Politecnico di Torino

ENGINEERING AND MANAGEMENT

LM-31

A.A. 2025/2026

July 2026

Analysis of the factors determining delays in the construction of healthcare PPPs

An empirical study of project- and country-level variables and their impact on the timely delivery of PPPs.

Supervisors:

Prof. Giulio Mangano, PhD, Associate Professor

Simone Preziosa, PhD Candidate

Candidate:

Arianna Raschieri

Table of Contents

Table of Tables

Table of Figures

Introduction and thesis overview

Delays are a recurring feature of major infrastructure delivery, and Public-Private Partnerships are no exception. In the EU context, the European Court of Auditors documents that, among the audited EU-supported PPPs, most completed projects experienced construction delays, eroding expected benefits and generating measurable inefficiencies in public spending (European Court of Auditors , 2018).

This thesis focuses on healthcare PPPs because schedule slippage in hospitals translates quickly into lost capacity and delayed service availability. Hospitals concentrate interface and commissioning risk (Mechanical, Electrical, and Plumbing systems, critical areas, Information and Communications Technology integration, and service decanting). Delays affect budgets and contractor performance but also the timing of usable capacity and the public value of the investment. (National Audit Office, 2020).

Research question.

The research question is: which observable project-level and country-level factors are associated with the probability that a hospital PPP is delivered late, relative to its planned construction baseline?

Empirical approach.

To answer this question, the thesis constructs a cross-country, project-level dataset of hospital PPPs from public and institutional sources and evaluates timeliness against the project baseline. The core outcome is a binary indicator: Late = 1 if realised construction duration exceeds planned construction duration, and Late = 0 otherwise (including on-time and early completion). This baseline-based definition aligns the empirical design with how audit and portfolio evidence operationalise schedule underperformance.

The core predictor set is parsimonious and relies on variables observable at project start: project scale and interface burden, baseline schedule scale, contractual design, and country context. Descriptive evidence profiles delay magnitude and distributions; the inferential question is addressed through a multivariate logistic regression on the probability of Late.

Contribution and relevance.

The thesis offers an interpretable delays-focused assessment for hospital PPPs. It links baseline schedule outcomes to a small set of observable predictors aligned with guidance and audit evidence, and it uses those patterns to inform the policy discussion on preparation discipline, contract architecture, and contract-management capability in healthcare PPP delivery.

Positioning relative to existing literature.

Scholarship on PPP performance has produced two complementary streams. The first, theoretical, has clarified the contractual logic of bundling design-build-finance-operate obligations under performance regimes and the trade-offs implicit in incomplete long-term contracts (Hart, 2003; Iossa & Martimort, 2015; Engel, Fischer & Galetovic, 2014). The second, empirical, has documented patterns of cost overruns, renegotiations, and schedule slippage primarily in transport (Guasch, 2004; Cruz & Marques, 2013; Soliño & Vassallo, 2009) and through aggregated audit and portfolio evidence (European Court of Auditors, 2018; National Audit Office, 2020; World Bank, 2024). Two reviews (Hodge and Greve 2007, 2017 for the broader PPP literature, and Roehrich, Lewis and George 2014 for healthcare PPPs) synthesise these strands but also flag the comparative scarcity of project-level cross-country empirical work focused on schedule outcomes in healthcare PPPs. The present thesis aims to address this gap by combining a project-level cross-country dataset of hospital PPPs with an audit-style baseline-based definition of delay and a parsimonious, pre-specified inferential design that maps observable predictors to the probability of late delivery.

Structure of the thesis.

The thesis is organised as follows. Chapter 1 contextualises PPP practice across sectors with focus on health, reviewing definitions, contract families, and the specific drivers of delay relevant to hospital infrastructure. Chapter 2 translates the delivery principles introduced in Chapter 1 into observable predictors and formalises the four working hypotheses. Chapter 3 sets out the statistical strategy: descriptive group comparisons, the core multivariate logit on Late, the pre-specified non-linearity in concession tenor, and the disciplined robustness plan. Chapter 4 reports the data construction, sample-flow rules, descriptive statistics, the core logistic model, and a pre-specified robustness architecture covering sample sensitivity, functional form, macro-proxy swaps, and country-control matrix variants. Chapter 5 discusses the results in light of the conceptual framework, identifies limitations and external-validity boundaries, and develops policy and managerial implications anchored to the empirical findings. The thesis closes with a focused conclusion that summarises the answer to the research question, the specific contributions of the work, and a structured set of future research directions.

Chapter 1 - Public-Private Partnerships: concepts, uses, sectors, and the health case

1.1. PPP features

PPP definition.

No single universally accepted definition of PPP exists. The PPP Reference Guide v3.0 defines a PPP as “a long-term contract between a private party and a government entity, for providing a public asset or service, in which the private party bears significant risk and management responsibility, and remuneration is linked to performance” (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017). The OECD similarly defines PPPs as “long term contractual arrangements between the government and a private partner whereby the latter delivers and funds public services using a capital asset, sharing the associated risks” (Organisation for Economic Co-operation and Development, 2012).

In traditional public procurement, government entities procure design and construction separately from the long-term operation of a public service. A Public Private Partnership (PPP) instead combines several phases of an asset’s life cycle (design, build, finance, operate, and maintain) into a single long-term contract with a private counterparty, often a special-purpose vehicle. The counterparty assumes material managerial responsibilities and risks, and its remuneration is tied to outputs and performance. This approach reflects the international practice described in the PPP Reference Guide v3.0 (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017). *PPPs are not merely “private finance” but performance-based service contracts. Whether they deliver value depends on risk transfer that is real, measurable, and enforceable.*

Over recent decades, governments have deployed PPPs across transport, health, education, water, and environmental services, with different intensities across countries and time cycles. The prudent way to characterise PPPs is sectoral rather than tied to rigid time windows, because volumes have been cyclical (e.g., pre-2007 expansion, post-crisis retrenchment, selective recoveries) and policy-driven. PPP prevalence is more accurately read as context-specific and shaped by policy cycles than as a constant. Reference overviews by Multilateral Development Banks adopt this caution (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017).

Legal and institutional framing.

No single binding global legal definition of a PPP exists. Practice is shaped instead by model laws, guidance, and standards. United Nations Commission on International Trade Law (UNCITRAL) issued the Legislative Guide on Privately Financed Infrastructure Projects (adopted in 2000, published 2001) and the Model Legislative Provisions (adopted in 2003), which many jurisdictions use as a point of departure (United Nations Commission on International Trade Law (UNCITRAL), 2004). Then, there is the Organisation for Economic Co-operation and Development's Recommendation on the Principles for Public Governance of PPPs that stresses governance, transparency, and fiscal sustainability (Organisation for Economic Co-operation and Development, 2012). Also, the World Trade Organisation's Government Procurement Agreement (GPA), even if it does not define PPPs, sets binding procurement principles, like non-discrimination and transparency, relevant to PPP tendering where GPA coverage applies (World Trade Organization, 2012).

In Europe, PPPs rest on three pillars: Directive 2014/24/EU on public procurement (European Union, 2014), Directive 2014/25/EU for utilities/special sectors (European Union, 2014), and Directive 2014/23/EU on concessions (European Union, 2014). None of them regulates PPPs as a category, but together they regulate the contracts on which PPPs are built.

On the accounting side, the Eurostat/EPEC (European PPP Expertise Centre) guide (2016) is crucial to clarify when PPPs are recorded on- or off-government balance sheet under ESA 2010, which is important for fiscal transparency but not a policy objective in itself (European PPP Expertise Centre; Eurostat, 2016). Under ESA 2010 practice, a PPP asset is on the government balance sheet when the public sector bears most construction risk and (in substance) availability or demand risk; it can be recorded off balance sheet only when the private partner genuinely carries construction risk and at least one of availability or demand risk. *The sequence should run the other way: start from service objectives, value for money, and affordability, and let the accounting classification follow, rather than designing the contract around the label.*

National examples show how states have responded to the same directives. The legislative overview below covers the countries most frequently encountered during this research as PPP locations. Italy consolidated PPP rules under the new Public Contracts Code (Italy, 2023). Portugal's PPP framework evolved through Decree-Laws 86/2003 and 141/2006, with a central

PPP unit established by Decree-Law 111/2012 (Portugal, 2003). The UK launched PFI in 1992 without a single "PPP statute," and announced the end of PFI/PF2 for new projects in 2018 (Treasury, 2018). In the U.S., PPPs are largely state-driven, with federal transport statutes (Congress, 1991) (Administration, 1998) (Administration, SAFETEA-LU – Summary Information, 2005) enabling instruments and programs. These examples show that PPP law is ecosystem-specific rather than one-size-fits-all.

PPP types: what the acronyms actually mean.

No single "PPP model" exists; practice spans a continuum of contract forms that differ by scope, duration, risk allocation, and remuneration.

O&M / Management contracts. A private operator runs and maintains an existing asset to specified service levels with deductions for under-performance. Only when contracts are genuinely performance-based, multi-year, and involve meaningful private obligations (including capex) do they approximate PPPs in the strict sense. In O&M and management contracts, operating-performance risk typically transfers to the private side, while demand and investment risk remain with the public sector. (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017). *O&M sits at the 'light' end of the PPP spectrum: it can strengthen operational discipline, but it is not, in itself, a financing model.*

DB (Design & Build). Single contract for design and construction, delivered "turnkey". Because DB excludes long-term operations and demand risk, it is a procurement method, not a PPP proper. DB remains a relevant benchmark for whole-life cost comparisons against DBFM/DBFOM structures (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017) (FHWA, Federal Highway Administration, 2016)

DBO (Design, Build, Operate). The private partner designs, builds, and operates; the public side often finances and retains ownership. Payments are service-based (availability/quality), with little or no financing risk on the private side. Authorities choose DBO when they want to internalise operational incentives without outsourcing the balance sheet (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017).

DBFM/DBFOM (Design, Build, Finance, (Operate), Maintain). The "full life-cycle" family. Revenues can be user-pay (economic exploitation rights typical of concessions) and/or government-pay through availability payments; in both cases, the contract provides for automatic deductions for unavailability or underperformance measured by KPIs and availability thresholds. Risk transfer focuses on construction (time/costs) and availability/operability; in user-pay configurations, demand risk is added. When well specified, these schemes align whole-life costing incentives: the private sector internalises capital costs, design choices and maintenance throughout the entire life cycle, because the quality of the asset directly affects payments and penalties (Federal Highway Administration, n.d.) (European PPP Expertise Centre, 2018) (World Bank, 2019). *Payment mechanisms are a policy lever, and the risk package has to be bankable from the outset; otherwise the contract is more likely to be reopened later.*

BOT (Build, Operate, Transfer) and variants: BOO (Build, Own, Operate), BOOT (Build, Own, Operate, Transfer), BLOT (Build, Lease, Operate, Transfer). Under BOT, the private side builds and operates the asset for a term, then hands it back; BOO retains private ownership; BOOT combines temporary ownership with transfer; BLOT layers a lease. These are output-based concession families, prevalent in greenfield transport and energy (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017).

Concessions (EU framing). EU law defines concessions as contracts where operational risk is transferred to the concessionaire, whose consideration is the right to exploit the service (possibly with public support). The legal test focuses on exposure to market/demand or availability risk that is not purely nominal (European Union, 2014). *Concessions range from pure user-pay structures to hybrid models that combine user revenues with public contributions.*

Institutional PPPs (IPPP). Joint-venture companies between public and private shareholders that execute procurement or concessions. The Commission’s Green Paper (2004) and the Interpretative Communication on IPPPs (2008) clarify how to select the private shareholder and award contracts lawfully. IPPPs are frequent in local public services, where corporate governance can support integrated delivery (European Commission, 2004) (European Commission, 2008).

Table 1 - Overview of PPP contract models

Model	Private role (core)	Who finances / owns?	Key risks to private	PPP level
O&M	Operate & maintain existing asset	Public finances & owns	Operational performance	Very light
DB	Design & build, turnkey	Public finances & owns	Construction (time, cost, quality)	Not a PPP (procurement)
DBO	Design, build & operate	Public usually finances & owns	Construction + operation/ availability	Intermediate
DBFM / DBFOM	Design, build, finance, (operate), maintain	Private finances during term; asset usually returns to public	Construction + availability; sometimes demand	Full PPP
BOT / BOO / BOOT / BLOT	Build & operate; ownership varies	Mainly private finance, especially greenfield	Construction + operation + demand	Concession-type PPP
Concession (EU)	Provide & manage service/works	Concessionaire may finance and (co-)own during term	Operational + real market/ availability risk	Concession (legal category)

IPPP	Joint venture delivers services / contracts	Mixed public–private shareholding	Shared via contracts & governance	Institutional PPP

1.2 Why governments use PPPs

Long-term service delivery and performance discipline.

Public–private partnerships are structured as long-term contracts in which a private counterparty (typically a special purpose vehicle) takes on clearly specified obligations of availability and quality and is remunerated against measurable performance. The appeal of the model lies in bundling design, construction, and life-cycle operations/maintenance so that incentives on whole-life costs are internalised rather than split across multiple contracts. In practice, this is less a matter of "outsourcing works" and more a mechanism to buy outputs under a performance regime. *Performance must be observable and enforceable, and risk transfer should support, not replace, management accountability (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017) (European PPP Expertise Centre, 2018).*

Strategic rationale and governance conditions.

When PPPs work, they combine private management and industrial know-how with output-based payment mechanisms to pursue value for money. Outcomes, however, depend on governance: credible project preparation, a like-for-like comparison with traditional procurement, risk allocation that mirrors capabilities, and contract management over the full term. Without these, potential gains tend to shrink or even reverse. The OECD's guidance is explicit that institutional capacity is a prerequisite, not an afterthought (Organisation for Economic Co-operation and Development, 2012).

Public finance perspective.

PPPs can accelerate delivery when budget headroom is tight, but they create multi-year fiscal commitments, availability payments, guarantees, and rebalancing mechanisms, that must be transparent and affordable. In the EU, the Eurostat/EPEC guide clarifies when PPP assets/liabilities are recorded on or off government balance sheet under ESA 2010; "off-balance-sheet" classification is a statistical outcome, not a policy goal. At programme level, ministries of finance should set affordability caps and assess fiscal risks up front. Private finance does not make projects cheaper per se; it reshapes cash flows and reallocates risks. (European PPP Expertise Centre; Eurostat, 2016) (World Bank, 2019)

What the evidence shows.

Audit evidence in Europe is mixed. The European Court of Auditors documents cases of delays, cost increases, and under-use where preparation and oversight were weak. This does not invalidate PPPs; it simply shows they are no shortcut and require strong public capabilities to capture value (European Court of Auditors, 2018).

Political-economy considerations.

Because PPPs alter the timing and visibility of costs, they can be attractive for political reasons unrelated to efficiency. Institutional safeguards (fiscal transparency, rigorous business cases, and independent gatekeeping) help align decisions with the public interest and mitigate pro-PPP bias. The Public-Private Partnerships Fiscal Risk Assessment Model, an Excel-based tool developed by the International Monetary Fund (IMF) and the World Bank (WB), and the OECD's governance recommendations are designed precisely for this purpose (World Bank, 2019) (Organisation for Economic Co-operation and Development, 2012).

Renegotiations and information asymmetries.

Long-term contracts face shocks: regulatory changes, demand shifts, and input-price and inflation swings. These often trigger renegotiations; the issue is not whether adjustments occur, but how they are governed. Good practice codifies ex-ante the triggers, data requirements, caps, and dispute-resolution routes so that rebalancing preserves incentives rather than rewarding poor performance. Guidance from EPEC and the G20 Global Infrastructure Hub provides practical templates for availability-based PPPs (European PPP Expertise Centre, 2018) (Global Infrastructure Hub, 2018).

1.3 Main areas of application of PPPs

Transport and major mobility infrastructure.

Transport has long been the core field for PPPs because assets lend themselves to measurable outputs (availability, safety, travel-time reliability) and to either user revenues or availability-based mechanisms. Both user-pay concessions and government-pay DBFM/DBFOM models are used, with risk transfer centred on construction and operational availability and, where relevant, demand (European PPP Expertise Centre, 2018) (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017). In the U.S., Denver's Eagle P3 is widely cited as the first commuter-rail PPP to integrate design–build, private finance, and long-term operations and maintenance under a 34-year agreement in which the public authority retains ownership and makes performance-linked availability payments to the private partner (U.S. Department of Transportation – Build America Bureau, 2024) (Federal Transit Administration, 2016). In Europe, the Dutch HSL-Zuid superstructure contract with Infraspeed is a DBFM with a 25-year maintenance period and stringent availability targets (≈99 percent) under an availability-payment regime (Railway Gazette International, 2001) (Netherlands Court of Audit (Algemene Rekenkamer), 2014). *Transport PPPs work best where outputs can be specified and verified at low cost; long-term O&M obligations are central to realising whole-life-cost efficiencies rather than one-off capex savings.*

Urban services and local infrastructure.

At municipal level, PPPs are frequently used in water and wastewater, solid-waste management, street lighting, public buildings (e.g., schools), and urban transit. The rationale rests less on tariff monetisation alone and more on embedding industrial know-how and performance incentives in long-term service delivery while keeping core policy levers (standards, tariffs, equity of access) with the public authority (European PPP Expertise Centre, 2018) (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017). International finance institutions have documented pipelines across these sectors, alongside guidance on preparation and procurement adapted to sub-national capacities (European PPP Expertise Centre, 2018). *At municipal level, outcomes tend to improve when the contract mirrors the service mandate, specifies outputs clearly, and embeds robust monitoring and change mechanisms that preserve incentives.*

Health sector.

In healthcare, three broad PPP families are observed: (i) service PPPs (clinical or diagnostic services), (ii) infrastructure PPPs (DBFM/DBFOM for facilities and non-clinical FM, with clinical services remaining public), and (iii) integrated models combining infrastructure and selected services (International Finance Corporation (IFC), 2023) (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017) (World Bank, 2023). International evaluations show mixed results: where project preparation is strong and performance regimes are credible, efficiency and quality improvements are reported; where these preconditions are weak, renegotiations and cost overruns are more frequent. The systematic review by Roehrich, Lewis and George (2014) of healthcare PPPs reaches a similar conclusion at the academic level: outcomes are highly heterogeneous and conditional on contextual capacity rather than on the PPP label itself. Earlier work by McKee, Edwards and Atun (2006) had already cautioned against treating hospital PPPs as a financing shortcut and emphasised that clinical and operational outcomes depend more on the design of the contractual interface between public and private services than on the financing structure (Independent Evaluation Group (World Bank Group), 2016) (European PPP Expertise Centre, 2018). *Two practical takeaways are relevant: scope discipline (what is, and what is not, in the PPP perimeter) and contract management capacity matter at least as much as financing structure in determining outcomes.*

1.4 Why health is a distinctive PPP case

Strategic salience and policy constraints.

Hospitals and clinical estates carry high social sensitivity: universal access, continuity of care, infection control, and clinical safety set a demanding baseline. Post-pandemic investment agendas emphasise resilience and capacity, yet major institutions stress that delivery quality and institutional strength govern sustainability (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017) (European PPP Expertise Centre, 2018) (European PPP Expertise Centre; Eurostat, 2016). In Europe, statistical classification under ESA 2010 may allow off-balance-sheet treatment in defined cases, but Eurostat/EPEC explicitly caution that accounting presentation cannot substitute for affordability analysis and transparent disclosure of long-term commitments (European PPP Expertise Centre; Eurostat, 2016). *In healthcare,* *PPPs are best treated as a whole-life delivery tool, not a budget workaround; fiscal space and governance need to be demonstrated ex-ante, not inferred from accounting classification.*

Contract architecture and role allocation.

Health PPPs are typically availability-based DBFM/DBFOM structures: the private partner designs, builds, finances, and maintains the facility (hard Facility Management, covering operation and maintenance of the physical estate and technical systems, and often specified soft Facility Management, covering non-technical services that ensure daily functionality and user comfort), while clinical services remain public or are separately contracted. Effective contracts use output-based specifications, testable KPIs, payment mechanisms with deductions and bonuses, variation procedures, step-in rights, and codified handback (World Bank, 2019). The UK Treasury standardisation (PFI/PF2) and EPEC guidance remain key reference points for clauses and lifecycle management; National Health Service (NHS) Health Building Notes (HBNs) complement these with clinical-estate design standards that inform output specifications and commissioning (HM Treasury, 2018) (European PPP Expertise Centre, 2014). *Output and KPI design must map to clinical workflows and infection-control regimes, and lifecycle handback needs measurable asset-condition benchmarks to avoid end-loaded risk.*

Regulatory and managerial complexity.

Hospitals combine complex Mechanical, Electrical, and Plumbing (MEP) systems, imaging and sterile services, Information and Communications Technology (ICT) integration (Electronic Health Record (EHR), cybersecurity), and clinical flows (clean/dirty separation). Commissioning extends beyond physical completion to systems integration and the decanting of services. EPEC's operational guidance stresses that contract-management capability should be set up before operations, with defined processes for monitoring, benchmarking and market-testing of FM, and controlled variations (European PPP Expertise Centre, 2014). *Many issues described as "PPP problems" are, in reality, client-side capacity problems: weak contract management can turn small frictions into delays and renegotiations.*

Fiscal risk and assessment tools.

From a public-finance perspective, the salient risks are multi-year payment obligations (availability charges), indexation, technology refresh, and renegotiation exposure.

The PPP Fiscal Risk Assessment Model (PFRAM 2.0) developed by the IMF and World Bank helps authorities identify explicit/implicit liabilities, test sensitivity to macro shocks (inflation, rates, FX), and set affordability caps (International Monetary Fund; World Bank Group, 2019) (World Bank Group, 2020). *From a policy perspective, private finance changes the timing and risk profile of public outlays; it does not eliminate them. Robust fiscal screening is a prerequisite, not an afterthought.*

Evidence from audits and recent programmes.

Independent audits in Europe report schedule slippage and cost pressure where preparation and governance were weak. The European Court of Auditors found widespread shortcomings in EU PPPs, with lengthy procurements and execution delays that eroded expected benefits (European Court of Auditors, 2018). In England, the National Audit Office's investigation into the Royal Liverpool and Midland Metropolitan hospitals after Carillion's collapse documented substantial lifecycle cost uplifts, from £746m to £1,063m (+42%) for Royal Liverpool and from £686m to £988m (+44%) for Midland Metropolitan, while the expected public-sector share of costs rose modestly in one case and fell slightly in the other due to private-sector losses and restructuring (National Audit Office, 2020). In New South Wales, the Auditor-General identified material shortcomings in the Northern Beaches Hospital PPP's preparation and operational governance, with consequences for integration and perceived quality (Audit Office of New South Wales, 2025). Large emerging-market programmes, such as Türkiye's City Hospitals, illustrate how credit enhancements (e.g., European Bank for Reconstruction and Development (EBRD) liquidity facilities) and political risk insurance (Multilateral Investment Guarantee Agency) can unlock financing for availability-based DBFM hospital campuses (European Bank for Reconstruction and Development, 2016) (Multilateral Investment Guarantee Agency (MIGA), 2025) (Global Infrastructure Hub, 2021). *Strong risk transfer can cushion the public balance sheet when contractors fail, but only if the client can step in quickly; in higher-risk macro contexts, bankability hinges on carefully engineered risk-mitigation frameworks.*

1.5 Delays in PPP projects: a widespread challenge

Premise.

Schedule slippage is a recurring feature of large infrastructure delivery, and PPPs are no exception. A special report by the European Court of Auditors (ECA) on EU-supported PPPs found that seven of nine completed projects audited experienced construction delays ranging from two to 52 months, and that one-third of the 12 projects reviewed required unusually long procurement phases, five to six-and-a-half years, before contract award. The same audit estimated about €1.5 billion of ineffective public spending linked to delays, cost increases, and under-use, with roughly €0.4 billion co-financed by EU funds (European Court of Auditors, 2018). *PPP delivery advantages do not materialise automatically: without strong preparation and governance, time overruns can quickly erode value for money.*

Generic drivers of delay in major projects.

Recent cross-portfolio evidence from the World Bank identifies a clear pattern: capacity and planning gaps on the public-sector side, design changes, and funding constraints dominate the proximate causes of delay. In a 2024 multi-dataset analysis of infrastructure contracts, implementing entities were responsible for delays in about 44% of cases, external stakeholders for 27%, and contractors for 16%. The longest average duration occurred in the bid-evaluation stage, and nearly half of contracts showed delay during execution; many "deferred" causes originated in weak preparation (e.g., incomplete designs) that surfaced downstream (World Bank, 2024). *The clock starts slipping early, during project preparation; procurement discipline and realistic baselines are central to schedule control.* The academic literature on megaprojects reaches the same conclusion through different evidence: Flyvbjerg (2014) documents systematic optimism bias in baselines and strategic misrepresentation in project preparation, while Guasch (2004) shows that renegotiations of infrastructure concessions concentrate on contracts where the bid-evaluation stage was weak. The convergence of audit, megaproject, and concession evidence is a recurring theme across the literature reviewed in this thesis.

Land, permitting, and safeguards.

Right-of-way, land acquisition, and environmental and social (E&S) approvals are consistent bottlenecks across regions. Inter-American Development Bank (IDB) evidence for airports in Latin America and the Caribbean reports average construction delays of roughly two years, with permitting and resettlement among the critical factors (Inter-American Development Bank (IDB), 2020). Multilateral safeguard frameworks codify how to plan and mitigate these risks: the World Bank's Environmental and Social Framework (ESS5) sets requirements for land acquisition and involuntary resettlement; the Asian Development Bank's Safeguard Policy

Statement and IDB's Environmental and Social Policy Framework play similar roles (The World Bank, 2017) (The World Bank, 2017) (Asian Development Bank (ADB), 2009) (Inter-American Development Bank (IDB), 2020). *Early stakeholder engagement and realistic E&S critical-path planning are not only compliance steps; they are also practical tools to protect the schedule.*

PPP-specific complexities: procurement and financial close.

