presence serves a function that the asynchronous channels cannot.

The value threshold adapts to each company's historical P50. Companies with sufficient transaction history use their own median as the threshold, falling back to the global €58 figure when no history is available. This per-company calibration avoids applying a global threshold to portfolios with systematically different value distributions.

5.4.3 SIBS Modifier

When the portfolio is initialised during an active SIBS window, the base cadence sequence is intensified. Up to two additional contacts are distributed across the campaign horizon, and the spacing between existing contacts is compressed by 20%. The campaign always begins at D+0 regardless of SIBS window status: the seasonal signal influences intensity, not the decision to initiate contact.

The modifier operates on the sequence as a whole: all contact offsets are compressed uniformly by 20% and up to two contacts are inserted at evenly spaced positions across the compressed horizon. Its purpose is to concentrate collection effort whenever the initialisation date falls inside a SIBS window. The operational expectation, consistent with the finding in Section 5.3.4 that IC lift varies substantially within each window, is that operators initiate during the high-lift sub-period of the window (June–July for the mid-year regime, January–February for the year-end regime) so that the intensified sequence lands on the productive months rather than on the low- or negative-lift months that also sit inside the window.

The Mar–Jul window returned by the engine reflects STL decomposition of the `WIDE_TS_CANAL` per-channel file (51 months, January 2022–March 2026), which is the file the live recommendation path consumes; the Jun–Oct window reported in Table 5.1 for All/`N.Compras` reflects STL on the `WIDE_TS` aggregate file (60-month training slice, January 2019–December 2023, evaluated against the held-out period from January 2024 onward). The two files capture aggregated payment activity at different granularities and therefore resolve partially different seasonal peaks for the same modality–metric pair.

Outside a SIBS window, the base cadence applies with a minimum inter-contact spacing of 3 days. These defaults are configurable per company.

5.4.4 Example Engine Output

To illustrate the engine's output concretely, this section presents two worked examples, both classified as Critical+Low Value. The first (Company 112) illustrates the portfolio profile and timing panels as rendered by the engine's User Interface (UI). The second (Company C62) contrasts inside-window and outside-window cadence sequences for the same cell classification under different initialisation dates.

Company 112: portfolio profile and timing panels. The portfolio comprises 231 documents with a median pending balance of €21.45 (P25 = €0, P75 = €178.60), median days overdue of 491, and 100% of documents over 60 days overdue. The company's historical P50 of €125 is used as the value-tier threshold; with a portfolio median of €21.45 well below this threshold, the portfolio is classified as Critical+Low Value. The engine is evaluated on 16 May 2026, which falls inside the active Mar–Jul SIBS window; the SIBS modifier therefore applies, expanding the base 9-contact/30-day sequence to 11 contacts over 24 days.

Figure 5.6 shows the portfolio profile panel, including the cadence cell classification, historical resolution rate, and VMS policy rationale.

Figure 5.7 shows the timing and dunning sequence panels, including the active SIBS window, the expected IC lift, and the full 11-step contact sequence with SIBS-motivated steps flagged.

Company C62: inside-window vs. outside-window sequences. Table 5.2 presents a second illustration using a larger portfolio (Company C62, 64,714 documents), contrasting the cadence sequences produced under two initialisation dates for the same Critical+Low classification. Case A is initialised on 16 May 2026, inside the active Mar–Jul SIBS window. Case B is a retrospective simulation showing what the engine would have recommended for a portfolio initialised on 1 November 2025, scored against the seasonal model fit on the full 2022–March 2026 dataset; in deployment the corresponding live model would have been fit only on data available at that initialisation date. At the simulated initialisation date, no SIBS window was active and the next window was 120 days away. The merged timeline below shows both sequences on a shared day axis; step indices are a combined reference and do not represent per-case step counts.

Case A demonstrates the SIBS modifier in action: the base 9-step sequence is expanded to 11 steps and compressed from 30 to 24 days, with the two additional contacts labelled as seasonally motivated. The expected RR for Case A is 1.47%, computed as the IC overall mean of 1.28% scaled by the +15.2% empirical lift averaged across the five Mar–Jul months ($1.28\% \times 1.152 \approx 1.47\%$). This is a conservative estimate: averaging over the full window dilutes the contribution of the high-lift sub-period identified in Section 5.3.3. Restricting the average to the June–July high-lift sub-period yields +33.9% lift and a corresponding RR of $1.28\% \times 1.339 \approx 1.71\%$. The 16 May 2026 initialisation date is chosen precisely so that the compressed 24-day sequence lands on those high-lift months rather than on the low-lift March–April portion of the window. The independent SIBS seasonal uplift of +32.0% on the same window confirms the direction from the macro signal but does not enter the RR formula. For Case B, the engine reports the same window

RR of 1.47% as the best available estimate for when that window next opens; outside the window, the expected rate depends on the month's IC lift; the annual unconditional mean is 1.28%, but individual months deviate substantially from this figure in both directions. Case B produces the base sequence unchanged, noting that the next favourable window is 120 days away.

5.4.5 Discussion

The engine produces recommendations that satisfy the encoded `ConstraintPolicy` by construction: sequences violating the active policy's frequency caps, mode-appropriate minimum spacing rule (3-day base floor or 1-day SIBS-modifier floor, depending on whether the modifier is engaged), or channel consent restrictions are excluded prior to output. The Constraint Compliance Rate (CCR) is therefore 1.00 across all generated recommendations with respect to the active constraint set, by design rather than by empirical observation. This guarantee is bounded by what the policy actually encodes; recommendations remain compliant with respect to the configured organisational and regulatory ceilings, not with respect to every conceivable legal obligation a deployment context might impose. Operators retain responsibility for ensuring the encoded policy reflects the applicable regulatory regime.

The engine is a decision-support system, not an autonomous actor. All recommendations are advisory and include explainability metadata linking each contact to the temporal signals, heuristic rules, and constraint activations that produced it. The operator retains full discretion over whether and how to act on the recommendation. The expected RR figures in the output are empirical averages from the IC dataset, not predictions for the individual case, and are clearly labelled as associational indicators in the UI.

The scope is population-level temporal alignment. The engine does not model individual debtor behavior, does not learn from feedback loops, and does not make causal claims about the effect of any specific contact on any specific payment. The `INSUFFICIENT_SIGNAL` flag visible in Figure 5.7 illustrates this: when the seasonal signal for a given modality is too weak to support a reliable window recommendation, the engine falls back to a conservative uniform cadence rather than producing a low-confidence recommendation.

5.5 Validation

5.5.1 Quantitative Validation: Retrospective Metrics

The quantitative validation assesses whether the engine's recommendations exhibit meaningful temporal alignment with observed payment and response patterns, following the evaluation framework defined in Section 4.4. All results are interpreted as associational. The evaluation addresses five objectives: alignment with historical payment intensity, alignment with empirical response patterns, outperformance of baseline timing strategies, constraint compliance, and temporal stability.

Alignment with historical payment intensity (macro layer). Across the eight modality–metric combinations with confidence above 0.5, the engine achieved a mean TAS of 0.485 and a mean precision of 1.164 on the validation period (2024–March 2026). The strongest result was All/-Val.Médio (Oct–Feb), with TAS = 0.636 and precision = 1.527, meaning 63.6% of annual payment volume in that series concentrates within a five-month window at a density 52.7% above uniform. The weakest qualifying result was Physical/N.Compras (Jun–Oct), with TAS = 0.387 and precision = 0.928, the only combination below the precision = 1.0 threshold, indicating that window is slightly under-concentrated relative to uniform. Online/Val.Médio fell below the confidence threshold entirely and produced no recommendation, consistent with the engine's conservative fallback design. These results confirm that the engine successfully identifies calendar windows with above-average payment concentration across the majority of analysed combinations.

Alignment with empirical response patterns (micro layer). The IC lift analysis provides an independent empirical signal. The five highest-lift months (July $+44.6\%$, June $+23.2\%$, January $+15.4\%$, May $+13.8\%$, February $+12.5\%$) all fall within at least one SIBS-recommended window, supporting the core assumption that population-level payment intensity is reflected in debtor responsiveness. The convergence is partial rather than complete: October (-35.1%), August (-19.2%), and November (-16.4%) produce strongly negative lift while sitting inside SIBS windows, confirming that the macro layer identifies seasons but cannot resolve within-season timing. The SIBS modifier addresses this by intensifying the sequence whenever the initialisation date sits inside a SIBS window, on the operational expectation that operators initiate during the high-lift sub-period (June–July or January–February) rather than during the low- or negative-lift remainder.

Outperformance of baseline strategies. The engine's mean TAS of 0.485 compares favourably against all three baselines: the uniform baseline (TAS = 0.42), the random baseline (mean TAS ≈ 0.42 by construction), and the expert heuristic (January–February, mean TAS ≈ 0.21). The expert heuristic represents a lower bound rather than a competitive alternative, as it targets a period of reduced consumer liquidity. The engine outperforms the uniform baseline in seven of the eight qualifying combinations; the only exception is Physical/N.Compras (precision = 0.928), the lowest-precision entry in the set. These results indicate that the engine captures genuine seasonal structure beyond what uniform or heuristic timing would achieve.

Constraint compliance. The CCR is 1.00 across all generated recommendations with respect to the encoded `ConstraintPolicy`, as the engine filters non-compliant cadence sequences prior to output. This result reflects the compliance-by-design architecture described in Section 5.4.5 rather than an empirical finding: no recommendation is surfaced to the operator unless it satisfies the active policy's frequency caps, mode-appropriate minimum spacing (3-day base floor or 1-day SIBS-modifier floor), and channel consent restrictions. The practical significance is that the operator receives only outputs compatible with the configured policy, without a separate review step. The guarantee is bounded by the policy itself: compliance with a particular regulatory regime is achieved by configuring the policy to encode that regime's constraints, not implied by the architecture alone.

Stability across temporal, sectoral, and geographic variation. The validation period (2024–March 2026) provides a held-out temporal evaluation of windows inferred from the 2022–2023 training slice. The eight qualifying windows show consistent confidence scores, with four entries at confidence ≥ 0.90, indicating that the seasonal patterns identified in the training period persist into the validation period. Sectoral and geographic disaggregation is available in the SIBS dataset but was not the primary focus of this evaluation; the reported results reflect the financial sector aggregate. Extension to sector- and region-specific windows is supported by the engine architecture and is identified as a direction for future work.

Summary. Table 5.3 consolidates the primary quantitative validation outcomes against the evaluation objectives defined in Section 4.4.

5.5.2 Perceived Utility Evaluation: Questionnaire Design

The retrospective metrics in Section 5.5.1 establish that the engine's recommendations are temporally aligned with observed payment and response patterns. They cannot, however, speak to how the resulting contact strategies are *perceived* by the people they affect. Under partial observability, and in the absence of a live controlled experiment (discussed as future work in Section 5.5.5), perceived utility provides a complementary, face-validity dimension of evaluation: whether the personalised strategies are judged as fairer, less intrusive, more appropriately timed, and more operationally useful than non-personalised alternatives, from both sides of the collection relationship.

Instruments and target populations. Two structured questionnaires were designed, one for each stakeholder perspective, each administered in parallel European Portuguese and English versions so that respondents could answer in their preferred language. The *debtor* instrument elicits the perspective of a hypothetical recipient of collection communication and was distributed to a broad convenience sample drawn from the academic community. The *manager* instrument elicits the perspective of collection practitioners (professionals with direct or supervisory responsibility over credit-recovery processes) and targets the operational network of the platform partner. The two populations differ sharply in size and accessibility: the debtor instrument can reach a large sample, whereas the manager instrument addresses a small, specialised population. This asymmetry is reflected in the analysis plan below, where the debtor instrument supports inferential testing and the manager instrument is treated as descriptive corroboration.

Strategy stimuli. Each respondent evaluates three contact strategies, presented as a gradient of the engine described in Section 5.4, in which each step adds one component:

- **Uniform:** contacts evenly spaced across a fixed horizon, with the same calendar maintained throughout the year, a fixed escalation point, and no profile- or season-dependent adaptation. This corresponds to the uniform baseline.
- **Segmented:** an escalation- and profile-adaptive cadence whose rhythm and channel mix intensify as the situation persists (introducing VMS at later stages) and according to portfolio urgency and value, but with no calendar-level seasonal targeting and the same schedule across the year. This corresponds to the decision layer (cadence matrix) without the SIBS modifier.
- **Engine:** the segmented cadence further concentrated in the periods of the year associated with higher historically observed payment intensity. This corresponds to the full engine (cadence matrix plus SIBS modifier).

The three strategies form a monotone design gradient (Uniform $\rightarrow$ Segmented $\rightarrow$ Engine). The experimental design isolates the contribution of each component by changing one element at a time: in both instruments, the only feature distinguishing Segmented from Uniform is profile- and escalation-adaptivity, and the only feature distinguishing Engine from Segmented is the seasonal concentration introduced by the SIBS modifier. The adjacent Engine-versus-Segmented contrast therefore isolates the perceived value of the seasonal layer specifically. The strategies were described to respondents in neutral, behaviourally framed language, deliberately balanced in valence so that no strategy was presented as normatively superior; descriptions referred only to how and when contacts are distributed, avoiding any framing of one approach as "manual" or "smart".

Instrument structure. The debtor instrument opens with a generic scenario (an overdue invoice more than 60 days past due, contactable by email, SMS, or voice message). The manager instrument fixes a concrete portfolio type, Critical and Low Value (debts over 60 days overdue with an average value below the company threshold), corresponding to the worked example in Section 5.4.4, so that practitioner responses are conditioned on a single well-defined portfolio. In both instruments, respondents evaluate each strategy independently on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree), and a final forced-choice item asks which strategy they would prefer to receive (debtor) or implement (manager), with an explicit "indifferent" option. The manager instrument additionally records, as anonymous context, the average number of contacts per debt process under the respondent's current organisational practice, providing a reference point against which the proposed cadences can be situated.

