Ports Output & Service KPIs — Integrated Data Plan (Excel + Yearbooks + Service PDFs) [Updated]

> One page to see **exactly what we can extract** from our three sources, with **units, granularity, coverage**, the **tidy tables** we'll build, and **how each variable plugs into the econometrics** (Y, M=ln K/L, Z, C).

---

0) Study spine (one-line reminder)

Question: Did competition/privatization raise **labor productivity (LP)** in container ops, and how much is **mediated by K/L**?
Units: Haifa & Ashdod (terminals where relevant).
Clocks: Bayport 2021-09; HCT Phase-A 2022-01 (STS 2022-09); Haifa privatization 2023-01.

---

1) What we can extract — by source (metrics • units • granularity • coverage)

A) Excel workbook — 'סה"כ מטענים – כולל מקשרים ימיים' (primary for monthly output)

What's in there (containers): - **TEU_total** — **TEU**, **monthly** (01..12 by year) by **port** and often **terminal** (e.g., Haifa legacy, "נמל המפרץ"/SIPG_Haifa, Ashdod legacy, "נמל הדרום"/HCT, Eilat).
- **TEU_full, TEU_empty** — **TEU**, **monthly** (composition).
- **TEU_import, TEU_export** — **TEU**, **monthly** *(if present on the sheet)*.
- **TEU_transshipment** — **TEU**, **monthly** *(if present)*.
- **Container units** — **boxes (units)**, typically **annual** or **period totals** (used for TEU/box mix).
- **Container tons** — **k-tons**, sometimes by load/unload (use for tons/TEU control).

Broader cargo & context (from other tabs/blocks in the workbook): - **Total cargo (all types)** — **k-tons**, **monthly** by port (and Haifa sub-entities where listed).
- **Loaded vs Unloaded** — **k-tons**, **monthly/annual**.

Coverage: At least **2018–2024 monthly** (earlier years may exist for some tables).
Why it matters: This is our **Tier-1 monthly output (Y)** with key **composition splits**.

B) Yearly Statistical Reports — שנתון סטטיסטי לשיט ונמלים (2013–2024)

What's in there (annual; cross-check & extras): - **TEU_total per port** — **TEU**, **annual** (cross-check/backfill).
- **Full/Empty; Import/Export** — **TEU**, **annual** (composition at annual level).
- **Vessel calls by cargo and/or ship class** — **count**, **annual** (mix/intensity).
- **Container units & container tons** — **units / k-tons**, **annual** (TEU/box, tons/TEU composition controls).

- **Selected service KPIs** (in some editions): waiting/berth/port times — **hours**, **annual**.
- **Definitions & metadata** — canonical **variable descriptions** and any **method changes**.

Coverage: Typically **2013–2024 annual**; some series go earlier.
Why it matters: Audit/backfill for TEU; **annual mix controls**; **data dictionary** for consistent definitions.

C) Service KPI PDFs (2024 – מדדי שירות בנמלי ישראל) — five cargo families

Files: Containers • General Cargo • Dry Bulk • Automated Facilities • Cargo Quantities Snapshot.
Granularity: Mostly **annual** by port/terminal, ≈**2014–2024**.

What's in there (by cargo family): - **Calls** (מס׳ ביקורים) — **count**, **annual**; sometimes with **ship-length bins** (mix).
- **Avg team allocation per ship** (הקצאת צוותים ממוצעת) — **teams/ship (level)**, **annual** (staffing intensity).
- **Waiting time / Time at berth / Time in port** — **hours**, **annual** (congestion/service quality).
- **Productivity per work-hour (containers)** — **TEU/hour**, **annual** (direct LP proxy).
- **Moves per crew/crane hour** — **moves/hour**, **annual** (operations LP).
- **Productivity per ship-hour** — **TEU/ship-hour**, **annual** (utilization × ops).
- **Crew responsiveness** — **%**, **annual** (staffing adequacy/utilization).
- **Cargo Quantities snapshot (latest year)** — by port: **containers (k-TEU, units, k-tons)**; **general/dry/auto facilities k-tons**.