Because PPP contracts integrate risk allocation, payment mechanisms, and step-in/handback provisions, market sounding and negotiation take longer than in conventional works tenders. The ECA documents procurement durations of up to 6.5 years in one-third of the 12 audited EU PPPs (European Court of Auditors (ECA), 2018). Targeted standardization (e.g., using the World Bank's Guidance on PPP Contractual Provisions and sector-specific risk-allocation matrices from the G20 Global Infrastructure Hub) reduces uncertainty for bidders and lenders and shortens the path to financial close (World Bank, 2019) (Global Infrastructure Hub (GI Hub), 2020) (Inter-American Development Bank (IDB), 2022). *The policy takeaway is to pre-agree "known-good" clause sets where possible and deviate only with explicit rationale.*

Renegotiations and post-close change.

Design drift, service-scope changes, and macro shocks often trigger contract variations and renegotiations. EPEC's guidance on managing PPPs throughout the contract life emphasises building a contract-management function early, with clear processes for variations, benchmarking and market testing, dispute resolution, and handback; the GI Hub provides practical playbooks on renegotiation that preserve performance incentives (European PPP Expertise Centre, 2014) (Global Infrastructure Hub (GI Hub), 2018-2020). *In this thesis, contract management is treated as a core capability rather than an afterthought, and change protocols should be codified before issues surface.*

Health-sector exposure.

Hospital PPPs are delay-prone because they combine complex MEP systems, infection-control pathways, high-tech equipment integration, and ICT interoperability with clinical commissioning and decanting. UK evidence following the collapse of a main contractor in two acute hospital Private Finance Initiative schemes shows multi-year delays and large construction cost uplifts. National Audit Office analysis finds multi-year delays and sharp increases in construction costs but indicates the budgetary shock borne by the public sector was limited in real terms (around 1%) because of contractual risk transfer, with significant losses (≥£603m) absorbed by investors and insurers (National Audit Office (NAO), 2020). *Risk transfer can cushion public budgets when things go wrong, but it does not make the hospital arrive sooner; timeliness still depends on robust preparation and day-to-day contract management.*

Macroeconomic and systemic shocks.

Pandemic-era practice notes documented pipeline disruptions, rebased timetables, and the use of force majeure or change-in-law mechanisms to adjust payment regimes, particularly for availability-based projects (The World Bank (PPP Resource Center), 2021). *A practical lesson is to design for resilience: define clear triggers and boundaries for relief events ex-ante, so adjustments under stress do not devolve into ad hoc bargaining.*

1.6 Variables that influence timely delivery in PPPs

Framing.

Across the PPP lifecycle, timeliness turns on three pillars: (i) credible project preparation, (ii) a clear institutional and contractual architecture, and (iii) disciplined contract management. The *PPP Reference Guide v3.0* situates tools and roles across these pillars, while the OECD Recommendation on the Governance of Infrastructure (2020) links on-time delivery to strategic planning, capacity, integrity, and transparency (Organisation for Economic Co-operation and Development, 2020) (World Bank; Asian Development Bank; Inter-American Development Bank, 2017). *These pillars can be proxied with observable variables at both project and country level.*

Internal factors: preparation quality and business case.

Where feasibility work, options analysis (including PSC/VfM), site due diligence, and baseline schedules are weak, delay risk is elevated. HM Treasury's *Green Book* and business case model (the "five-case" approach) provide a replicable standard for defining need, options, commercial structure, affordability, and management arrangements before tender (HM Treasury, 2020) (HM Treasury, 2018). *Variables that proxy preparation quality, such as* *reliance on availability-based DBFM/DBFOM with standard clauses and a realistic planned duration, should in principle be associated with fewer slippages.*

Payment mechanism design and risk allocation.

Ambiguity in payment deductions, availability thresholds, or change-order rules often translates into disputes and re-sequencing during construction. The World Bank's contractual guidance and the Global Infrastructure Hub risk-allocation matrices (including the hospital matrix) capture "typical" allocations for construction, demand (if any), and availability risk by sector (The World Bank, 2019) (Global Infrastructure Hub (GI Hub), 2018-2020). *Projects aligned with tested allocation patterns should close faster and face fewer renegotiations, while outlier allocations require higher capacity and negotiation time to remain bankable.*

Standardization and contract management.

EPEC's *Managing PPPs during their Contract Life* recommends standardised clauses for variations, step-in, handback, indexation, and dispute resolution, alongside a staffed contract-management team with data dashboards and KPI audits (European PPP Expertise Centre (EPEC) - European Investment Bank, 2014). The Infrastructure and Projects Authority's *Project Routemap* adds a diagnostic view on governance and procurement capability for complex programs (Infrastructure and Projects Authority (IPA), 2022). *In empirical terms, sponsor capacity (e.g., specialised PPP units, contract-management plans) is a plausible predictor of timely delivery.*

Counterparty capacity and supply chain.

The World Bank's 2024 study shows public-sector implementing entities as the most frequent locus of delay, followed by external stakeholders and contractors (World Bank, 2024). *This aligns with the intuition that the number of SPV partners, JV structure, and main-contractor health can matter for schedule.*

Financial structure and macro risk.

Time to financial close reflects market conditions (rates, inflation, liquidity), hedging availability, and clause bankability. The IMF–World Bank Public-Private Partnerships Fiscal Risk Assessment Model tool formalizes fiscal commitments and sensitivity to macro shocks, which, if unrecognized, can cause pauses and re-openings at close or mid-construction

(International Monetary Fund; World Bank Group, 2019). *Practically, affordability caps and stress tests reduce later re-profiling that would delay works.*

External factors: regulatory clarity and institutional quality.

Overlaps between PPP law, procurement statutes, and sector regulation complicate approvals and tender timelines. In the EU, the Eurostat/EPEC guide clarifies ESA-2010 statistical treatment, reducing decision-time spent on on/off-balance debates (Eurostat; European PPP Expertise Centre (EPEC), 2016). At country level, better governance scores (e.g., WGI Government Effectiveness, Regulatory Quality) tend to coincide with cleaner decision paths. (The World Bank, 2024) (The World Bank, 2024).

Health-sector specifics.

Hospitals concentrate interface risk: decanting, commissioning of critical areas (theatres, ICUs), and integration of imaging, gases, and ICT. In availability-based DBFM/DBFOM models, the private partner holds availability risk, but clinical operations usually remain public; mis-specified KPIs or unclear responsibilities at the "clinical-FM" boundary create rework and schedule creep. Sectoral matrices from the GI Hub and operational guidance from EPEC help set realistic KPIs and deductions that align with clinical realities (European PPP Expertise Centre (EPEC) - European Investment Bank, 2014) (Global Infrastructure Hub (GI Hub), 2018-2020). *Specify outputs in operational terms clinicians recognize; and sequence commissioning and ICT integration in the baseline program.*

1.7 Chapter conclusions and bridge to the empirical analysis

Synthesis.

This chapter positioned PPPs as long-term, performance-linked contracts that bundle design, construction, finance, and maintenance to align private incentives with public service objectives. We reviewed generic and sector-specific (health) sources of delay and showed how governance, preparation quality, risk allocation, and contract management shape schedule outcomes. The evidence is mixed: PPPs can deliver strong service performance once in operation, yet poorly prepared or weakly governed projects slip and renegotiate, with fiscal and social costs (European PPP Expertise Centre, 2014) (European Court of Auditors (ECA), 2018) (World Bank, 2019). *The overarching message is not "PPP or not," but "PPP done right":*

rigorous ex-ante testing of Value for Money, clear payment/penalty logic, and capabilities to manage the contract over decades.

Research question and hypotheses.

Against this backdrop, the thesis asks whether identifiable project-level and country-level variables systematically explain the probability of delay in hospital PPPs (Late = 1). Building on the preceding framework, the working hypotheses are:

- H1 (Institutions). Higher institutional quality (proxied by WGI Government Effectiveness and Regulatory Quality) is associated with a lower likelihood of delay (The World Bank, 2024) (The World Bank, 2024)

- H2 (Scale/complexity). Larger schemes (Total Investment, Hospital Beds) and more complex governance (SPV number of partners) exhibit higher schedule risk, all else equal.

- H3 (Tenor). The Concession Period is expected to have a non-linear association with delay (e.g., very short or very long tenors may increase delay risk due to selection/bankability effects) (Global Infrastructure Hub (GI Hub), 2018-2020) (The World Bank, 2019).

- H4 (Development level). Projects in high-income economies should, on average and conditional on controls, display a lower probability of delay, reflecting deeper supply chains and administrative capacity (an empirical claim to be tested rather than assumed) (WDI/WBG classification; GDP/GDP per capita used as continuous controls).

Chapter 2 - Theoretical frame and variables: evidence, definitions and hypotheses

2.1 From delivery principles to observable variables

Public–private partnerships are designed to deliver services under a long-term, performance-linked contract where the private partner assumes material risk and management responsibility (The World Bank, 2017). Chapter 1 showed that delays tend to emerge when preparation, governance, and contract management are weak, so Chapter 2 translates those delivery principles into observable predictors of the Late Indicator (1/0) used later (European Court of Auditors, 2018). *Chapter 2's purpose is to convert the PPP logic into measurable pretictors that plausibly move delay risk; the dependent variable is built from planned vs actual duration.*

Preparation quality and credible baselines.

Public guidance treats realistic baselines as a core control tool because weak appraisal, incomplete scope definition, and optimistic schedules create downstream redesign and re-sequencing (HM Treasury, 2013) (HM Treasury, 2013). In this thesis, timeliness is operationalised as "Planned Total Construction Duration" and "Actual Total Construction Duration," which allow assessing whether delivery outperformed, matched, or missed the baseline (World Bank, 2024). In PPP settings, baseline credibility is also tied to whether the contract tenor is consistent with whole-life obligations and handback logic, making Concession Period a meaningful observable correlate of delivery discipline (World Bank, 2019). *Delays are measured against baselines, so baseline-related fields must be central; concession tenor is a "design" variable that can proxy lifecycle discipline and selection effects.*

Contract architecture, payment logic, and risk allocation.

Institutional guidance stresses that ambiguous payment deductions, performance regimes, and change/variation rules can generate disputes and delays, especially in availability-based projects. The economic rationale for this design discipline is captured by the incomplete-contracts framework: when contracts cannot specify all future contingencies, the structure of payment incentives and the locus of residual control rights determine whether parties internalise whole-life costs or open the door to costly renegotiations (Hart, 2003; Iossa & Martimort, 2015; Engel, Fischer & Galetovic, 2014). This theoretical lens motivates the operational variables tested in this thesis (concession tenor as a composite proxy for lifecycle and handback discipline, scale variables as proxies for interface complexity) (World Bank, 2019). Risk-allocation toolkits emphasise that "typical" allocations (construction + availability, and sometimes demand) are used as bankable starting points, while bespoke allocations raise negotiation burden and execution risk (Global Infrastructure Hub (GI Hub), 2020). Contract-form categories are retained for descriptive profiling (e.g., ANOVA/χ^2 patterns) but excluded from the multivariate logit to keep the main model aligned with design constraints (The World Bank, 2017). *Contract form matters, but it sits in descriptives; the inferential core focuses on continuous proxies that are consistently available across countries.*

Contract management capacity and governance as a measurable proxy.

PPP performance depends on the authority's ability to monitor service levels, apply deductions, manage variations, and enforce handback, which is why contract-management capacity is treated as a first-order success condition in PPP guidance (European PPP Expertise Centre, 2014). When project-level capacity variables are not directly observed, cross-country governance indicators are commonly used as proxies for implementation capability and regulatory credibility (World Bank, 2024). WGI Government Effectiveness and WGI Regulatory Quality enter as country-level predictors for H1, consistent with Chapter 1's argument that institutions shape procurement discipline and post-award oversight (The World Bank, 2024). *Governance in PPPs is an empirical proxy for the public side's ability to run complex long contracts; WGI supports both measurement and uncertainty bounds.*

Macro constraints and development level.

Macro capacity affects procurement and execution through supply-chain depth, administrative resources, and fiscal management constraints, so a development proxy is warranted even in a project-level model (Organisation for Economic Co-operation and Development, 2012). WDI GDP per capita is a standard development proxy in cross-country work and is defined consistently across economies, which supports comparability in multi-country PPP samples (The World Bank, 2024). D*evelopment matters but must be controlled; GDP per capita is usually the cleanest single macro proxy for "capacity" in cross-country regressions.*

Health-sector scale and interface complexity: why beds and investment are central.

Health PPPs concentrate technical interfaces (Mechanical, Electrical, and Plumbing systems, commissioning, clinical adjacency constraints, Information and Communications Technology integration), which increases schedule exposure relative to simpler public-building assets (European Investment Bank – European PPP Expertise Centre , 2014). Project scale is proxied by Total Investment and Hospital Capacity (beds), which reflect both engineering magnitude and operational commissioning scope (The World Bank, 2024). These variables operationalise Chapter 1's "complexity channel" without requiring subjective project narratives that cannot be standardised globally (World Bank Group; Asian Development Bank (ADB); Inter-American Development Bank (IDB/IADB), 2017). *In hospitals, "bigger" often means more interfaces and longer critical paths; Hospital Capacity (beds) and capex are defensible, observable scale proxies across countries.* The choice of beds as the operational scale proxy is partly forced by the data: alternative measures that would map more directly to interface complexity (such as gross floor area, number of operating theatres, ICU bed share, or MEP/IT system count) are not consistently reported across the public and institutional sources used to construct this dataset. Beds are the closest standardised, cross-country measure of clinical-operational scale; they are an imperfect but defensible proxy for interface burden in a multi-source PPP sample, and the reader should interpret the H2 specification accordingly.

Private consortium complexity: SPV partners as a coordination proxy.

Multi-party SPVs can increase coordination costs, governance friction, and interface risk, especially when decision rights are split across several equity partners (The World Bank (World Bank Group), 2019). The SPV Number of Partners is used in grouped form for descriptive/ANOVA comparisons rather than as a continuous predictor in the baseline logit, aligning with Chapter 1's emphasis on governance and interfaces rather than "PPP labels" (Global Infrastructure Hub (GI Hub); Allen & Overy, 2020). *The SPV structure is an "interface variable", while partner count is a simple, comparable proxy for coordination burden.*

2.2 Hypotheses and expected signs (Late = 1)

The hypotheses formalise the Chapter 1 framework into testable statements about the probability that a project is Late, defined against planned duration (World Bank, 2024). All continuous predictors are Z-standardised so that odds ratios are interpretable per one standard-deviation change and are directly comparable across variables measured in different units (National Institute of Standards and Technology, 2020). *Hypotheses are written on Late (binary) to match the modelling plan adopted in this thesis.*

- H1 (Institutions). Higher institutional quality is associated with a lower probability of delay because it supports clearer procurement, better oversight, and stronger enforcement capacity (Organisation for Economic Co-operation and Development, 2012). Worldwide Governance Indicators Government Effectiveness and Worldwide Governance Indicators Regulatory Quality are expected to have **negative** effects on Late (The World Bank (International Bank for Reconstruction and Development), 2024). *Institutions are modelled via Worldwide Governance Indicators; the expected sign is negative.*

- H2 (Scale/complexity). Larger and more complex hospital projects should face higher schedule risk because they embed more interfaces, approvals, and commissioning tasks (European PPP Expertise Centre, 2014). Total Investment and Hospital Capacity (beds) are expected to **increase** Late probability.

- H3 (Tenor / non-linearity). The association between concession period and the probability of delay is non-linear (non-monotone), reflecting selection/bankability and complexity mechanisms at different tenor ranges (The World Bank, 2019). Concession Period is modelled with a **squared** term to test a non-linear association with Late (Global Infrastructure Hub (GI Hub), 2020). *Tenor effect is not assumed linear; Concession and Concession^2 are pre-specified.*

- H4 (Development level). Higher-income contexts are expected to show lower delay probability on average because administrative capacity and supply-chain depth tend to be stronger. (Organisation for Economic Co-operation and Development, 2012). GDP per capita (preferred) or GDP (robustness) is expected to have a **negative** association with Late once governance and project characteristics are controlled for (The World Bank, 2024). *Development is a control channel to be tested, not assumed as a narrative fact.*

2.3 Operational definitions and variable mapping

This section defines outcomes and predictors exactly as used later so that the empirical chapter reads as an implementation of the theory rather than a new set of choices. Delay is measured against planned baselines because audit and portfolio evidence treats baseline slippage as the practical definition of schedule underperformance (European Court of Auditors (ECA), 2018).

The operational definitions lock the research design; "Late" is a baseline-based concept, not a subjective assessment.

Outcomes (timeliness).

- Planned Total Construction Duration (Months). Baseline construction program used to define expected schedule at award/close.
- Actual Total Construction Duration (Months). Realised duration to completion/commissioning used to assess schedule performance.
- Project Delay (Months) = Actual − Planned. Absolute slippage used for descriptive profiling, not the core logit.
- Project Delay (%) = [(Actual − Planned)/Planned]. Relative slippage used for descriptive summaries to compare projects with different planned baselines.
- Late Indicator (1/0). Late = 1 if Actual > Planned, else 0 (early completion remains Late = 0), consistent with baseline-based delay definitions in procurement/execution evidence (The World Bank, 2024).

Core regression predictors (continuous Xs used in the logit).

- Concession Period (Years). Tenor capturing lifecycle obligations and handback structure, entered with a squared term to test non-linearity.
- Total Investment (USD millions). Scale proxy capturing capex magnitude and (indirectly) technical/managerial complexity in hospital delivery.
- Hospital Capacity (Beds). Clinical scale proxy reflecting operational commissioning burden and interface density.
- Planned Total Construction Duration (Months) (same column as in outcomes). In addition to defining the baseline, Planned duration is treated as an ex-ante schedule scale/control in the regression specifications.

Country-level predictors and controls (continuous)

- WGI Government Effectiveness. Proxy for public-sector implementation and service delivery capacity, used as an institutions channel.
- WGI Regulatory Quality. Proxy for the credibility of the regulatory environment relevant to private participation and contractual enforcement.
- Development level. GDP per Capita (Current US$), preferred, or GDP (Current US$), used as development proxy for macro capacity and resources, included one at a time to preserve parsimony.

Descriptive / group-based variables (NOT continuous; used for ANOVA/CHI or stratification).

- SPV – Number of Partners (grouped factor). Coded as categorical bins: ≤2, 3, 4, ≥5. Used for ANOVA-style comparisons / descriptive profiling, not as a continuous X.
- Contract form groups (categorical). Used for descriptive distributions and CHI/ANOVA checks; not part of the "continuous-X core logit".
- Country development status (categorical) and Developed dummy. Used for grouping/robustness checks rather than as a continuous covariate.

2.4 Measurement choices and transformations (Minitab-aligned).

To avoid repetition across chapters, all measurement choices and modelling transformations (z-standardisation, quadratic concession specification, parsimony, diagnostics, capped removals, and the restricted robustness set) are defined once in Chapter 3. Chapter 4 then applies these pre-specified choices to the dataset and reports results in the same order as the empirical design.

Chapter 3 - Empirical strategy and statistical approach

3.1 Purpose of statistical analysis in this thesis.

Empirical research relies on observed data to describe patterns and to evaluate whether such patterns are consistent with the conceptual framework. In this thesis, statistical analysis serves two complementary purposes. First, it provides a descriptive profile of schedule outcomes and contextual contrasts across key project groupings. Second, it supports an inferential model aligned with the research question, framed as timeliness risk in healthcare PPP projects.

A central constraint is the limited inferential sample size and the heterogeneity of project and country contexts. The empirical strategy is deliberately structured to remain interpretable and stable. The design is on two layers: (i) a descriptive layer focused on continuous delay measures (delay magnitude), and (ii) a core inferential layer focused on a binary outcome capturing the probability that a project is Late. Chapter 4 details operational definitions, sample construction rules, and variable measurement choices; this chapter focuses on the statistical logic of the tools used and on how results are interpreted.

3.2 Descriptive, explanatory, and predictive modelling: what applies here.

Statistical models can be used for different objectives, descriptive, explanatory, or predictive. (Shmueli G. , 2010) Descriptive modelling aims to summarise and represent the data clearly, often through tables, graphical tools, and simple group contrasts. Explanatory modelling applies statistical models to test whether observed associations are consistent with theoretically motivated channels. Predictive modelling focuses on forecasting or predicting outcomes for new observations and is not the primary objective of this thesis. (Shmueli G. , 2010)

This distinction clarifies the role of each statistical tool used. Continuous measures of slippage (such as delay_months and delay_pct) are primarily used for descriptive profiling and for group-based comparisons that contextualise the data. In contrast, the main explanatory question is whether key channels discussed in Chapters 1-2 are systematically associated with the probability that a project is Late, conditional on other factors. Since the inferential target is binary, the coherent modelling framework for the core analysis is logistic regression (The Pennsylvania State University, s.d.).

3.3 Descriptive group evidence on delay magnitude: one-way ANOVA and Kruskal-Wallis

To complement the binary Late analysis, the thesis reports group-based comparisons on delay magnitude, focusing on delay_pct as the primary continuous measure for comparability across projects with different planned durations. Two categorical groupings are prioritised because they map directly into the conceptual channels developed in Chapters 1–2:

- Contract-form grouping (project_type_3_group), interpreted as a proxy for contractual architecture and interface complexity (risk allocation and coordination channels).
- Country development status (developed_1_0), interpreted as a broad proxy for baseline context capacity.

These analyses are interpreted as descriptive evidence: they document whether delay magnitude differs across groups within the observed sample, but they are not presented as causal identification.

One-way ANOVA (mean-based group comparison).

One-way ANOVA is used to test a global hypothesis on mean differences in delay magnitude across group categories. Let y_i denote the continuous delay measure (delay_pct) for project i, and let μ_g be the mean of group $g = 1, \dots, k$. The null hypothesis is $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$, against the alternative that at least one group mean differs (NIST/SEMATECH, s.d.).

ANOVA is appropriate when the dependent variable is continuous and the independent grouping variable is categorical with two or more levels. In this thesis, ANOVA is used as a compact, global comparison tool, avoiding an emphasis on numerous pairwise comparisons that can inflate Type I error, particularly in small, unbalanced samples (NIST/SEMATECH, s.d.).

ANOVA relies on (i) independence of observations, (ii) approximately equal within-group variances, and (iii) residual behaviour that is not extremely inconsistent with a mean-based comparison (NIST/SEMATECH, s.d.). Given that delay measures can be skewed and may exhibit heavy right tails, ANOVA results are treated as descriptive and are complemented by the robustness checks described below, In Chapter 4, ANOVA outputs are reported in a transparent set: group Ns, group means (and medians for context), the F statistic with degrees of freedom and p-value, and a compact measure of explained variance R^2 (Levine & Hullett, 2002).

Equal-variance check (Levene) and distribution-robust confirmation (Kruskal-Wallis).

To screen for heteroscedasticity that could distort mean-based inference, the thesis reports an equal-variance check using Levene's test (NIST/SEMATECH, s.d.). This test evaluates whether within-group dispersion differs across categories; it is used here as an assumption screen rather than as a substantive result.

Because delay measures can be non-normal and influenced by outliers, the thesis also reports Kruskal-Wallis as a distribution-robust confirmation. Kruskal-Wallis tests whether group distributions (often summarised in terms of central tendency on ranks) differ systematically without relying on mean-based normality assumptions (NIST/SEMATECH, s.d.). Kruskal-Wallis H statistic (with degrees of freedom and p-value), along with group medians and mean ranks, are reported in Chapter 4. This design ensures that descriptive group evidence remains informative while being robust to skewness and extreme tails.

Interpretation discipline. Across both ANOVA and Kruskal-Wallis, results are interpreted as evidence (or lack of evidence) of group differences within the observed sample. Non-significant results are reported as "no statistical evidence of differences detectable with this sample", rather than as proof of equality (Wasserstein, Ronald L.; Lazar, Nicole A. , 2016).

3.4 Core inferential model: logistic regression for Late

The core inferential object is the conditional probability of being Late. For project i, the model is $P(Late_i = 1 \mid X_i) = \frac{1}{1+\exp[-(\beta_0 + X_i'\beta)]}$, where $Late_i$ is a binary indicator and X_i is the vector of predictors. In this thesis, Late is a binary indicator defined from planned versus actual construction duration (see Chapter 4 for operational construction). Logistic regression is the coherent GLM for binary outcomes because it models the log-odds of the event and constrains fitted probabilities to the $[0, 1]$ interval (The Pennsylvania State University, s.d.).

Parsimony and standardisation (OR per one standard deviation).

Given the limited number of projects with non-missing planned and realised durations, the baseline specification is kept deliberately parsimonious to reduce the risk of overfitting and unstable estimates (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). Alternative specifications and variable swaps are confined to the pre-specified robustness section.

Continuous predictors are Z-standardised prior to estimation (Gelman, 2008). Standardisation serves two purposes: (i) it makes predictors comparable despite heterogeneous units (years, USD, beds, indices), and (ii) it allows effect sizes to be reported consistently as odds ratios (OR) for a one standard deviation increase in each predictor.