The two instruments measure different constructs, appropriate to each perspective. The debtor constructs (Table 5.4) concern intrusiveness and rhythm, temporal adequacy, equity and relationship, response intention, and overall appropriateness. The manager constructs (Table 5.5) concern operational efficacy, cadence and escalation, compliance and transparency, decision utility, and overall calibration. The full instruments are reproduced in Appendix B.

Bias controls. Because the same respondent evaluates all three strategies, the design is within-subject, which removes between-respondent variability and improves statistical efficiency. To guard against label- and order-related bias, the letter assigned to each strategy is randomised independently per respondent, so that any strategy is equally likely to appear under any label; the displayed letter therefore carries no fixed identity (and indeed differs between the two instruments). Strategy identity is reconstructed at analysis time from a stored per-submission mapping, ensuring that respondents cannot infer which strategy is the system's own recommendation. The single reverse-coded debtor item is reversed prior to aggregation.

The Portuguese and English versions are direct translations of the same instrument, with identical item ordering, scale, and per-respondent label randomisation; responses from both are pooled, with strategy identity recovered uniformly from the stored mapping regardless of language.

Analysis plan and statistical power. The primary analysis treats the three strategies as a within-subject factor. For each construct, the omnibus hypothesis that perceived utility differs across the three strategies is tested with the Friedman test, with the Uniform $\rightarrow$ Segmented $\rightarrow$ Engine ordering as the hypothesised direction. Pairwise contrasts are examined with the Wilcoxon signed-rank test, applying a Bonferroni correction across the three pairwise comparisons. A Monte Carlo power analysis (within-subject design, assumed intra-respondent correlation of 0.5, Likert responses generated from a latent gradient; full procedure in Appendix C) indicated that approximately 45 respondents yield 0.80 power to detect a small omnibus effect (Kendall's $W \approx 0.09$, corresponding to a medium adjacent matched-pairs effect of $r \approx 0.30$) in the Friedman test at $\alpha = 0.05$, and that the Uniform-versus-Engine contrast is comparably well powered. Resolving the adjacent Engine-versus-Segmented contrast, the smallest step in the gradient, is substantially more demanding (on the order of 150 respondents); the study is therefore powered for the omnibus comparison and the Uniform-versus-Engine contrast, with the adjacent contrast reported descriptively where the sample is insufficient. The manager instrument, addressing a small specialised population, is not powered for inferential testing and is reported using medians and response distributions only.

Ethics. Participation in both instruments was voluntary and anonymous, with no collection of personal or identifying data, consistent with the ethical considerations set out in Section 4.5. Respondents were informed of the academic purpose of the study and that there were no correct or incorrect answers.

5.5.3 Questionnaire Results

The debtor instrument was completed by 46 respondents and the manager instrument by 5. Internal consistency of the ten-item debtor scale was good across all three strategies (Cronbach's $\alpha =$

.80, .83, and .86 for Uniform, Segmented, and Engine respectively), supporting aggregation of the items into the constructs of Table 5.4. Because the strategies differ by design on a single dimension, results are reported at the construct level (Table 5.6, Figure 5.8); the ten-item composite is discussed separately below.

Temporal adequacy is the pre-specified primary construct: it measures whether contacts fall at times of year when the respondent feels able to deal with financial matters, which is precisely the dimension the SIBS modifier acts upon. On this construct the strategies differ strongly (Friedman $\chi^2(2) = 14.45$, $p < .001$, Kendall's $W = .16$), and the difference survives Holm correction across the five constructs ($p_{\text{Holm}} = .004$). The Engine is rated well above both baselines and clears the neutral midpoint, while the two non-seasonal strategies both fall below it (means 2.92, 2.88, and 3.70 for Uniform, Segmented, and Engine respectively) and are statistically indistinguishable from one another. Post-hoc Wilcoxon signed-rank tests with Bonferroni correction confirm both Engine contrasts: Engine versus Uniform (mean difference +0.78, $p_{\text{Bonf}} = .008$, rank-biserial $r = .56$) and Engine versus Segmented (mean difference +0.82, $p_{\text{Bonf}} = .005$, $r = .55$), with Uniform versus Segmented non-significant ($p_{\text{Bonf}} = 1.00$; raw $p = .62$). Notably, the adjacent Engine-versus-Segmented contrast, which the power analysis (Appendix C) indicated could not be resolved at this sample size on a global measure, is resolved here on the construct where the seasonal layer's effect is concentrated.

None of the remaining four constructs shows a significant difference across strategies (all $p_{\text{Holm}} = .498$):[1] intrusiveness and rhythm ($\chi^2(2) = 3.64$, $p = .16$), equity and relationship (3.26, $p = .20$), response intention (4.17, $p = .12$), and global evaluation (2.72, $p = .26$). The seasonal concentration therefore produces no detectable increase in perceived intrusiveness, nor any deterioration in perceived equity or relationship.

The ten-item composite is, by construction, uninformative here. Averaging all ten items blends the two temporal-adequacy items with eight items on which the strategies do not differ, so each respondent's three composite scores converge and the within-subject ranking becomes close to random ($\chi^2(2) = 0.32$, $p = .85$, $W = .003$). The composite is reported only to make this dilution explicit; the construct-level analysis is the substantive result.

The forced-choice preference item (Figure 5.9) asked which strategy respondents would prefer to receive. Segmented was the modal choice (23 of 46), with Uniform and Engine roughly tied behind it (11 and 10) and two respondents indifferent. The Engine's perceived advantage is thus specific to temporal fit and does not extend to an overall preference for receiving it, a point taken up in Section 5.5.4.

The manager instrument, addressing a small specialised population ($n = 5$), is reported descriptively (Figure 5.10). The direction favours the personalised strategies: median composite ratings were 3.83 (Engine), 3.75 (Segmented), and 3.50 (Uniform); the Engine was the modal implementation choice (2 of 5); and its clearest advantage was on operational efficacy (median 4.0

[1] The four secondary constructs share the same Holm-adjusted value because Holm's step-down procedure enforces monotone non-decreasing adjusted p-values. The smallest raw p among the secondaries is Response intention's $p = .1246$; multiplied by its Holm rank ($k - r + 1 = 4$) it yields .498, which then becomes the running maximum that all subsequent (larger) adjusted values are floored to.

versus 3.0 for Uniform). Reported current practice ranged from one to ten contacts per process, with three respondents not specifying. These figures are indicative only and are not subjected to inferential testing.

5.5.4 Discussion

The perceived-utility study yields a single, narrow finding: of the five debtor constructs, the strategies separate on exactly one, temporal adequacy, and the separating strategy is the Engine. In other words, debtors do not see the engine's internals, yet they report that its contacts fall at times of year when they feel more able to act on financial matters, which is exactly the dimension the SIBS modifier targets. The effect is statistically robust (surviving both Holm correction and Bonferroni-corrected post-hoc tests) and substantial in magnitude (roughly four fifths of a scale point above either baseline).

Equally important is what does not move. The seasonal concentration introduced by the Engine does not register as more intrusive, less fair, or more damaging to the creditor relationship: the intrusiveness, equity, and response-intention constructs are statistically indistinguishable across the three strategies. This is the perceptual evidence for the second clause of the hypothesis, that timing personalisation can be achieved without increasing perceived intrusiveness or eroding the debtor's relationship with the creditor. The Engine improves perceived temporal fit without producing a statistically detectable deterioration in the measured intrusiveness, equity, response-intention, or relationship constructs.

Two caveats keep this in perspective. First, the forced-choice preference does not track the temporal-adequacy result: debtors modally prefer to receive the Segmented strategy, and the Engine does not lead on global evaluation. This divergence is less of a contradiction than it first looks. Personalisation has been studied most extensively in marketing and e-commerce contexts [9], where its aim is to raise the relevance and appeal of solicited contact, so a preference to receive a given strategy reflects how well it fits the recipient. Debt collection inverts that premise: the contact is unsolicited and the recipient would, all else equal, prefer not to be contacted at all, so additional contact is liable to provoke reactance [5]. A preference for the strategy that contacts least is therefore in part a preference against contact volume rather than a judgement on timing, which is precisely why a strategy can be judged better aligned with one's financial capacity over the year yet not be the one a respondent would choose to receive. Read this way, the modal preference for Segmented and the temporal-adequacy advantage of the Engine are not in tension; they answer different questions. Second, and more fundamentally, the study measures stated perception, not behaviour. It corroborates the alignment and acceptability claims of this dissertation; it does not, and cannot, establish that the recommendations cause faster resolution. This distinction runs through the whole thesis and motivates the controlled experiment proposed in Section 5.5.5.

The manager evidence points in the same direction but is descriptive only. With five respondents the instrument cannot support inference; the favourable median ratings and the modal implementation preference for the Engine are recorded as qualitative corroboration from practitioners rather than as a test.

Several limitations bound these conclusions. The debtor sample is a convenience sample responding to a hypothetical scenario rather than an active arrears situation, so the ratings capture anticipated rather than experienced reaction. This bounds the external validity of the absolute rating levels, which may not transfer to people actually in arrears; it bears less on the within-subject contrast, however, since each respondent rated all three strategies and so served as their own control, which makes the relative ordering across strategies considerably more robust to individual miscalibration than the absolute levels are. The per-respondent randomisation of strategy labels further removes strategy-position confounds, but two design features remain uncontrolled: item order is fixed across respondents, admitting possible fatigue or practice effects, and the three strategies are evaluated side by side within each item, which encourages comparative rather than absolute judgement. Finally, the adjacent Engine-versus-Segmented contrast is resolved here only on the temporal-adequacy construct, where the effect is large; on a global measure it remains underpowered at this sample size, as anticipated by the power analysis, and its resolution on behavioural data is left to the experiment of Section 5.5.5.

5.5.5 Future Work: A/B Test Proposal

The validation presented in this chapter is retrospective and perceptual: it establishes that the engine's recommendations align with historical payment and response patterns, and that the personalised strategies are perceived favourably, but it does not establish a causal effect of the recommendations on payment outcomes. The natural next step, beyond the scope of this dissertation, is a prospective controlled experiment.

A live A/B test would randomly assign comparable portfolios, or documents within a portfolio, to either the engine's recommended cadence or the incumbent collection strategy, holding portfolio composition, value, and arrears stage balanced across arms. The primary outcome would be the realised payment rate within a fixed observation window, with secondary outcomes including time-to-payment and total contact volume. Such a design would, for the first time in this work, support a causal comparison rather than an associational one, and would allow the Engine-versus-Segmented question, only partially answered by the perceived-utility study, to be resolved on behavioural data.

A deployment of this kind carries ethical and regulatory obligations that must be addressed before any live trial. Randomising debtors to contact strategies raises questions of equitable treatment; the experiment would need to ensure that no arm applies contact in excess of regulatory frequency caps or prohibited windows, that both arms remain compliant by construction, and that the trial is subject to oversight, informed organisational consent, and a pre-registered analysis plan. Within those bounds, a controlled trial would convert the temporal-alignment evidence assembled here into a direct estimate of operational effect.