Coverage: Ports show **Haifa legacy**, **SIPG/Bayport**, **Ashdod legacy**, **HCT/Southport**, **Eilat**, **מספנות ישראל** (Israel Shipyards) where applicable.
Why it matters: (i) **Build L** from TEU per work-hour; (ii) **new outcomes** for mechanism tests; (iii) **congestion & mix controls**.

---

2) How each feeds our variables (Y, M=ln K/L, K, L, Z, C)

Output Y (primary)

- **Monthly TEU_total** from Excel → `ln(TEU_total)` at **port×month**.
- **Optional Y variants**: TEU_full, TEU_import, TEU_transship as alternative outcomes/controls.

Labor L (proxy, monthly)

- From **Service PDFs (containers):** take **prod_per_work_hour_teu (TEU/hour) by port×year**.
  Compute **annual hours**: `L_year = TEU_year / prod_per_work_hour_teu_year`.
  Downscale to **monthly**: `L_month = L_year × (TEU_month / TEU_year)` (or weight by ship-hours).
- **Extras**: avg team allocation & calls can refine within-year distribution (sensitivity checks).

Capital K and mediator M = ln(K/L)

- **K_index (monthly)** (built elsewhere): STS in service, yard cranes, deep-draft berth meters (automation in Z at baseline).
- **M**: `ln(K_index) − ln(L_hours_monthly_from_ops)`.

Instruments Z (first stage for M)

- Dated engineering milestones: STS commissioning, yard crane commissioning, dredging/depth completion, OCR/TOS live.
- (Not supplied by these three sources; compiled from engineering/news/admin records.)

Controls C

- **Composition:** full share, import share, transshipment share (Excel); **TEU/box** & **tons/TEU** (Yearbook/Quantities).
- **Congestion & utilization:** wait/berth/port times; crew responsiveness; **calls** & **ship-length mix** (Service PDFs).
- **Macro load:** total cargo tons monthly (Excel).
- **FE & shocks:** month FE, COVID/war pulses, spillover (other-port-post), seasonality.

---

3) Tidy datasets to produce (exact schemas)

3.1 From Excel → `Data/Output/TEU/`

1) `teu_monthly_by_port.tsv`
`month \t port \t terminal \t teu_total`

2) `teu_full_empty_monthly.tsv`
`month \t port \t terminal \t teu_full \t teu_empty`

3) `teu_import_export_monthly.tsv` *(if present)*
`month \t port \t terminal \t teu_import \t teu_export`

4) `teu_transship_monthly.tsv` *(if present)*
`month \t port \t terminal \t teu_transship`

5) `total_cargo_tons_monthly.tsv` *(context)*
`month \t port \t sub_entity \t tons_k_total \t tons_k_loaded \t tons_k_unloaded`

> **Heb → Eng mapping (canonical):** `חיפה→Haifa`, `נמל המפרץ→SIPG_Haifa`, `אשדוד→Ashdod`, `נמל הדרום→HCT`, `אילת→Eilat`, `סה"כ נמלים→Ports_total`.

3.2 From Service PDFs → `Data/Output/ServiceKPIs/`

- `service_kpis_annual.tsv`
  `year \t port \t cargo_type \t metric \t value`
  where `cargo_type ∈ {containers, general, bulk, automated}` and `metric ∈ {calls, avg_team_allocation, wait_time_h, berth_time_h, time_in_port_h, prod_per_work_hour_teu, moves_per_crew_hour, prod_per_ship_hour_teu, crew_responsiveness_pct}`.

- `container_mix_annual.tsv` (from Cargo Quantities &/or Yearbook)
  `year \t port \t teu_total \t container_units \t container_tons_k`.

- *(Optional if visible)* `calls_by_ship_length_annual.tsv`
  `year \t port \t cargo_type \t ship_length_bin_m \t calls`.