Coefficients are reported as odds ratios ($\mathrm{OR}_j = \exp(\beta_j)$) and interpreted as multiplicative changes in the odds of being Late associated with a one standard deviation increase in a predictor, holding other covariates constant. Confidence intervals and p-values are reported, but interpretation remains anchored in conditional association rather than causal claims, especially given correlated country-level controls (University of California, Los Angeles (UCLA) – Office of Advanced Research Computing (OARC), s.d.) (The Pennsylvania State University, s.d.).

Statistical-power constraints (events-per-variable).

The inferential sample contains 22 Late events out of 43 projects in the baseline (21 in the adjusted no35 sample), against a core specification with one intercept and eight standardised predictors plus a quadratic term, for a total of nine to ten estimated parameters. The resulting events-per-variable (EPV) ratio is approximately 2.3, materially below the conventional threshold of 10 (Peduzzi et al., 1996) and substantially below more recent recommendations for stable maximum-likelihood estimation in logistic regression (van Smeden et al., 2019; Riley et al., 2020). Estimates should be read as conditional associations under high statistical uncertainty, not as structural effects. Robustness checks (Section 4.6) document the directional stability of the main signal under controlled perturbations, but they cannot compensate for limited power; this thesis privileges qualitative coefficient patterns over precise magnitude claims.

Two further design choices warrant explicit motivation. First, the logit link is preferred over probit because the resulting coefficients map directly into odds ratios, the dominant effect-size metric in applied PPP and infrastructure research; substantive conclusions are unaffected by the choice. Second, the capped-removals rule allows at most two diagnostic removals: this cap is set ex-ante to limit the practical risk of iterative cleaning driven by significance rather than diagnostics, while still permitting a transparent sensitivity test for the most extreme cases.

Pre-specified non-linearity in concession tenor.

To operationalise the pre-specified non-linearity in concession tenor, the baseline logit includes both the standardised linear term and its square. Tenor is Z-standardised as $Z\_concession_i = (concession\_period\_years_i - \bar{c})/s_c$, and the quadratic term is constructed as $Z\_concession\_sq_i = (Z\_concession_i)^2$. The binary logit is evaluated through the linear predictor $\eta_i = \beta_0 + \beta_1 Z\_concession_i + \beta_2 Z\_concession\_sq_i + X_i'\gamma$, and mapped into predicted probabilities using $p_i = 1/(1 + \exp{(-\eta_i)})$, which constrains fitted values to the [0, 1] interval. (Rundel, 2013)

The sign of the squared term determines curvature: $\beta_2 > 0$ implies a convex (U-shaped) pattern, whereas $\beta_2 < 0$ implies a concave (inverted-U) pattern. The turning point on the z-scale can be summarised by $z^{\backslash *} = -\beta_1/(2\beta_2)$) (Henry, 2024). For visual interpretation, predicted probabilities are computed along a grid of tenor values while holding all other standardised covariates at zero (i.e., their sample means) (Torres-Reyna, 2011).

Comparing the linear and quadratic specifications supports treating concession tenor as a non-linear channel: adding $Z^2_{concession}$ improves fit and yields an interpretable curvature. Substantively, the estimated pattern in this sample is concave, with the highest predicted delay risk around an interior tenor close to the sample mean, rather than monotonically increasing/decreasing across tenor. *This specification allows the association between contract duration and delay risk to be non-monotone, consistent with theoretical arguments related to selection, bankability constraints, and project complexity at different tenor ranges. Functional-form sensitivity (quadratic versus linear-only) is treated as a controlled robustness check and is reported in Chapter 4.*

Limited adjustments: capped removals (maximum two).

If diagnostics flag highly influential observations, a capped adjustment strategy (remove at most two observations) limits "specification shopping" while allowing a transparent sensitivity check. Each removal is documented with the diagnostic trigger and with stability checks on key ORs and fit metrics, consistent with audit-style emphasis on documenting performance issues and deviations (European Court of Auditors , 2018). *Removals are a documented robustness step; stability, not significance, is the goal.*

Robustness plan (pre-specified swaps and skewness handling).

Robustness checks are restricted to controlled swaps aligned with the database: GDP per capita (current US$) (preferred) versus GDP (current US$), and (if needed) the use of LOG(GDP) to address right-skewness in scale variables (The World Bank, 2024) (International Monetary Fund; World Bank Group, 2019). These checks are consistent with fiscal-risk tool guidance that explicitly includes macro shock and sensitivity analysis for PPP portfolios variables (The World Bank, 2024) (International Monetary Fund; World Bank Group, 2019).

3.5 Diagnostics and robustness plan (pre-specified and audit-style)

To make modelling choices auditable and to protect inference from a small number of extreme observations, the thesis includes diagnostics and limited adjustments, combined with controlled robustness checks. Influence screening is used to identify observations that disproportionately affect parameter estimates (National Institute of Standards and Technology (NIST), 2023). When an observation is flagged as highly influential, the thesis applies a capped-removals rule (maximum two removals) to limit specification shopping while allowing transparent sensitivity assessment. The main adjusted specification excludes Project ID 35 under this rule (Chapter 4 reports the trigger and consequences).

A pre-emptive identification caveat applies to the country-level block. GDP per capita, WGI Government Effectiveness, and WGI Regulatory Quality move together strongly across the project sample (the Chapter 4 correlation matrix and VIFs make this explicit), so in a small-N cross-country setting the data cannot decompose this shared signal into independent "institutional" and "developmental" effects. Throughout the chapter, the country block is treated as a jointly identified country-context proxy: individual coefficients are reported and discussed conditionally on the chosen parametrisation, but the substantive claims rest on the stability of the project-level signals (notably the concession non-linearity), not on the magnitude of any specific institutional or macro coefficient. Robustness checks are restricted to theoretically motivated swaps aligned with the database structure:

- baseline versus adjusted sample (N = 43 vs N = 42),
- functional form sensitivity for concession tenor (quadratic versus linear-only),
- macro proxy swap (GDP per capita versus GDP),

- stability under collinearity using alternative country-control blocks.

Interpretation remains disciplined: coefficients for institutional and macro variables are read as conditional associations in a correlated control matrix, whereas the pre-specified concession non-linearity is treated as a core structural element whose stability and interpretability are central (The Pennsylvania State University, s.d.). Chapter 4 reports results in the same order as this empirical design: descriptive profiling and group evidence first, followed by the core logit estimates and the restricted robustness architecture.

Chapter 4 - Data and results: operational definitions, sample design and empirical findings

4.1 Dataset and sample definition

Unit of analysis and scope.

The unit of analysis is the hospital PPP project, with one record per *Project Identifier (project_id)*. The master database initially includes 56 projects and is designed to support a cross-country assessment of timely delivery in hospital PPPs. Because hospital PPP information is not standardised globally, project records exhibit heterogeneous completeness across fields and across delivery stages. A subset of projects was explicitly reported as "Under Construction" at the time of dataset compilation (around May-June 2025), implying that realised construction duration is not yet observable for those projects. These records are retained in the master dataset for traceability but are excluded from outcome construction and any inference requiring an observed completion duration. *The dataset is project-level and global, but inferential analysis must rely on projects with observed completion information; "Under Construction" projects are part of the database footprint but cannot enter timeliness outcomes.*

Data sources and merging logic.

The database integrates two information blocks: project-level variables and country-level indicators. Project-level information (planned and actual construction duration, concession tenor, total investment, and hospital capacity) was compiled using a mixed-source strategy, combining specialised PPP repositories with complementary public and local sources (including institutional publications and press reporting) when needed to complete key fields. Given that planned and actual dates may differ across sources in definition or reporting, schedule fields were subject to a conservative credibility screen to avoid outcome misclassification.

Country-level covariates were collected consistently from World Bank datasets: governance indicators from the Worldwide Governance Indicators (WGI) Government Effectiveness and Regulatory Quality and macro controls from the World Development Indicators (WDI) GDP and GDP per capita, both in current US$. Country-level indicators are aligned to the project start year, defined as the year of the actual construction start date, so that covariates capture the institutional and macro context prevailing at the onset of delivery rather than ex-post conditions.

Core inferential sample: inclusion/exclusion rules.

The core inferential sample is defined by transparent inclusion criteria that ensure both construct validity of the dependent variable and comparability across projects. A project is included in the baseline inferential sample only if both *planned_construction_months* and *actual_construction_months* are observed and sufficiently reliable to be treated as consistent across sources, since these fields are necessary to construct *Late* and the continuous delay measures. Starting from 56 projects, exclusions were applied as follows. First, five projects were explicitly labelled Under Construction at dataset compilation; as completion is not observed, realised duration cannot be established and outcomes cannot be constructed. Second, two projects were excluded because hospital capacity (*hospital_beds*) could not be determined with sufficient clarity, which would compromise the scale/complexity channel in the core specification. Third, six projects were excluded when schedule information (planned and/or actual start/end) was judged inconsistent or uncertain across sources, creating a material risk that measurement error would dominate the outcome definition.

After applying these rules, the baseline inferential sample comprises N = 43 projects with complete and credible planned and realised duration data. The main adjusted specification used for inference excludes one observation (Project ID 35) identified as highly influential in diagnostic screening, yielding N = 42 (Excluded Project ID 35). The country coverage for the inferential sample spans 14 developed, 28 developing, and 1 least developed country contexts, and the start-year alignment window runs from 2005 to 2018. The 42 projects in the main adjusted sample are distributed across a small number of unique countries; country frequencies are concentrated in a few jurisdictions (notably Spain, the United Kingdom, Italy, Australia, Canada and Portugal), with the remaining projects scattered across countries with one or two observations each. Country-level statistical power is therefore concentrated in a handful of jurisdictions, and country-level coefficients should be read as conditional associations within an effectively clustered design. Classical standard errors may underestimate uncertainty in the country block, although the substantive caution embedded in the chapter, treating institutional and macro coefficients as conditional allocations of shared cross-country variation, is consistent with this clustering structure.

Because Late is mechanically defined from planned vs actual duration, data credibility screens are not optional: they protect the dependent variable from misclassification. The thesis privileges a smaller but internally consistent inferential sample (N=43; main adjusted N=42) over broader but noisy coverage.

Table 2 - Sample flow and analysis samples

Sample step	**Criterion / reason**	**N (projects)**
Master database	Hospital PPP projects collected in the master dataset	56
Excluded	Projects explicitly labelled "Under Construction" at dataset compilation (≈ May–June 2025), hence no observed realised duration	5
Excluded	Missing / unclear hospital capacity (beds) information	2
Excluded	Uncertain or inconsistent schedule information across sources (planned and/or actual dates/durations not reliable)	6
Baseline inferential sample	Projects with credible Planned and Actual construction duration (outcomes constructible)	43
Main adjusted sample (Excluded Project ID 35)	Baseline inferential sample excluding Project ID 35 flagged as highly influential in diagnostics	42

Data harmonisation and reporting conventions.

All schedule variables are harmonised in months. Timeliness is evaluated against the planned baseline:

> Late = 1 when *actual_construction_months* > *planned_construction_months* and Late = 0 otherwise.

The zero category includes both on-time and early completion. Monetary values for project scale (total_investment_usd_m) are expressed in USD (millions); where original sources reported investment in local currency, amounts were converted to USD using the annual exchange rate of the actual project start year, to preserve temporal coherence. Country macro indicators (GDP and GDP per capita) are taken directly in current US$ as provided by WDI. All descriptive statistics and model outputs are reported to two decimals, including odds ratios, confidence intervals, p-values, and model fit/discrimination metrics. This convention ensures consistent comparability across sections and tables. Harmonisation choices are fixed up front (months, USD, current US$, two decimals) so that later results read as implementation rather than ad-hoc adjustments.

4.2 Variable construction and measurement choices

Outcome variables.

As defined in Section 4.1, timeliness is operationalised through a baseline-based comparison of realised vs planned construction duration. The core dependent variable is Late (1 if actual_construction_months exceeds planned_construction_months, 0 otherwise), complemented by continuous delay measures to describe slippage magnitude. To complement the binary outcome with a magnitude-based description of slippage, two continuous delay measures are computed:

- delay_months = *actual_construction_months − planned_construction_months*
- delay_pct = *(actual_construction_months − planned_construction_months) / planned_construction_months*

These continuous measures are used for descriptive profiling of the distribution of delays and (where reported) contextual comparisons. The core inferential analysis, however, is conducted on Late (1/0) to match the research question and hypotheses formulated in Chapter 2. A construct overlap requires explicit acknowledgement: Late is mechanically defined from the comparison of Actual and Planned construction duration (Late = 1 if Actual > Planned), and Planned construction duration is also retained as a baseline-scale predictor in the core logit. The model estimates the conditional probability that Actual exceeds Planned given Planned itself (and other covariates); this is a defensible specification, because Planned captures programme-scale information correlated with interface burden (longer planned programmes typically reflect larger and more complex deliveries) and the residual variation in Actual that is independent of Planned remains substantively meaningful. The choice is not without trade-offs, however: an alternative specification that excludes Planned from the predictor set is conceptually cleaner and would isolate the effect of project- and country-level characteristics on the probability of overrun without absorbing schedule scale. The construct overlap is documented here so that the OR of Planned (3.22, $p = 0.11$) is read as a baseline-scale association and not as a stand-alone driver of delay; an excluding-Planned sensitivity check is a natural extension on a larger sample.

Predictor set used in the core model.

The baseline specification uses a single multivariate binary logistic regression with a set of continuous predictors, aligned with the channels developed in Chapters 1–2 (institutions, scale/complexity, baseline realism, and macro capacity). The core predictor set includes:

- Concession tenor (*concession_period_years*)
- Project scale (*total_investment_usd_m*, *hospital_beds*)
- Baseline schedule control (*planned_construction_months*)
- Institutional quality (*wgi_ge_est*, *wgi_rq_lb_90pct*)
- Macro capacity control (mutually exclusive): GDP per capita (*gdp_per_capita_current_usd*) in the core model, with GDP (*gdp_current_usd*) used only in robustness checks

All continuous predictors entering the logistic model are z-standardised and the concession non-linearity is implemented via Z_concession_sq = $(\text{Z\_concession})^2$, following the specification described in Chapter 3.4. Section 4.4 reports the corresponding estimates and interpretation.

Table 3 - Variable dictionary and analysis role (revised, no leakage)

Field (dataset / Minitab)	**Description**	**Unit**	**Construction / Transform**	**Role**
project_id	Project identifier	–	Unique key	Identifier
planned_construction_months	Planned total construction duration	Months	Raw field	Outcome input / baseline control
actual_construction_months	Actual total construction duration	Months	Raw field	Outcome input
late	Late indicator (1/0)	Binary	1 if Actual > Planned; else 0	Primary outcome
delay_months	Absolute delay	Months	Actual − Planned	Descriptive outcome
delay_pct	Relative delay	Percent	(Actual−Planned)/Planned	Descriptive outcome
concession_period_years	Concession period	Years	Raw field	Core predictor (tenor)
total_investment_usd_m	Total investment	USD (millions)	Converted to USD using annual FX rate	Core predictor (scale)

hospital_beds	Hospital capacity	Beds	Raw field	Core predictor (scale/complexity)
wgi_ge_est	WGI Government Effectiveness	Index	World Bank WGI	Core predictor (institutions)
wgi_rq_lb_90pct	WGI Regulatory Quality (LB 90% CI)	Index	World Bank WGI	Core predictor (institutions)
gdp_per_capita_current_usd	GDP per capita (current US$)	USD	World Bank WDI	Core macro (preferred)
gdp_current_usd	GDP (current US$)	USD	World Bank WDI	Robustness-only macro
Z_concession	Standardised concession	SD units	Z-score of concession_period_years	Core model regressor
Z_concession_sq	Quadratic tenor term	SD units²	(Z_concession)²	Core model regressor
Z_invest	Standardised investment	SD units	Z-score of total_investment_usd_m	Core model regressor

Z_beds	Standardised beds	SD units	Z-score of hospital_beds	Core model regressor
Z_planned	Standardised planned duration	SD units	Z-score of planned_construction_months	Core model regressor
Z_wgi_ge	Standardised WGI GE	SD units	Z-score of wgi_ge_est	Core model regressor
Z_wgi_rq	Standardised WGI RQ	SD units	Z-score of wgi_rq_lb_90pct	Core model regressor
Z_gdppc	Standardised GDP per capita	SD units	Z-score of gdp_per_capita_current_usd	Core model regressor
Z_gdp	Standardised GDP	SD units	Z-score of gdp_current_usd	Robustness-only regressor

Table 3 reports only variables that are used in the core analysis, robustness swap, or diagnostics; ex-post or unused fields are documented separately in the Appendix for transparency. Appendix Table 13 summarises these variables and states why each was retained in the dataset but excluded from the core specification, and where it is used (if anywhere).

Figure 1 - Concession tenor distribution (years), baseline inferential sample (N = 43)

The boxplot summarises the distribution of concession tenor in the baseline inferential sample (N = 43). The median is 30.00 years, the mean is 28.86 and the standard deviation is 5.41, with an interquartile range roughly between 25 and 33 years, indicating moderate dispersion around a long-term tenor structure.

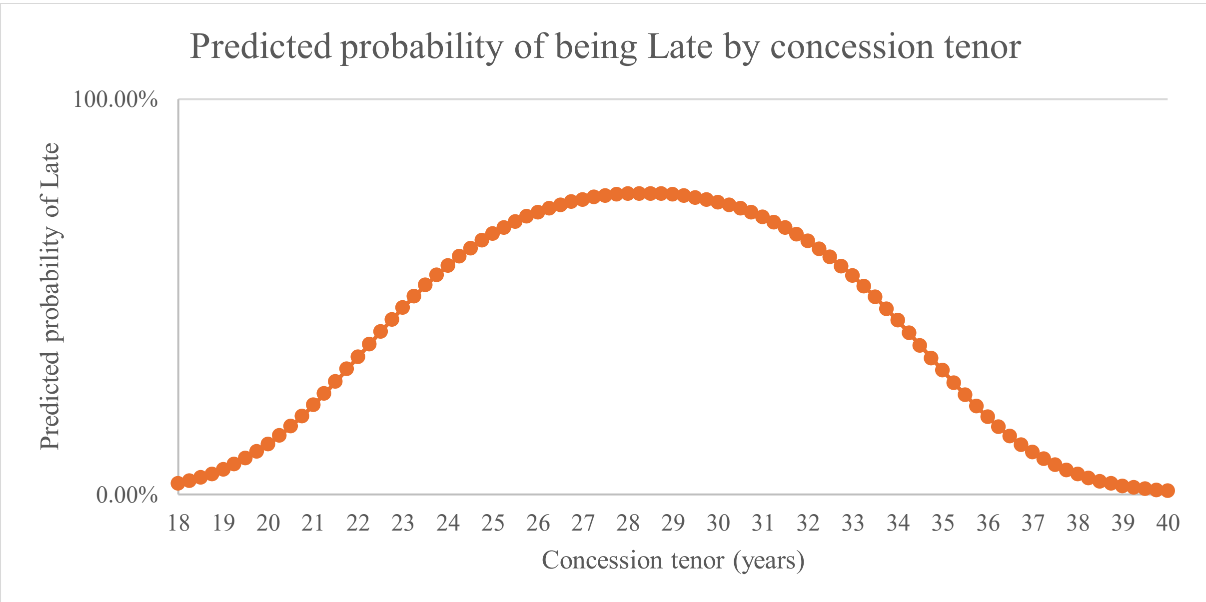


Figure 2 - Predicted probability of being Late by concession tenor

The curve plots predicted probabilities from the baseline binary logit including a quadratic tenor term. Concession tenor is varied across a grid of values, while all other standardised *covariates are held at zero (their sample mean). The estimated pattern is concave (inverted-U), with predicted delay risk peaking around an intermediate tenor close to the sample mean.*

4.3 Descriptive statistics (core sample)

Descriptive profile of the inferential sample.

This section provides a compact quantitative profile of the baseline inferential sample (N = 43) used to estimate the core logistic regression. The sample exhibits substantial heterogeneity in project scale and context, consistent with the cross-country scope. Planned construction duration averages 38.42 months (median 37.00), while realised duration averages 44.95 months (median 40.00), indicating that completion time is typically longer than the baseline schedule. In absolute terms, the median delay is 1.00 month and the interquartile range is 10.00 months, while the relative delay measure (*delay_pct*, stored as a proportion) has a mean of 0.17 (≈ 17%) and a median of 0.03 (≈ 3%). The probability of late delivery is material: 22 out of 43 projects are Late (51.16%).

Project characteristics show wide dispersion. Concession tenor is centred around long-term arrangements (mean 28.86 years, median 30.00). Scale proxies are highly dispersed: total investment averages 598.02 USD m (median 399.80 USD m) and hospital capacity averages 771 beds (median 651), with broad interquartile ranges. Country-level context also varies: GDP per capita averages 29,275.47 current US$ (median 32,459.92) and governance indicators are relatively high on average (WGI Regulatory Quality mean 76.00; WGI Government Effectiveness mean 1.18), while still displaying meaningful dispersion across projects.

The inferential sample (N=43) combines wide cross-project heterogeneity with a non-trivial late prevalence (51.16%), which supports the use of a multivariate model rather than purely descriptive comparisons.

Table 4 - Descriptive statistics (baseline inferential sample, N = 43)

(All values reported to two decimals; IQR = Q3 − Q1. delay_pct is stored as a proportion: 0.17 corresponds to 17%.)

Variable	N	Mean	SD	Median	IQR

concession_period_years	43	28.86	5.41	30.00	7.50
total_investment_usd_m	43	598.02	592.61	399.80	363.63
planned_construction_months	43	38.42	18.19	37.00	15.00
actual_construction_months	43	44.95	22.42	40.00	24.50
delay_pct	43	0.17	0.33	0.03	0.25
delay_months	43	6.53	12.46	1.00	10.00
hospital_beds	43	771.40	703.09	651.00	546.00
gdp_per_capita_current_usd	43	29275.47	13079.92	32459.92	21059.67
wgi_rq_lb_90pct	43	76.00	17.58	78.43	19.88
wgi_ge_est	43	1.18	0.65	1.50	0.79
Late indicator (1 = Late)	43	**0.51**	0.51	1.00	1.00

Table 4 summarises the baseline inferential sample used throughout the empirical analysis; dispersion is substantial in both project scale and country context, while Late prevalence is close to one-half of projects.

Minimal visuals aligned to the research question.

To keep the descriptive layer aligned with the research question (timeliness risk), only two visuals are reported. First, the distribution of planned vs actual construction duration is shown to illustrate baseline realism against realised delivery. Second, the overall Late share is reported as a single summary measure (without stratification), to avoid introducing group narratives that are not part of the core inferential design.

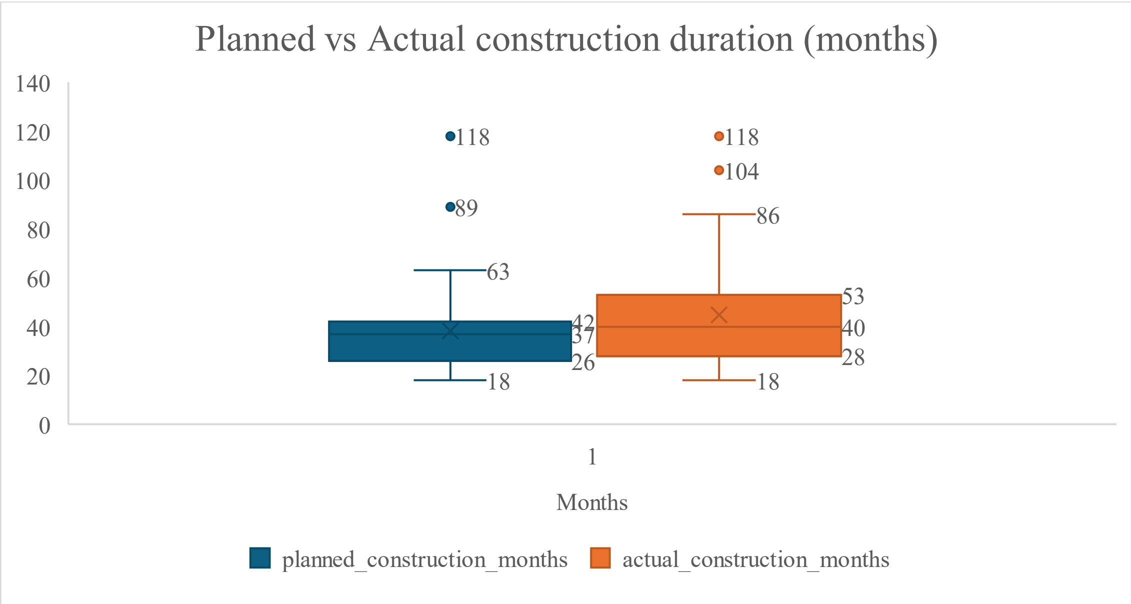


Figure 3 - Planned vs actual construction duration (months), baseline inferential sample (N = 43)

Boxplots compare planned and realised construction duration (months) for the baseline inferential sample. The distribution of actual duration is shifted upward relative to planned duration, consistent with a positive delay distribution.

