Transferability of load-level measurement and determinants of segregation value: multi-case evidence from ship and port dismantling

Gabriel Gouveia Magalhães; Fernanda Araujo Baião; Paula Maçaira — Pontifical Catholic University of Rio de Janeiro (PUC-Rio) — gabrielgouveiamagalhaes@gmail.com

Thematic track: Technical and Technological Aspects (interface with Environmental Aspects and Circular Economy)

English translation of the Portuguese version revised on 15 October 2026.

Introduction. In current Brazilian practice, inventories of recoverable steel in dismantling operations are estimated from LDT (lightweight displacement tonnage) or manufacturers' specifications, and are rarely measured by load. Since traceability, the circular economy and end-of-life decisions rely on these inventories, poor inventory quality affects subsequent stages. The Hong Kong International Convention (IMO, 2009), in force since 2025, requires material inventories, chain of custody and auditing of the physical destination of steel; in parallel, digital MRV is discussed as a measurement and audit layer for circularity chains (World Bank, 2022). The decommissioning literature focuses on options and impacts and on multicriteria analysis of strategies (Martins et al., 2020; Önal et al., 2023), with little attention to inventory production as a load-level measurement problem. An earlier study by the authors (Magalhães and Baião, 2026) proposed and demonstrated a load-level physical and financial measurement protocol (MRV-P) in a single Ship-to-Shore crane case, identifying replication in another asset class as a research priority. This paper addresses that priority: it applies the same protocol to a second case, involving two passenger vessels, and contrasts the two cases to determine which aspects of the protocol are invariant across asset classes and which depend on material composition.

Objective. To assess the transferability of the MRV-P protocol across distinct asset classes and examine Proposition P1, treated here as a directional and falsifiable proposition: assets with a larger mass fraction of high-grade structural categories tend, ceteris paribus, to yield a higher economic premium from segregation relative to unsegregated sales. The two cases provide preliminary evidence consistent with P1, rather than confirmation of a monotonic relationship. In addition to transferability, the study examines whether physical and financial reconciliation reveals record-integrity defects that are not visible in aggregate controls. The subsidiary questions are: (i) does load-level measurement support an auditable inventory in a second asset class? and (ii) does the curve's closure error distinguish record quality across different operations?

Methodology. An instrumental multiple-case study (Yin, 2018) is adopted, using replication logic along two dimensions. Replication is literal with respect to the procedure: the same MRV-P protocol should operate in both asset classes. It is theoretical with respect to the economic outcome: identical values are not expected, but rather divergence conditioned by material composition. Case 1 is the authors' reference result (figures reproduced, not recomputed); Case 2 provides the new, fully audited primary evidence. Case 1 — Ship-to-Shore (STS) crane, Brazilian port terminal, 2021: 808.0 t recovered, recorded in 107 dispatch entries and checked against gate and buyer weighing records; four categories (scrap, box sections, beams and pipes) and three buyers. Of the 107 entries, 34 are individually recorded loads; 22 of the 24 structural loads have third-party documentation. The audited core (n = 34) accounts for approximately 66% of revenue, allowing results to be reported in three strata (n = 107; n = 34; and n = 24, restricted to structural sections). Case 2 — Vital Brazil and Itapuca passenger ferries, shipyard in Guanabara Bay, 2020: 962 t in total, of which 762 t of metal was sold in 123 ferrous loads, predominantly low-grade hull steel (approximately 91% of the metallic mass). Data are recorded at one row per load (df_cycles), with weight, unit price, category, destination, date and three documentary markers (weighbridge ticket, manifest and invoice). One author's prior

affiliation with the operator is disclosed; bias is mitigated through load-level documentary markers, buyer anonymization and separation between commercial execution and analysis. The measurement layer links measurement (M — certified weighing that establishes the ground truth w_i), reporting (R — load-level attributes) and verification (V — reconciliation of the cumulative physical and financial curve). The segregation premium is $\Delta = R_obs - (\Sigma w_i \cdot p_ref)$; the closure error (ε) is calculated for the physical and financial components. The reference price p_ref is constructed for each case from disclosed external sources: an international quotation (CFR Turkey) in Case 1, consistent with the market for structural sections, and domestic hull scrap in Case 2. Case-specific references avoid confusing the segregation premium with differences in grade and market timing; an additional test using a common reference for both cases serves as a robustness check. Sensitivity of the sign is assessed using six p_ref scenarios in Case 1.