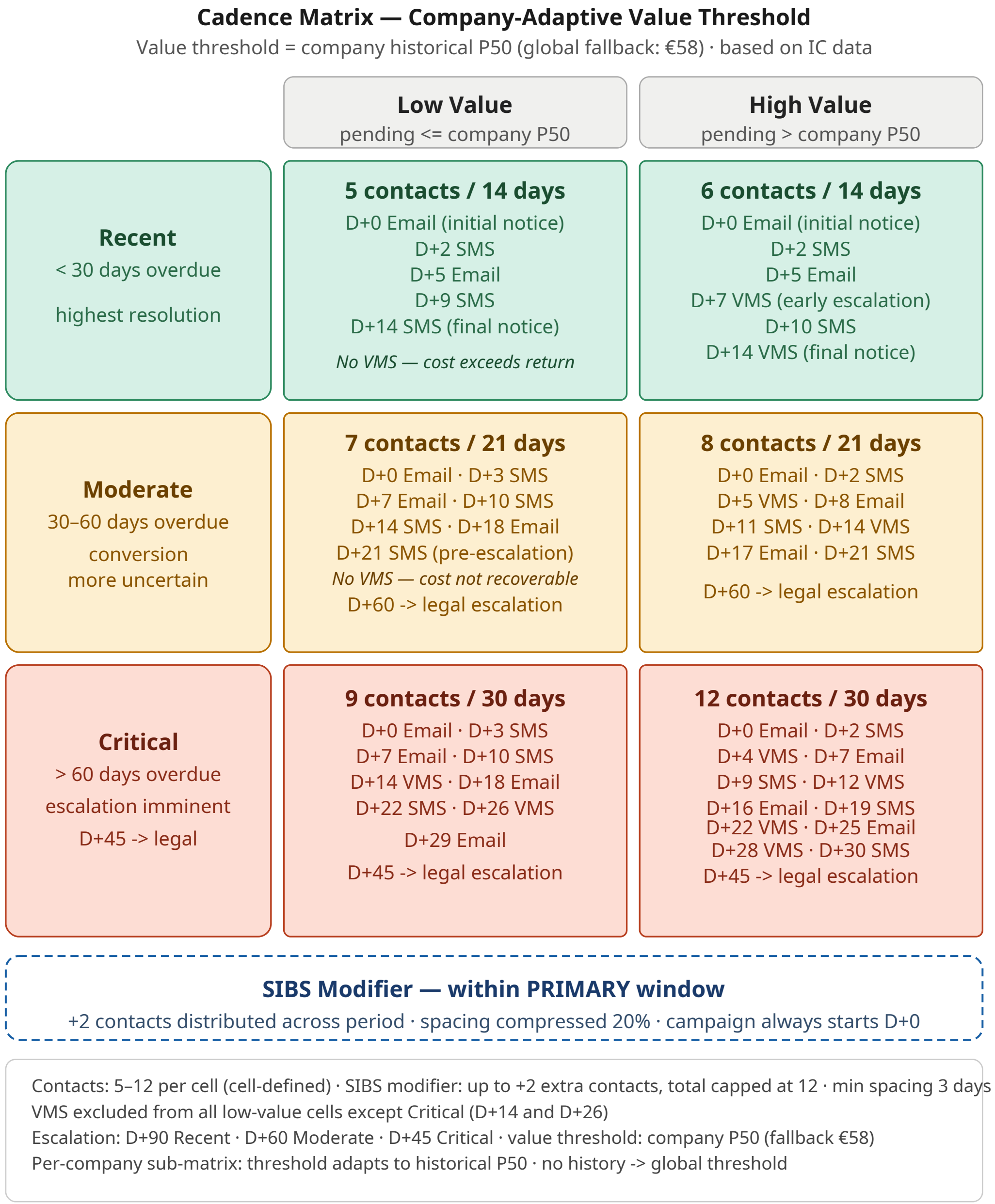

Figure 5.5: Cadence matrix mapping portfolio urgency (rows: Recent, Moderate, Critical) and value tier (columns: Low Value, High Value) to a contact sequence. Each cell shows the number of contacts, the day-offset sequence, the channel mix, and the escalation threshold. VMS is excluded from Low Value cells except Critical, where it is introduced from D+14 onward. The lower panel summarises the demonstration `ConstraintPolicy` and the SIBS modifier. The "min spacing 3 days" value applies to the base sequences produced without the SIBS modifier; when the modifier is engaged, the 20% compression and inserted seasonally-motivated contacts can reduce the effective floor to 1 day (visible in the worked example of Table 5.2), and the applicable policy floor is 1 day. Stricter organisational or regulatory policies can be encoded through `ConstraintPolicy`, in which case the same filtering layer would return sparser or stretched schedules without changing the cell selection logic.

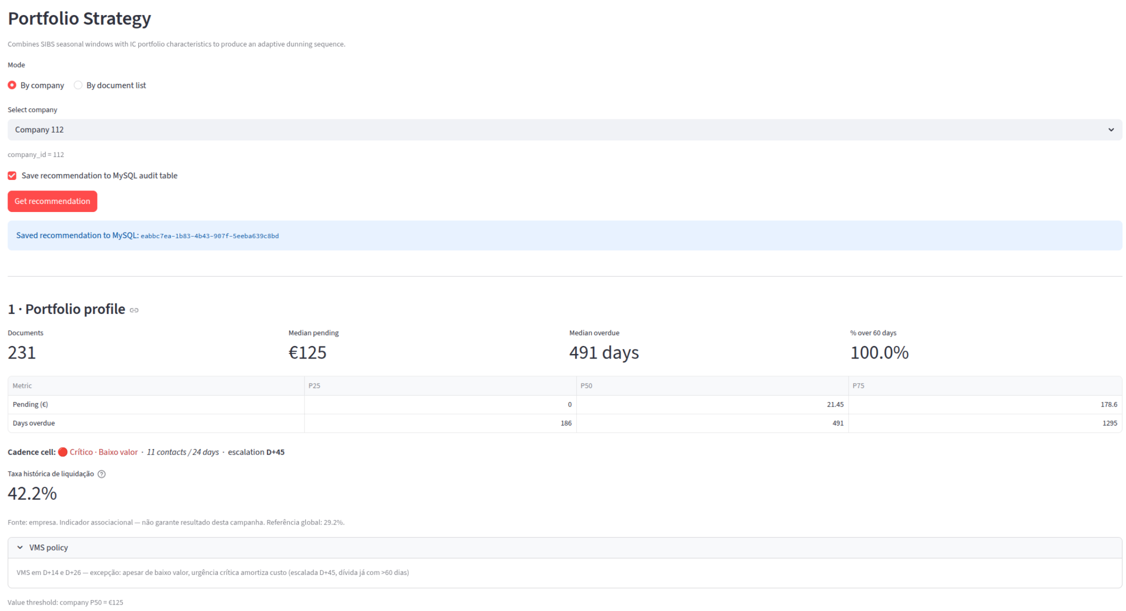

Figure 5.6: Portfolio profile panel for Company 112 (anonymised), showing 231 documents, median 491 days overdue, 100% over 60 days. The engine classifies the portfolio as Critical+Low Value. The "median pending €125" headline tile in this panel is a labelling artefact: it reports the company's historical P50 (used as the value-tier threshold), not the portfolio's own median pending balance, which is €21.45 as shown in the distribution table directly below the tile (P25 = €0, P50 = €21.45, P75 = €178.60). The Critical+Low classification is unaffected: €21.45 sits below the €125 threshold. The SIBS modifier is active (Mar–Jul window), expanding the base sequence from 9 to 11 contacts and compressing the horizon from 30 to 24 days (escalation D+45). The historical resolution rate of 42.2% is shown as an associational indicator alongside the global reference of 29.2%. The VMS policy exception for Critical+Low (introduced at D+14 and D+26) is displayed in the expandable panel below. Interface shown in Portuguese, the operational language of the platform partner.

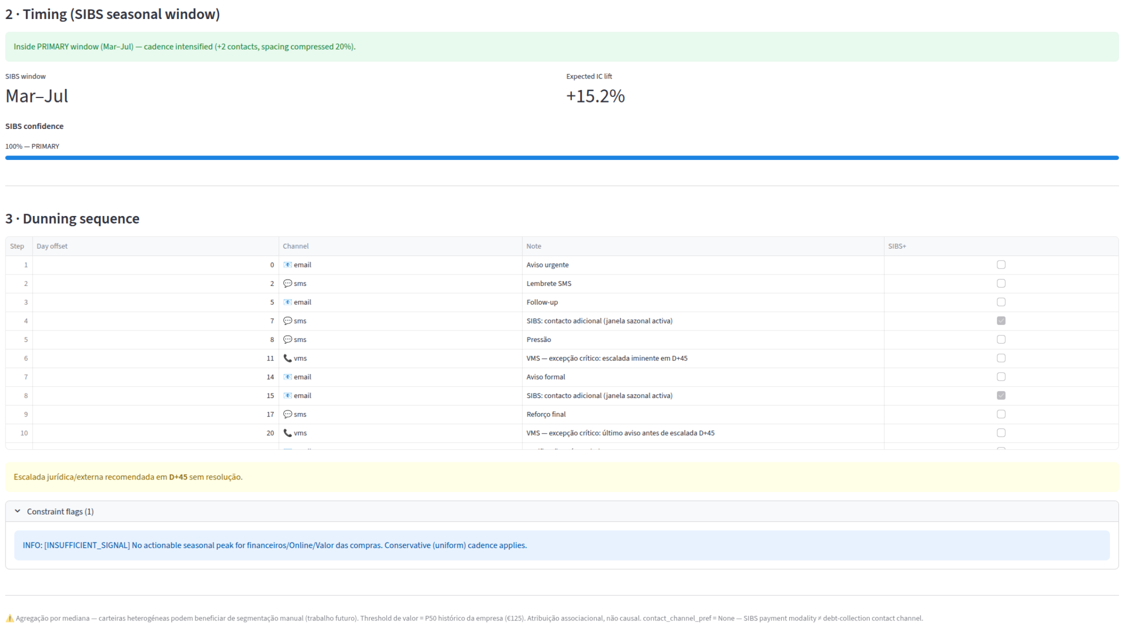

Figure 5.7: Timing and dunning sequence panels for the same portfolio. The SIBS modifier is active (Mar–Jul window, confidence 100%), producing an expected IC lift of +15.2%. The base 9-step sequence is expanded to 11 steps and compressed from 30 to 24 days; steps 4 and 8 are flagged as SIBS-motivated additional contacts. The constraint flag at the bottom reports an `INSUFFICIENT_SIGNAL` for the Online/Val.Compras combination, indicating that a conservative uniform cadence applies for that modality. The disclaimer at the foot of the page notes the associational nature of the output and that the SIBS payment modality differs from the debt-collection contact channel. Interface shown in Portuguese, the operational language of the platform partner.

Table 5.2: Engine output for an anonymised portfolio (Company C62, 64,714 documents) under two temporal conditions. Case A is initialised inside the Mar–Jul SIBS window (16 May 2026, confidence 1.00); Case B is initialised outside the window (1 November 2025, next window 120 days away). Both cases produce a Critical+Low classification with escalation at D+45. The sequences shown illustrate the demonstration `ConstraintPolicy` configuration used in this work; stricter organisational policies would produce sparser schedules through the same filtering mechanism without changing the urgency/value classification or the seasonal modifier logic.

Case A, May 2026 (inside SIBS window), 11 contacts / 24 days			
Step	**Day**	**Ch.**	**Action**
1	D+0	Email	Urgent notice
2	D+2	SMS	Reminder
3	D+5	Email	Follow-up
4	D+7	SMS	*SIBS: extra contact*
5	D+8	SMS	Pressure
6	D+11	VMS	Critical exception
7	D+14	Email	Formal notice
8	D+15	Email	*SIBS: extra contact*
9	D+17	SMS	Final reinforcement
10	D+20	VMS	Final warning
11	D+23	Email	Pre-escalation
Escalation D+45	Expected Response Rate (RR): 1.47%		
Case B, November 2025 (outside window), 9 contacts / 30 days			
Step	**Day**	**Ch.**	**Action**
1	D+0	Email	Urgent notice
2	D+3	SMS	Reminder
3	D+7	Email	Follow-up
4	D+10	SMS	Pressure
5	D+14	VMS	Critical exception
6	D+18	Email	Formal notice
7	D+22	SMS	Final reinforcement
8	D+26	VMS	Final warning
9	D+29	Email	Pre-escalation
Escalation D+45	Expected RR: see note*		

*The engine reports 1.47% as the best available estimate for when the next Mar–Jul window opens (120 days away); for the campaign window itself (1 November 2025 onward, outside any active SIBS window), the expected RR depends on the month's IC lift, with an annual unconditional mean of 1.28% and substantial month-to-month deviation in both directions.

Table 5.3: Quantitative validation summary. Each evaluation objective from Section 4.4 is assessed against the results reported in Sections 5.2–5.4. All results are associational.

Evaluation objective	**Metric**	**Result**
Alignment with payment intensity	Mean TAS; Precision	TAS = 0.485; Precision = 1.164 (8 combinations, confidence $>$ 0.5)
Alignment with response patterns	IC lift convergence	5/5 highest-lift months inside SIBS windows; high-lift sub-period pattern (Jun–Jul, Jan–Feb) confirmed
Outperformance of baselines	TAS vs. uniform / random / expert	Engine $>$ uniform in 7/8 combinations; expert baseline TAS $\approx$ 0.21
Constraint compliance	CCR	1.00 with respect to encoded `ConstraintPolicy`
Temporal stability	Confidence scores on held-out period	4/8 entries at confidence $\geq$ 0.90

Table 5.4: Constructs measured by the *debtor* questionnaire, with abbreviated example items. All items use a five-point Likert scale and are answered independently for each strategy, except the final forced-choice preference item. Item 1 is reverse-coded.

Construct	**Example item (abbreviated)**
Intrusiveness & rhythm	"This contact strategy seems excessive" (reverse-coded); "the rhythm seems appropriate to the seriousness of the situation"; "it gives me enough time to act before the next contact".
Temporal adequacy	"Contacts occur at times of year when I feel more able to deal with financial matters"; "it seems to take my financial availability over the year into account".
Equity & relationship	"This strategy seems fair and proportional"; "it preserves a balanced relationship with the creditor"; "my rights as a debtor are respected".
Response intention	"I would feel more motivated to deal with the situation than if I received no contact at all".
Global evaluation	"Overall, I consider this strategy appropriate for the situation"; forced-choice strategy preference.

Table 5.5: Constructs measured by the *manager* questionnaire, with abbreviated example items. Items are evaluated for each strategy against a fixed Critical+Low-Value portfolio. A final forced-choice item records which strategy the respondent would implement.

Construct	**Example item (abbreviated)**
Operational efficacy	"Appropriate for this portfolio's urgency and value profile"; "contact intensity is proportional to the non-payment risk"; "the calendar accounts for seasonal variation in debtors' payment capacity".
Cadence & escalation	"Compatible with my team's operational capacity"; "channel progression and escalation timing suit this portfolio type"; "the minimum interval avoids reflexive debtor reactance"; "it meets my organisation's minimum contact KPIs".
Compliance & transparency	"It appears to respect the applicable regulatory limits on contact frequency and channels"; "I could explain its logic to an auditor or supervisor without difficulty".
Decision utility	"It provides useful information to support my operational decisions"; "I would use it as a starting point for defining the portfolio approach".
Global evaluation	"Overall, a well-calibrated collection strategy for this portfolio type"; forced-choice implementation preference.

Table 5.6: Debtor perceived-utility ratings by construct and strategy ($n = 46$). Values are means of per-respondent construct scores on a five-point Likert scale (higher is more favourable; the single reverse-coded item is reversed before aggregation). Omnibus tests are Friedman; p_{Holm} applies Holm-Bonferroni across the five constructs. The composite row is reported for context and is not part of the construct family.