3.3 From Yearbooks → `Data/Output/Yearbook/`

- `yearbook_teu_annual.tsv`
  `year \t port \t teu_total \t teu_full \t teu_empty \t teu_import \t teu_export`
  (as available).
- `yearbook_calls_annual.tsv`
  `year \t port \t cargo_type \t calls_total` (and by ship class if published).
- `yearbook_container_mix_annual.tsv`
  `year \t port \t container_units \t container_tons_k`.
- `definitions.md`
  Notes on variable definitions and any changes across editions.

3.4 Derived labor series → `Data/L_proxy/`

- `L_hours_monthly_from_ops.tsv`
  Built via: (1) compute `L_year = TEU_year / prod_per_work_hour_teu_year`; (2) allocate to months by TEU share (or ship-hour share); (3) record allocation rule and QA.

---

4) How it plugs into the econometrics

- **Primary Y:** `ln(TEU_total)` (or `ln(TEU_total/L_hours)` once L is built).
- **Operational Y (mechanisms):** `prod_per_work_hour_teu`, `prod_per_ship_hour_teu`, `moves_per_crew_hour`, `wait_time_h`, `berth_time_h`, `time_in_port_h`.
- **Mediator:** `M = ln(K_index) − ln(L_hours_monthly_from_ops)`.
- **First stage (Z → M):** use engineering milestones; include treatment dummies (competition, privatization) and controls.
- **Controls:** composition, congestion/utilization, vessel mix, macro load, FE/shocks.
- **Design:** Sun–Abraham / Callaway–Sant'Anna ES; stacked ES; local RDiT; IV-mediation.

---

5) QA & reconciliation

- **Within-Excel:** months → quarters/annual; `Ports_total` equals sum of ports; terminal open-date checks (zeros before opening).
- **Across sources:** sum months → Yearbook annual; spot-check against CBS (quarterly) where applicable.
- **Event sanity:** first positive TEU for **SIPG** ≈ 2021-09; **HCT** ≈ 2022-01/09 (STS 2022-09). Investigate deviations.

---

6) What we still need (open gaps)

- **Monthly L (ground truth):** direct **hours** or **overtime** series by port/terminal (if obtainable).
- **Monthly ship-hours at berth** (for alternative L or utilization weights).
- **Crane availability/rates** (to enrich K_index and ops controls).
- **Dredging/depth monthly** (to solidify deep-draft meter stock component).
- **Exact terminal first-call dates** (Ashdod P21 East confirmation for container ops clock).

---

7) Next concrete steps

1) Parse Excel container tabs → emit **four TEU TSVs** + **total_cargo_tons_monthly.tsv**.
2) Parse Service PDFs → emit **service_kpis_annual.tsv**, **container_mix_annual.tsv**, *(optionally) calls_by_ship_length_annual.tsv*.
3) Build **L_hours_monthly_from_ops.tsv** from TEU & `prod_per_work_hour_teu`; document allocation rule and sensitivities.
4) Aggregate Yearbook tables → `yearbook_*` TSVs and `definitions.md`.
5) Join into `panel_monthly.tsv`: TEU, composition, **L_hours**, **K_index**, treatments, controls.
6) Plot ES for **TEU**, **service times**, **ops LP**; run first-stage (Z → ln K/L), then main ES & IV-mediation.

Appendix — Glossary & mapping

- **Y (output):** ln TEU (monthly); variants: ops LP, service times.
- **M (mediator):** ln K/L.
- **K (capital services):** STS, yard cranes, deep-draft meters (weighted index).
- **L (labor):** monthly hours (proxy from TEU per work-hour).
- **Z (instruments):** commissioning milestones (STS/yard/dredging/OCR-TOS).
- **C (controls):** composition, congestion, vessel mix, macro load, FE/shocks.
- **Ports:** `חיפה=Haifa`, `נמל המפרץ=SIPG_Haifa`, `אשדוד=Ashdod`, `נמל הדרום=HCT`, `אילת=Eilat`, `מספנות ישראל=Israel_Shipyards`.