LP Data Map & Build Contract — Econ H191

This document is a **drop-in briefing** for any assistant or collaborator so they can immediately understand the data sources and how we build labor productivity (LP) series for Israel's port reform project. It covers **where the data live**, **how to read them**, **what's port vs terminal**, **the pre/post-reform split**, and **the six LP series** we maintain.

---

0) Scope & Conventions

- **Ports in scope:** `Ashdod`, `Haifa`. (Ignore **Eilat** everywhere.)
- **Terminals (post-reform):** `Ashdod-HCT` (ACH/TIL), `Ashdod-Legacy`, `Haifa-SIPG` (Bayport), `Haifa-Legacy`.
- **Granularities:**
- **Monthly** series for **ports** up to the reform cutover (pre-reform).
- **Quarterly** series for **terminals** after the reform cutover (post-reform), because TEU are available only quarterly at the terminal level.
- **Date parsing:**
- `Month-Year` like `03-2020` → `(year=2020, month=3)`.
- `MonthIndex` like `202003` → `(year=2020, month=3)`.
- Quarter labels: `{Q1,Q2,Q3,Q4}` with quarter-end months `{3,6,9,12}`.
- **Units:**
- Tons in the "tons" file are **thousands of tons** (`tons_k`). Multiply by **1,000** to obtain **tons**.
- TEU values use `TEU` when present; if missing, use `TEU_thousands × 1,000`.

---

1) Data Sources (no intermediate "normalization" required)

A) Monthly tons (ports and terminals)

Path: `Data/Output/monthly_output_by_1000_tons_ports_and_terminals.tsv`
Columns: `PortOrTerminal`, `Month-Year`, `tons_k`

Semantics: - `PortOrTerminal = Ashdod` or `Haifa` → **port-level** monthly tons. - `PortOrTerminal = Ashdod HCT` or `Haifa SIPG` → **terminal-level** monthly tons. - `PortOrTerminal = All Ports` → ignore.

Rule to avoid double-counting: - For a given **port-month**, if terminal rows exist (e.g., `Haifa SIPG`, `Haifa-Legacy`), **sum terminals** to obtain the **port tons**. Otherwise, use the single port row.

B) Mixed-frequency TEU (monthly *and* quarterly in one file)

Path: `Data/Output/teu_monthly_plus_quarterly_by_port.tsv`
Columns: `Port`, `Period`, `Freq`, `Year`, `MonthIndex`, `TEU_thousands`, `TEU`

Semantics: - `Freq = Monthly` & `Port ∈ {Ashdod, Haifa}` → **pre-reform monthly TEU** at the **port** level. - `Freq = Quarterly` & `Port` equals one of: - `Haifa` → **Haifa-Legacy** terminal quarterly TEU. - `Haifa SIPG` → **Haifa-SIPG (Bayport)** terminal quarterly TEU. - `Ashdod` → **Ashdod-Legacy** terminal quarterly TEU. - `Ashdod HCT` → **Ashdod-HCT (ACH/TIL)** terminal quarterly TEU.

Port-quarter TEU (for w at the port level, post-reform): - Compute `TEU_port_q = Σ_terminal TEU_terminal_q` (Legacy + Private entrant for the same port and quarter).

C) Terminal×month Labor & Π (L_Proxy)

Path: `Data/L_proxy/L_Proxy.tsv`
Key columns: `port`, `terminal`, `year`, `month`, `quarter`, `L_hours_i_m`, `Pi_teu_per_hour_i_y`, plus useful mix fields (`TEU_i_m`, `share_i_p_q`, etc.).

Semantics: - Grain is **terminal × month** and already contains: - **Annual Π_{iy}** (`Pi_teu_per_hour_i_y`) — terminal-year intrinsic productivity in TEU per hour. - **Monthly labor hours** (`L_hours_i_m`). - **Mix fields** (e.g., `share_i_p_q`) helpful for Π mix base. - **Port labor** at month: sum `L_hours_i_m` across terminals of that port-month.

---

2) What is “port” vs “terminal” in each file?

- **Tons file:**
- `Ashdod`, `Haifa` → **port**; `Ashdod HCT`, `Haifa SIPG` → **terminal**.
- **TEU file (Quarterly):** the `Port` value names the **terminal** in post-reform:
- `Haifa` ↔ Haifa-Legacy; `Haifa SIPG` ↔ Haifa-SIPG; `Ashdod` ↔ Ashdod-Legacy; `Ashdod HCT` ↔ Ashdod-HCT.
- **TEU file (Monthly):** `Port ∈ {Ashdod, Haifa}` with `Freq=Monthly` ↔ **port** (pre-reform only).

Always ignore rows for **Eilat** and **All Ports**.

---

3) LP Construction (math stays the same)

3.1 Notation

- `tons_{p,m}`: port-month tons (sum terminals if present; else port row), from tons file × 1,000.
- `TEU_{p,m}`: port-month TEU, pre-reform only (from TEU file, monthly).
- `TEU_{i,q}`: terminal-quarter TEU (from TEU file, quarterly).

- `TEU_{p,q} = Σ_i TEU_{i,q}` (sum of terminal quarterlies for a port).
- Π_{iy}: annual intrinsic productivity for terminal i in year y (from L_Proxy).
- `shares_{i,p,q}`: terminal shares within (port, quarter), usually from L_Proxy (`share_i_p_q`) or computed from terminal TEU.