Construct	Unif.	Seg.	Eng.	$\chi^2(2)$	p	p_{Holm}	W
Intrusiveness & rhythm	3.58	3.47	3.37	3.64	.16	.498	.040
Temporal adequacy†	2.92	2.88	**3.70**	14.45	$< .001$	.004	.157
Equity & relationship	3.67	3.65	3.39	3.26	.20	.498	.035
Response intention	3.65	4.07	3.65	4.17	.12	.498	.045
Global evaluation	3.57	3.96	3.65	2.72	.26	.498	.030
Composite (10 items)	3.48	3.52	3.50	0.32	.85	n/a	.003

† Pre-specified primary construct (the dimension the SIBS modifier targets).

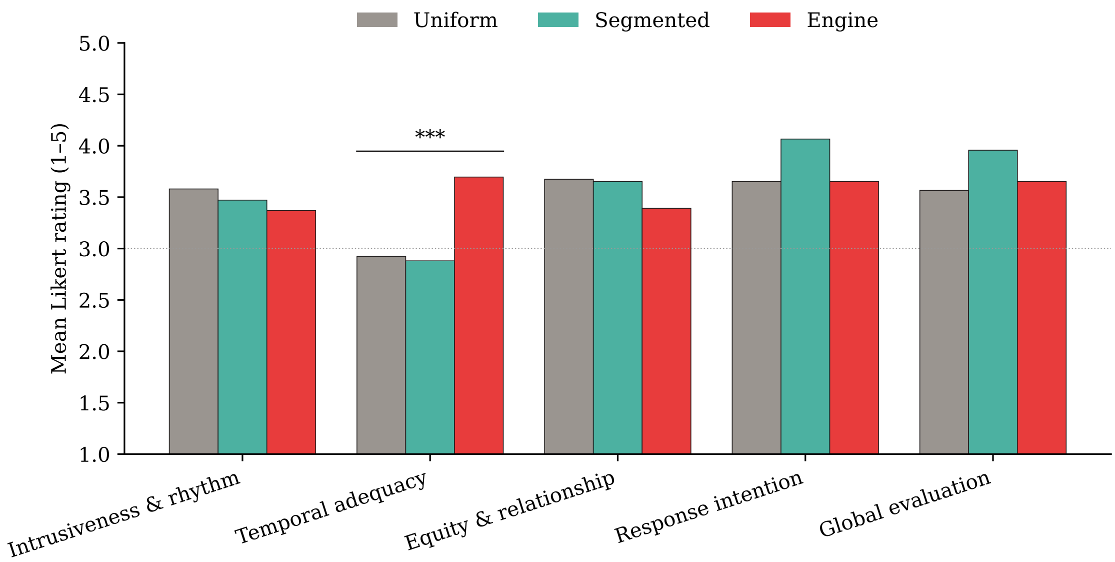

Figure 5.8: Debtor perceived-utility ratings by construct and strategy ($n = 46$). Bars are mean Likert ratings. The bracketed construct is the only one significant after Holm correction across the five constructs; the stars report its Friedman omnibus p ($^{***}p < .001$; Holm-adjusted $p = .004$). The three strategies separate only on temporal adequacy, the construct targeted by the seasonal (SIBS) layer.

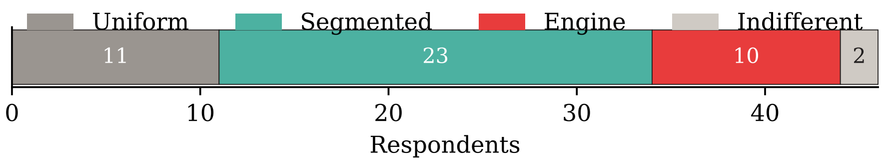

Figure 5.9: Debtor forced-choice preference: the strategy each respondent would prefer to receive ($n = 46$). Segmented is modal; Uniform and Engine are roughly tied behind it.

Figure 5.10: Manager perceived-utility ratings by construct and strategy ($n = 5$, descriptive). Median Likert ratings; no inferential test is applied given the sample size.

5.6 Chapter Summary

This chapter assembled the empirical basis for the engine and evaluated it from complementary angles. At the macro level, the SIBS data confirmed two stable seasonal regimes in payment activity, a mid-year peak and a year-end peak, differentiated by payment modality and value. At the micro level, the operational records from the platform partner exhibited monthly response patterns whose high-lift sub-periods (June–July and January–February) sit inside the macro regimes, and payments followed contact rapidly (a median of 2.6 days), reinforcing the temporal logic of the approach. The engine integrates these two signals: it identifies the seasonal regime from population-level structure and the productive sub-period from operational response, and translates them into cadence and timing recommendations produced within the encoded `ConstraintPolicy`, so that every recommendation is compliant against that policy and auditable by construction.

The engine was then evaluated on two fronts. Quantitatively, its recommended windows were benchmarked against uniform, random, and expert-heuristic baselines on a held-out period, outperforming the uniform baseline in seven of eight combinations and corroborated by the independent lift analysis on its high-lift entry months. Perceptually, a within-subject questionnaire study showed that debtors perceive the full engine as significantly better aligned with their financial capacity across the year than either non-seasonal baseline, with no accompanying increase in perceived intrusiveness or loss of fairness, while a small descriptive manager sample pointed in the same direction. Together these results establish temporal alignment and perceived acceptability; the causal question of whether the recommendations accelerate payment is left, by the partial-observability conditions of the setting, to the prospective experiment outlined in Section 5.5.5.

Chapter 6

Conclusions

6.1 Summary of the Work

This dissertation set out to address a persistent gap in digitally mediated debt collection: the absence of an integrated, timing-aware decision-support system that remains interpretable, auditable, and compliant under conditions of partial observability. Collection practice continues to rely on static schedules and manual heuristics, applying the same contact calendar to every debtor regardless of when recipients are most able to act, and although predictive models for default risk and repayment propensity are well studied, their integration into the operational decisions that govern contact timing, cadence, and sequencing remains fragmented.

In response, the work designed, implemented, and evaluated a decision-support engine that personalises the timing and cadence of collection communications through *temporal alignment* rather than causal optimisation. The engine follows a four-layer processing pipeline: ingestion and preprocessing of heterogeneous payment data; extraction of temporal structure through seasonal-trend decomposition and peak detection; constraint-filtered decision logic that proposes contact windows and cadences within regulatory and behavioural limits; and an output layer providing windows, confidence levels, traceability, and an audit trail. The two temporal-feature stages constitute the dual analytical layer (macro from SIBS, micro from IC) that feeds the shared decision layer; the four-pipeline, two-analytical-layer, and three-component framings (macro, micro, decision) used elsewhere in the document refer to the same system at different levels of granularity. It was evaluated on two independent sources: 87 months of aggregated payment data from SIBS MULTIBANCO, capturing population-level seasonality, and operational message and payment records from the IC platform, providing empirical response patterns.

The central empirical finding is that the two data layers are complementary rather than redundant: the macro layer identifies the seasonal regime, and the micro layer identifies the productive sub-period within that regime. This division of labour, established in Chapter 5, is what the engine operationalises through the urgency–value cadence matrix and the SIBS modifier.

6.2 Revisiting the Research Questions

The work was structured around six research questions formulated in Chapter 1. Each is revisited below in light of the results.

RQ1: Informative features for high-responsiveness windows. The macro-level analysis (Section 5.2) found that the most informative features derived from aggregated payment dynamics were the calendar month in combination with the payment modality (physical versus online) and the transaction metric (purchase count, total value, or average value). Seasonal-trend decomposition of these dimensions yielded stable, interpretable structure: two distinct seasonal regimes were identified, a mid-year peak (approximately May–October) most pronounced in physical and purchase-count series, and a year-end peak (approximately October–March) more strongly associated with average-value series. Modality emerged as a substantive discriminator, with online activity inverting the physical pattern and concentrating at year end, indicating that a single population-level window is insufficient to characterise all payment behaviour.

RQ2: Inferring effective contact windows under partial observability. The engine infers windows by decomposing each series into trend, seasonal, and residual components, detecting peaks and valleys in the seasonal component, and proposing calendar windows scored by a confidence measure reflecting signal strength and stability (Section 5.2.3). Across the eight modality–metric combinations with sufficient confidence, the recommended windows achieved a mean TAS of 0.485 and a mean precision of 1.164 on a held-out validation period, confirming that the inferred windows align with periods of higher observed payment intensity. Consistent with the conditions of partial observability, these windows are interpreted as aligned with observed payment intensity, not as causal drivers of repayment.

RQ3: Adapting cadence and sequencing under behavioural and regulatory limits. Cadence is produced as a constraint-compliant schedule rather than an unconstrained optimum (Section 5.4). The 3×2 cadence matrix adapts contact intensity, channel mix, and escalation timing to portfolio urgency and value, ranging from five contacts over fourteen days for the least urgent cases to twelve contacts over thirty days for the most critical, while a minimum inter-contact spacing and an explicit frequency cap mitigate the behavioural reactance documented in the literature. The SIBS modifier intensifies the sequence only when contact opens in a favourable window, with operators expected to initiate during the window's high-lift sub-period so that the intensified cadence lands on the productive months rather than on the low-lift remainder.

RQ4: Heuristic decision models under ethical, operational, and legal constraints. The decision layer is a transparent, rule-based mechanism rather than an opaque learned policy. It demonstrates that heuristic models, informed by empirical signals, can support adaptive communication strategies while keeping every recommendation traceable to the temporal signals, rules, and constraint activations that produced it. Because non-compliant sequences are filtered prior to output, the constraint compliance rate is 1.00 with respect to the encoded `ConstraintPolicy` by construction, so compliance is treated as a design requirement rather than a post-hoc check, and each recommendation carries explainability metadata supporting audit and operator oversight.

RQ5: Evaluation under partial observability. In the absence of fully attributed action–outcome data, the engine was evaluated through a combination of retrospective temporal proxies and perceived-utility assessment (Section 5.5). Quantitatively, the recommended windows were benchmarked against uniform, random, and expert-heuristic baselines on a held-out period, outperforming the uniform baseline in seven of eight combinations, and the independent IC lift analysis corroborated the macro signal on its high-lift entry months. This was complemented by a within-subject perceived-utility study across debtor and manager instruments. The evaluation establishes alignment and perceived acceptability rather than causal effect; the limitations of this approach, and the controlled experiment that would overcome them, are discussed below and in Section 5.5.5.

RQ6: How regulatory constraints shape the feasible policy space. Regulatory and organisational constraints were modelled as hard limits defining the feasible recommendation space rather than as soft penalties. The work shows that such constraints, frequency caps, minimum spacing, channel and consent restrictions, and prohibited windows, can be embedded directly in the decision logic so that the space of producible recommendations is, by construction, the space of recommendations compatible with the configured policy. The practical consequence is that compliance and deployability are not in tension with the engine's output: the operator receives only actionable recommendations compatible with the configured regulatory and organisational policy, and the seasonal signal modulates intensity within those bounds rather than overriding them.

6.3 Revisiting the Hypothesis

The hypothesis stated that personalising communication timing and cadence, based on temporal and behavioural metadata, improves temporal alignment with periods of higher observed payment responsiveness in overdue debt recovery cases, and may support faster resolution subject to prospective validation, without increasing perceived intrusiveness or deteriorating the debtor's relationship with the creditor.

The evidence assembled in this dissertation supports the hypothesis in its *alignment* dimension while appropriately qualifying its *causal* dimension. The engine demonstrably aligns contact with periods of higher observed payment intensity and response: the recommended windows concentrate payment volume well above uniform expectation, the independent micro-level lift analysis converges with the macro windows on their high-lift sub-periods (June–July, January–February), and rapid payment following contact (a median of 2.6 days) is consistent with the temporal logic of the approach. On the intrusiveness side, the engine enforces frequency caps and minimum spacing by construction, so the personalised strategies do not increase contact volume relative to a comparably configured baseline, and the perceived-utility study corroborates this on the debtor side: the full engine registered no detectable increase in perceived intrusiveness and no detectable deterioration in perceived equity or relationship relative to either non-seasonal baseline, while being perceived as significantly better aligned with debtors' financial capacity across the year (Section 5.5.3).

What the work does not, and given partial observability cannot, establish is the causal claim that the recommendations *cause* faster resolution. All retrospective results are associational, and the perceived-utility study measures stated perception rather than behaviour. The hypothesis is therefore supported as a claim about temporal alignment and perceived acceptability, and is left open as a causal claim pending the prospective experiment outlined as future work.

6.4 Contributions

The detailed scientific and practical contributions are enumerated in Section 4.6. In summary, the dissertation delivers a dual-layer analytical framework that combines macro-level seasonal structure with micro-level empirical response patterns to support timing-aware decisions without relying on causal attribution; an evaluation methodology centred on temporal-alignment metrics and retrospective benchmarking, suitable for environments with limited action–outcome observability; and a deployable, auditable, compliance-by-design decision-support engine that translates these signals into concrete, integration-ready timing and cadence recommendations. Beyond the artefact, the work contributes an empirical characterisation of seasonal payment behaviour and collection response across time, modalities, and channels, drawn from two independent real-world data sources, and sets out the key idea behind the engine's design: the difference between identifying the season and identifying the productive moment within it.

Conference Paper for the EPIA Conference on Artificial Intelligence. A further contribution of this work is the preparation of a conference paper based on the methodological core of this dissertation. The paper formulates contact-timing decisions as a problem of decision-making under partial observability, and presents the family of alignment metrics: Temporal Alignment Score, coverage and precision, developed to evaluate such decisions when outcomes cannot be attributed to individual actions. The two-layer engine is presented as the instantiation of the formulation, and the empirical evaluation against uniform, random and expert baselines as its demonstration. The manuscript is intended to be submitted to the next edition of the EPIA conference when the submission process opens, within the thematic track on intelligent decision support systems.