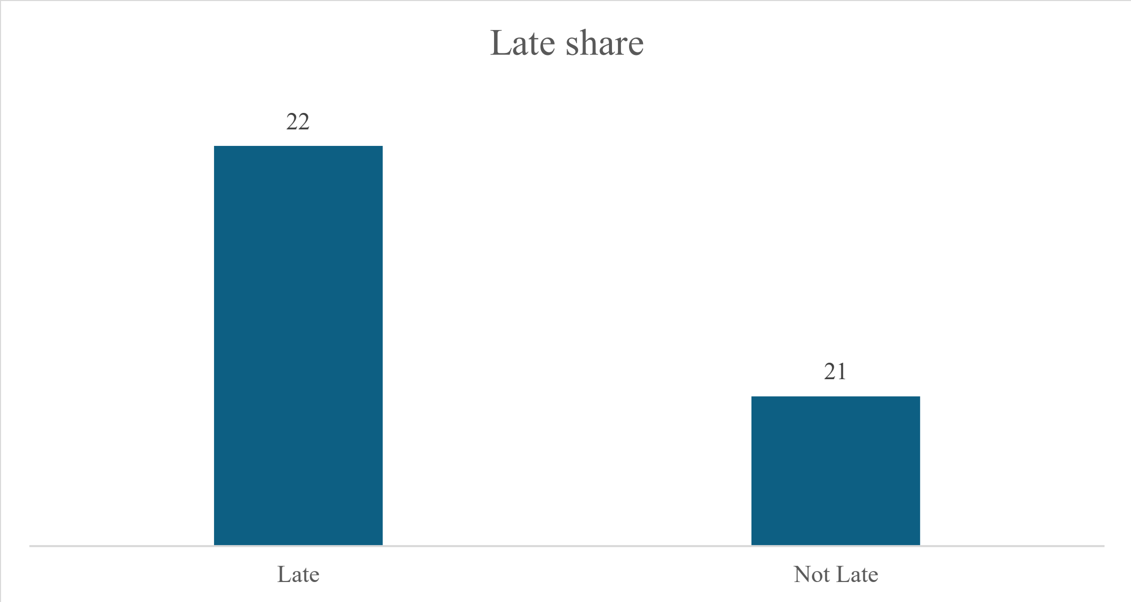


Figure 4 - Late share (overall), baseline inferential sample (N = 43).

Overall, 22 out of 43 projects (51.16%) are classified as Late. This prevalence motivates modelling timeliness as a binary risk outcome in the core logistic regression.

Group-based comparisons on delay magnitude (delay_pct).

Contract-form groupings are examined as descriptive evidence on delay magnitude. Using delay_pct (stored as a proportion and reported as percentages for readability), both mean-based and distribution-robust comparisons indicate no statistically detectable differences across contract-form categories in this sample. The one-way ANOVA does not reject equality of mean delay_pct across groups ($F(2,40)=1.62$, $p=0.211$; $N=43$). An equal-variance screen does not indicate heteroscedasticity (Levene $p=0.255$), supporting the standard ANOVA specification. The Kruskal-Wallis test (adjusted for ties) does not reject equality across group distributions ($H=4.52$, $df=2$, $p=0.104$).

Descriptively, DBFM exhibits higher delay magnitude (median delay_pct ≈ 10.4%) than DBFO and Other (both medians ≈ 0%), and its mean rank is also higher; however, group differences are imprecisely estimated given small and unbalanced group sizes (notably the "Other" category). These comparisons are interpreted as contextual descriptive evidence and do not alter the core inferential analysis of Late.

Table 5 - Delay magnitude by contract form (project_type_3_group)

project_type_3_group	N	Mean	SE Mean	StDev	Minimum	Median	Maximum
DBFM Group	24	0.25	0.08	0.40	0.00	0.10	1.81
DBFO Group	12	0.09	0.05	0.18	-0.11	0.00	0.55
Other	7	0.05	0.05	0.12	-0.03	0.00	0.32

Table 6 - Inferential tests for differences in delay magnitude across contract forms (delay_pct)

Test	Statistic(s) to report	df	p-value
One-way ANOVA	F = 1.62	(2, 40)	0.211
Levene (equal variances)	Levene = 1.42	-	0.255
Kruskal–Wallis	H = 4.52	2	0.104

To document distributional features that can distort mean-based inference in small and unbalanced groups, Figure 5a reports ANOVA residual diagnostics for delay_pct across contract-form groups. The plots show a heavy right tail driven by a small number of extreme positive residuals, implying departures from normality and sensitivity of mean-based inference to outliers. ANOVA is interpreted descriptively and complemented with the rank-based Kruskal–Wallis test.

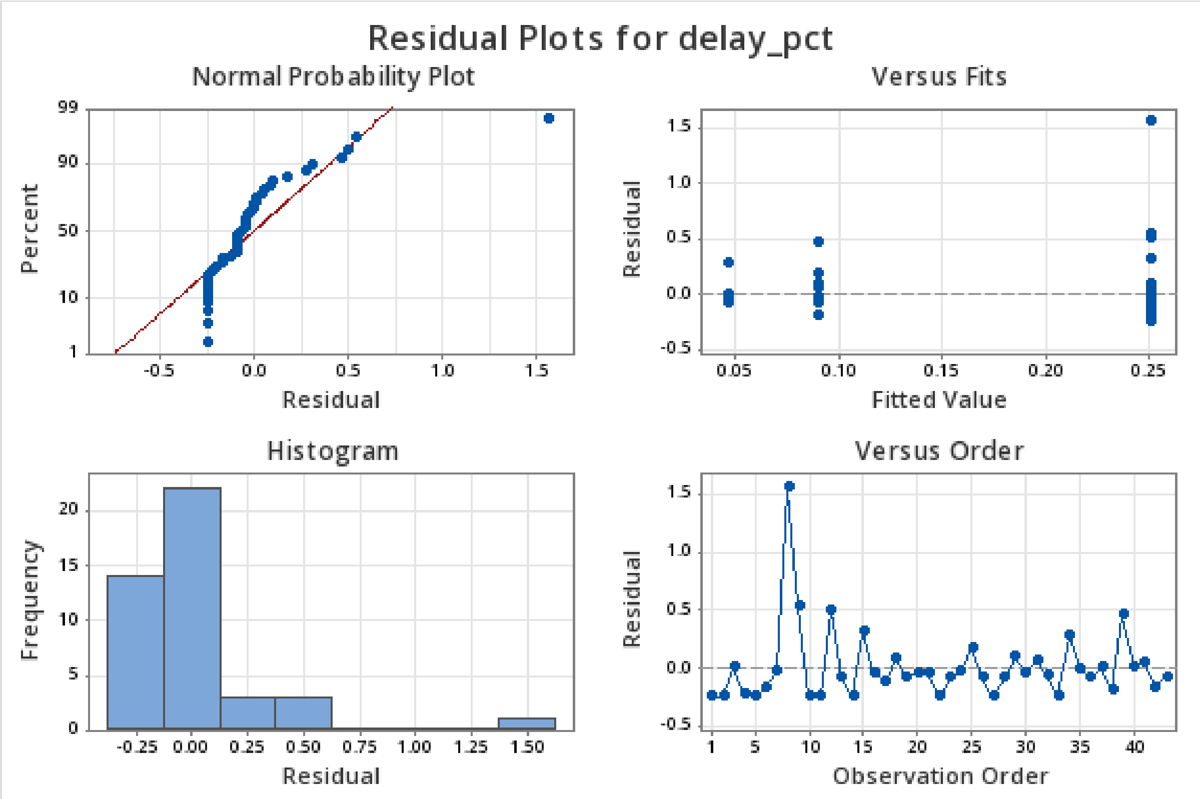


Figure 5 - ANOVA residual diagnostics for delay_pct (contract form groups).

The diagnostics suggest that the average delay can be pulled upward by a few extreme overruns. Group distributions largely overlap, and mean differences are sensitive to a small number of extreme overruns, in line with the non-significant ANOVA ($p = 0.211$) and Kruskal-Wallis ($p = 0.104$). To complement the mean-based tests, Figure 5 reports residual diagnostics (assumption screen), while Figure 6 visualises the distribution of delay_pct across contract-form groups, highlighting dispersion and outliers.

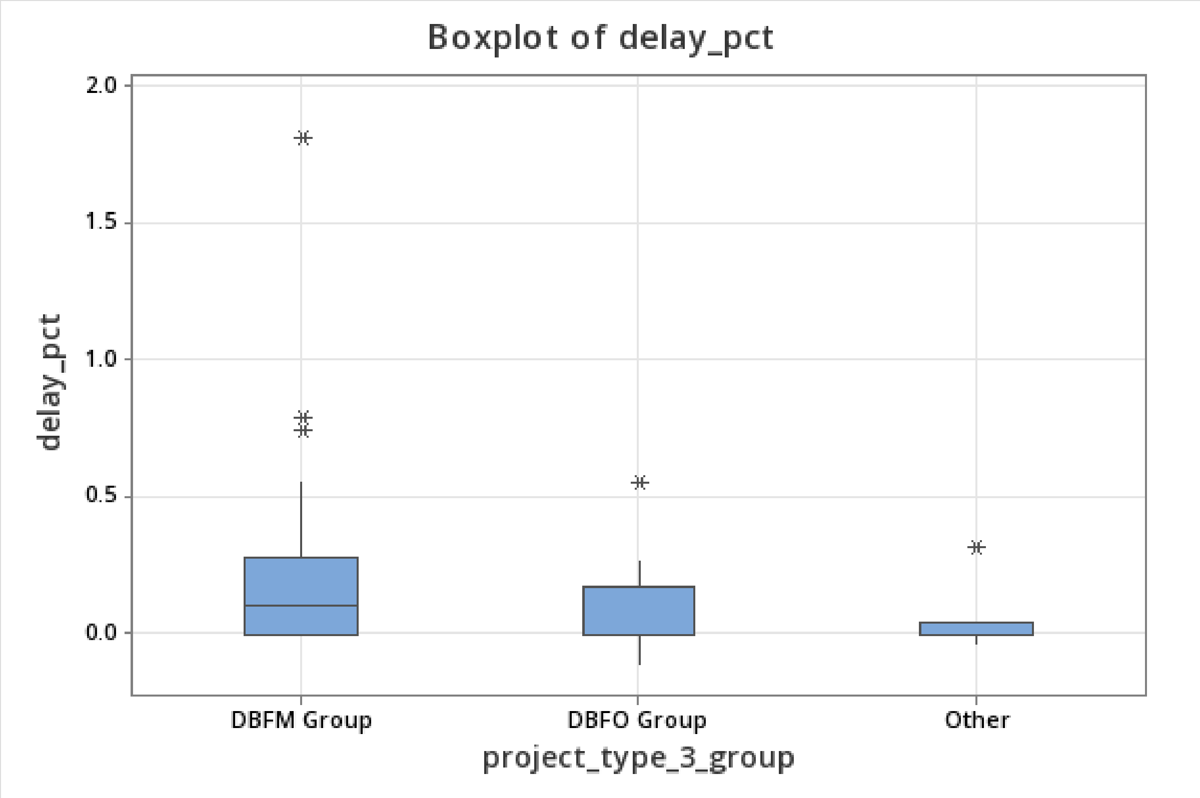


Figure 6 - Boxplot of delay_pct by contract form (project_type_3_group).

The distributions largely overlap, consistent with the non-significant ANOVA (p = 0.211) and Kruskal-Wallis evidence (p = 0.104). DBFM shows a higher median delay magnitude, but a small number of extreme overruns appears to drive mean differences.

Delay magnitude is also compared across development status as a descriptive proxy for broad context capacity. The one-way ANOVA does not reject equality of mean delay_pct across the two groups (F(1,40)=2.14, p=0.151; N=42), and the equal-variance check does not indicate statistically detectable heteroscedasticity (Levene p=0.150). The distribution-robust Kruskal-Wallis test provides a consistent conclusion (adjusted for ties: H=0.09, df=1, p=0.770), with very similar mean ranks across groups.

While mean delay_pct is higher in the developed group, the median pattern and rank-based evidence indicate no systematic shift in the centre of the distribution, suggesting that mean differences may be influenced by skewness and extreme values rather than a stable group effect. One observation is excluded from this comparison due to missing development classification. *No statistical evidence of differences in delay magnitude by development status is detectable with this sample; the results are treated as descriptive context rather than as causal identification.*

Table 7 - Delay magnitude by development status (developed_1_0)

developed_1_0	**N**	**Mean**	**SE Mean**	**StDev**	**Minimum**	**Median**	**Maximum**
0	28	0.12	0.03	0.17	-0.11	0.05	0.56
1	14	0.28	0.14	0.52	0.00	0.01	1.81

One observation excluded due to missing classification: overall N=42.

Table 8 - Inferential tests for differences in delay magnitude by development status (delay_pct)

Test	**Statistic(s) to report**	**df**	**p-value**
One-way ANOVA	F = 2.14	(1, 40)	0.151
Levene (equal variances)	Levene = 2.16	-	0.150
Kruskal–Wallis	H = 0.09	1	0.770

These group comparisons are descriptive and complement the core logit analysis on Late.

Figure 7 reports ANOVA residual diagnostics for *delay_pct* by development status. Residuals are concentrated near zero but exhibit a pronounced right tail with an extreme observation, indicating non-normality and potential leverage on the mean. This supports placing emphasis on median/rank-based evidence (Kruskal–Wallis) alongside ANOVA.

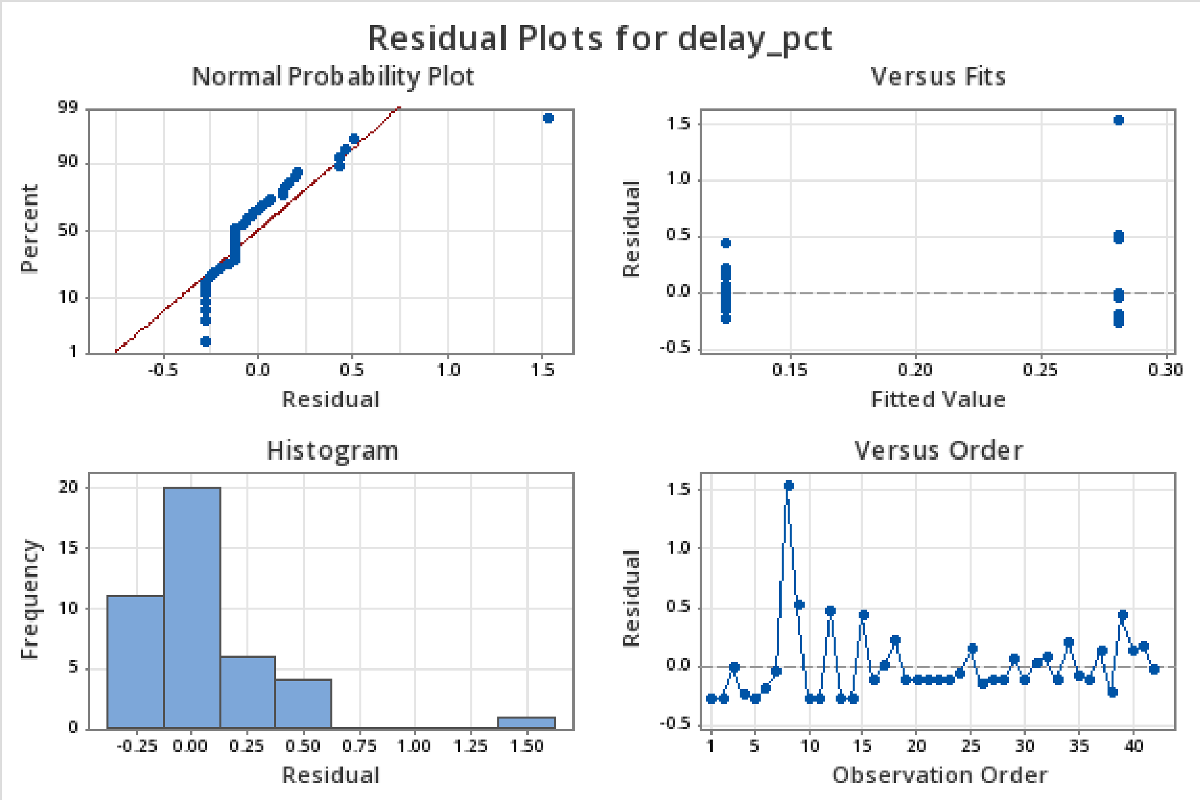


Figure 7 - ANOVA residual diagnostics for delay_pct (development status).

The plots help interpret why mean delay_pct differences may look larger than median differences: tail observations disproportionately affect the mean, while the central mass of residuals remains close to zero. This is consistent with the non-significant ANOVA ($p = 0.151$) and Kruskal–Wallis evidence ($p = 0.770$).

Figure 7 reports residual diagnostics (assumption screen), and Figure 8 provides a distributional view of delay_pct by development status to interpret mean vs median contrasts.

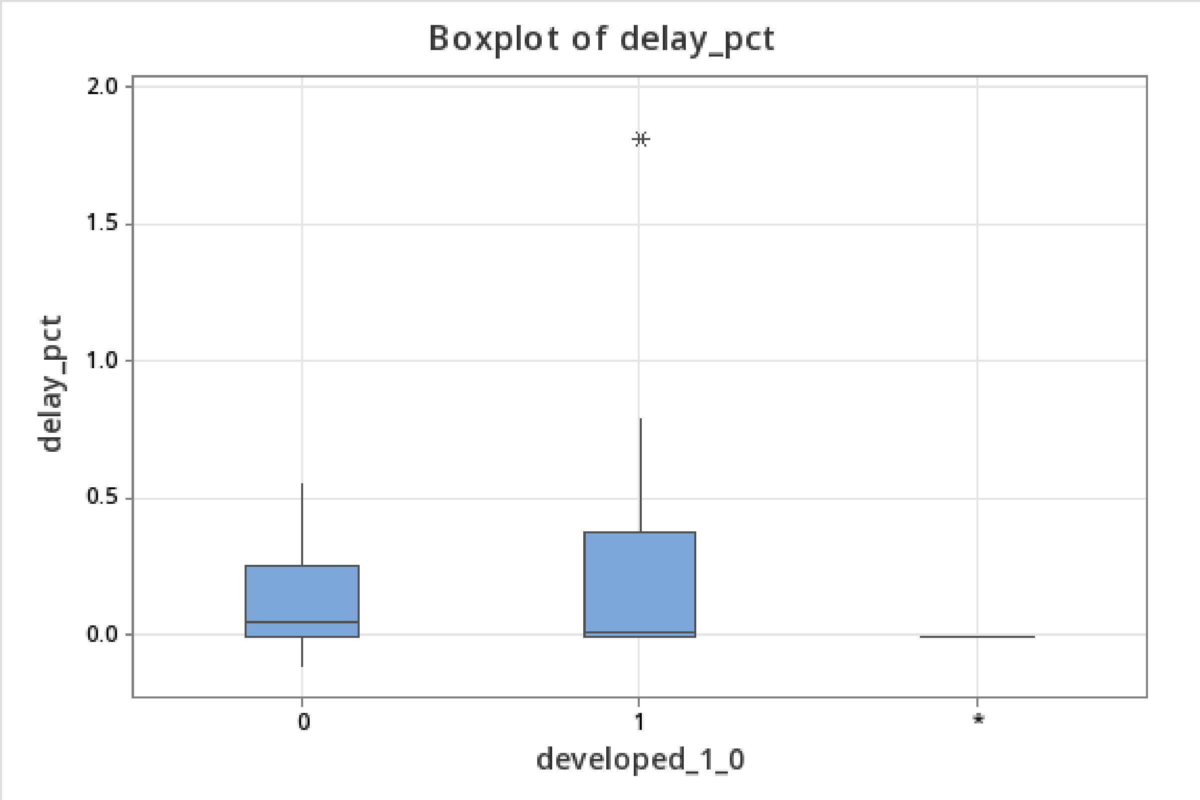


Figure 8 - Boxplot of delay_pct by development status (developed_1_0).

Notes: delay_pct is expressed as a proportion (1 = 100%). Boxes represent the interquartile range with median line; whiskers extend to 1.5×IQR; markers denote outliers. One observation with missing group classification is excluded from group tests (N = 42).

Medians are similar while dispersion differs, suggesting that tail observations rather than a systematic shift drive mean differences. This is consistent with Kruskal–Wallis (p = 0.770) and the non-significant ANOVA (p = 0.151).

4.4 Core inferential model: multivariate binary logistic regression (Late)

Model specification: primary inferential design.

The primary inferential model is a multivariate binary logistic regression where the dependent variable is Late (1/0). The model estimates the probability that a hospital PPP project exceeds its planned construction duration, conditional on a parsimonious set of project-level and country-level covariates. The predictor set is restricted to continuous covariates and estimated on the main adjusted sample (excluded project ID 35, N=42), to preserve parsimony and comparability across observations. All predictors included in the core model are continuous and Z-standardised prior to estimation. Reported effects are odds ratios per one standard deviation increase in each predictor, holding other covariates constant. To avoid post-treatment bias and information leakage, Actual duration is not used as a predictor, since it is mechanically linked to the outcome definition. *The core inferential strategy is one single multivariate logit on Late, with effects interpreted as OR-per-1SD, and with no ex-post predictors included.*

The core model includes the following Z-standardised predictors:

- Concession period (Z) and Concession period squared (Z^2)
- Total investment (USD m, Z)
- Planned construction duration (months, Z)
- Hospital capacity (beds, Z)
- GDP per capita (current US$, Z)
- WGI Regulatory Quality (Z)
- WGI Government Effectiveness (Z)

The covariate set is fixed, continuous-only, and time-aligned to project start to keep the model comparable across countries. Ex-post variables (e.g., Z_actual) are excluded to avoid leakage.

Estimation sample and reporting outputs.

The core model is estimated on the main adjusted sample excluding Project ID 35 (N = 42), which is treated as the primary inferential specification. Results are reported as odds ratios (OR) with 95% confidence intervals and p-values, rounded to two decimals.

Model fit and discrimination are summarised through:

- AIC = 61.40**,** BIC = 77.04
- Deviance R^2 = 25.46% (adj. 11.72%**)**
- Hosmer-Lemeshow test p = 0.45
- ROC AUC = 0.82 (Figure 9)

Inference is reported (OR, CI, p) and complemented with standard fit/discrimination diagnostics (AIC/BIC, HL, AUC).

Table 9 - Core logistic regression results (GDP per capita; main adjusted sample "no35", N = 42)

Predictor	**OR**	**95% CI**	**p-value**
Concession period (Z)	0.51	[0.12 ; 2.23]	0.37
Concession period squared (Z^2)	0.16	[0.04 ; 0.71]	0.02
Total investment, USD m (Z)	0.33	[0.08 ; 1.41]	0.14
Planned construction duration, months (Z)	3.22	[0.75 ; 13.73]	0.11
Hospital capacity, beds (Z)	3.17	[0.67 ; 14.99]	0.15
GDP per capita, current US$ (Z)	0.11	[0.01 ; 1.62]	0.11
WGI Regulatory Quality (Z)	36.97	[1.15 ; 1187.02]	0.04
WGI Government Effectiveness (Z)	0.51	[0.04 ; 6.54]	0.61

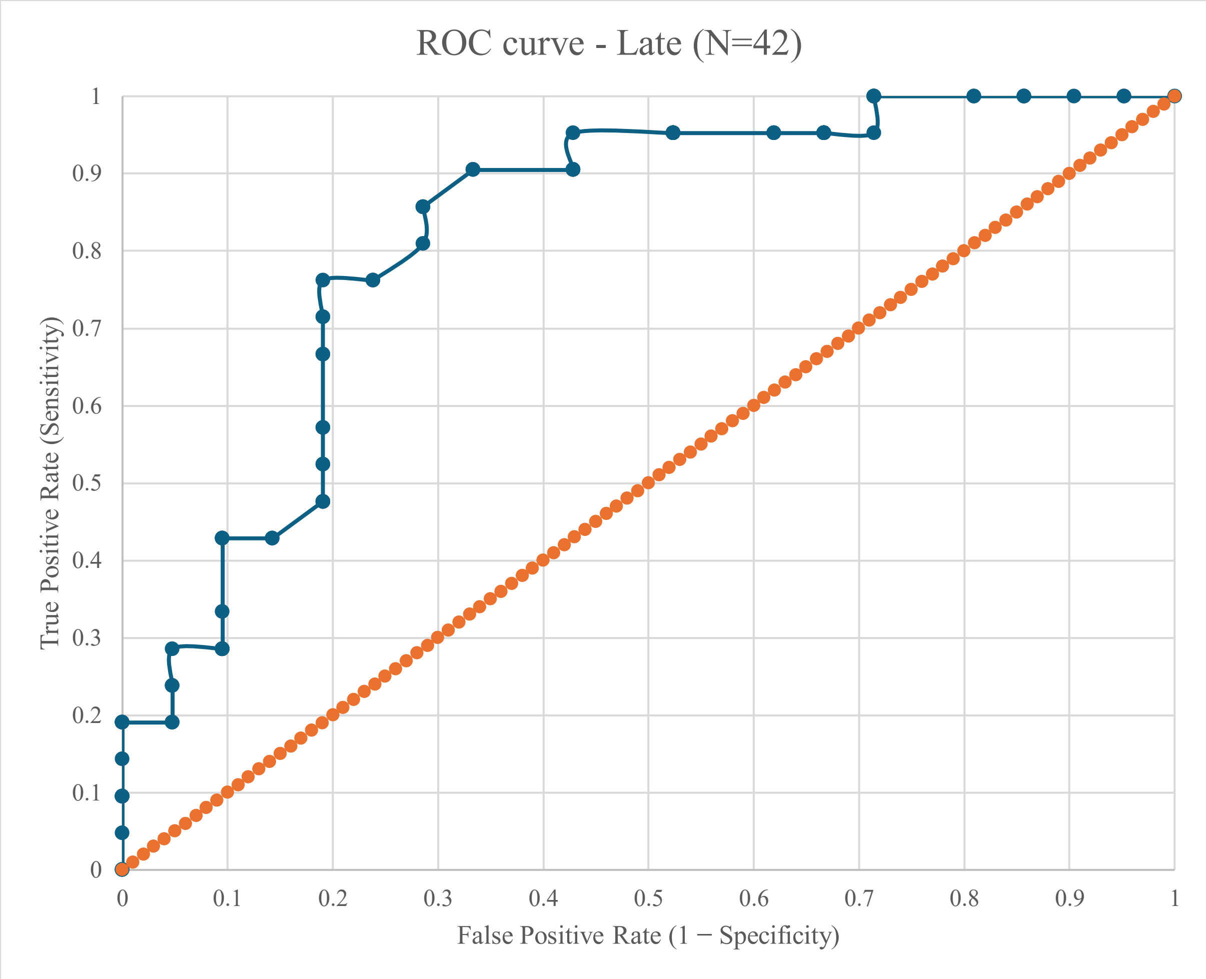


Figure 9 - ROC curve (Core model; AUC = 0.82)

The ROC curve for the main adjusted specification (no35, N = 42) yields an AUC of 0.82, indicating good in-sample discrimination between Late and non-Late projects. Calibration is acceptable based on the Hosmer-Lemeshow test (p = 0.45), which does not suggest systematic miscalibration. Given the small sample size and the cross-country setting, these performance metrics are interpreted as descriptive of in-sample fit rather than as out-of-sample predictive guarantees. *Performance is used as a plausibility check on the specification, while substantive interpretation focuses on coefficient patterns and robustness.*

Main coefficients.

