

Supplementary Appendix to “Zombie Capacity That Won’t Clear: Behavioral Anchoring and Rigid Supply in Offshore Drilling, 2002–2025”

Kang Lixue (terreck)

1 Appendix: Data-Construction Flow, Variable Dictionary, Bai-Perron Setting, and External Data Gaps

2 Appendix A. Data Construction Pipeline (reproducible steps)

The entire paper regenerates from raw 10-K filings via `make_all` in the project repository. The pipeline:

1. **PROBE.** One filing per contractor per era is sampled to reverse-engineer the disclosure schema (E1/E2/E3 layout detection).
2. **PARSE.** A single parser ingests `.txt` fixed-width SGML (E1), `.htm` expanded tables (E2), and inline-XBRL-as-HTML (E3), normalizing rig-class against the controlled vocabulary of DDL Q1.
3. **INGEST.** All 110 collected 10-K filings (per our EDGAR retrieval; the auditable unit is the 1,174-row de-duplicated table) are parsed into long operating cells.
4. **LINEAGE.** Cells are keyed on (`group`, `report_date`, `rig_class`) *after* corporate-identity reconciliation (§3.3 maps: VAL=Ensco+Rowan; ATW→RDC; HERC↔PGN; ESV/PACD standalone). The 2011 RIG jacking column-misalignment fix is pinned in the parser.
5. **DEDUP + FLAG.** Duplicate keys are removed (610 candidate → 477 breakeven-valid), and every missing value is tagged `INDETERMINATE` / `pending_collection` — never coerced.
6. **BUILD.** `Batch3_final_test_table.csv` ($1,174 \times 35$, 0 dup keys) and the P2 matrix are emitted.

3 Appendix B. Variable Dictionary

Variable	Type	Definition
<code>breakeven_gap_quarterly</code>	float	dayrate – OPEX-derived breakeven
<code>utilization</code>	float	contracted utilization %
<code>diseq_flag_v2</code>	bool	price<0 OR util<80 for $\geq 2q$
<code>price_diseq / qty_diseq</code>	bool	component signals
<code>k_break_within_4q_v2</code>	bool	Bai-Perron break within 4q
<code>G_credit_regime</code>	cat	normal/expansion/stress/severe
<code>zombie_credit_stressed_candidate</code>	bool	credit condition TRUE
<code>zombie_status</code>	cat	INDETERMINATE / INDETERMINATE-BUT-HEALTHY
<code>B_retired_ge_15pct</code>	cat	NEEDS_EXTERNAL_RIGCOUNT / INDETERMINATE-BUT-INDUSTRY-ST
<code>data_gap_flag</code>	bool	any missing core field

4 Appendix C. Bai-Perron Settings

- Algorithm: Bai & Perron (1998, 2003) global OLS with `max breaks = 5`. - Trimming: **0.15** (minimum segment = 15% of T). - Break count selected by **BIC**. - Series: contractor dayrate, deflated by an offshore CAPEX index (real, not nominal). - Behavioral read: a break is "behavioral" only if it survives in dayrate residuals after controlling for Brent oil price. - H2 auxiliary: prior-cycle peak dayrate; structural downshift if rate stays $\geq 15\%$ below peak (i.e., $\geq 85\%$) for 12 of 24 months post-break; sensitivity at 80% / 90%.

All definitions are identical to those used in the main text and are reproduced here for reference.

5 Appendix D. External Data Gaps and Sources

- **GFC price fill (RIG)**: external 10-K / fleet exhibits, 2008–09 dayrate \$228k $\rightarrow$ \$256k. - **PGN / SDRL / VAL external price**: company disclosures, investor presentations, bankruptcy filings, and financial press (2014–2017). Directional corroboration only — not a continuous panel. - **Industry rig-count proxy**: Baker Hughes-style count not in local panel; external proxy (91 rows) returns TRUE 15 / FALSE 11 / INDETERMINATE 65; does **not** support $\geq 15\%$ industry retirement. - **Credit source gap**: SLOOS / OAS feed ends 2026Q3; 5 rows (2026Q4) pending.

6 Appendix E. Missingness Audit (selection-bias check)

We report the count of INDETERMINATE / `pending_collection` cells per firm for the four core operational variables, computed directly from `Batch3_final_test_table.csv` (1,174 rows). This is the audit referenced in §3.6 and §4.5.

Firm	Rows	breakeven_gap_q	utilization	price_diseq_v2	diseq_flag_v2
ATW	19	19	19	19	19
DO	44	5	0	5	1
NE	239	236	239	236	239
PGN	45	45	45	45	45
RDC	96	91	16	91	39
RIG	462	32	68	32	43
SDRL	43	43	0	43	34
VAL	226	226	0	226	117
Total	1,174	697	387	697	537

The pattern is not random. Breakeven / price data are *entirely* missing for the four INDETERMINATE firms (ATW, NE, PGN, VAL) and for SDRL, while the only firms with complete price data (RIG, DO) are the large, continuously-disclosing contractors. Consequently the *resolved* sample — on which every disequilibrium count in the main text rests — over-represents well-disclosed, healthier contractors. This is a **selection bias** (missing-not-at-random). If distress reduces disclosure completeness, the cross-firm statistics that show disequilibrium concentrated in non-crisis credit regimes could be biased toward the null of no credit effect. We cannot sign this bias from the parsed table; we therefore flag it as the principal limitation of the non-credit-origin inference

and rest that inference primarily on the RIG-internal decoupling (which uses only RIG's complete panel).