The paper is motivated by the observation that the methodological problem addressed in this thesis is not specific to debt collection. Any domain that combines an aggregate population-level behavioural signal, an operational record of contacts, and hard constraints on contact frequency faces the same difficulty: the quantity of interest is a contrast between schedules, which is never directly observed, and the attribution of outcomes to actions is a modelling choice rather than an observation. By separating what can be estimated from aggregated data from what cannot, and by proposing metrics that are computable under that restriction, the paper aims to contribute to the wider discussion on how temporal decisions can be evaluated honestly when causal identification is unavailable. It also reports an empirical finding of independent interest, namely that detected seasonal windows are internally heterogeneous, so that naive window-based targeting captures periods of negative return. The prepared manuscript is included in Appendix D.

6.5 Limitations

The principal limitations follow directly from the scope and the conditions under which the engine operates, and are detailed in Sections 5.3.5 and 5.5.4.

First, all results are associational. The attribution methodology links payments to preceding messages within a policy-defined window but cannot establish causation, and the reported response rates therefore represent an upper bound on any causal effect. Second, the analysis operates at monthly granularity and at the population level: the engine infers calendar-level windows, not intra-day timing, and does not model individual debtor behaviour or learn from individual action–outcome feedback. Third, the macro and micro datasets differ in span and origin (87 months of aggregated national payment data versus roughly eighteen months of operational records from a single platform), so their convergence is evidence of a shared signal rather than of identical populations. Fourth, the perceived-utility evaluation measures self-reported perception rather than behaviour and draws its debtor sample from a convenience population responding to a hypothetical scenario, which bounds the external validity of the absolute acceptability levels more than it does the within-subject contrast on which the analysis rests. At the achieved sample of 46 debtor respondents the study was powered for the omnibus comparison but not for the adjacent Engine-versus-Segmented contrast on a global measure, which the power analysis (Appendix C) indicated would require on the order of 130 to 170 respondents depending on the multiplicity correction; that contrast was nonetheless resolved on the temporal-adequacy construct, where the seasonal effect is concentrated (Section 5.5.3). The manager instrument, addressing a small specialised population, is descriptive only. These limitations bound the strength of the claims but do not undermine the central finding of temporal alignment, and several of them define the directions for future work below. A fifth direction is **replication on an in-arrears population**. The debtor perceived-utility study reported here draws on a convenience sample responding to a hypothetical scenario; replicating it among individuals actually in arrears, sampled through a collaborating creditor under appropriate ethical oversight, would test whether the temporal-adequacy advantage persists among respondents experiencing the situation rather than imagining it, and would supply the external validity that the present design cannot.

6.6 Future Work

Several extensions follow naturally from the work and the limitations above.

The most consequential is a **prospective controlled experiment**. As detailed in Section 5.5.5, a live A/B test that randomly assigns comparable portfolios to the engine's recommended cadence or the incumbent strategy, with realised payment rate as the primary outcome, would convert the associational evidence assembled here into a direct causal estimate, and would in particular resolve the perceived value of the seasonal layer on behavioural rather than perceptual data.

A second direction is **finer-grained segmentation**. The current engine classifies portfolios into a fixed urgency–value matrix; automatic sub-segmentation, for example by clustering on days overdue and pending value prior to classification, would allow the cadence to adapt to structure within a portfolio rather than treating it as homogeneous.

A third direction is **learning thresholds and signals from data**. The value and urgency thresholds are presently set by company medians and fixed cut-offs; learning these per company from historical conversion data, and extending window inference to sector- and region-specific series already available in the SIBS data, would sharpen the recommendations without departing from the interpretable design.

A fourth direction concerns **individual-level extension**. The architecture was explicitly designed to be extensible: when timestamped individual interaction logs become available, the engine could incorporate individual debtor scoring and, in time, policy learning from action–outcome feedback, moving beyond the population-level, partially observable setting that defines its current scope. A related operational extension is an automatic write-off rule that flags debts above a configurable age and failed-contact count for discard, concentrating effort where recovery remains plausible.

6.7 Closing Remarks

This dissertation began from the observation that debt collection is not only a financial process but a relational one, and that technological innovation should enhance, rather than erode, the quality of interaction between creditors and debtors. The engine developed here pursues that aim in a deliberately limited way: it does not claim to predict who will pay or to optimise an unobservable outcome, but only to align contact with the moments when payment is historically most likely, and to do so within strict limits on frequency, channel, and intrusiveness. By showing that such alignment is measurable, that population-level seasonal signals and operational response patterns can be combined into compliant and explainable recommendations, and that those recommendations can be produced in a deployable form, the work is a step toward debt collection that is both more efficient and more respectful of the people involved.

References

[1] N. Abe et al. “Optimizing debt collections using constrained reinforcement learning”. In: *Proc. 16th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*. 2010, pp. 75–84.

[2] A. P. Appel et al. “Optimize Cash Collection: Use Machine Learning to Predict Invoice Payment”. In: *arXiv preprint* (2019). arXiv: 1912.10828.

[3] Robert B. Barsky and Jeffrey A. Miron. “The Seasonal Cycle and the Business Cycle”. In: *Journal of Political Economy* 97.3 (1989), pp. 503–534. DOI: 10.1086/261614.

[4] Anthony Graham Bellotti et al. “Forecasting recovery rates on non-performing loans with machine learning”. In: *International Journal of Forecasting* 37 (1 2021), pp. 428–444. DOI: 10.1016/j.ijforecast.2020.06.009.

[5] Sharon S. Brehm and Jack W. Brehm. *Psychological Reactance: A Theory of Freedom and Control*. New York: Academic Press, 1981.

[6] Ximena Cadena and Antoinette Schoar. *Remembering to pay? Reminders vs. financial incentives for loan payments*. Tech. rep. 17020. National Bureau of Economic Research, 2011.

[7] John Y. Campbell. “Household finance”. In: *Journal of Finance* 61.4 (2006), pp. 1553–1604.

[8] Andrés Carrera and Marco E. Benalcázar. “Optimizing the Collection Process in Credit Risk Management: A Comparison of Machine Learning Techniques for Predicting Payment Probability at Different Stages of Arrears”. In: *Journal of Risk and Financial Management* 18 (11 2025). DOI: 10.3390/jrfm18110630.

[9] Shobhana Chandra et al. “Personalization in personalized marketing: Trends and ways forward”. In: *Psychology & Marketing* 39.8 (2022), pp. 1529–1562. DOI: 10.1002/mar.21670.

[10] Naveed Chehrazi and Thomas A. Weber. “Dynamic valuation of delinquent credit-card accounts”. In: *Management Science* 61 (12 2015), pp. 3077–3096. DOI: 10.1287/mnsc.2015.2203. URL: http://econspace.net/papers/Chehrazi-Weber-MS-2015.pdf.

[11] Robert B. Cleveland et al. “STL: A Seasonal-Trend Decomposition Procedure Based on Loess”. In: *Journal of Official Statistics* 6.1 (1990), pp. 3–73.

[12] Consumer Financial Protection Bureau. *Fair Debt Collection Practices Act: CFPB Annual Report 2024*. Tech. rep. Consumer Financial Protection Bureau, Sept. 2024. URL: https://files.consumerfinance.gov/f/documents/cfpb_fdcpa-2024-annual-report_2024-09.pdf.

[13] *Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC*. Official Journal of the European Union, L 337, 23 December 2015, p. 35–127. Revised Payment Services Directive (PSD2). 2015.

[14] *Directive (EU) 2021/2167 of the European Parliament and of the Council of 24 November 2021 on credit servicers and credit purchasers, and amending Directives 2008/48/EC and 2014/17/EU*. Official Journal of the European Union, L 438, 8 December 2021, p. 1–37. Credit Servicers and Credit Purchasers Directive (also known as the NPL Servicing Directive). 2021.

[15] *Directive (EU) 2023/2225 of the European Parliament and of the Council of 18 October 2023 on credit agreements for consumers and repealing Directive 2008/48/EC*. Official Journal of the European Union, L series, 30 October 2023. Consumer Credit Directive II (CCD2). 2023.

[16] Ekrem Duman, Nazli Mehrabi, and Çağrı Çakır. "Collection Model Development to Predict Which Customers Will Pay". In: 2025, pp. 237–241. ISBN: 9798331513702. DOI: 10.1109/SPML66318.2025.11199754.

[17] Minou Ghaffari, Maxime Kaniewicz, and Stephan Stricker. "Personalized Communication Strategies: Towards a New Debtor Typology Framework". In: *Psychology and Behavioral Sciences* 10 (6 2021), pp. 256–268. DOI: 10.11648/j.pbs.20211006.20. URL: https://doi.org/10.11648/j.pbs.20211006.20.

[18] David Graeber. *Debt: The First 5,000 Years*. Brooklyn, NY: Melville House, 2011.

[19] David B. Gross and Nicholas S. Souleles. "Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data". In: *Quarterly Journal of Economics* 117.1 (2002), pp. 149–185.

[20] *Guidelines on management of non-performing and forborne exposures (EBA/GL/2018/06)*. Tech. rep. EBA Guidelines on arrears management and forbearance practices. European Banking Authority, Oct. 2018.

[21] Philip Halsey et al. "Analyzing the Role of Digital Communication Channels in Debt Collection". In: *2021 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE, Apr. 2021, pp. 1–6. ISBN: 978-1-6654-1250-6. DOI: 10.1109/SIEDS52267.2021.9483751. URL: https://ieeexplore.ieee.org/document/9483751.

[22] Harvard Business Review Analytic Services. *Modernizing Debt Collection through AI and Emotional Intelligence*. Teleperformance. 2021. URL: https://www.tp.com/media/d5vead4z/harvard-business-review-analytic-services-modernizing-debt-collection-through-ai-and-ei.pdf.

[23] *IEEE Xplore*. https://ieeexplore.ieee.org/. Accessed: 2025-12-14.

[24] Rafał Jankowski and Andrzej Paliński. "Debt Collection Model for Mass Receivables Based on Decision Rules—A Path to Efficiency and Sustainability". In: *Sustainability (Switzerland)* 16 (14 2024). DOI: 10.3390/su16145885.

[25] Dean Karlan, Melanie Morten, and Jonathan Zinman. "A personal touch in text messaging can improve microloan repayment". In: *Behavioral Science & Policy* 1.2 (2015), pp. 25–31.

[26] Benjamin J. Keys and Jialan Wang. "Minimum payments and debt paydown in consumer credit cards". In: *Journal of Financial Economics* 131.3 (2019), pp. 528–548.

[27] Jongmyoung Kim and Pilsung Kang. "Late payment prediction models for fair allocation of customer contact lists to call center agents". In: *Decision Support Systems* 85 (2016), pp. 84–101. DOI: 10.1016/j.dss.2016.03.002.

[28] Yeongmin Kim et al. "Predict Sequential Credit Card Delinquency with VaDE-Seq2Seq". In: *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*. 2021, pp. 1159–1164. ISBN: 9781479938407; 0780307208; 9781424427949; 0780385667; 9781665410205; 9781457706523; 9781665442077; 9780780307209; 9781728145693; 1424401003. DOI: 10.1109/SMC52423.2021.9659039.

[29] Theresa Kuchler. "Sticking to your plan: The role of present bias for credit card paydown". In: *Journal of Financial Economics* 121.3 (2016), pp. 533–555.

[30] L.E.K. Consulting. *Europe's Debt Collection Managers Must Respond to New Tech-Led Competitors*. 2023. URL: https://www.lek.com/insights/ei/europes-debt-collection-managers-must-respond-new-tech-led-competitors.

[31] Wenhui Liu et al. "Adversarial Conservative Alternating Q-Learning for Credit Card Debt Collection". In: *IEEE Transactions on Knowledge and Data Engineering* 37 (4 2025), pp. 1542–1555. DOI: 10.1109/TKDE.2025.3528219.

[32] Roger C. Mayer, James H. Davis, and F. David Schoorman. "An Integrative Model of Organizational Trust". In: *Academy of Management Review* 20.3 (1995), pp. 709–734. DOI: 10.5465/amr.1995.9508080335.

[33] José M.R. Murteira and Mário Augusto. "Hurdle models of repayment behaviour in personal loan contracts". In: *Empirical Economics* 53 (2 2017), pp. 641–667. DOI: 10.1007/s00181-016-1140-2.

[34] M. G. Muzzarelli. *The medieval invention of emancipating credit*. Socioeco.org. [Accessed: Oct. 4, 2025]. 2006. URL: https://www.socioeco.org/bdf_fiche-document-254_de.html.

[35] Bruno S. Nascimento, Denis Maia, and Lucas Almada. *Applying Machine Learning to Improve Collection and to Reduce Write-Offs in Utilities*. 2021. DOI: 10.1109/AIBT53261.2021.00017.

[36] Sezi Çevik Onar et al. *A decision support system to optimize debt collection assignments*. Vol. 1029. 2020, pp. 178–187. ISBN: 9783319604855; 9783319276427; 9783319419343; 9783319232034; 9783319938844; 9783642330414; 9783319262833; 9788132220084; 9783642375019; 9783030026820. DOI: 10.1007/978-3-030-23756-1_23.

[37] Matthew J. Page et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews". In: *BMJ* 372 (2021), n71. DOI: 10.1136/bmj.n71. URL: https://www.bmj.com/content/372/bmj.n71.

[38] Dibu John Philip, Nandan Sudarsanam, and Ravindran Balaraman. "Improved insights on financial health through partially constrained hidden Markov model clustering on loan repayment data". In: *Data Base for Advances in Information Systems* 49 (3 2018), pp. 98–113. DOI: 10.1145/3242734.3242741.