Table 9 highlights one clear result: the pre-specified concession non-linearity is the strongest and most interpretable signal in the core specification. The quadratic term is statistically informative (OR = 0.16, 95% CI [0.04; 0.71], p = 0.02), while the linear concession term is imprecise (OR = 0.51, p = 0.37). Because the squared regressor captures distance from the mean tenor in either direction, the negative coefficient on Z_concession_sq implies a concave (inverted-U) association between concession tenor and Late risk under this specification: predicted Late is highest around an interior tenor and declines as tenor moves away from that interior region. This interpretation aligns with the tenor-based predicted-probability curve reported earlier (Figure 2), which shows Late risk peaking close to the sample's typical concession length.

Other project-level coefficients show economically plausible directions but wide uncertainty intervals in this small cross-country sample. Planned construction duration (OR = 3.22, p = 0.11) and hospital capacity (OR = 3.17, p = 0.15) both point toward higher Late odds as baseline schedules lengthen and hospitals become larger, consistent with interface complexity and longer critical paths in acute facilities. Total investment enters with an OR below one (OR = 0.33, p = 0.14), in the direction opposite to that expected under H2: the data do not support the "scale increases delay" channel for the capex component. The opposite direction may reflect ex-ante selection (larger capex projects subject to more rigorous preparation, tighter lender scrutiny, and more conservative baselines) rather than a true scale effect; the estimate is too imprecise to distinguish this from sampling noise, and is not interpreted as a substantive finding. GDP per capita has the expected negative direction (OR = 0.11, p = 0.11), suggesting lower Late odds in higher-capacity contexts, yet the confidence interval is wide, reflecting limited power.

Institutional and macro coefficients are interpreted cautiously due to strong collinearity and large uncertainty. Pairwise correlations among the country-level covariates (Table 10) are very high ($r \approx 0.83$–0.94). Consistent with this overlap, variance inflation factors are elevated for the same block (VIF = 10.45 for Z_gdppc, 16.83 for Z_wgi_rq, 10.49 for Z_wgi_ge), indicating severe multicollinearity.

Table 10 - Correlation matrix among the country-level covariates

	Z_gdppc	**Z_wgi_rq**	**Z_wgi_ge**
Z_gdppc	1.00	0.83	0.88
Z_wgi_rq	0.83	1.00	0.94

Z_wgi_ge	0.88	0.94	1.00

A direct implication is that standard errors for macro/institution predictors are inflated and coefficient magnitudes may become unstable. This is particularly relevant for Regulatory Quality, which is statistically significant (OR = 36.97; p = 0.04) but extremely imprecise (CI spanning several orders of magnitude). In this setting, the model is effectively allocating shared cross-country variation across highly correlated predictors; the institutional and macro coefficients are treated as suggestive conditional associations rather than definitive structural effects. By contrast, the pre-specified concession non-linearity remains the most interpretable coefficient signal in the core design.

The pattern observed for Regulatory Quality (OR = 36.97 with a 95% CI of [1.15; 1187.02], spanning more than three orders of magnitude) is a textbook signature of sparse-data bias under high collinearity and limited events-per-variable (Greenland, Mansournia & Altman, 2016). Maximum-likelihood estimates of logistic regression coefficients become unstable in the presence of correlated predictors and few effective informational degrees of freedom: the point estimate moves toward implausible magnitudes while the confidence interval widens to cover almost the entire admissible range. The OR magnitude is not interpreted substantively, and the apparent statistical significance is read as a small-sample artefact of the joint estimation rather than as evidence of a stand-alone institutional effect. Penalised estimation (Firth's correction) or shrinkage methods would be the natural mitigation in larger samples, but they are not adopted here because they would alter the pre-specified inferential design; the chapter instead documents the issue transparently and conditions all institutional and macro inference on the country-control matrix in Section 4.6.

4.5 Diagnostic and limited adjustments: capped removals

Diagnostics inspected and identification of influential observations.

Given the small multi-country inferential sample, model stability was assessed through standard logistic-regression diagnostics stored from the estimation output. The analysis inspected standardised deviance residuals (to detect outcome-level misfit), leverage (HI, to detect unusual covariate patterns), and influence measures (Cook's distance and ΔDeviance, to detect observations that disproportionately affect parameter estimates).

In the baseline estimation, Minitab flagged a small subset of observations as unusual. Project ID 35 was identified as the most consequential case because it combined (i) a large standardised deviance residual (SDEVRES = 2.17, flagged "R"), (ii) high leverage (HI = 0.43), and (iii) materially non-negligible influence indicators (COOK = 0.41, DDEV = 4.80). In practical terms, this observation is both difficult to fit (residual-based signal) and structurally influential (leverage/influence-based signal), which motivates a limited adjustment to test stability.

Diagnostics were used to identify observations that are simultaneously hard-to-fit and influential; Project ID 35 stands out because it combines large residual, high leverage, and sizeable influence metrics.

Capped-removals rule and definition of the main adjusted specification.

To avoid iterative "cleaning" driven by statistical significance, the thesis adopts a capped-removals rule: at most two observations can be removed in total, and any removal must be justified by clear diagnostic triggers and documented by Project ID. As a single capped adjustment, the analysis excludes Project ID 35, producing the main adjusted sample (no35, N = 42) used for the core model. No further removals are applied in the main analysis. *The adjustment is not used to maximise significance but to assess stability under a minimal, documented change. The main results remain anchored to one primary model (no35), preventing model proliferation in a small-N setting.*

4.6 Robustness and sensitivity checks

This section reports the pre-specified robustness checks designed to assess whether the chapter's core message is stable under a small set of controlled perturbations. Robustness is evaluated with two complementary criteria: (i) stability of coefficient patterns (sign and order of magnitude, especially for pre-specified terms) and (ii) stability of model-level performance (AIC/AUC and a basic calibration check), with applied logistic-regression reporting conventions.

Given the small multi-country sample and the high collinearity among country-level controls, robustness checks are not used to "improve significance" but to document whether interpretation depends on (a) one influential observation, (b) a functional-form choice for concession tenor, (c) the selected macro proxy (GDP per capita vs GDP), and (d) the way the institutional/macro block is parameterised.

Robustness checks included in 4.6 (as pre-specified):

1. N = 43 and N = 42 (baseline sample vs "no35" adjusted sample)
2. Concession² and Concession (linear-only) (functional-form sensitivity)
3. GDP vs GDP per capita (macro proxy swap)
4. Country-control matrix (Core: GDPpc + RQ + GE and reduced/alternative blocks)

Check 1: Baseline and adjusted sample: N = 43 and N = 42 ("no35").

As documented in Section 4.5, Project ID 35 is the only observation that simultaneously triggers residual-, leverage-, and influence-based flags. Under the capped-removals discipline, it is excluded to form the main adjusted sample ("no35", N = 42), which is used as the core inferential specification for the robustness checks below. The impact of the adjustment is assessed by comparing the baseline model estimated on N = 43 with the same modelling architecture estimated on N = 42. Without changing the covariate set, performance metrics improve modestly: AUC increases from 0.77 to 0.82 and AIC decreases from 66.35 to 61.40. Calibration remains acceptable in both cases (Hosmer–Lemeshow p = 0.82 in the baseline; p = 0.45 in the no35 specification).

This is reported as a sensitivity step, not as an assertion that the excluded project is substantively invalid: it documents that baseline fit/discrimination were disproportionately affected by one high-leverage observation in a small-N cross-country dataset. The "no35" model reduces dependence on a single influential observation and delivers improved fit/discrimination while preserving the pre-specified modelling design.

Table 11 - Baseline vs Adjusted sample (N=43 vs N=42).

N	**#Late**	**#non-Late**	**AIC**	**AUC**	**HL p-value**
43	22	21	66.35	0.77	0.82
42	21	21	61.40	0.82	0.45

N	OR(Z_concession_sq)	OR(GDPPC)	OR(WGI RQ)	OR(WGI GE)
43	0.27	0.22	25.36	0.22
42	0.16	0.11	36.97	0.51

Check 2: Concession functional form: Concession² (core) and Concession-only (linear).

The core specification models concession tenor allowing for non-linearity by including both the linear standardised term and its quadratic transformation on the z-scale, $Z_{\text{concession\_sq}} = (Z_{\text{concession}})^2$. This choice reflects the idea that the relationship between contract duration and delivery risk may not be monotone: very short tenors can be associated with tight delivery constraints, while very long tenors may reflect complex contractual structures and exposure to renegotiation and governance frictions.

To test whether the chapter's conclusion depends on this functional-form assumption, a linear-only specification (including $Z_{\text{concession}}$ but excluding $Z_{\text{concession\_sq}}$) is compared to the quadratic specification on the same adjusted sample (no35, $N = 42$). Results are reported in Table 12.

The contrast is sharp. The linear-only model exhibits weak discrimination (AUC = 0.59) and a higher information criterion (AIC = 70.49), while the quadratic specification shows substantially improved fit and predictive separation (AUC = 0.82; AIC = 61.40). Calibration remains acceptable in both cases (HL p-values 0.769 vs 0.449), suggesting that the performance gain is not obtained at the expense of obvious miscalibration. In addition, the estimated linear tenor effect changes direction and remains imprecise across the two models (OR($Z_{\text{concession}}$) = 1.19 in the linear-only specification vs 0.51 in the quadratic specification), consistent with the idea that a single linear slope is an inadequate summary of the tenor-Late relationship once curvature is allowed.

This robustness check indicates that concession tenor contains meaningful information for Late only when the functional form allows curvature. For interpretation, the thesis places greater weight on the quadratic specification and on the implied non-linear pattern (as visualised in the predicted-probability plot), rather than on the linear term alone.

The concession channel is not well captured by a linear-only effect: adding the pre-specified squared term markedly improves model fit and discrimination, supporting a non-linear interpretation of tenor.

Table 12 - Concession functional form check (linear vs quadratic)

	Linear-only specification	Quadratic specification
N	42	42
AIC	70.49	61.40
AUC	0.59	0.82
HL p-value	0.77	0.45
OR(Z_concession)	1.19	0.51

Check 3: Macro proxy swap: GDP per capita and GDP (current US$).

As a controlled robustness check, the macro development control is swapped from GDP per capita (current US$) (core) to GDP (current US$), keeping all other modelling choices unchanged (same outcome, Z-standardisation strategy, quadratic construction for concession tenor, country institutional controls, and the same adjusted sample no35, N = 42). The goal is not coefficient-by-coefficient "significance stability" (which is unlikely under a correlated country block), but stability of the key design channel and of model-level behaviour.

Table 13 - Robustness checks (core adjusted sample, N = 42): macro swap

	N	AIC	AICc	AUC	HL p-value
GDPpc	42	61.40	67.02	0.8231	0.45
GDP	42	63.3	68.93	0.7959	0.915

Table 14 - Robustness checks (core adjusted sample, N = 42): macro swap, OR

	GDPpc	GDP
OR(Z_concession)	0.51 (0.12 ; 2.23)	0.44 (0.12 ; 1.57)
OR(Z_concession_sq)	0.16 (0.04 ; 0.71)	0.33 (0.11 ; 0.98)
OR(Z_gdp)	-	2.43 (0.62 ; 9.54)
OR(Z_gdppc)	0.11 (0.007 ; 1.62)	-
OR(Z_wgi_rq)	36.83 (1.15 ; 1177.48)	2.75 (0.13 ; 56.85)
OR(Z_wgi_ge)	0.51 (0.04 ; 6.59)	0.38 (0.03 ; 4.43)

Table 13 shows that overall performance is not overturned by the macro proxy swap. Discrimination remains in the same range (AUC = 0.823 for GDPpc vs 0.796 for GDP), and information criteria remain comparable (AIC = 61.40 vs 63.30; AICc = 67.02 vs 68.93). Calibration is acceptable in both cases, with the Hosmer-Lemeshow test becoming more favourable under the GDP specification (HL p = 0.45 vs 0.915). Given the small-N cross-country setting, these metrics are interpreted as descriptive in-sample plausibility checks rather than as out-of-sample predictive guarantees.

The key design channel remains interpretable under the swap. The quadratic concession term is below one in both specifications: OR(Z_concession_sq) = 0.1588 [0.0357; 0.7069] under GDPpc and 0.3347 [0.1138; 0.9843] under GDP; supporting the same qualitative conclusion that concession tenor is better captured through a non-linear (curvature) specification than through a single linear slope. The linear concession term remains imprecise in both cases (GDPpc: 0.512 [0.1178; 2.2258]; GDP: 0.436 [0.1209; 1.5732]), consistent with the functional-form discussion.

In contrast, country-level coefficients are more sensitive to the choice of macro proxy, consistent with high overlap among macro and institutional measures. Replacing GDPpc with GDP moves the macro coefficient substantially (GDPpc: OR(Z_gdppc) = 0.1086 [0.0073; 1.6164] vs GDP: OR(Z_gdp) = 2.4324 [0.6202; 9.5399]) and is accompanied by sizable changes in the institutional terms (WGI RQ: 36.83 [1.15; 1177.48] under GDPpc vs 2.75 [0.13; 56.85] under GDP; WGI GE: 0.51 [0.04; 6.59] vs 0.38 [0.03; 4.43]). Confidence intervals remain wide, reinforcing that these macro/institution coefficients should be treated as conditional associations rather than definitive stand-alone effects in this small, collinear country block.

The macro proxy swap does not overturn model-level performance and preserves the qualitative design conclusion (concession non-linearity). It mainly highlights that country-level coefficients are less stable and less precisely estimated than the pre-specified concession non-linearity in a small, correlated cross-country sample.

Check 4: Country-control matrix: Core and reduced blocks (stability under collinearity).

As documented in Section 4.4 (Table 10), the country-level block is strongly collinear: GDP per capita and WGI indicators capture overlapping cross-country variation in development and institutional quality. In a small-N setting, this overlap inflates standard errors and makes coefficient magnitudes sensitive to how the shared signal is allocated across correlated controls. The objective of this check is not to “select” the best macro/institution proxy, but to document whether the chapter’s substantive message is stable under alternative, pre-defined parametrisations of the country block.

To make the limitation transparent, the thesis estimates a compact matrix of five specifications on the same adjusted sample (no35, N = 42), keeping the full project-level structure unchanged (continuous-only, Z-standardised predictors, and the pre-specified concession non-linearity) while varying only the country controls: (M1) the core block GDPpc + WGI RQ + WGI GE; (M2) GDPpc-only; (M3) WGI-only (RQ + GE); and the "split" variants (M4) RQ-only and (M5) GE-only. Throughout, interpretation is explicitly conditional: under multicollinearity, macro and institutional coefficients should be treated as conditional associations within a correlated control set; instability is expected and limits stand-alone interpretation of GDP/WGI magnitudes. The robustness claim rests on whether the pre-specified concession non-linearity remains the most consistent and interpretable signal across the matrix.

Table 15 - Country-control matrix (core adjusted sample, no35, N = 42).

	AIC	**AICc**	**AUC**	**HL p-value**
M1: OR(GDPpc), OR(RQ), OR(GE)	61.40	67.02	0.8231	0.45
M2: OR(GDPpc)	64.73	68.03	0.737	0.50
M3: OR(RQ), OR(GE)	62.96	67.33	0.7823	0.07
M4: OR(RQ)	62.67	65.97	0.7619	0.42
M5: OR(GE)	64.45	67.75	0.746	0.35

Table 16 - Country-control matrix (core adjusted sample, OR and 95% CI values, no35, N = 42).

	M1: OR(GDPpc) OR(RQ) OR(GE)	**M2: OR(GDPpc)**	**M3: OR(RQ) OR(GE)**	**M4: OR(RQ)**	**M5: OR(GE)**
OR(Z_concession)	0.51 (0.12,2.23)	0.90 (0.37,2.22)	0.61 (0.21,1.80)	0.57 (0.20,1.60)	0.74 (0.28,1.94)
OR(Z_concession_sq)	0.16 (0.04,0.71)	0.42 (0.18,0.98)	0.29 (0.09,0.91)	0.31 (0.10,0.94)	0.39 (0.15,0.99)
OR(Z_gdppc)	0.11 (0.01,1.62)	0.88 (0.30,2.57)	-	-	-
OR(Z_wgi_rq)	36.83 (1.15,1177.5)	-	9.88 (0.77,127.18)	2.45 (0.67,9.29)	-
OR(Z_wgi_ge)	0.51 (0.04,6.59)	-	0.23 (0.02,2.23)	-	1.38 (0.46,4.16)

Table 15 reports AIC/AICc, ROC AUC and Hosmer–Lemeshow p-values across the matrix. Performance remains broadly comparable, with AUC ranging from 0.737 to 0.823 and AIC from 61.40 to 64.73.

The core specification (M1) achieves the best combined fit/discrimination (AIC = 61.40; AUC = 0.823), while GDPpc-only (M2) shows the weakest discrimination (AUC = 0.737), consistent with information loss when institutional measures are excluded. Calibration is generally acceptable (HL p-values between 0.35 and 0.50), with the WGI-only specification (M3) showing more borderline calibration (HL p = 0.07).

These diagnostics indicate that the modelling architecture remains plausible under alternative country blocks and that no single parametrisation mechanically drives overall fit.

The coefficient patterns align with the expected multicollinearity mechanism. Country-level odds ratios vary materially across block definitions and exhibit wide uncertainty intervals, while the concession non-linearity remains comparatively stable.

- Design channel (concession non-linearity). The quadratic concession term is below one across all specifications: OR(Z_concession_sq) ranges from 0.1588 to 0.4182 and remains bounded away from one in most cases (e.g., M1: 0.1588 [0.0357; 0.7069]; M2: 0.4182 [0.1791; 0.9766]; M3: 0.2948 [0.0953; 0.9123]; M4: 0.3066 [0.1000; 0.9402]; M5: 0.3864 [0.1496; 0.9985]). This stability in direction and magnitude supports the interpretation already developed in the functional-form section: Late risk is not well summarised by a single linear tenor slope, and the most interpretable signal is the pre-specified curvature captured by the squared term. The linear concession term remains imprecise and modestly sensitive to block definition, as expected in a specification where curvature is the primary content.
- Country block (GDP/WGI). In contrast, the macro/institution terms show marked instability in magnitude and precision as the block is re-parameterised. GDP per capita shifts substantially between the core model and the GDP-only model (M1: OR(Z_gdppc) = 0.1086 [0.0073; 1.6164] versus M2: 0.8854 [0.3048; 2.5721]), illustrating how the GDP coefficient depends on whether institutional measures are allowed to absorb part of the shared cross-country signal. Regulatory Quality is a clear example of collinearity-driven volatility: it is very large and extremely imprecise in the core block (M1: OR(Z_wgi_rq) = 36.83 [1.15; 1177.48]) and changes materially once GDPpc is removed or the WGI block is split (M3: 9.88 [0.77; 127.18]; M4: 2.49 [0.67; 9.29]). Government Effectiveness also changes in magnitude and can cross the unit threshold depending on the specification (M1: 0.51 [0.04; 6.59]; M3: 0.23 [0.02; 2.23]; M5: 1.38 [0.46; 4.16]). These movements are consistent with the model allocating highly correlated country-level variation differently across alternative blocks, and they reinforce the caution against stand-alone substantive claims on the macro/institution coefficients.

Across alternative country-control blocks, the strongest and most interpretable signal is the pre-specified concession non-linearity; macro/institution coefficients are treated cautiously because their magnitudes depend on how shared cross-country variation is allocated under collinearity.

ROC curves across country-control blocks (M1-M5) are reported in Appendix A.2.

Compact synthesis of robustness evidence.

First, the capped adjustment ($N = 43 \rightarrow N = 42$) improves overall fit and in-sample discrimination (AIC 66.35 → 61.40; AUC 0.77 → 0.82) without altering the modelling architecture, indicating that baseline performance was disproportionately influenced by a single high-leverage observation rather than by substantive re-specification. Second, functional-form sensitivity reinforces the central design conclusion: concession tenor is materially more informative when modelled non-linearly. The quadratic specification dominates the linear-only alternative on the same adjusted sample (AIC 61.40 vs 70.49; AUC 0.82 vs 0.59), supporting the curvature interpretation already suggested by the predicted-probability plot. Third, the macro proxy swap (GDP per capita vs GDP) leaves model-level behaviour broadly comparable (AUC 0.823 vs 0.796; AIC 61.40 vs 63.30; calibration improves in HL terms), while making clear that the macro/institution block is sensitive to correlated proxies: the macro and WGI odds ratios shift and remain imprecisely estimated, consistent with the overlap between development and governance measures. Finally, the country-control matrix makes the collinearity consequences explicit and quantifiable: alternative country blocks yield similar performance ranges (AUC 0.737-0.823; AIC 61.40-64.73) but materially different, and often very wide, country-level confidence intervals, whereas the pre-specified concession non-linearity remains directionally stable across specifications ($OR(Z\_concession\_sq) < 1$ throughout).

Robustness resides less in the exact magnitudes of GDP/WGI coefficients, which are best interpreted as conditional allocations of shared cross-country variation, and more in the repeated emergence of the concession non-linearity as the most stable and interpretable signal in the chapter's core design.

Chapter 5 - Results, discussion, policy implications, and conclusions

5.1 Results: descriptive and inferential

Dataset scope, sample construction.

The unit of analysis is the hospital PPP project. The master database contains 56 projects, but inference relies on projects where both planned and realised construction duration are observed and credible, because Late is mechanically constructed from those fields.

To protect construct validity, three exclusion rules reduce noise rather than maximise sample size: Under Construction projects (no completion observed), projects with missing/unclear beds, and projects with uncertain or inconsistent schedule information across sources. The resulting samples are:

- Baseline inferential sample: N = 43
- Main adjusted sample (without project ID 35): N = 42, excluding Project ID 35 identified as highly influential under a capped-removals discipline.

Country coverage spans 14 developed, 28 developing, and 1 least-developed context, with country covariates aligned to the project start year and an alignment window from 2005 to 2018.

Descriptive profile of delays and prevalence of Late.

The descriptive evidence indicates that schedules are frequently missed. Planned construction duration averages 38.42 months (median 37), while realised duration averages 44.95 months (median 40). Delay magnitude is right-skewed: mean delay_pct ≈ 0.17 (≈17%) but median ≈ 0.03 (≈3%), consistent with a distribution driven by tail events.

Late prevalence is substantial: 22 out of 43 projects (51.16%) are classified as Late. *This prevalence motivates modelling timeliness as a binary risk outcome, rather than relying on group contrasts in delay magnitude.*

Descriptive group evidence on delay magnitude (context).

(a) Contract-form groupings and delay_pct.

Delay magnitude differs descriptively across contract labels (e.g., DBFM median delay_pct ≈ 10.4% vs near-zero medians in DBFO/Other), but inferential group tests do not detect statistically reliable differences in this sample (ANOVA $p = 0.211$; Kruskal-Wallis $p = 0.104$). *This is treated as contextual evidence only.*

(b) Development status groupings and delay_pct.

Similarly, mean delay_pct differs between developed and non-developed groups, but medians are close and dispersion differs; both ANOVA ($p = 0.151$) and Kruskal–Wallis ($p = 0.770$) do not support systematic group separation. *The pattern is consistent with tails driving mean contrasts rather than a uniform shift.*

(c) Why these group comparisons are not escalated.

The thesis design treats grouping as descriptive context. The core inferential claim is produced by a single multivariate model on Late with continuous covariates (parsimony, comparability, reduced over-parameterisation in small N).

Core inferential model results (main claim of the empirical chapters).

(a) Model recap to interpret ORs.

The dependent variable is Late (1/0). Predictors are continuous-only and z-standardised, so odds ratios are interpreted per one standard deviation increase, holding others constant. Country covariates are aligned to the project start year; ex-post variables (e.g., actual duration) are excluded to avoid leakage. Model reporting includes AIC/BIC, Deviance R^2, Hosmer-Lemeshow, and ROC/AUC.

(b) Central empirical result: concession tenor non-linearity.

Among pre-specified channels, the quadratic concession term is the most stable signal across robustness specifications, with OR($Z\_concession^2$) consistently below one (range: 0.16-0.42 across model variants). The associated p-value lies near conventional thresholds in the core specification (OR = 0.16, $p = 0.02$, N = 42), and the magnitude varies materially between specifications, so the curvature is best interpreted as a qualitatively coherent pattern rather than a precisely estimated effect. *The implied pattern is concave (inverted-U), consistent with the predicted probability curve (Figure 2), where Late risk peaks around an intermediate tenor* *close to the sample mean and declines as tenor deviates from that interior region.* The implied concave shape is consistent with several alternative mechanisms (selection into different contracting strategies, bankability constraints at the tenor extremes, leverage of a small number of projects in tenor tails); the data do not allow disentangling them.