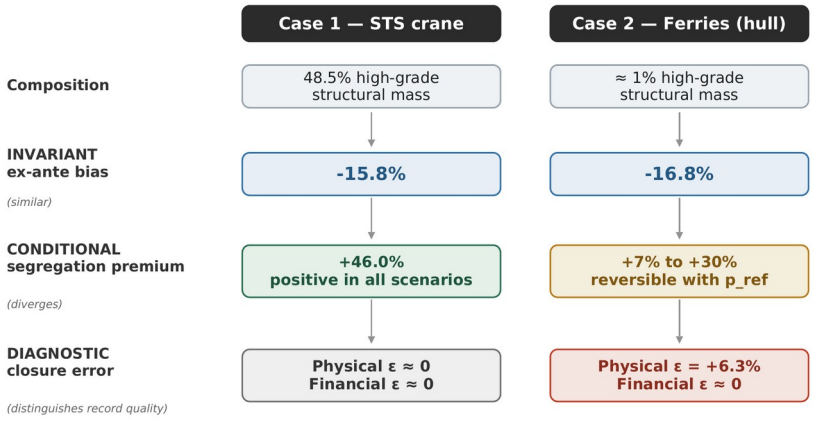

Figure 1 — Structure of the findings across the two cases, at three levels: ex-ante bias is replicated (invariant); the segregation premium diverges with material composition (conditional); and the closure error distinguishes record quality (diagnostic).

Results. The ground truth established by load-level weighing exceeded pre-operation estimates in both cases: the field estimate understated mass by 15.8% in Case 1 (680 t versus 808 t) and by 16.8% in Case 2 (800 t versus 962 t). The segregation premium differed between cases (Figure 1). In Case 1, it was +BRL 556,931 (+46.0%) at the base price and remained positive in all six p_ref scenarios assessed (+21.6% to +71.7%), with approximately 81% of the value attributable to the structural box-section category. In the stratum restricted to structural sections (n = 24), the arithmetic premium accounts for 85.1% of the total. In Case 2, the premium was smaller and sensitive to the reference: +30% at p_ref ≈ BRL 0.65/kg, +7% at BRL 0.79/kg, zero at ≈ BRL 0.85/kg and -7.1% at BRL 0.91/kg. The mass fraction of high-grade structural material is 48.5% in Case 1 and approximately 1% in Case 2. Under the same price reference applied to both cases, the ordering Δ(Case 1) > Δ(Case 2) held throughout the assessed range.

In Case 1, the cumulative physical and financial curve of the operational dataset reconciles with the declared total (ε ≈ 0 for both mass and revenue), and aggregate mass is confirmed by third-party weighing. In Case 2, the financial component reconciled (ε ≈ 0), but the physical component showed +6.3%. Investigation of the discrepancy identified an outdated summary cell and a duplicate entry in the source spreadsheet; certified load-level weighing showed no error. Table 1 summarizes the two cases.

Table 1 — Comparative summary of the cases.

Dimension	Case 1 — STS	Case 2 — Ferries
Class / period	Port crane / 2021	Passenger vessels / 2020
Mass; loads	808 t; 107 entries	962 t (762 t metallic); 123 ferrous loads
High-grade structural fraction	48.5%	≈ 1%
Ex-ante estimate (field)	-15.8%	-16.8%
Segregation premium Δ	+46.0% (positive in all scenarios)	+7% to +30% (reversible)
Physical / financial ε	≈ 0 / ≈ 0	+6.3% / ≈ 0

Discussion. Interpreted through replication logic (Yin, 2018), the results separate invariant from conditional findings. Literal replication is supported: the same load-level measurement procedure operated in both asset classes and produced auditable inventories in both. Ex-ante bias is the invariant element: in the two cases analyzed, underestimation of comparable magnitude (-15.8% and -16.8%) was observed despite the differences between asset classes. The segregation premium is the conditional element: it diverged in the direction predicted by material composition, being high in the asset with the larger structural fraction (Case 1) and modest and reversible in the asset dominated by low-grade hull steel (Case 2). The two cases therefore provide preliminary evidence consistent with P1. They do not establish full monotonicity or a complete causal function: two contrasting composition points establish a direction, while the relationship remains a conjecture to be tested in a multi-asset panel. The common-reference test reduces the risk that the contrast is an artifact of the p_ref convention, without eliminating other sources of endogeneity.