[39] Michał Przybyłek et al. "Towards a smart debt collection system: A Design Science Research approach". In: *Journal of Big Data* 12.193 (2025). DOI: 10.1186/s40537-025-01252-0. URL: https://link.springer.com/article/10.1186/s40537-025-01252-0.

[40] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Official Journal of the European Union, L 119, 4 May 2016, p. 1–88. General Data Protection Regulation (GDPR). 2016.

[41] Denise M. Rousseau et al. "Not So Different After All: A Cross-Discipline View of Trust". In: *Academy of Management Review* 23.3 (1998), pp. 393–404. DOI: 10.5465/amr.1998.926617.

[42] Abel Sancarlos et al. "Towards a data-driven debt collection strategy based on an advanced machine learning framework". In: (Nov. 2023).

[43] Catalina Sánchez, Sebastián Maldonado, and Carla Vairetti. "Improving debt collection via contact center information: A predictive analytics framework". In: *Decision Support Systems* 159 (Aug. 2022). ISSN: 01679236. DOI: 10.1016/j.dss.2022.113812.

[44] A. Saulitis. "Personalized messaging enhances hospital debt collection while prosocial appeals fail: Evidence from a field experiment". In: *Digital Health* 10 (2024). DOI: 10.1177/20552076241277035.

[45] Andris Saulı¯ tis. "Nudging debtors with non-performing loans: Evidence from three field experiments". In: *Journal of Behavioral and Experimental Finance* 37 (Mar. 2023). ISSN: 22146369. DOI: 10.1016/j.jbef.2022.100776.

[46] *Scopus*. https://www.scopus.com/. Accessed: 2025-12-14.

[47] Keerthana Sivamayilvelan et al. "Flexible recommendation for optimizing the debt collection process based on customer risk using deep reinforcement learning". In: *Expert Systems with Applications* 256 (2024), p. 124951. DOI: 10.1016/j.eswa.2024.124951.

[48] Melvin Stephens. "3rd of tha month: Do Social Security recipients smooth consumption between checks?" In: *American Economic Review* 93.1 (2003), pp. 406–422.

[49] Richard H. Thaler. "Mental accounting and consumer choice". In: *Marketing Science* 4.3 (1985), pp. 199–214.

[50] Richard H. Thaler. "Mental accounting matters". In: *Journal of Behavioral Decision Making* 12.3 (1999), pp. 183–206.

[51] Zihao Wang et al. "Two-stage behavior cloning for spoken dialogue system in debt collection". In: *IJCAI International Joint Conference on Artificial Intelligence*. Vol. 2021-January. 2020, pp. 4633–4639. ISBN: 9780999241196; 9781577357384; 9781956792003; 9780999241165; 9780999241141; 9781577355120; 9781577354260; 9780769536156; 0934613346; 9780999241127.

[52] S. Zeng et al. "Using predictive analysis to improve invoice-to-cash collection". In: *Proc. 14th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*. 2008, pp. 1043–1050.

[53] Jonathan Zinman. "Household debt: Facts, puzzles, theories, and policies". In: *Annual Review of Economics* 7 (2015), pp. 251–276.

Appendix A

Supplementary Material

A.1 Additional Predictive Modeling Results

Table A.1: Predictive performance of payment propensity models across different feature selection techniques, adapted from [35].

Model	All Features		Feature Importance		Chi-square		PCA	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Payment Propensity 30	0.9222	0.8696	0.9187	0.8660	0.9073	0.8576	0.9116	0.8616
Payment Propensity 60	0.9122	0.8640	0.9086	0.8619	0.9036	0.8612	0.9036	0.8585
Payment Propensity 90	0.8955	0.8439	0.8908	0.8405	0.8862	0.8396	0.8886	0.8391
Spontaneous Payment	0.9001	0.8903	0.8891	0.8861	0.8870	0.8826	0.8851	0.8832
Collection Effectiveness	0.8692	0.8295	0.8645	0.8264	0.8217	0.7946	0.8311	0.7971

Appendix B

Questionnaire Instruments

This appendix reproduces the two perceived-utility instruments described in Section 5.5.2. Each instrument was administered in parallel European Portuguese and English versions with identical structure, item ordering, and five-point Likert scale (1 = strongly disagree to 5 = strongly agree); the per-respondent randomisation of strategy labels and the recovery of strategy identity at analysis time are described in Section 5.5.2. The Portuguese versions are reproduced in full below; the English versions are direct translations with no structural differences. Strategy descriptions are shown in one representative label ordering; the displayed letters were randomised independently per respondent and carry no fixed identity.

B.1 Debtor Questionnaire

Scenario presented to respondents. Respondents were asked to imagine holding an overdue debt with a company (for example, an unpaid service invoice more than 60 days overdue), contactable by email, SMS, or voice message, and to evaluate each of three contact strategies independently, based on how they would feel on receiving it.

Strategy descriptions.

Uniform Regular contacts over several weeks, evenly spaced, by email or SMS; the same calendar is maintained throughout the year; a constant, predictable rhythm until resolution.

Segmented Contacts over several weeks by email or SMS, with a voice message possible at later stages; the calendar is maintained across the year, but the rhythm intensifies if the situation persists.

Engine Contacts by email or SMS, with a voice message possible at later stages; the rhythm is reinforced in the periods of the year where, based on historical patterns, higher payment activity is observed; rhythm and channels adapt to both the situation profile and the time of year.

Items. Each item below was answered independently for each of the three strategies on the five-point Likert scale, except the final forced-choice preference item.

1. This contact strategy seems excessive to me. *(reverse-coded)*
2. The rhythm of this contact strategy seems appropriate to the seriousness of the situation.
3. This strategy gives me enough time to act before receiving a new contact.
4. The contacts in this strategy occur at times of year when I feel more able to deal with financial matters.
5. This strategy seems to take my financial availability over the year into account.
6. This contact strategy seems fair and proportional to the situation.
7. This approach preserves a balanced relationship between me and the creditor company.
8. I would feel that my rights as a debtor are respected with this strategy.
9. If I received this contact strategy, I would feel more motivated to deal with the situation than if I received no contact at all.
10. Overall, I consider this contact strategy appropriate for the situation described.
11. *Forced choice:* Which of the strategies would you prefer to receive, were you in the situation described? (Uniform / Segmented / Engine / Indifferent)

B.2 Manager Questionnaire

Target respondents and scenario. This instrument was directed at professionals with direct or supervisory responsibility over credit-recovery processes or the management of overdue portfolios. Respondents were asked to evaluate three strategies generated for a collection portfolio classified as Critical and Low Value (debts more than 60 days overdue with an average value below the company threshold), based on their operational experience.

Strategy descriptions.

Uniform A regular sequence of uniformly spaced contacts over roughly 30 days, with a constant and predictable interval; email and SMS applied consistently; a single calendar applied identically to all portfolios throughout the year; a fixed escalation point (day 90) for all portfolios.

Segmented A sequence adjusted to the portfolio profile (urgency and value) over roughly 30 days, with more critical portfolios receiving a more intense cadence; email and SMS, with VMS at later stages; immediate start; the same calendar across the year; escalation point adjusted to the portfolio profile.

Engine A sequence adjusted to the portfolio profile over 24 to 30 days, with intensity modulated by the active seasonal window; email, SMS, and VMS (introduced from D+14 in critical low-value portfolios); immediate start; intensity reinforced when the period coincides with windows of historically higher payment activity; minimum spacing of 3 days; escalation adjusted to profile (for example, D+45 for critical portfolios), with the value threshold set by the company's historical median.

Items. Each item below was answered independently for each of the three strategies on the five-point Likert scale, except the final forced-choice item.

1. This strategy is appropriate for the urgency and value profile of this portfolio.
2. The contact intensity is proportional to the non-payment risk of this portfolio.
3. The calendar of this strategy accounts for seasonal variation in debtors' payment capacity.
4. This strategy is compatible with my team's operational capacity.
5. The channel progression and escalation timing are appropriate for this type of portfolio.
6. The minimum interval between contacts is sufficient to avoid reflexive negative debtor reactance.
7. With this strategy, I meet my organisation's minimum contact KPIs.
8. Based on the description, this strategy appears to respect the applicable regulatory limits on contact frequency and channels.
9. I could explain the logic of this strategy to an auditor or supervisor without difficulty.
10. This strategy provides useful information to support my operational decisions.
11. I would use this strategy as a starting point for defining the portfolio approach.
12. Overall, this is a well-calibrated collection strategy for this type of portfolio.
13. *Forced choice:* Which of the strategies would you implement for this portfolio? (Uniform / Segmented / Engine / Indifferent)

Context item (anonymous). Respondents were additionally asked, for context, how many contacts per debt process their organisation's current collection strategy involves on average (1–5 / 6–10 / 11–15 / more than 15 / not applicable or unsure).

Appendix C

Questionnaire Power Analysis

This appendix details the Monte Carlo power analysis referenced in Section 5.5.2, which informed the target sample size for the debtor questionnaire.

C.1 Design and Data-Generating Model

The debtor instrument is a within-subject design in which each respondent rates all three strategies (Uniform, Segmented, Engine) on each item. Because no closed-form power expression exists for the Friedman test under ordinal responses, power was estimated by simulation. For a respondent i and strategy j, a latent rating was generated as

$$y_{ij}^{*} = b_i + \mu_j + \varepsilon_{ij}, \qquad b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2),$$

where b_i is a respondent-level random effect capturing the tendency of a respondent to rate all strategies consistently high or low, and ε_{ij} is residual noise. The strategy means $\mu = (0, \delta, 2\delta)$ encode the hypothesised monotone gradient Uniform $\rightarrow$ Segmented $\rightarrow$ Engine, with equal spacing δ between adjacent strategies. The latent value was then discretised into the five-point Likert scale via fixed thresholds. With $\sigma_b = \sigma_e = 1$, the intra-respondent correlation (ICC) is $\sigma_b^2/(\sigma_b^2 + \sigma_e^2) = 0.5$, a moderate value consistent with repeated subjective ratings of related stimuli.

C.2 Procedure

For each combination of sample size n and gradient magnitude δ, 2,000 datasets were simulated. Each dataset was analysed with the Friedman test (omnibus) and with Wilcoxon signed-rank tests for the pairwise contrasts, at $\alpha = 0.05$. Power was estimated as the proportion of simulations returning a significant result. Effect sizes were recorded on the observed (discretised) data so they correspond to what would be computed from real responses: Kendall's W for the omnibus test, and the matched-pairs $r = Z/\sqrt{n}$ for the adjacent Engine-versus-Segmented contrast, the smallest step in the gradient and therefore the binding constraint on sample size. The Bonferroni-corrected variant applies $\alpha/3$ to account for the three pairwise comparisons.

C.3 Results

Table C.1 reports the approximate sample sizes required to reach 0.80 power for each test, under a medium adjacent effect (matched-pairs $r \approx 0.30$, corresponding to $\delta = 0.40$, observed Kendall's $W \approx 0.09$) and a solidly medium adjacent effect ($r \approx 0.37$).

Table C.1: Approximate sample size for 0.80 power at $\alpha = 0.05$, by test and assumed effect size, from 2,000 simulations per cell. The omnibus test is inexpensive, and the Uniform-versus-Engine contrast is similarly well powered at $n \approx 50$ (uncorrected); resolving the adjacent Engine-versus-Segmented contrast is substantially more demanding.

Test / claim	**Medium** ($r \approx 0.30$)	**Solid-medium** ($r \approx 0.37$)
Friedman omnibus (a gradient exists)	$n \approx 45$	$n \approx 30$
Wilcoxon, Uniform vs. Engine (uncorrected)	$n \approx 50$	$n \approx 35$
Wilcoxon, Engine vs. Segmented (uncorrected)	$n \approx 130$	$n \approx 80$
Wilcoxon, Engine vs. Segmented (Bonferroni)	$n \approx 170$	$n \approx 100$

C.4 Interpretation

Three conclusions follow. First, the omnibus hypothesis that perceived utility differs across the three strategies is well powered at a modest sample ($n \approx 45$ for a medium effect). Second, the Uniform-versus-Engine contrast, the largest gap in the gradient, is comparably inexpensive and is well powered at $n \approx 50$. Third, the adjacent Engine-versus-Segmented contrast, which isolates the perceived value of the seasonal modifier specifically, is far more demanding (on the order of 150 respondents under Bonferroni correction), because it requires detecting a small difference between two similar strategies. The study was therefore designed to target a sample sufficient for the omnibus comparison and the Uniform-versus-Engine contrast, with the adjacent contrast reported descriptively where the achieved sample was insufficient to support an inferential conclusion. A null adjacent result at the achieved sample size is therefore uninformative rather than evidence of no difference.

The manager instrument, addressing a small specialised population, was not powered for inferential testing and is reported using medians and response distributions only.

Appendix D

EPIA Conference Paper

Evaluating Contact-Timing Decisions Under Partial Observability: Alignment Metrics and a Two-Layer Engine for Debt Collection

João António Teixeira Coelho[1], Pedro Mendes[2], and Luís Paulo Reis[1]

[1] University of Porto, Porto, Portugal
[2] Invisible Cloud

Abstract. Contact optimization in debt collection is usually framed around who should be contacted and whether a contact should happen at all, because those decisions can be evaluated against observable outcomes. By contrast, deciding when to contact and how frequently to do so remains underexplored, largely because the target of evaluation is temporal contrast rather than a directly observable state. We address this gap with a two-part contribution: alignment metrics for evaluating contact timing under partial observability, and a two-layer decision engine that combines seasonal decomposition on aggregate signals with operational filtering at the account level. Across two independent datasets, the proposed configurations converge on the mid-year peak. Mean temporal alignment reaches 0.485 against 0.420 for uniform and random baselines and 0.210 for an expert-inspired fixed calendar; seven of eight emitted combinations outperform the uniform and random baselines, and all eight exceed the expert-inspired one. These results support the method as a decision-support mechanism for temporal alignment under observational constraints, while stopping short of causal claims about intervention effects.