(c) Scale/complexity and baseline schedule signals.

Planned duration and beds point in economically plausible directions (higher Late odds as baseline duration and facility scale increase), but confidence intervals are wide (e.g., planned duration OR ≈ 3.22, $p = 0.11$; beds OR ≈ 3.17, $p = 0.15$). *Total investment is not decisive (OR < 1, $p = 0.14$), and the thesis treats interpretations like selection into better-prepared structures or conservative scheduling as plausible but not claimable.*

(d) Country-level block: institutions and development proxies under collinearity.

Country variables are conceptually relevant, but empirically difficult to separate: GDP per capita and WGI indicators are highly correlated (pairwise r ≈ 0.83–0.94) and VIFs are severe (≈10–17). This implies macro/institution coefficients behave as conditional allocations of shared cross-country signal and are expected to be unstable.

Model-level performance is good in-sample (AIC 61.40; AUC 0.82; HL $p = 0.45$), but the thesis explicitly treats these as plausibility checks rather than promises of external predictive performance.

Diagnostics and robustness synthesis.

(a) Why the model without project ID 35 is primary.

Project ID 35 is flagged as both hard-to-fit and influential (large deviance residual, high leverage, meaningful Cook's distance / ΔDeviance). The capped-removals rule is transparent: maximum of two, only one used, making the model without project ID 35 the main specification.

(b) Robustness checks (pre-specified order).

Across controlled perturbations (baseline vs adjusted sample, quadratic vs linear tenor, GDPpc ↔ GDP swap, alternative country-control blocks) the model-level behaviour remains consistent with one key message: the concession-tenor curvature is comparatively stable, while macro/institution coefficients shift materially under correlated controls. *The quadratic specification dominates the linear-only alternative (AIC 61.40 vs 70.49; AUC 0.82 vs 0.59), reinforcing that the curvature is not an artefact of one functional form choice.*

(c) What the results answer.

Within the disciplined design, the evidence supports three defensible statements:

1. The most stable correlational signal for Late risk is the non-linear concession tenor specification.
2. Scale/baseline schedule proxies are directionally coherent but imprecisely estimated in small N.
3. Country-level development/institution proxies are conceptually important but not stably separable statistically under collinearity.

5.2 Critical discussion and results interpretation

Reconnect to the conceptual framework (Chapters 1, 2).

Chapters 1, 2 frame PPP delivery performance as conditional on preparation quality, enforceable payment mechanisms, risk allocation aligned with capability, and contract-management capacity over the life cycle. In health, this logic is amplified by interface and commissioning burdens: hospitals integrate complex MEP systems, ICT and cybersecurity, infection-control flows, and commissioning processes that extend beyond physical completion.

The empirical chapters operationalise this framework through observable correlates at (or before) project start (tenor design, baseline schedule scale, facility scale, and country context) while keeping modelling choices auditable.

Discussion by hypothesis (H1–H4).

H3 (Tenor non-linearity): the thesis' strongest channel.

The concession curvature aligns with PPP design logic where tenor is not a neutral "duration choice" but a composite proxy for lifecycle obligations, handback discipline, and bankability/selection mechanisms. The evidence suggests that modelling tenor linearly misses information: Late risk concentrates around an interior region, while moving away from that region is associated with lower risk in the fitted pattern. Interpreted cautiously, the implication is not "longer is better" or "shorter is better," but that *tenor structure matters in a non-monotonic way*, consistent with the idea that different tenor regimes may correspond to different contracting strategies and capability requirements.

Several alternative mechanisms are consistent with the observed concave shape and the data cannot adjudicate among them.

(a) Selection: longer tenors may be assigned to projects with stronger sponsors, more sophisticated lender consortia, or more disciplined preparation, mechanically lowering Late risk at the long-tenor end and producing the descending right tail of the curve.
(b) Confounding with scale: shorter tenors may co-occur with simpler, smaller hospitals where commissioning interfaces are less demanding, lowering Late risk at the short-tenor end through a scale channel rather than through tenor itself.
(c) Sample composition and leverage: the tenor extremes contain a small number of projects, so the curve's tails are partly driven by leverage; the curvature could in part reflect a sampling artefact rather than a structural pattern.
(d) Bankability and contracting capability: very long tenors require deeper capital markets and stronger client capacity, which could attract a self-selected pool of more capable counterparties at the upper end.

The curvature is therefore treated as direction-of-effect evidence for the population of hospital PPPs reachable through this dataset, not as identification of the mechanism; isolating any of (a)-(d) would require a substantially larger sample or design features (instruments, matched comparisons) not available here.

H2 (Scale/complexity: beds, investment, planned duration): plausible mechanisms, limited precision.

The direction of planned duration and beds is coherent with hospital delivery realities: larger acute facilities typically imply more interfaces (MEP integration, clinical adjacencies, commissioning complexity) and longer critical paths, increasing the probability of missing baselines. However, the thesis remains disciplined: wide CIs mean the data are consistent with both moderate and large effects, so these channels are best interpreted as *plausible* rather than precisely estimated.

Investment is deliberately treated with caution. In PPPs, larger capex can coexist with better preparation, stronger consortiums, tighter lender scrutiny, and more conservative baselines, so the sign is not a pure "complexity" indicator. The observed OR < 1 is not escalated beyond a hypothesis-consistent discussion of selection/structuring, explicitly marked as non-identifiable in this dataset.

H1 (Institutions: WGI GE / RQ): conceptually central, empirically conditional.

Institutional quality should matter via regulatory clarity, credible enforcement, procurement competence, and public-side delivery capacity, channels emphasised throughout the PPP guidance reviewed in Chapters 1, 2. Empirically, however, governance and development proxies overlap heavily (high correlations, high VIFs), so coefficients are volatile and read as conditional associations. This does not negate the institutional hypothesis; it limits what can be inferred about *separate* institutional effects in a small cross-country sample where governance and income move together.

H4 (Development level: GDPpc / GDP): consistent direction, not a stand-alone claim.

GDP per capita's direction is consistent with the "macro capacity / supply-chain depth" intuition (lower Late odds in higher-capacity contexts), but it is imprecise and not stably separable from governance measures. The disciplined conclusion is about shared country-context signal rather than a clean decomposition into "development vs institutions."

What the thesis does not find.

Two "non-findings" are informative when interpreted properly.

(i) Descriptive tests do not detect statistically reliable delay_pct differences across contract-form labels. This is consistent with at least three explanations that do not contradict PPP theory: contract labels may be too coarse cross-country, the sample may be underpowered, and timeliness may depend more on preparation and management than on nominal form.

(ii) Development-status grouping does not produce systematic delay_pct differences; dispersion and tail events dominate. This strengthens the modelling stance: if country context is to be discussed, it should be through continuous controls (with explicit collinearity discipline), not through coarse group dummies.

Triangulation with audit and guidance evidence.

The empirical patterns are broadly consistent with the failure modes emphasised in audit and governance evidence: delays where preparation and oversight are weak, and the importance of enforceable performance regimes and contract-management capability. Triangulation is used to contextualise the *type* of mechanisms, not to upgrade associational estimates into causal claims. The two systematic reviews of the academic literature (Hodge and Greve (2007, 2017) on PPPs broadly, and Roehrich, Lewis and George (2014) on healthcare PPPs specifically) reach a converging high-level conclusion: the PPP performance debate is empirically unresolved, with outcomes that vary widely across contexts and that depend more on preparation quality and contract management than on the PPP form itself. The thesis' finding that the most stable correlational signal is a non-monotone tenor effect, while institutional/development decomposition is not statistically separable, is consistent with this academic stance: tenor structure operates as a composite proxy for design-stage decisions and capability, whereas country-level capacity should be treated as a joint context signal in small samples.

5.3 Limitations and generalisability

Data and measurement limitations.

Late depends on planned vs actual durations; strict credibility screens reduce N by design to protect against outcome misclassification. Under Construction projects remain for traceability but cannot enter timeliness inference. Cross-source definitional inconsistencies (completion vs commissioning) remain a limitation even under conservative screening. Two further measurement limitations apply. First, the binary Late definition treats a one-day overrun as equivalent to a multi-year overrun, discarding magnitude information that is preserved in delay_pct; this is a deliberate choice consistent with audit-style reporting, but it means the model cannot distinguish marginal late deliveries from severely late ones. Replicating the analysis with delay_pct under a tobit or gamma specification, and with alternative thresholds (Late = 1 if delay_pct > 5% or > 10%) is a natural extension on a larger sample. Second, country-level WGI and WDI series are aligned to the project start year (2005-2018), but construction itself spans roughly three to four additional years, during which the institutional and macro context can shift; in addition, the WGI series are periodically recalibrated, so the values for an early year read in 2024 are not strictly identical to those that would have been read in real time. Both points argue for treating country-level coefficients as broadly indicative of the "country context at start" rather than as precise contemporaneous descriptors of the construction window.

Sample and modelling constraints.

N = 43/42 implies limited power and wide CIs. The continuous-only model avoids over-parameterisation but necessarily omits rich categorical detail from the core logit. Influence sensitivity exists and is handled via capped removals rather than iterative cleaning.

Country-level collinearity.

Development and governance are strongly correlated; VIFs are high, so macro/institution coefficients are conditional and unstable. This limits cross-country decomposition claims.

Scope limitations.

The thesis does not model procurement duration, renegotiations, or post-close governance events because these are not measured consistently in the dataset; it also avoids confusion-matrix threshold reporting by design.

Implications for external validity.

The strongest generalisable message concerns which signals remain stable under controlled robustness (especially tenor curvature), not precise parameter magnitudes for macro/institution elasticities.

5.4 Policy and managerial implications

The implications below are anchored where possible to the empirical evidence reported in Chapter 4: the asymmetric distribution of delay magnitudes (mean delay_pct ≈ 17%, median ≈ 3%, with substantial right-tail mass), the 51% Late prevalence in the inferential sample, the directional coherence of beds and planned-duration coefficients, and the non-monotone tenor signal that survives robustness checks. Where the empirical evidence is not decisive (institutional decomposition, magnitude of country effects), the implications below are framed as guidance-aligned rather than as findings-driven, and that distinction is made explicit in each subsection.

Project preparation and credible baselines.

Timeliness starts before signature. Because the outcome is baseline-referenced, baseline realism is not only a measurement requirement but a substantive channel: weak preparation drives redesign and re-sequencing, raising Late risk. Policy implication: treat schedule baselines as a governed deliverable (scope maturity, permitting critical path, commissioning plan), not as an optimistic planning artefact. The empirical asymmetry of the delay distribution in the sample (mean delay_pct ≈ 17%, median ≈ 3%) reinforces this point: the average is pulled up by a small number of severe overruns rather than by a uniform shift, so baseline-management interventions should focus on the tail risk of major slippage (multi-month or multi-year overruns) rather than on a generic “all baselines slip a little” assumption. Tail-risk thinking, applied to scope maturity and permitting critical-path checks at the gateway stages, is a more proportionate response than uniform contingency. (Flyvbjerg, 2014) (European PPP Expertise Centre, 2014)

Contract architecture: payment mechanisms, KPIs, enforceability.

The guidance sources converge on practical levers: output-based specifications, testable KPIs, availability payments with structured deductions/bonuses, disciplined variation procedures, step-in rights, and codified handback. For hospitals, KPI design must match clinical workflows and infection-control constraints; mis-specified boundaries between FM and clinical operations invite rework and delays. (World Bank – Public-Private Partnership Legal Resource Center (PPPLRC), 2017) (European PPP Expertise Centre (EPEC), 2011)

Risk allocation and bankability discipline.

Risk-allocation tools stress “typical” allocations as bankable starting points; bespoke patterns increase negotiation burden and execution risk and require higher public capability to manage. This complements the empirical tenor result: tenor likely co-moves with structuring choices and bankability regimes, so standardisation and disciplined deviations are time-risk management tools, not bureaucratic preferences. (HM Treasury, 2007) (Hart, 2003)

Contract management capacity as an operational prerequisite.

EPEC-style guidance frames contract management as a staffed function with monitoring routines, KPI audits, benchmarking/market testing (where applicable), and controlled variations. The implication is operational: risk transfer can protect budgets during contractor distress, but it does not automatically protect timeliness; only governance capacity does. (European PPP Expertise Centre (EPEC), 2014)

Health-sector operational safeguards.

Hospitals are not simply large buildings: complexity is driven by MEP intensity, ICT integration and cybersecurity, commissioning that extends beyond physical completion, and service-continuity and decanting constraints. Beds are a defensible proxy for interface density in a harmonised global dataset, even if imperfect. (Cartlidge, 2017) (World Health Organization, 2015)

Fiscal sustainability and transparency (PFRAM + ESA 2010).

A core separation is maintained: ESA 2010 accounting classification (on/off balance sheet) is not a policy objective; affordability and fiscal risk assessment are the gatekeeping tasks. PFRAM-style assessment operationalises this through identification of explicit/implicit liabilities, stress tests (inflation, rates, FX), and affordability caps; private finance reshapes cash flows and risk, but does not remove costs. (International Monetary Fund, 2016) (Eurostat, 2019)

Governance and consortium complexity.

The dataset includes SPV partner count as a governance and interface proxy, but it is excluded from the core model because of reliability and parsimony constraints. The implication remains conceptual rather than findings-driven (the variable is descriptive only): multi-party governance and coordination complexity should be anticipated through clear decision rights, variation governance, and escalation routes, especially in hospital commissioning phases. Yescombe's (2017) treatment of SPV governance and Engel, Fischer and Galetovic's (2014) contractual analysis converge on the same point: the locus of decision-making within the SPV materially shapes how disputes are resolved during construction, particularly when sub-contractor distress (cf. Carillion) hits availability-based deals. (Yescombe, 2017)

5.5 Conclusions

Direct answer to the research question.

The research question asks which observable project-level and country-level factors are associated with the probability that a hospital PPP is delivered late, relative to its planned construction baseline. The evidence supports a layered, qualified answer.

(i) The most directionally stable signal in the data is the pre-specified non-linear effect of concession tenor on Late risk, with a concave (inverted-U) pattern in which predicted delay risk is highest around an intermediate tenor (close to the sample mean of about 29 years) and lower at the tenor extremes; the magnitude is not robustly estimated, but the direction is consistent across robustness specifications.

(ii) Scale and baseline-schedule proxies (hospital beds and planned construction duration) point in the expected positive direction (larger and longer baselines associated with higher Late odds), but confidence intervals are wide and the data cannot rule out either modest or large effects.

(iii) Total investment enters with the opposite of the expected sign and remains imprecise; the data do not support a "capex scale increases delay" channel and are also consistent with selection (better-prepared, more conservatively scheduled large-capex projects).

(iv) Country-level development and institutional proxies are conceptually meaningful but not statistically separable in this sample because of severe collinearity (pairwise correlations 0.83–0.94, VIFs 10–17): the country block should be read as a joint "country context" signal rather than as a decomposition into independent institutional and development effects. The disciplined message is that, conditional on this sample and on the pre-specified design, tenor curvature is the only consistently interpretable mechanism, while the other channels are best read as direction-of-effect evidence under high statistical uncertainty.

What the thesis contributes.

The thesis offers four specific contributions beyond replicating known stylised facts about PPP delivery. First, it builds and documents an auditable, multi-source, project-level cross-country dataset of hospital PPPs with explicit credibility screens for the schedule fields used to construct the baseline-based Late outcome, transparently trading sample breadth for outcome reliability. Second, it implements a pre-specified, parsimonious inferential design that maps audit-style timeliness directly onto observable predictors at project start, with a quadratic concession term as a substantive (not exploratory) feature; this design avoids the typical small-sample temptation of post-hoc specification search. Third, it identifies the concession-tenor non-linearity as the most directionally stable correlational signal in the sample, while transparently characterising the country-level block as a non-identifiable joint signal under severe collinearity, a methodological clarification that is itself a contribution, since it constrains the kinds of stand-alone institutional or developmental claims that can be made on small cross-country PPP samples. Fourth, by quantifying construct overlap (Planned as both baseline and predictor), sparse-data bias (in the Regulatory Quality coefficient), and country clustering, the thesis turns the limitations of cross-country PPP datasets into an operational checklist for future empirical work.

- an auditable, multi-source dataset construction with explicit credibility screens for the schedule fields used to construct the Late outcome (smaller but reliable sample privileged over larger but noisy coverage);
- a pre-specified inferential design (quadratic concession, capped-removals rule, restricted robustness sequence) that limits specification search and produces a single primary model with documented sensitivity rather than a shifting set of "preferred" estimates;
- an explicit characterisation of the country-level block as a jointly identified "country context" signal under severe collinearity (with documented sparse-data bias on Regulatory Quality and country clustering), which delimits the substantive claims that can responsibly be made on a small cross-country PPP sample.

Future research directions.

Three priorities follow from the design and the limitations identified. (i) Expand the dataset by harmonising outcome definitions (completion vs commissioning vs availability date) and by integrating additional national PPP units and IFI repositories, in order to lift the events-per-variable ratio above conventional thresholds and reduce sparse-data bias on the country block. (ii) Add structured project-level proxies that map directly to mechanisms discussed but not measured here (procurement duration, renegotiation and variation intensity, contract-management staffing and turnover, payment-mechanism strictness, and handback enforceability) coded systematically from contract documents and audit reports rather than from press reporting. (iii) Re-run the analysis with two pre-specified extensions: a continuous outcome (delay_pct) modelled with a tobit or gamma specification to recover magnitude information lost in the binary Late, and a sensitivity check that excludes Planned construction duration from the predictor set to address the construct overlap documented in Section 4.2. Consistent with the methodology adopted here, these extensions should preserve leakage avoidance, pre-specify robustness, and match model complexity to sample size.

Acronym table

Acronym	**Expanded**
ADB	Asian Development Bank
AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
ANOVA	Analysis of Variance
AUC	Area Under the (ROC) Curve
BIC	Bayesian Information Criterion
BLOT	Build, Lease, Operate, Transfer
BOO	Build, Own, Operate
BOOT	Build, Own, Operate, Transfer
BOT	Build, Operate, Transfer
CHI	Chi-square test
CI	Confidence Interval
DB	Design & Build

DBFM	Design, Build, Finance, Maintain
DBFO	Design, Build, Finance, Operate
DBFOM	Design, Build, Finance, Operate, Maintain
DBO	Design, Build, Operate
EBRD	European Bank for Reconstruction and Development
ECA	European Court of Auditors
EIB	European Investment Bank
EPEC	European PPP Expertise Centre
E&S	Environmental and Social
ESA 2010	European System of Accounts 2010
ESS5	Environmental and Social Standard 5 (World Bank ESF)
FHWA	Federal Highway Administration
FM	Facility Management
FX	Foreign Exchange
GE	Government Effectiveness

GI Hub	Global Infrastructure Hub
GLM	Generalized Linear Model
GPA	Government Procurement Agreement (WTO)
EHR	Electronic Health Record
HL	Hosmer-Lemeshow (goodness-of-fit) test
IADB	Inter-American Development Bank
ICT	Information and Communications Technology
IDB	Inter-American Development Bank
IFC	International Finance Corporation
IMF	International Monetary Fund
IPA	Infrastructure and Projects Authority
IPPP	Institutional Public-Private Partnerships
IQR	Interquartile Range
KPI	Key Performance Indicator
LDC	Least Developed Country

MEP	Mechanical, Electrical, and Plumbing
MIGA	Multilateral Investment Guarantee Agency
NAO	National Audit Office
NHS	National Health Service
NIST	National Institute of Standards and Technology
O&M	Operations and Maintenance
OARC	Office of Advanced Research Computing
OR	Odds Ratio
PF2	Private Finance 2
PFI	Private Finance Initiative
PFRAM	Public-Private Partnerships Fiscal Risk Assessment Model
PPP	Public-Private Partnership
PPPLRC	Public-Private Partnership Legal Resource Center
PSC	Public Sector Comparator
ROC	Receiver Operating Characteristic

RQ	Regulatory Quality
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users
SD	Standard Deviation
SEMATECH	Semiconductor Manufacturing Technology (consortium)
SPV	Special-purpose vehicle
UCLA	University of California, Los Angeles
UNCITRAL	United Nations Commission on International Trade Law
VIF	Variance Inflation Factor
WB	World Bank
WBG	World Bank Group
WDI	World Development Indicators
WGI	Worldwide Governance Indicators

Appendix

A.1 Data dictionary extensions and variables not used in the core model.

This appendix documents variables that were available in the master dataset or generated during the Minitab workflow but were not included in the core inferential specification. The purpose is transparency: in a multi-source, cross-country PPP dataset, it is common to retain auxiliary fields for traceability, exploratory checks, or descriptive profiling. However, the thesis adopts a parsimonious empirical strategy centred on a single pre-specified multivariate logistic model with ex-ante predictors only. Certain fields are excluded for principled reasons, including (i) avoiding post-outcome information leakage (e.g., any transformation of realised duration), (ii) limiting categorical over-parameterisation in a small-N setting, and (iii) maintaining a controlled robustness plan restricted to the GDP per capita ↔ GDP swap. *The appendix increases transparency without altering the empirical design: variables are excluded based on clear methodological criteria (leakage, parsimony, and pre-specified robustness boundaries), so the core model remains coherent and replicable.*

The core results can be interpreted as driven by predictors observable at (or before) project start, rather than by ex-post correlates of completion. Documenting "available but not used" variables reduces ambiguity about dataset scope and prevents misreading exploratory artefacts (e.g., classification cutoffs) as part of inferential claims.

Table 17 - Available variables not used in the core inferential model (transparency table)

Field	Description	Why available	Why not used in the core model	Where (if anywhere) used
Z_actual	Z-score of *actual_construction_months*	Created during standardisation workflow	Ex-post information ("leakage"): using realised duration as a predictor would mechanically	Not used

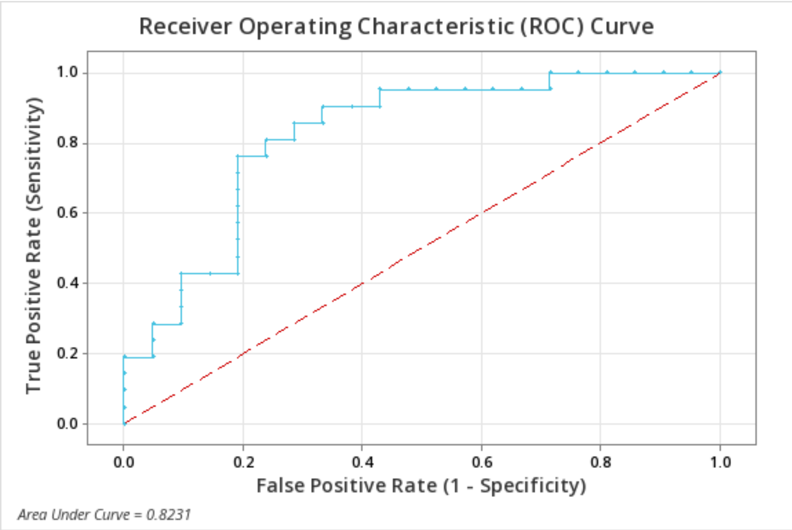
			embed the outcome definition	
spv_partners	Number of SPV equity partners	Collected as governance/interface proxy	Not included to preserve a continuous-only, core model and because partner counts are not consistently reliable across sources	Optional descriptive grouping only (if reported)
project_type_5_cat_anova	Project type (five categories)	Contract-form classification	Categorical and potentially uneven across countries; excluded from core model to avoid over-parameterisation in small N	Descriptive profiling / optional contextual comparison
project_type_5_group	Same 5-group project type	Harmonised grouping	Same reason as above	Descriptive only

project_type_3_group	Project type (three groups)	Aggregated grouping	Same reason as above	Descriptive only
country_dev_status_group	Development status (LDC/Developing/Developed)	Country context classification	Redundant with continuous macro controls (GDPpc/GDP) and would add categorical parameters in small N	Descriptive profiling (sample composition)
developed_1_0	Developed dummy	Simplified grouping	Same reason as above	Descriptive profiling only
construction_length_anova	Planned duration bins (grouped)	Prepared for potential group comparisons	Grouping discards information; core model uses continuous planned duration	Optional contextual comparison only
log_gdp	log(GDP), current US$	Prepared as potential skewness handling	Not adopted because robustness plan is restricted to GDPpc ↔ GDP swap; log-transform	Not used (unless explicitly reported in appendix as exploratory)

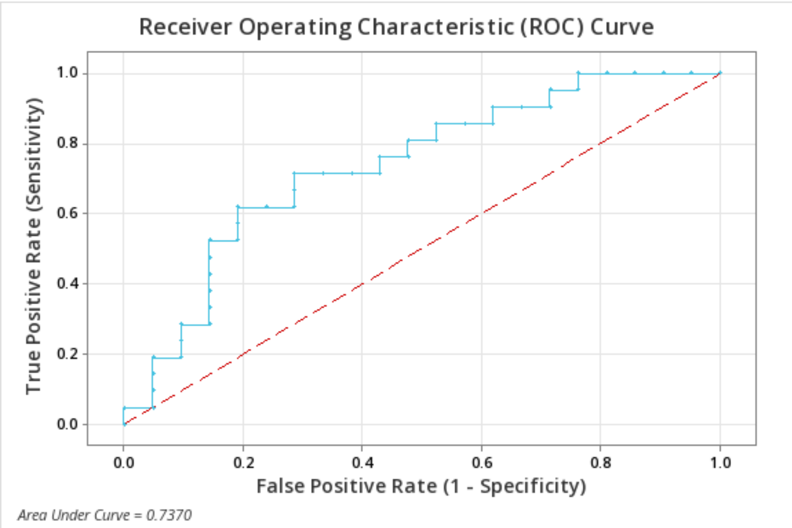
			not required for core inference	
pred_class	Predicted class from fitted prob	Classification workflow	No confusion-matrix reporting; cutoff would be arbitrary	Not used
tp / fp / fn / tn	Confusion matrix counts	Classification workflow	Not reported (no confusion matrix section)	Not used
tpr / fpr	Sensitivity / false positive rate	Classification workflow	Same as above	Not used
auc_segment	AUC for segments/subsets	Optional segmentation	No segment analysis; core AUC reported for the main model	Not used

A.2 ROC curves across country-control blocks (M1-M5).