The closure error warrants separate consideration. In Case 2, financial reconciliation closed while physical reconciliation did not, and investigation attributed the difference to a record defect — an outdated summary cell and a duplicate entry — rather than to measurement error. The protocol exposed an inconsistency that aggregate controls based on totals would be unlikely to reveal. This finding suggests that MRV-P does more than measure mass and revenue: it also acts as a diagnostic mechanism for record integrity. This is evidence from one case, not proof of universal validity, but it points to a use of closure error that extends beyond material accounting.

Regarding validity, the construct is anchored in weighing and third-party documents, with the asymmetry in Case 2 and its physical error openly reported. External validity is analytical, not statistical: propositions transfer through theoretical replication, not sampling. In sustainability terms, the effect operates through a simple mechanism: granular measurement reduces the loss of information on material origin and category; this information supports segregation; segregation preserves material identity; preserved identity enables a more suitable destination; and a more suitable destination retains value in the circular economy. The load-level record (ticket, manifest and invoice) produces evidence compatible with the chain of custody required by the Hong Kong Convention (IMO, 2009) and creates conditions for greater circular value retention, without attributing any unmeasured environmental benefit here.

Complementary exploratory analysis (data science). The granularity of load-level records enables the use of data science techniques as a secondary contribution subordinate to the MRV-P multiple-case study. As an exploratory exercise using 25 actual dismantling projects from the authors' records, a linear regression baseline was compared with Random Forest and XGBoost. The target was the weight of steel recovered per project, and predictors were operational and structural attributes (cutting hours, energy, number of trips, manual estimate and area). The comparison indicated a better fit for ensemble methods than for the linear baseline. Given the

small sample, the fact that fitting was performed on the same sample without out-of-sample validation, and the strong correlation between the manual estimate used as a predictor and the target (risk of leakage), the results are strictly illustrative: no claim is made to operational predictive capability or generalization. Formalization remains a future research priority.

Conclusions. The findings can be summarized in four points: the protocol transfers, ex-ante bias is replicated, the premium diverges and reconciliation provides a diagnosis. The two cases provide evidence that the load-level measurement procedure is transferable across distinct asset classes. The economic return from segregation, however, does not transfer at a constant rate: the results are consistent with the hypothesis that the premium depends on material composition. The closure error adds a diagnostic function for record integrity and quality. Full monotonicity, general causality, statistical generalization and automatic transferability to complex offshore assets remain beyond the scope of this evidence. The proposed next steps are a multi-asset panel with statistical inference and incorporation of actual offshore components. Formalizing load-level predictive modeling, with out-of-sample validation and leakage control, and process mining using formal event logs are future research priorities subordinate to the present contribution.

Keywords: Ship and port dismantling. Load-level physical and financial measurement. Steel circular economy. Multiple-case study. Data science.

References

IMO. Hong Kong International Convention for the Safe and Environmentally Sound Recycling of Ships. London: IMO, 2009.

MAGALHÃES, G. G.; BAIÃO, F. A. Mensuração físico-financeira de aço recuperado em desmantelamento naval-portuário: evidência empírica e camada MRV auditável [Physical and financial measurement of steel recovered in ship and port dismantling: empirical evidence and an auditable MRV layer]. SBPO, 2026.

MARTINS, I. D. et al. A review of the multicriteria decision analysis applied to oil and gas decommissioning problems. Ocean & Coastal Management, vol. 184, 105000, 2020.

ÖNAL, M.; KARAKAYA, E.; SONER, M. Life cycle assessment of ship recycling alternatives. Journal of Cleaner Production, vol. 384, 135623, 2023.

WORLD BANK. Digital Measurement, Reporting and Verification (D-MRV). Washington, DC, 2022.

YIN, R. K. Case Study Research and Applications. 6th ed. Thousand Oaks: Sage, 2018.