Keywords: Temporal alignment · Partial observability · Decision support · Debt collection · Seasonal decomposition

1 Introduction

Debt-collection decision systems are often organized around three questions: who to contact, whether a contact should occur, and when or how often a contact should be made. The first two are comparatively well studied because their targets can be tied to observable states, such as repayment, response, or delinquency transitions. The third question remains materially less developed, not because it is operationally unimportant, but because its target is harder to define and evaluate. Recent work has substantially improved repayment scoring, stage-aware prioritization, and operational rule extraction, but mostly around observable repayment states rather than timing contrasts [1–5]. Timing is fundamentally contrastive: a decision is only meaningful relative to alternative moments in which a contact could have been made.

Under partial observability, that contrast cannot be directly recovered from historical logs. We may observe outcomes and contact traces, but we do not observe the counterfactual effect of contacting the same account at a different moment. This makes contact-timing evaluation susceptible to confounding and to overly optimistic interpretations of seasonal regularities. As a result, organizations often fall back on fixed calendars, heuristics, or global seasonality assumptions that are difficult to validate in a principled way, despite the growing use of digital channels and sequencing logic in collection workflows [6–8].

This paper argues that timing decisions can still be evaluated usefully under partial observability if the objective is framed as *alignment* rather than causality. Instead of claiming that a given contact caused a better outcome, we ask whether the windows selected by a timing engine capture a disproportionate share of observed payment intensity relative to their calendar span, and whether that alignment is consistent with independently observed operational lift. This framing yields a practical evaluation layer for timing support while making the inferential boundary explicit.

The contributions are threefold:

- We formalize contact-timing evaluation under partial observability and introduce a set of complementary alignment metrics tailored to contrastive temporal decisions.
- We propose a two-layer engine that combines aggregate-level seasonal signal extraction with account-level operational constraints and attribution rules.
- We show, on two independent datasets, that the resulting configurations consistently improve alignment relative to baseline strategies, with seven of eight combinations outperforming the uniform and random baselines, and all eight exceeding the expert-inspired baseline.

The scope of the paper is therefore deliberately narrow. We study temporal alignment as a decision-support problem. We do not claim causal identification of contact effects, and we treat randomized experimentation as the appropriate next step for that objective.

2 Related Work

Prior work in debt collection and adjacent credit operations has largely focused on risk scoring, repayment propensity estimation, and prioritization under observable outcomes. These approaches are effective for ranking accounts or identifying high-risk segments, but they do not directly answer whether a contact should occur at one time rather than another. In this sense, they address selection over entities more than selection over time [1–3, 5].

A second stream of literature studies action selection through reinforcement learning and related sequential decision methods. These approaches are appealing because they can encode repeated interaction, budget constraints, and delayed rewards. However, they inherit the same attribution difficulty that affects timing evaluation more broadly: the reward assigned to a contact policy depends

Table 1. Positioning of the proposed work against adjacent literature.

Stream	Typical focus	Limitation for timing evaluation
Risk and propensity scoring	Which accounts are likely to repay or default	Does not directly evaluate temporal contrast
Sequential action selection / RL	Policy optimization under repeated interaction	Typically assumes attribution or reward definition is already resolved
Temporal evaluation under partial observability	When and how often to act	Sparse explicit treatment in the debt-collection setting

on assumptions about how outcomes are generated and observed. In much of the applied literature, that attribution problem is handled implicitly rather than evaluated directly [8–10].

The gap is therefore not the absence of optimization models, but the absence of evaluation mechanisms for temporal decisions when only observational traces are available. Our position is that a useful intermediate step is to evaluate alignment against continuous payment intensity, while being explicit that alignment is not causality. This framing is also consistent with prior observations that payment outcomes in collection are jointly shaped by channel exposure, sequencing, and behavioral response rather than by isolated touches [6, 11].

3 Problem Formulation

We define partial observability here as the setting in which the realized data contain contact events, operational account descriptors, and outcome traces, but not the counterfactual outcomes that would have resulted from alternative contact timings for the same account. Under this definition, the effect of timing cannot be identified from logs alone without additional assumptions, a difficulty that mirrors the attribution problem already noted in operational collection analytics and channel-effect studies [6–8].

Formally, let a_i denote a delivered contact and y_j a recorded payment, and let s denote a contact schedule. The quantity of interest for a timing decision is the contrast

$$\tau(s, s') = \mathbb{E}[Y \mid \mathrm{do}(s)] - \mathbb{E}[Y \mid \mathrm{do}(s')]\,, \tag{1}$$

between two schedules applied to the same portfolio in the same period. Only one schedule is ever realised, so Eq. 1 is counterfactual by construction. A second obstacle compounds the first: the mapping from payments to contacts is not observed but constructed. Writing $\mathcal{Y}$ for all recorded payments and $\mathcal{Y}_{\mathrm{attr}}$ for the subset attributable to a contact under a stated attribution rule,

$$\mathcal{Y}_{\mathrm{attr}} \subsetneq \mathcal{Y}, \qquad \frac{|\mathcal{Y} \setminus \mathcal{Y}_{\mathrm{attr}}|}{|\mathcal{Y}|} > 0, \tag{2}$$

so the measurement itself depends on a modelling choice. We refer to this pair of conditions, an unobservable contrast and a constructed, incomplete attribution, as partial observability. Neither is a data-collection defect that better instrumentation would remove: the first is definitional, and the second reflects that debtors pay for reasons unrelated to any specific contact. We do not estimate τ; we estimate alignment, defined next.

3.1 Alignment Metrics

Let $\mathcal{M} = \{1, \ldots, 12\}$ denote calendar months, treated cyclically, and let $P(m) \geq 0$ denote aggregate payment intensity observed in month m over the evaluation period, measured either as a transaction count or as a monetary total. A *contact window* is a contiguous subset $W \subseteq \mathcal{M}$, contiguity understood modulo 12 so that windows may cross the year boundary.

The *Temporal Alignment Score* of a window is the share of annual payment intensity falling inside it,

$$\mathrm{TAS}(W) = \frac{\sum_{m \in W} P(m)}{\sum_{m \in \mathcal{M}} P(m)} \in [0, 1]. \tag{3}$$

Coverage is the share of the calendar the window occupies,

$$\mathrm{Cov}(W) = \frac{|W|}{|\mathcal{M}|}, \tag{4}$$

and *precision* is their ratio,

$$\mathrm{Prec}(W) = \frac{\mathrm{TAS}(W)}{\mathrm{Cov}(W)}. \tag{5}$$

Precision admits a direct reading: $\mathrm{Prec}(W) = 1$ means the window captures exactly the share of payment intensity its duration would imply, and $\mathrm{Prec}(W) > 1$ means intensity is concentrated within it. All three quantities are computable from aggregated series and require no individual-level attribution, which is what makes them applicable under partial observability.

3.2 Why the Triple Is Necessary

TAS must not be reported alone, since it is trivially maximised by enlarging the window: $\mathrm{TAS}(\mathcal{M}) = 1$ by construction, at $\mathrm{Prec}(\mathcal{M}) = 1$, which carries no information. Conversely, precision alone favours degenerate windows: a single high-intensity month may attain high precision while leaving most of the year uncovered, which is not actionable for an operation that must run continuously. The pair encodes an explicit trade-off between how much intensity is captured and how concentrated the capture is, and we additionally impose a hard cap of nine months on window length so that near-annual windows are not returned.

3.3 Selection Criterion versus Evaluation Metric

Because TAS is computed on a held-out period, it is an *evaluation* metric and is not available at decision time. Selecting a window by TAS would amount to selecting on the test set. Selection must therefore use a quantity estimable from the training period alone. We use a confidence score combining the relative amplitude of a detected seasonal peak with the fraction of observed years in which the pattern recurred,

$$\mathrm{conf}(W) = \mathrm{clip}\left(\tfrac{1}{2} \cdot \frac{u(W)}{u_0} + \tfrac{1}{2} \cdot \frac{n_{\mathrm{sup}}(W)}{n_{\mathrm{years}}},\ 0,\ 1\right), \tag{6}$$

where $u(W)$ is the relative uplift of the window over the seasonal mean, u_0 a normalising constant, and $n_{\mathrm{sup}}(W)$ the number of observed years in which the window mean exceeded that year's mean. The engine abstains when $\mathrm{conf}(W)$ falls below a threshold, emitting an explicit insufficient-signal outcome rather than a window. We regard the distinction between a criterion available *ex ante* and a metric available only *ex post* as essential to reporting such systems honestly.

The optimization target in this paper is improved alignment relative to baselines under explicit operational constraints. We do not optimize or claim causal uplift, treatment effect, or policy optimality in the counterfactual sense.

4 The Two-Layer Engine

The proposed engine decomposes the timing problem into a macro layer and a micro layer. The macro layer operates on aggregate repayment-related signals and identifies candidate windows of interest. The micro layer filters and ranks operational opportunities within those windows under business constraints.

4.1 Macro Layer

At the macro level, the input is an aggregate signal derived from SIBS-linked payment activity over time. We apply seasonal-trend decomposition using Loess (STL) to separate recurring seasonal structure from trend and residual components [12]. Candidate peaks are then detected on the seasonally informative component, and windows are selected using a confidence-based rule. The macro motivation is that repayment-related activity is embedded in broader seasonal and calendar effects rather than being temporally uniform [13].

The macro signal is not used as a proxy for purchasing behaviour but as an indicator of *liquidity*. Household-finance research establishes that income cycles, including wages, pensions and social transfers, produce predictable fluctuations in disposable liquidity, and that consumption and payment activity are synchronised with those cycles. Settling a debt requires the same condition as completing a purchase, so aggregate payment activity and debt repayment are effects of a common cause rather than one being a proxy for the other. We note further that the series used are filtered to the financial sector, and that the assumption is not taken on trust: the operational lift analysis of Section 6, computed exclusively from collection records that never enter the macro pipeline, independently identifies the same productive months within the mid-year regime.

This layer is designed to answer a narrow question: which periods exhibit population-level temporal patterns that are plausibly favorable for contact? It does not allocate specific accounts or channels. Its output is a set of candidate windows and associated confidence scores. The macro dataset comprises 294 CSV files covering 87 months from January 2019 to March 2026, aggregated at monthly level at source and segmented by sector, modality, and geography; the present study uses the financial-sector slice only.

4.2 Micro Layer

The micro layer operates on proprietary operational data containing account status, channel information, assignment rules, and historical interaction traces. It determines which accounts are admissible for contact within the macro-selected windows and applies business filters before ranking or forwarding cases. This reflects the practical separation between predictive scorecards and operational optimization engines discussed in the collection literature [8, 14].

This layer can incorporate channel availability, assignment consistency, and lift-related heuristics, but the coupling to the macro layer remains minimal: an account-period pair either falls inside an admissible window or it does not. Importantly, payment modality is not used to infer the communication channel of a contact event, which avoids the strong attribution assumptions that often blur channel studies in collection settings [6].

The deployable artefact is a cadence matrix over two dimensions: urgency and value. Urgency has three levels: recent (< 30 days), moderate (30–60 days), and critical (> 60 days), while value is split into low and high relative to the firm's historical median pending amount, with a global fallback of 58 EUR when firm-level history is unavailable. The resulting 3×2 matrix assigns between 5 and 12 contacts per horizon, concretely 5/14d, 6/14d, 7/21d, 8/21d, 9/30d, and 12/30d across the six cells, with escalation horizons of D+90 for recent cases, D+60 for moderate cases, and D+45 for critical cases.

A seasonal modifier is then applied. When a segment falls inside the primary macro window, the engine adds two contacts subject to a hard ceiling of 12, compresses spacing by 20%, and inserts the extra actions at one third and two thirds of the horizon. The coupling is binary: `in_window` is membership of the calendar month in the primary seasonal window. The macro signal does not choose channel or cadence form; it only gates whether the seasonal modifier is active. This separation is deliberate, and we state it explicitly because the SIBS payment modality is not used to infer the contact channel.

4.3 Constraint-First Coupling

The combined system is formulated as a constraint-first engine. Operational rules define the admissible action space before any timing preference is applied. This avoids the common failure mode in which a scoring layer suggests actions that are infeasible, policy-incompatible, or semantically invalid. Constraint-first design is especially important in debt collection because capacity, escalation, and frequency constraints are structurally part of the decision space rather than optional post-processing [4, 8, 14].

The engine also abstains explicitly. When confidence falls below the threshold of 0.50, it emits an insufficient-signal outcome rather than forcing a window; this occurred in one of the nine modality–metric combinations. At the micro level, the constraint-compliance rate (CCR) is 1.00 by construction: this is a design property of the admissibility layer, not an empirical performance result. The encoded `ConstraintPolicy` covers frequency caps, minimum spacing, channel and consent restrictions, and prohibited windows.

We also distinguish between evaluation tuning and production fitting. The evaluation configuration is trained up to December 2023 and scored on a subsequent hold-out period, which is the only way to obtain an honest TAS. The production configuration consumes all historical data available and therefore has no out-of-sample period against which it can be scored. Reporting TAS for the production fit would amount to evaluating on the training period. This distinction explains why the evaluated window and the deployed window may differ while still belonging to the same seasonal regime.