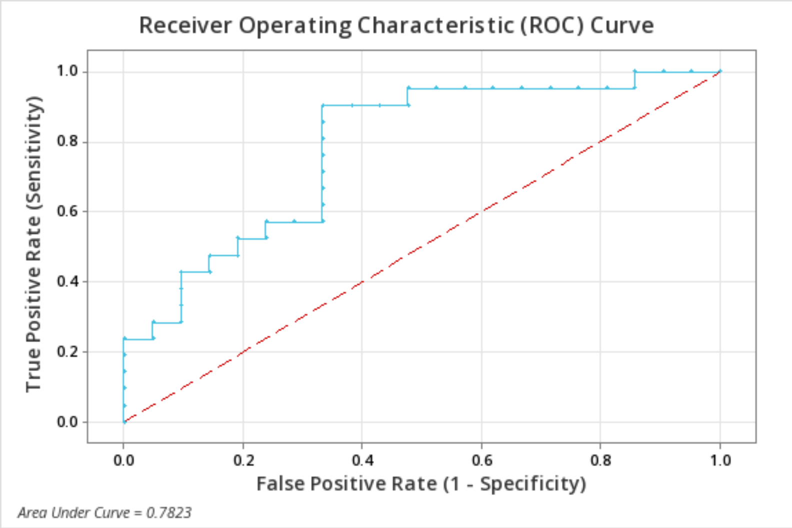
M1: OR(GDPpc), OR(RQ), OR(GE)



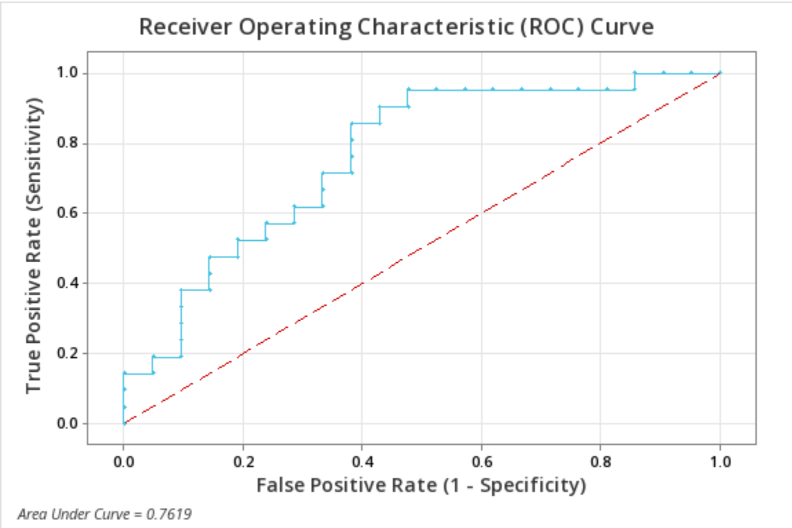
M2: OR(GDPpc)



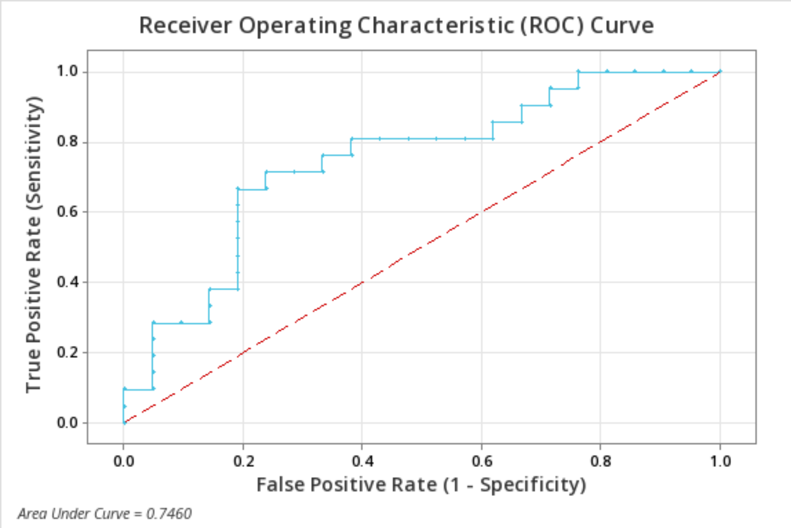
M3: OR(RQ), OR(GE)



M4: OR(RQ)



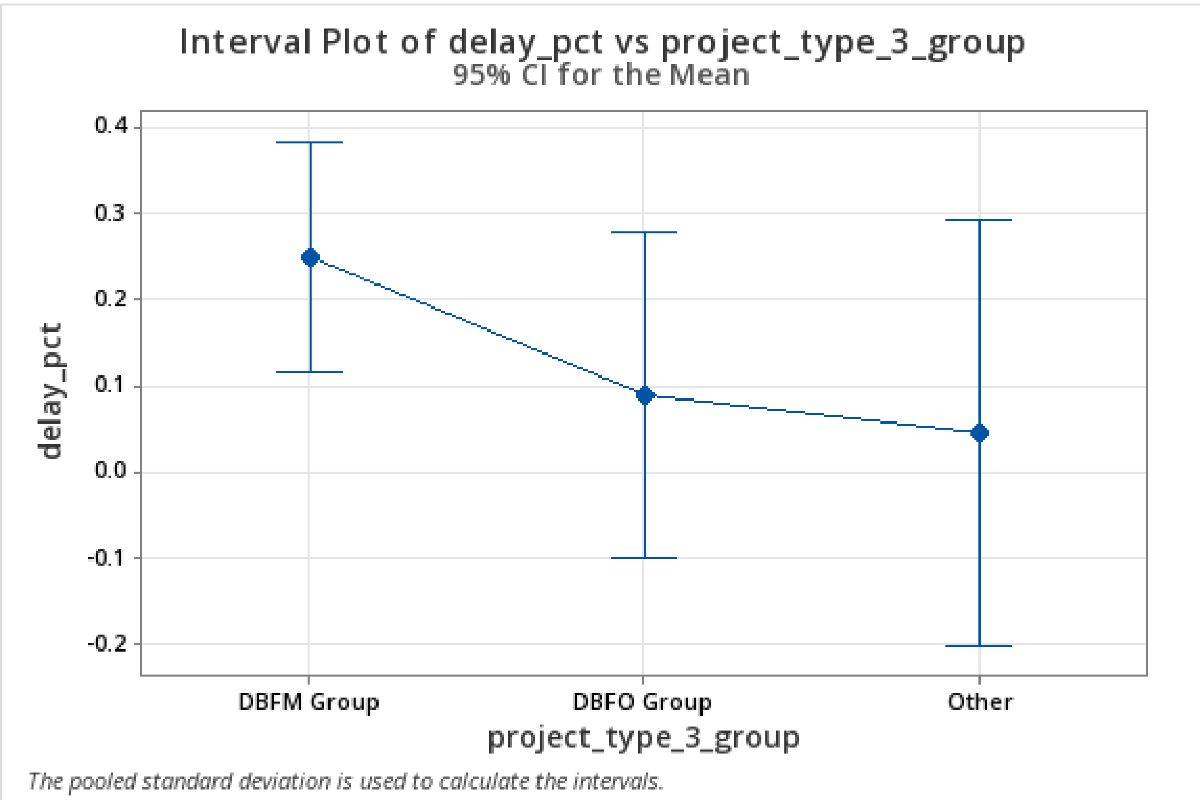
M5: OR(GE)

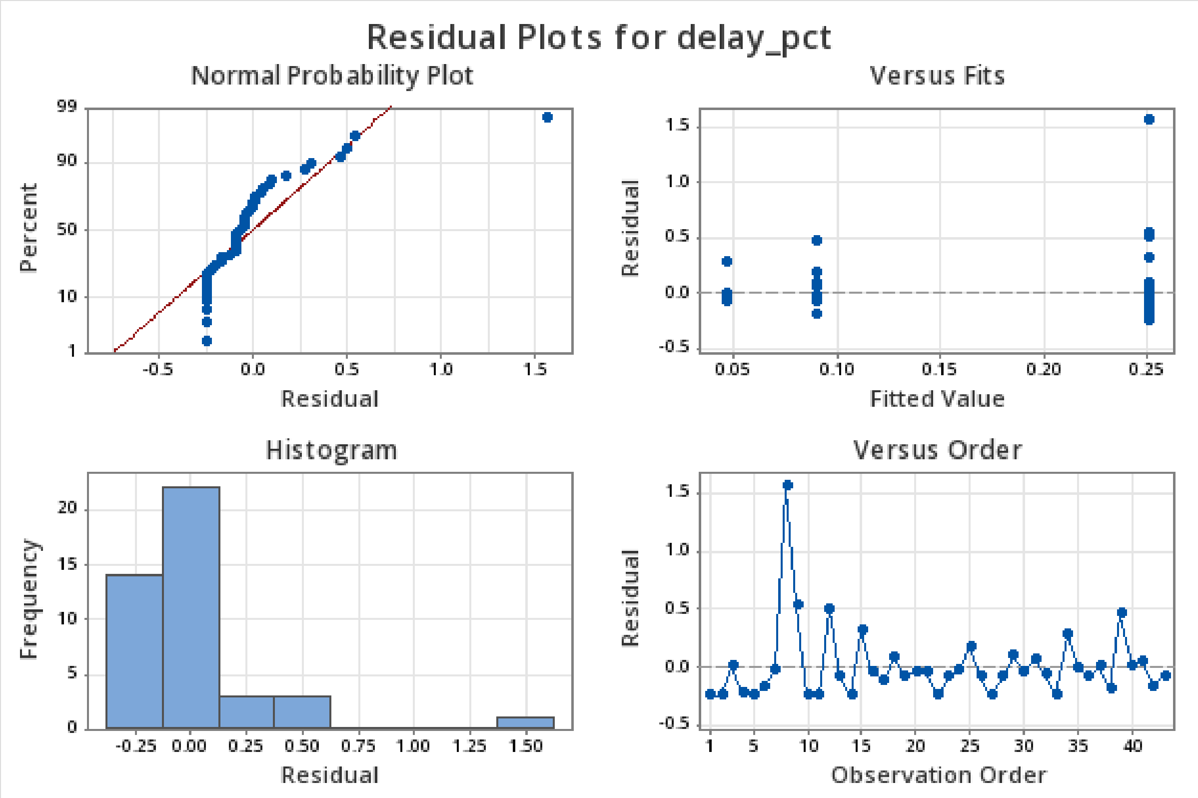


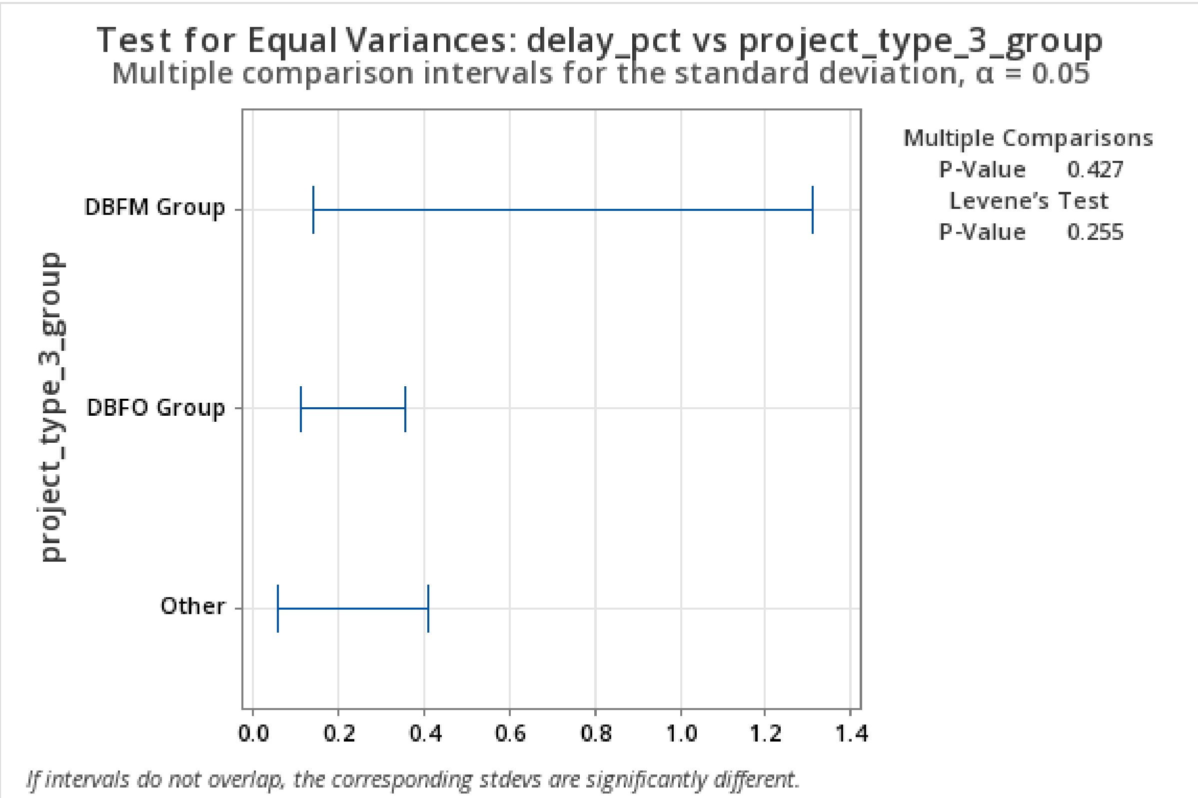
A.3 Residual plots (ANOVA diagnostics)

Residual diagnostics are reported to document distributional features (e.g., skewness/outliers) motivating the inclusion of rank-based tests. Diagnostics indicate departures from normality and the presence of influential observations; Kruskal-Wallis results are reported alongside ANOVA in the main text.

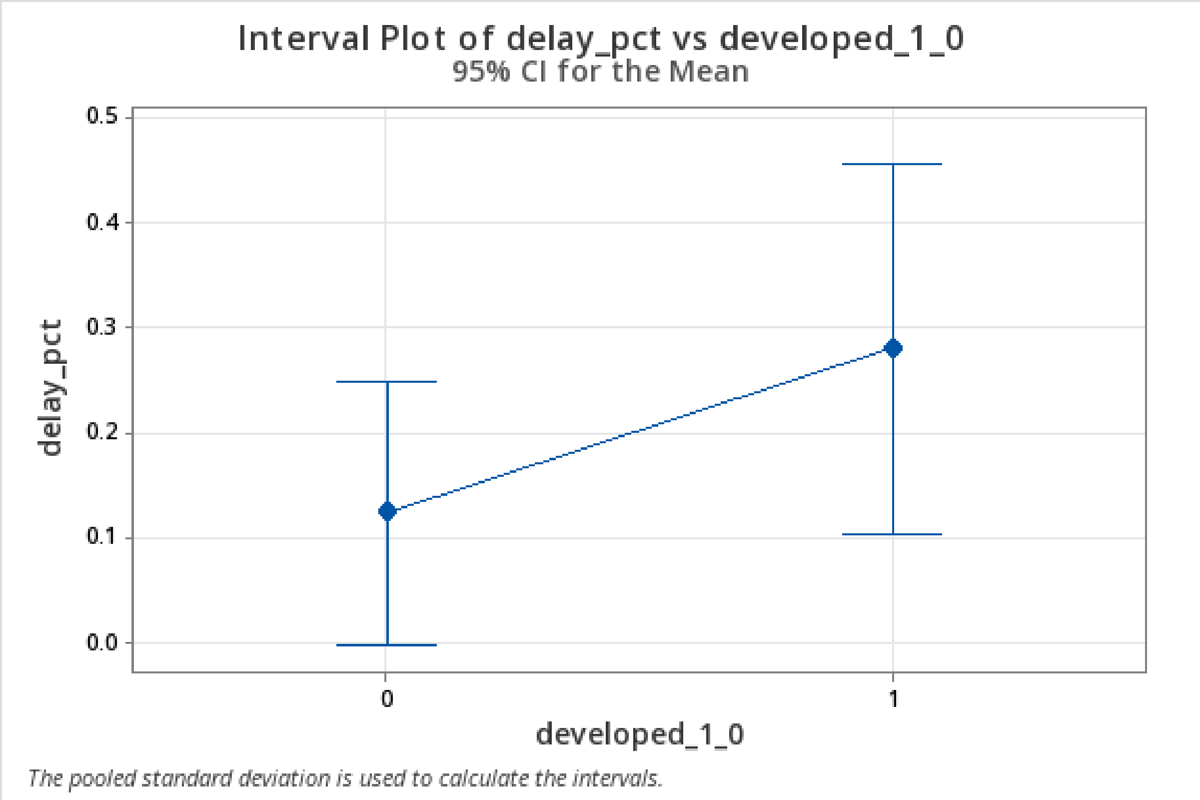
ANOVA residual diagnostics for delay_pct (contract form groups).

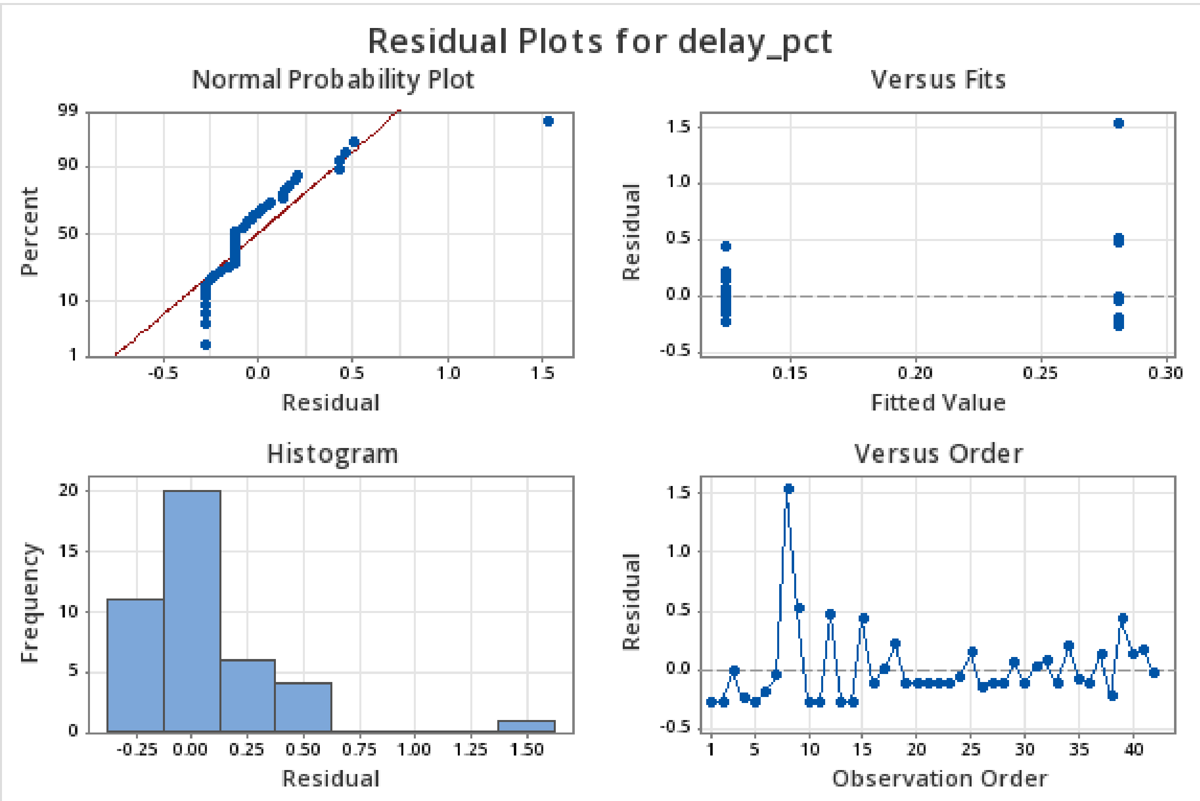


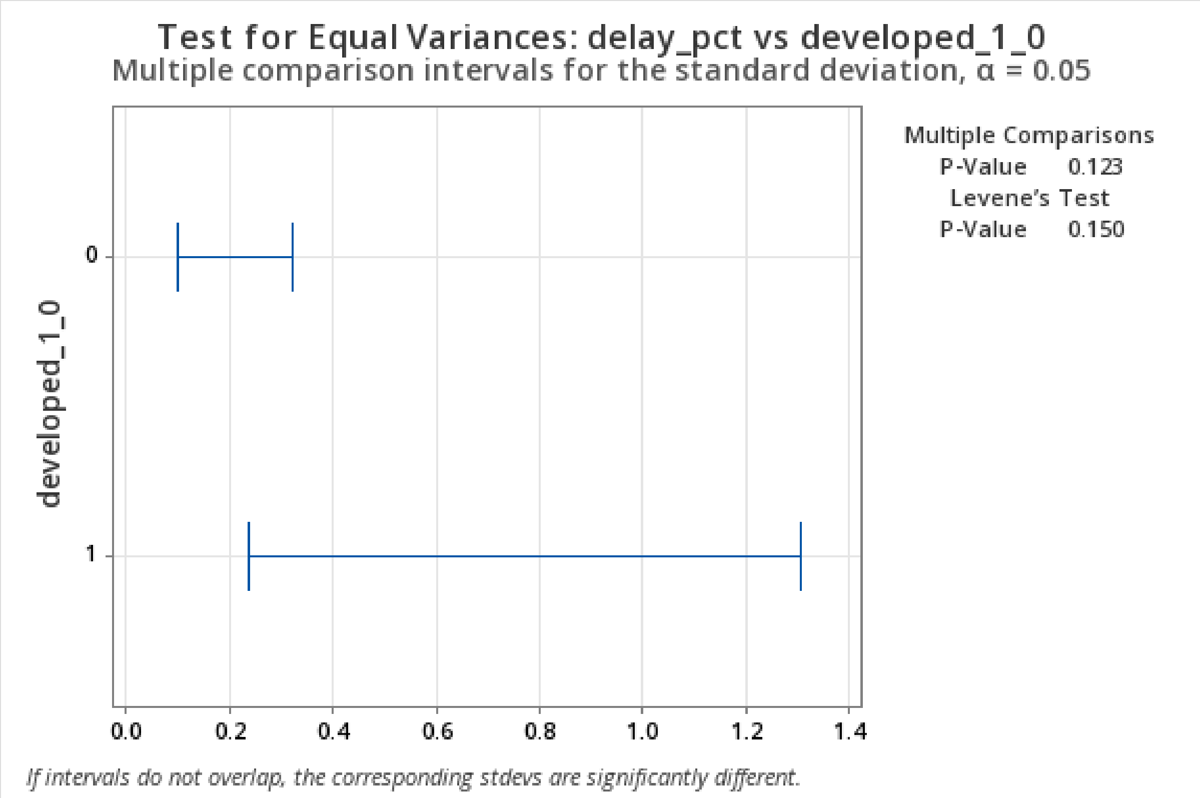




ANOVA residual diagnostics for delay_pct (development status).


Interval Plot of delay_pct vs developed_1_0
95% CI for the Mean
delay_pct
0.5
0.4
0.3
0.2
0.1
0.0
0
1
developed_1_0
The pooled standard deviation is used to calculate the intervals.


Residual Plots for delay_pct
Normal Probability Plot
Percent
Residual
Versus Fits
Residual
Fitted Value
Histogram
Frequency
Residual
Versus Order
Residual
Observation Order



Test for Equal Variances: delay_pct vs developed_1_0
Multiple comparison intervals for the standard deviation, α = 0.05
Multiple Comparisons
P-Value 0.123
Levene's Test
P-Value 0.150
developed_1_0
0
1
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
If intervals do not overlap, the corresponding stdevs are significantly different.