3.2 Monthly port LP (pre-reform)

1) **Monthly ratio:** `r_{p,m} = tons_{p,m} / TEU_{p,m}`. 2) **Winsorize & rebase** within `(port, year)`: - Winsorize `r` at 1–99%. - Rebase to mean 1 per `(port, year)` → `w_{p,m}`. 3) **Π mix base (month):** quarter-constant terminal shares applied to annual Π_{iy} → `Π_{p,m}`. 4) **LP:** `LP_port,m = w_{p,m} × Π_{p,m}`.

3.3 Quarterly terminal LP (post-reform)

1) **Quarterly ratio at port level:**
`r_{p,q} = (Σ_{m∈q} tons_{p,m}) / TEU_{p,q}` where `TEU_{p,q} = Σ_i TEU_{i,q}`. 2) **Winsorize & rebase** within `(port, year)` to get `w_{p,q}`. 3) **Terminal LP:** `LP_{i,q} = w_{p(i),q} × Π_{iy}` (use Π_{iy} from L_Proxy; month→year mapping is straightforward).

> The identity diagnostic `LP_id = TEU_port,m / Σ_i L_hours_i,m` is optional and used only for QA.

---

4) The Six LP Series We Maintain

1) **Haifa (port) — monthly**
Range: 2018-01 → 2021-08
Inputs: tons_{p,m}, TEU_{p,m} (monthly); Π from L_Proxy
Output: `LP_port_m (Haifa)`

2) **Haifa-Legacy (terminal) — quarterly**
Range: 2021-Q3 → 2024-Q4
Inputs: TEU_{Haifa,q} (quarterly terminal), port-level tons (summed quarterly), Π_{iy}
Output: `LP_i_q (Haifa-Legacy)`

3) **Haifa-SIPG / Bayport (terminal) — quarterly**
Range: 2021-Q3 → 2024-Q4
Inputs: TEU_{Haifa SIPG,q}, port-level tons (quarterly), Π_{iy}
Output: `LP_i_q (Haifa-SIPG)`

4) **Ashdod (port) — monthly**
Range: 2018-01 → 2021-08
Inputs: tons_{p,m}, TEU_{p,m} (monthly); Π from L_Proxy
Output: `LP_port_m (Ashdod)`

5) **Ashdod-Legacy (terminal) — quarterly**
Range: 2021-Q3 → 2024-Q4
Inputs: TEU_{Ashdod,q} (terminal quarterly), port-level tons (quarterly), Π_{iy}
Output: `LP_i_q (Ashdod-Legacy)`

6) **Ashdod-HCT / ACH / TIL (terminal) — quarterly**
Range: 2021-Q3 → 2024-Q4
Inputs: TEU_{Ashdod HCT,q}, port-level tons (quarterly), Π_{iy}
Output: `LP_i_q (Ashdod-HCT)`

> **Important:** For quarterly **w**, always use **port-level** TEU (`Σ_i TEU_{i,q}`) and **port-level** quarter tons (`Σ_m tons_{p,m}` over the three months in the quarter). Terminal LP multiplies the **port's** `w_{p,q}` by the **terminal's** Π_{iy}.

---

5) Practical Rules & Edge Cases

- **Eilat & All Ports:** drop always.
- **Tons precedence:** when terminal tons exist for a port-month, use the **sum of terminal tons**; else use the **port row**.
- **Zeros & negatives:** treat TEU$\leq$0 or tons$\leq$0 as missing for the `r` ratio; avoid divide-by-zero.
- **DTypes:** keep `year`, `month` as `Int64` (nullable), ratios as `float64`.
- **Keys:**
- Port-month tables: `(port, year, month)` unique.
- Terminal-quarter tables: `(port, terminal, year, quarter)` unique.
- **Winsorization:** default `[1%, 99%]` within `(port, year)`; rebase to mean 1 (same group).
- **QA to run after build:** 1) **Key uniqueness** at the stated grains. 2) **Annual preservation:** `mean_m(LP_port) ≈ mean_m(Π_port)` within `(port, year)` where Π exists. 3) **Coverage report:** counts by series, NA rates for `w`, `Π`, `LP`. 4) **Post-reform check:** for each `(port, year)`, verify quarterly TEU exist for both terminals or flag.

---

6) Output Artifacts (suggested)

- `LP_Haifa_port_month.tsv`
- `LP_Ashdod_port_month.tsv`
- `LP_Haifa_Legacy_quarter.tsv`
- `LP_Haifa_SIPG_quarter.tsv`
- `LP_Ashdod_Legacy_quarter.tsv`
- `LP_Ashdod_HCT_quarter.tsv`
- `LP_panel_long.tsv` (optional convenience: stack the six into one long panel with `series_id`, `freq ∈ {M,Q}`)
- `qa_lp_report.tsv` (coverage + checks)

Each row should include: `series_id`, `port`, `terminal` (if any), `year`, `month` (or `quarter`), `month_index` (if monthly), `w`, `w_source` (port monthly or port quarterly), `Pi`, `LP`, and any provenance fields (`tons_source` when applicable).

---

7) One-page Mental Model

- **Before reform:** ports are single production units → monthly TEU and tons exist → compute `w_{p,m}` at month; mix-adjust via Π; output **monthly port LP**.
- **After reform:** ports split into terminals, and TEU become **quarterly** per terminal → compute `w_{p,q}` at the **port** level (quarterly tons / port-quarter TEU) → multiply by each terminal's Π_{iy} → output **quarterly terminal LP**.

This keeps the **economics** constant (LP = `w × Π`) while respecting the **measurement** changes introduced by reform.