5 Experimental Setup

The empirical study uses two datasets. The macro dataset is reproducible and constructed from aggregate temporal signals. The micro dataset is proprietary and contains operational records needed to instantiate account-level constraints and admissibility rules. Their joint use allows the timing engine to be evaluated through independent yet connected observational views.

The macro split is temporal, with training up to December 2023 and validation from January 2024 onward. The operational dataset covers 18 months from November 2024 to April 2026 across 30 firms, with 937,756 messages sent and 502,818 delivered. Of the 258,953 payments registered in the period, 233,506 could be linked to a document; the remaining 25,447 (9.8%) were orphaned and excluded from response-rate estimation. A payment is attributed to a message when it falls within the attribution window of that message for the same document, the window extending to the next workflow step with a 7-day fallback for terminal steps. Under this rule the overall response rate is 1.28%, and the median contact-to-payment delay is 2.6 days.

We use temporal splits rather than cross-validation. This choice reflects the deployment setting more faithfully because timing policies are applied forward in time and should be judged on future periods rather than on shuffled folds

Table 2. Detected windows per modality–metric combination, scored on the held-out period. One of the nine combinations (Online/Val.Médio) fell below the confidence threshold and was not emitted. Mean is over combinations with confidence above the threshold.

Modality	Metric	TAS	Precision	Confidence
All	Val.Médio	0.636	1.527	1.00
All	N.Compras	0.574	1.378	1.00
All	Val.Compras	0.477	1.144	0.56
Physical	Val.Médio	0.448	1.075	1.00
Physical	Val.Compras	0.440	1.055	0.77
Physical	N.Compras	0.387	0.928	0.90
Online	N.Compras	0.461	1.106	0.69
Online	Val.Compras	0.457	1.096	0.63
Mean		0.485	1.164	n/a
Uniform baseline		0.420	1.000	n/a
Random baseline		0.420	1.000	n/a
Expert-inspired (Jan–Feb)		0.210	1.260	n/a

that blur temporal dependence. Baselines include uniform scheduling, random selection, and an expert-inspired policy. Confidence-based tie breaking is held fixed and reported explicitly. This evaluation design is aligned with the forward-looking nature of operational collection deployment and with prior evidence that behavior and repayment propensity vary by arrears stage, interaction history, and timing context [3, 5, 14].

Three combinations saturate at confidence 1.00 (All/Val.Médio, All/N.Compras, and Physical/Val.Médio). Because the confidence score is bounded above, saturation is structural rather than accidental. In the current implementation the residual ordering among saturated candidates is deterministic but not semantically motivated; TAS is never used in selection. This preserves the evaluation logic of Section 3, at the cost of leaving the final ordering among equally confident candidates only weakly motivated semantically.

From a data-governance standpoint, the macro data are aggregated at source and contain no individual-level records. The operational data are anonymised in the ETL stage, with names, contact details, and tax identifiers removed. The system issues recommendations at segment level with a human in the loop, and is therefore positioned outside fully automated individual decision-making. The operational dataset is proprietary and non-redistributable; the macro pipeline is reproducible.

6 Results

Across the eight emitted combinations, mean TAS is 0.485 against 0.420 for the uniform and random baselines and 0.210 for the expert-inspired January–

Table 3. Worked example of the deployable output for an anonymised portfolio with 64,714 documents. The classification, channels, and escalation remain fixed; the seasonal state changes only the intensity and compression of the cadence.

Condition	Inside window	Outside window
Portfolio size	64,714 documents	64,714 documents
Classification	Critical + Low Value	Critical + Low Value
Escalation	D+45	D+45
Cadence	11 contacts / 24 days	9 contacts / 30 days
Seasonal effect	+2 contacts, 20% compression	none
Expected response rate	1.47%	month-dependent (1.28% annual mean)

February window. Mean precision is 1.164. Seven of eight combinations outperform the uniform and random baselines; all eight exceed the expert-inspired baseline. This result holds across TAS and precision jointly, which makes it less likely to be an artefact of trivially broad window selection. Relative to static heuristics, this is consistent with the broader literature showing that one-size-fits-all collection policies tend to underperform more adaptive and context-sensitive decision rules [4, 8, 15].

The expert-inspired baseline attains a precision of 1.26, above the engine mean of 1.164, while capturing only 21% of annual intensity in a two-month window. This is a concrete instance of why precision must be read jointly with TAS: concentration alone favours windows too narrow to sustain a continuous operation.

The unfavourable case should be reported explicitly. Physical/N.Compras is the only emitted combination below the uniform baseline, with TAS 0.387 and precision 0.928. Its inclusion is important because it shows the method is not being presented selectively: the engine can identify regimes that are not strong enough to beat a calendar-uniform benchmark even after clearing the confidence threshold.

The convergence claim must also be stated narrowly. The macro and micro layers converge on the mid-year peak, not on all months. Operationally, monthly lift is measured against the global response-rate mean of 1.28%. June shows +23.2% lift and July +44.6%, yielding a mean of +33.9% for the June–July window. By contrast, October is at –35.1% and November at –16.4%, despite belonging to a coarse year-end seasonal grouping that a naive policy might treat as homogeneous. This is precisely the kind of internal heterogeneity the two-layer design is intended to expose.

January is the important divergence. The macro signal classifies it as low liquidity, while the micro layer shows +15.4% operational lift. We interpret this not as a failure but as evidence that the available payer pool is small in January while those who do respond convert well. The two readings are therefore compatible: one concerns broad liquidity conditions, the other response conditional on contact. The engine accordingly signals windows that contain January rather than forcing full cross-layer agreement month by month.

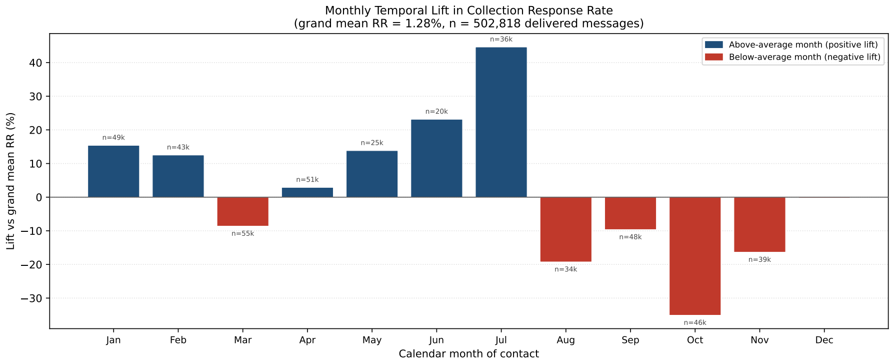

Fig. 1. Monthly operational lift in response rate relative to the global mean of 1.28%. The figure makes the internal heterogeneity of broad seasonal regimes visible: June and July are strongly positive, while October and November are negative despite belonging to the same coarse year-end grouping.

Transportability across geographic units was not assessed; the architecture supports it and the data permit it, and we identify it as the strongest available test of external validity.

7 Discussion

The proposed method has implications beyond debt collection. Any domain that combines a population-level temporal signal, operational logs, and hard action constraints may benefit from the same separation between aggregate window selection and local admissibility filtering. Examples include outreach, retention, and service-reminder systems where timing matters but counterfactual effects are difficult to observe directly.

The results also illustrate a practical risk of naive seasonal targeting: high-level seasonality labels can hide months or sub-periods with materially negative expected alignment. Treating such periods as uniformly favorable can degrade operational performance even when the broader seasonal narrative looks plausible, much as fixed escalation ladders or uniform communication rules can hide meaningful heterogeneity across debtor populations and stages [3–5].

The main modeling limitation on the macro side is the use of additive STL. In our data, residual variability grows with the level of the series, which suggests that a multiplicative decomposition, or an additive decomposition on the log scale, may fit better. The additive specification was retained for robustness to outliers and interpretability. Re-estimating in logarithmic scale and checking whether detected windows shift is a concrete refinement identified but not performed here.

The main limitation remains causal identification. The present study does not establish that contacting within the selected windows causes better outcomes. It establishes that the selected windows align more closely with favorable observed dynamics than competing baselines do. Randomized A/B testing remains the appropriate next step for resolving that boundary, particularly given prior experimental evidence that communication effects may reverse under poorly calibrated timing or pressure [11, 16].

8 Conclusion and Future Work

This paper introduces a practical framework for evaluating contact-timing decisions under partial observability through alignment metrics and a two-layer decision engine that couples population-level temporal signals with account-level operational constraints.

Three next steps are immediate. First, randomized A/B testing is required to move from alignment to causal effect estimation. Second, subsegmentation may improve timing specificity within heterogeneous debtor populations. Third, geographic transfer tests across independent units can assess how robust the engine is to operational and regional variation.

References

[1] S. Zeng et al. "Using predictive analysis to improve invoice-to-cash collection". In: *Proc. 14th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*. 2008, pp. 1043–1050.

[2] A. P. Appel et al. "Optimize Cash Collection: Use Machine Learning to Predict Invoice Payment". In: *arXiv preprint* (2019). arXiv: `1912.10828`.

[3] Bruno S. Nascimento, Denis Maia, and Lucas Almada. *Applying Machine Learning to Improve Collection and to Reduce Write-Offs in Utilities.* 2021. DOI: `10.1109/AIBT53261.2021.00017`.

[4] Rafał Jankowski and Andrzej Paliński. "Debt Collection Model for Mass Receivables Based on Decision Rules—A Path to Efficiency and Sustainability". In: *Sustainability (Switzerland)* 16 (14 2024). DOI: `10.3390/su16145885`.

[5] Andrés Carrera and Marco E. Benalcázar. "Optimizing the Collection Process in Credit Risk Management: A Comparison of Machine Learning Techniques for Predicting Payment Probability at Different Stages of Arrears". In: *Journal of Risk and Financial Management* 18 (11 2025). DOI: `10.3390/jrfm18110630`.

[6] Philip Halsey et al. "Analyzing the Role of Digital Communication Channels in Debt Collection". In: *2021 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE, Apr. 2021, pp. 1–6. ISBN: 978-1-6654-1250-6. DOI: `10.1109/SIEDS52267.2021.9483751`. URL: `https://ieeexplore.ieee.org/document/9483751`.

[7] Catalina Sánchez, Sebastián Maldonado, and Carla Vairetti. "Improving debt collection via contact center information: A predictive analytics framework". In: *Decision Support Systems* 159 (Aug. 2022). ISSN: 01679236. DOI: 10.1016/j.dss.2022.113812.

[8] Wenhui Liu et al. "Adversarial Conservative Alternating Q-Learning for Credit Card Debt Collection". In: *IEEE Transactions on Knowledge and Data Engineering* 37 (4 2025), pp. 1542–1555. DOI: 10.1109/TKDE.2025.3528219.

[9] N. Abe et al. "Optimizing debt collections using constrained reinforcement learning". In: *Proc. 16th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*. 2010, pp. 75–84.

[10] Zihao Wang et al. "Two-stage behavior cloning for spoken dialogue system in debt collection". In: *IJCAI International Joint Conference on Artificial Intelligence*. Vol. 2021-January. 2020, pp. 4633–4639. ISBN: 9780999241196; 9781577357384; 9781956792003; 9780999241165; 9780999241141; 9781577355120; 9781577354260; 9780769536156; 0934613346; 9780999241127.

[11] Andris Saulitis. "Nudging debtors with non-performing loans: Evidence from three field experiments". In: *Journal of Behavioral and Experimental Finance* 37 (Mar. 2023). ISSN: 22146369. DOI: 10.1016/j.jbef.2022.100776.

[12] Robert B. Cleveland et al. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess". In: *Journal of Official Statistics* 6.1 (1990), pp. 3–73.

[13] Robert B. Barsky and Jeffrey A. Miron. "The Seasonal Cycle and the Business Cycle". In: *Journal of Political Economy* 97.3 (1989), pp. 503–534. DOI: 10.1086/261614.

[14] Ekrem Duman, Nazli Mehrabi, and Çağrı Çakır. "Collection Model Development to Predict Which Customers Will Pay". In: 2025, pp. 237–241. ISBN: 9798331513702. DOI: 10.1109/SPML66318.2025.11199754.

[15] Minou Ghaffari, Maxime Kaniewicz, and Stephan Stricker. "Personalized Communication Strategies: Towards a New Debtor Typology Framework". In: *Psychology and Behavioral Sciences* 10 (6 2021), pp. 256–268. DOI: 10.11648/j.pbs.20211006.20. URL: https://doi.org/10.11648/j.pbs.20211006.20.

[16] A. Saulitis. "Personalized messaging enhances hospital debt collection while prosocial appeals fail: Evidence from a field experiment". In: *Digital Health* 10 (2024). DOI: 10.1177/20552076241277035.

Appendix E

Note on AI Usage

Artificial Intelligence tools were used throughout the preparation of this dissertation as supporting instruments for writing refinement, technical assistance, and quality control. Their use was limited to assisting the development of my own original work and did not replace the formulation of the research problem, the design of the methodology, the interpretation of results, or the final academic judgement presented in this dissertation.

In the writing process, AI tools were used to improve the clarity, structure, and readability of text originally written by the author. This included suggestions for rephrasing, improving transitions between sections, reducing ambiguity, and aligning the language with an academic style. All substantive claims, arguments, references, and interpretations were reviewed and validated by the author before inclusion.

AI tools were also used during the development of the experimental engine, particularly for debugging code, identifying possible causes of implementation errors, suggesting alternative coding approaches, and improving the readability of scripts. The resulting code was manually inspected, tested, and adapted by the author to ensure that it was consistent with the intended methodology and produced valid outputs.

No AI-generated text, code, or analytical result was included without authorial review. The responsibility for the content, implementation choices, results, and conclusions of this dissertation remains entirely with the author.