Bibliography

World Bank – Public-Private Partnership Legal Resource Center (PPPLRC). (2017). *Public–private partnerships reference guide (Version 3.0).* From World Bank Group: https://ppp.worldbank.org/sites/default/files/2024-08/PPP%20Reference%20Guide%20Version%203.pdf

United Nations Commission on International Trade Law (UNCITRAL). (2004). *Model Legislative Provisions on Privately Financed Infrastructure Projects.* From United Nations – UNCITRAL: https://uncitral.un.org/sites/uncitral.un.org/files/media-documents/uncitral/en/03-90621_ebook.pdf

Organisation for Economic Co-operation and Development. (2012, May 4). *Recommendation of the Council on Principles for Public Governance of Public-Private Partnerships.* From OECD Legal Instruments: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0392

World Trade Organization. (2012). *Agreement on Government Procurement (GPA 2012).* From World Trade Organization: https://www.wto.org/english/docs_e/legal_e/rev-gpr-94_01_e.pdf

European Union. (2014). *Directive 2014/23/EU on the award of concession contracts.* From EUR-Lex: https://eur-lex.europa.eu/eli/dir/2014/23/oj/eng

European Union. (2014). *Directive 2014/24/EU on public procurement.* From EUR-Lex: https://eur-lex.europa.eu/eli/dir/2014/24/oj/eng

European Union. (2014). *Directive 2014/25/EU on procurement by entities operating in the water, energy, transport and postal services sectors.* From EUR-Lex: https://eur-lex.europa.eu/eli/dir/2014/25/oj/eng

European PPP Expertise Centre; Eurostat. (2016). *A guide to the statistical treatment of public-private partnerships (PPPs) under ESA 2010.* From European Investment Bank – EPEC: https://www.eib.org/attachments/thematic/epec_eurostat_statistical_guide_en.pdf

Italy. (2023). *Legislative Decree No. 36/2023, Public Contracts Code (see Art. 174 – PPP).* From Normattiva: https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2023-03-31;36

Portugal. (2003). *Decreto-Lei n.º 86/2003 (Parcerias público-privadas).* From Diário da República: https://diariodarepublica.pt/dr/detalhe/decreto-lei/86-2003-240646

Treasury, H. (2018). *Private Finance Initiative (PFI) and Private Finance 2 (PF2): Budget 2018 brief.* From GOV.UK: https://www.gov.uk/government/publications/private-finance-initiative-pfi-and-private-finance-2-pf2-budget-2018-brief

Congress, U. (1991). *Intermodal Surface Transportation Efficiency Act of 1991 (H.R. 2950).* From Congress.gov: https://www.congress.gov/bill/102nd-congress/house-bill/2950

Administration, F. H. (1998). *TEA-21 – A Summary – An Overview.* From Federal Highway Administration (FHWA): https://www.fhwa.dot.gov/tea21/sumover.htm

Administration, F. H. (2005). *SAFETEA-LU – Summary Information.* From Federal Highway Administration (FHWA): https://www.fhwa.dot.gov/safetealu/summary.htm

Federal Highway Administration. (n.d.). *Availability payment DBFOM concessions.* From Federal Highway Administration: https://www.fhwa.dot.gov/ipd/p3/defined/new_build_facilities/dbfom_availability_payment_concessions.aspx

European PPP Expertise Centre. (2018). *The guide to guidance: How to prepare, procure and deliver PPP projects.* From European Investment Bank (EIB): https://www.eib.org/files/publications/epec_guide_to_ppp_en.pdf

World Bank. (2019). *Guidance on PPP contractual provisions (2019 edition).* From World Bank Group: https://ppp.worldbank.org/sites/default/files/2021-03/Guidance%20on%20PPP%20Contractual%20Provisions_2019%20edition.pdf
European Commission. (2004). *Green Paper on public-private partnerships and Community law on public contracts and concessions (COM(2004) 327 final).* From EUR-Lex: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52004DC0327
European Commission. (2008). *Commission interpretative communication on the application of Community law on public procurement and concessions to institutionalised public-private partnerships (IPPP) (2008/C 91/02).* From EUR-Lex: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj%3AJOC_2008_091_R_0004_01
European Court of Auditors. (2018). *Public-Private Partnerships in the EU: Widespread shortcomings and limited benefits (Special Report 09/2018).* From European Court of Auditors: https://www.eca.europa.eu/lists/ecadocuments/sr18_09/sr_ppp_en.pdf
FHWA, Federal Highway Administration. (2016). *Design-Build: program review and lessons learned.* From Federal Highway Administration: https://www.fhwa.dot.gov/construction/contracts/acm/2016_designbuild_report.pdf
Global Infrastructure Hub. (2018). *Managing PPP contracts after financial close (Summary).* From Global Infrastructure Hub: https://content.gihub.org/live/media/1495/gih_managingppp_summary.pdf
HM Treasury. (2018). *Private Finance Initiative (PFI) and Private Finance 2 (PF2): Budget 2018 brief.* From GOV.UK: https://www.gov.uk/government/publications/private-finance-initiative-pfi-and-private-finance-2-pf2-budget-2018-brief
Organisation for Economic Co-operation and Development. (2020). *Recommendation of the Council on the Governance of Infrastructure (OECD/LEGAL/0460).* From OECD Legal Instruments: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0460
U.S. Department of Transportation – Build America Bureau. (2024). *Eagle P3 Project (Denver, CO).* From U.S. Department of Transportation – Build America: https://www.transportation.gov/buildamerica/projects/project-highlights/eagle-p3-project-denver-co
Federal Transit Administration. (2016). *Denver Eagle.* From Federal Transit Administration: https://www.transit.dot.gov/funding/funding-finance-resources/private-sector-participation/denver-eagle
Railway Gazette International. (2001). *Infraspeed wins HSL-Zuid PPP.* From Railway Gazette International: https://www.railwaygazette.com/news/infraspeed-wins-hsl-zuid-ppp/28368.article
Netherlands Court of Audit (Algemene Rekenkamer). (2014). *Hogesnelheidslijn-Zuid: een rapportage in beeld.* From Netherlands Court of Audit: https://www.rekenkamer.nl/binaries/rekenkamer/documenten/rapporten/2014/07/01/hogesnelheidslijn-zuid-een-rapportage-in-beeld/Rapport_Hogesnelheidslijn.pdf
International Finance Corporation (IFC). (2023). *PPP Insights: Can PPPs expand access to affordable diagnostic services in emerging markets?* From International Finance Corporation: https://www.ifc.org/en/insights-reports/2023/can-ppp-s-expand-access-to-affordable-diagnostics
World Bank. (2023). *Public-private partnership and contracting out projects for imaging and laboratory diagnostics and their potential contribution to universal health coverage in low- and middle-income countries: Lessons learned.* From World Bank Group: https://documents1.worldbank.org/curated/en/099228212222328226/pdf/IDU13028bee71430614487187b41f19dfa71b870.pdf

Independent Evaluation Group (World Bank Group). (2016). *Public-Private Partnerships in Health: World Bank Group engagement in health PPPs - An IEG synthesis report.* From Independent Evaluation Group (World Bank Group): https://ieg.worldbankgroup.org/evaluations/public-private-partnerships-health

European PPP Expertise Centre. (2014). *The Guide to Guidance: How to Prepare, Procure and Deliver PPP Projects.* From UTAP (Portugal) – official host copy: https://www.utap.gov.pt/Publicacoes_interna/guide-to-guidance-en.pdf

International Monetary Fund; World Bank Group. (2019). *PPP Fiscal Risk Assessment Model (PFRAM 2.0): User Manual.* From World Bank – PPPLRC: https://ppp.worldbank.org/sites/default/files/2024-06/PFRAM_2_0_User_Manual_September_2019_6714932.pdf

World Bank Group. (2020). *Public-Private Partnerships Fiscal Risk Assessment Model (2019): Version 2.0.* From World Bank – PPPLRC: https://ppp.worldbank.org/library/public-private-partnerships-fiscal-risk-assessment-model-pfram-2019-version-2-0

European Court of Auditors. (2018). *Public-Private Partnerships in the EU: Widespread shortcomings and limited benefits (Special Report 09/2018).* From European Court of Auditors: https://www.eca.europa.eu/lists/ecadocuments/sr18_09/sr_ppp_en.pdf

National Audit Office. (2020). *Investigation into the rescue of Carillion's PFI hospital contracts (HC 23).* From National Audit Office: https://www.nao.org.uk/wp-content/uploads/2020/01/Investigation-into-the-rescue-of-Carillions-PFI-hospital-contracts.pdf

Audit Office of New South Wales. (2025). *Northern Beaches Hospital - Auditor-General's Report to Parliament 2025.* From Audit Office of New South Wales: https://www.parliament.nsw.gov.au/tp/files/190779/Tabling%20report%20Northern%20Beaches%20Hospital.pdf

European Bank for Reconstruction and Development. (2016). *Unique financing model as EBRD backs Elazig hospital PPP project in Turkey.* From EBRD: https://www.ebrd.com/home/news-and-events/news/2016/unique-financing-model-as-ebrd-backs-elazig-hospital-ppp-project-in-turkey.html

Multilateral Investment Guarantee Agency (MIGA). (2025). *Building Health Resilience: How MIGA Enabled Turkey's Elazig Hospital PPP (case study).* From MIGA: https://www.miga.org/sites/default/files/2025-03/Building%20Health%20Resilience%20How%20MIGA%20Enabled%20Turkey%27s%20Elazig%20Hospital%20PPP.pdf

Global Infrastructure Hub. (2021). *Knowledge Hub – Case studies (Türkiye).* From Global Infrastructure Hub: https://www.gihub.org/knowledge-hub/?country=tuerkiye&orderBy=rel&resourceType=case-studies

World Bank. (2024). *Drivers of Delays in Procurement of Infrastructure Projects.* From World Bank: https://documents1.worldbank.org/curated/en/099062824153516589/pdf/P173110138cdf70ef18cd21fb78946336ef.pdf

Inter-American Development Bank (IDB). (2020). *Public-Private Partnerships in Airports in Latin America and the Caribbean: Main Figures and Trends in the Sector.* From IDB Publications: https://publications.iadb.org/publications/english/document/Public-Private-Partnerships-in-Airports-in-Latin-America-and-the-Caribbean-Main-Figures-and-Trends-in-the-Sector.pdf

The World Bank. (2017). *Rethinking Infrastructure in Latin America and the Caribbean: Spending Better to Achieve More.* From World Bank Documents: https://documents1.worldbank.org/curated/en/676711491563967405/pdf/114110-REVISED-Rethinking-Infrastructure-Low-Res.pdf

The World Bank. (2017). *The World Bank Environmental and Social Framework.* From World Bank Documents: https://documents1.worldbank.org/curated/en/383011492423734099/pdf/The-World-Bank-Environmental-and-Social-Framework.pdf

Asian Development Bank (ADB). (2009). *Safeguard Policy Statement.* From Asian Development Bank: https://www.adb.org/sites/default/files/institutional-document/32056/safeguard-policy-statement-june2009.pdf

Inter-American Development Bank (IDB). (2020). *Environmental and Social Policy Framework.* From IDB Publications: https://publications.iadb.org/publications/english/document/Environmental-and-Social-Policy-Framework.pdf

European Court of Auditors (ECA). (2018). *Special Report 09/2018: Public-Private Partnerships in the EU – Widespread shortcomings and limited benefits.* From European Court of Auditors: https://www.eca.europa.eu/en/publications/SR18_09

Global Infrastructure Hub (GI Hub). (2020). *PPP Risk Allocation Tool – 2019 Edition.* From Global Infrastructure Hub: https://www.gihub.org/resources/publications/ppp-risk-allocation-tool-2019-edition/

Inter-American Development Bank (IDB). (2022). *Risk Matrix and PPP Contract Standardization: Best Practice and Gap Analysis in Brazil (PDF).* From IDB Publications: https://publications.iadb.org/publications/english/document/Risk-Matrix-and-PPP-Contract-Standardization-Best-Practice-and-Gap-Analysis-in-Brazil.pdf

Global Infrastructure Hub (GI Hub). (2018-2020). *Managing PPP Contracts.* From Global Infrastructure Hub: https://managingppp.gihub.org

National Audit Office (NAO). (2020). *Investigation into the rescue of Carillion's PFI hospital contracts.* From National Audit Office: https://www.nao.org.uk/wp-content/uploads/2020/01/Investigation-into-the-rescue-of-Carillions-PFI-hospital-contracts.pdf

The World Bank (PPP Resource Center). (2021). *COVID-19 and PPPs: Practice Note 2 – Projects in Preparation and Pipelines.* From World Bank – PPP Resource Center: https://ppp.worldbank.org/sites/default/files/2022-06/COVID-19_and_PPP_Practice_Note_2_-_Projects_in_preparation_and_pipelines_0.pdf

World Bank; Asian Development Bank; Inter-American Development Bank. (2017). *Public-Private Partnerships Reference Guide Version 3.0.* From PPP Knowledge Lab (World Bank): https://ppp.worldbank.org/library/ppp-reference-guide-3-0-full-version

HM Treasury. (2020). *The Green Book and accompanying guidance and documents.* From GOV.UK: https://www.gov.uk/government/collections/the-green-book-and-accompanying-guidance-and-documents

The World Bank. (2019). *Guidance on PPP Contractual Provisions (2019 edition) (PDF).* From World Bank – PPP Knowledge Lab: https://ppp.worldbank.org/sites/default/files/2021-03/Guidance%20on%20PPP%20Contractual%20Provisions_2019%20edition.pdf

Infrastructure and Projects Authority (IPA). (2022). *Project Routemap – Setting up projects for success.* From GOV.UK: https://www.gov.uk/government/publications/improving-infrastructure-delivery-project-initiation-routemap

European PPP Expertise Centre (EPEC) - European Investment Bank. (2014). *Managing PPPs during their contract life: Guidance for sound management (PDF).* From European Investment Bank / EPEC: https://www.eib.org/files/epec/epec_managing_ppps_en.pdf

Eurostat; European PPP Expertise Centre (EPEC). (2016). *A Guide to the Statistical Treatment of PPPs.* From Publications Office of the European Union / EIB–EPEC:

https://op.europa.eu/en/publication-detail/-/publication/ea7c19f7-9146-11e8-8bc1-01aa75ed71a1/language-en
The World Bank. (2024). *Worldwide Governance Indicators – Government Effectiveness (DataBank).* From The World Bank – DataBank: https://databank.worldbank.org/databases/governance-effectiveness
The World Bank. (2024). *Regulatory Quality: Percentile Rank.* From The World Bank – Data: https://data.worldbank.org/indicator/RQ.PER.RNK
The World Bank. (2017). *The World Bank Environmental and Social Framework (ESS incl. ESS5) (PDF).* From World Bank Documents: https://documents1.worldbank.org/curated/en/383011492423734099/pdf/The-World-Bank-Environmental-and-Social-Framework.pdf
Inter-American Development Bank (IDB). (2020). *Environmental and Social Policy Framework (PDF).* From IDB Publications: https://publications.iadb.org/publications/english/document/Environmental-and-Social-Policy-Framework.pdf
HM Treasury. (2024, May 16). *Business case guidance for projects and programmes*. From GOV.UK: https://www.gov.uk/government/publications/business-case-guidance-for-projects-and-programmes
The World Bank (International Bank for Reconstruction and Development). (2024, November 5). *The Worldwide Governance Indicators: Methodology and 2024 Update*. From World Bank – Documents & Reports: hhttps://documents.worldbank.org/en/publication/documents-reports/documentdetail/099005210162424110
The World Bank. (2024, November 5). *Worldwide Governance Indicators 2024 Update: Changes in Underlying Source Data*. From World Bank: https://www.worldbank.org/content/dam/sites/govindicators/doc/wgisourcedatarevisions.pdf
Organisation for Economic Co-operation and Development. (2012, May 4). *Recommendation of the Council on Principles for Public Governance of Public-Private Partnerships (OECD/LEGAL/0392)*. From OECD Legal Instruments: https://legalinstruments.oecd.org/public/doc/275/275.en.pdf
The World Bank. (2024, May 30). *GDP per capita (current US$)*. From World Bank Data360 (World Development Indicators metadata): https://data360files.worldbank.org/data360-data/metadata/WB_WDI/WB_WDI_NY_GDP_PCAP_CD.pdf
European Investment Bank – European PPP Expertise Centre . (2014, March 11). *Managing PPPs during their contract life: Guidance for sound management*. From European Investment Bank: https://www.eib.org/files/epec/epec_managing_ppps_en.pdf
The World Bank. (2024, October 31). *Drivers of Delays in Procurement of Infrastructure Projects*. From World Bank – Documents & Reports: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099062824153516589/p173110138cdf70ef18cd21fb78946336ef
World Bank Group; Asian Development Bank (ADB); Inter-American Development Bank (IDB/IADB). (2017, April 27). *PPP Reference Guide 3.0 (Full version)*. From World Bank PPP Knowledge Lab (Public Private Partnership): https://ppp.worldbank.org/library/ppp-reference-guide-3-0-full-version
The World Bank (World Bank Group). (2019, January 1). *Guidance on PPP Contractual Provisions (2019)*. From World Bank PPP Knowledge Lab (Public Private Partnership): https://ppp.worldbank.org/library/guidance-ppp-contractual-provisions-2019

Global Infrastructure Hub (GI Hub); Allen & Overy. (2020, July 21). *PPP Risk Allocation Tool 2019 Edition*. From Global Infrastructure Hub: https://www.gihub.org/resources/publications/ppp-risk-allocation-tool-2019-edition/

National Institute of Standards and Technology. (2020). *Comparing Instruments (NIST Technical Note 2106)*. From NIST Technical Series Publications (National Institute of Standards and Technology): https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.2106.pdf

National Institute of Standards and Technology (NIST). (2020). *Comparing Instruments (NIST Technical Note 2106)*. From NIST Technical Series Publications: https://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.2106.pdf

University of California, Los Angeles (UCLA) – Office of Advanced Research Computing (OARC). (n.d.). *FAQ: How do I interpret odds ratios in logistic regression?* From UCLA OARC Stats: https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/

University of California, Los Angeles (UCLA) – Office of Advanced Research Computing (OARC). (n.d.). *Logistic Regression | SPSS Annotated Output*. From UCLA OARC Stats: https://stats.oarc.ucla.edu/spss/output/logistic-regression

European Court of Auditors . (2018 , March 8). *Public Private Partnerships in the EU: Widespread shortcomings and limited benefits (Special Report 09/2018)*. From European Court of Auditors : https://www.eca.europa.eu/lists/ecadocuments/sr18_09/sr_ppp_en.pdf

HM Treasury. (2013, April 18). *The Green Book (2022)*. From GOV.UK: https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2020

HM Treasury. (2013, April 21). *Green Book supplementary guidance*. From GOV.UK: https://www.gov.uk/government/publications/green-book-supplementary-guidance-environment

World Bank. (2024). *The Worldwide Governance Indicators: Methodology and 2024 Update (Policy Research Working Paper 10952)*. From documents.worldbank.org: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099005210162424110

National Institute of Standards and Technology (NIST). (2012, January 27). *ZSCORE*. From NIST/ITL Dataplot: https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/zscore.htm

Torelli, N., & Pappadà, R. (2019-2020). *Elementi di Inferenza Statistica* (Vol. 2). Trieste, Trieste, Italy: Università degli Studi di Trieste, DEAMS.

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science, 25*(3), 289-310.

University of California, Riverside. (2025, July 8). *Introduction to the General Linear Model*. From University of California, Riverside: https://gradquant.ucr.edu/sites/default/files/2019-02/Introduction-to-the-General-Linear-Model.pdf

Rundel, C. (2013, 04 15). *Lecture 20 – Logistic Regression*. From Statistics 102 (STA 102), Duke University (Department of Statistical Science): https://www2.stat.duke.edu/courses/Spring13/sta102.001/Lec/Lec20.pdf

Henry, K. L. (2024). *Modeling Non-Linear Relationships (Module 15)*. From Foundations in Data Science (PSY 652), Colorado State University: https://psy652.colostate.edu/modules/m15-nonlinear_effects.html

Torres-Reyna, O. (2011, 01). *Predicted Probabilities and Marginal Effects After (Ordered) Logit/Probit models using margins in Stata (v. 1.0)*. From Princeton University: https://www.princeton.edu/~otorres/Margins.pdf

Shmueli, G. (2010, 08). *To Explain or to Predict?* From Project Euclid: https://projecteuclid.org/journals/statistical-science/volume-25/issue-3/To-Explain-or-to-Predict/10.1214/10-STS330.full

The Pennsylvania State University. (n.d.). *Logistic Regression (STAT 501 lesson)*. From Penn State Online: STAT 501: https://online.stat.psu.edu/stat501/lesson/13/13.2

NIST/SEMATECH. (n.d.). *The 1-way ANOVA model and assumptions*. From NIST/SEMATECH e-Handbook of Statistical Methods: https://www.itl.nist.gov/div898/handbook/prc/section4/prc432.htm

NIST/SEMATECH. (n.d.). *How can we make multiple comparisons?* From NIST/SEMATECH e-Handbook of Statistical Methods: https://www.itl.nist.gov/div898/handbook/prc/section4/prc47.htm

NIST/SEMATECH. (n.d.). *Levene Test for Equality of Variances*. From NIST/SEMATECH e-Handbook of Statistical Methods: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm

NIST/SEMATECH. (n.d.). *Kruskal–Wallis test for comparing populations with unknown distributions*. From NIST/SEMATECH e-Handbook of Statistical Methods.

Wasserstein, Ronald L.; Lazar, Nicole A. . (2016, 01 29). *The ASA's Statement on p-Values: Context, Process, and Purpose*. From Yale University (Department of Statistics): https://www.stat.yale.edu/~jtc5/238/readings/P%20values%20ASA%20Statement%20with%20commentary.pdf

Levine, T. R., & Hullett, C. R. (2002). *Eta squared, partial eta squared, and misreporting of effect size in communication research*. From Communication Monographs: https://www.toposbooks.gr/behavioralstats/Levine_Hullett_2002.pdf

Gelman, A. (2008). *Scaling regression inputs by dividing by two standard deviations*. From Columbia University: https://sites.stat.columbia.edu/gelman/research/published/standardizing7.pdf

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). *A simulation study of the number of events per variable in logistic regression analysis*. From PubMed: https://pubmed.ncbi.nlm.nih.gov/8970487/

National Institute of Standards and Technology (NIST). (2023, 12 4). *Regression Diagnostics*. From NIST Dataplot Reference Manual: https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/regrdiag.htm

The Pennsylvania State University. (n.d.). *Multicollinearity (STAT 501 lesson)*. From Penn State Online: STAT 501: https://online.stat.psu.edu/stat501/book/export/html/981

Flyvbjerg, B. (2014). *What You Should Know about Megaprojects and Why: An Overview*. From Project Management Journal (SAGE Journals): https://journals.sagepub.com/doi/10.1002/pmj.21409

European PPP Expertise Centre (EPEC). (2011). *PPPs in Healthcare*. From European Investment Bank: https://www.eib.org/en/publications/epec-ppps-in-healthcare

HM Treasury. (2007). *Standardisation of PFI Contracts (SoPC4)*. From UK Government: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/221556/standardisation_of_pfi_contracts_4.pdf

Hart, O. (2003). *Incomplete Contracts and Public Ownership: Remarks, and an Application to Public-Private Partnerships*. From The Economic Journal (Oxford Academic): https://academic.oup.com/ej/article/113/486/C69/5085160

European PPP Expertise Centre (EPEC). (2014). *Managing PPP Contracts: A Guide for Public Authorities*. From European Investment Bank: https://www.eib.org/en/publications/epec-managing-ppp-contracts

Cartlidge, D. (2017). *New Aspects of Quantity Surveying Practice*. From Routledge: https://www.routledge.com/New-Aspects-of-Quantity-Surveying-Practice/Cartlidge/p/book/9781138694388

World Health Organization. (2015). *Safe Hospitals in Emergencies and Disasters*. From WHO: https://www.who.int/publications/i/item/safe-hospitals-in-emergencies-and-disasters

International Monetary Fund. (2016). *The Public-Private Partnership Fiscal Risk Assessment Model (PFRAM)*. From IMF: https://www.imf.org/en/Publications/TNM/Issues/2016/12/31/Analyzing-and-Managing-Fiscal-Risks-Best-Practices-44453

Eurostat. (2019). *Manual on Government Deficit and Debt – Implementation of ESA 2010*. From European Commission: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-19-007

Yescombe, E. R. (2017). *Public-Private Partnerships: Principles of Policy and Finance*. From Elsevier (Butterworth-Heinemann): https://www.elsevier.com/books/public-private-partnerships/yescombe/978-0-12-815300-3

Academic references

Engel, E., Fischer, R., & Galetovic, A. (2014). The Economics of Public-Private Partnerships: A Basic Guide.. Cambridge University Press.

Hodge, G. A., & Greve, C. (2007). Public–private partnerships: An international performance review.. Public Administration Review, 67(3), 545–558.

Hodge, G. A., & Greve, C. (2017). On public–private partnership performance: A contemporary review.. Public Works Management & Policy, 22(1), 55–78.

Iossa, E., & Martimort, D. (2015). The simple microeconomics of public-private partnerships.. Journal of Public Economic Theory, 17(1), 4–48.

Roehrich, J. K., Lewis, M. A., & George, G. (2014). Are public–private partnerships a healthy option? A systematic literature review.. Social Science & Medicine, 113, 110–119.

Guasch, J. L. (2004). Granting and Renegotiating Infrastructure Concessions: Doing It Right.. World Bank Institute Development Studies, World Bank.

Cruz, C. O., & Marques, R. C. (2013). Endogenous determinants for renegotiating concessions: Evidence from local infrastructure.. Local Government Studies, 39(3), 352–374.

Soliño, A. S., & Vassallo, J. M. (2009). Using public-private partnerships to expand subways: Madrid–Barajas international airport case study.. Journal of Management in Engineering, 25(1), 21–28.

Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight.. BMJ (Clinical Research Ed.), 352, i1981.

van Smeden, M., Moons, K. G. M., de Groot, J. A. H., Collins, G. S., Altman, D. G., Eijkemans, M. J. C., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria.. Statistical Methods in Medical Research, 28(8), 2455–2474.

Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., & Collins, G. S. (2020). Minimum sample size for developing a multivariable prediction model: Part II – binary and time-to-event outcomes.. Statistics in Medicine, 39(2), 276–297.

McKee, M., Edwards, N., & Atun, R. (2006). Public-private partnerships for hospitals. Bulletin of the World Health Organization, 84(11), 890–896.