

# GALILEO

## Technical Measurement Framework

Galileo fits Hill saturation curves with geometric adstock, detects cross-channel interactions via a 3-method statistical cascade, optimizes budget allocation, and generates credibility-audited reports. It is designed for analysts who need a fast first-pass MMM or a structured second opinion on existing models.

## The Measurement Problem

Marketing mix modeling attempts to decompose revenue into channel contributions using observational data. The core challenges are well-known: multicollinearity between correlated channel spends, confounding from seasonality and external factors, systematic bias in platform self-reported attribution (typically 40–80% overclaim vs. incrementality benchmarks), and small sample sizes (most advertisers have 52–104 weeks of clean data).

Galileo does not claim to solve these problems. It applies a structured approach to manage them, with explicit uncertainty quantification and automated quality gates.

## Model Architecture

### Response Function

Each channel is modeled as:  $\text{response}(t) = \text{beta} \times \text{Hill}(\text{adstock}(\text{spend}(t)))$ . Geometric adstock captures temporal carryover ( $\lambda$  in  $[0,1]$ ). Hill saturation ( $y = x^S / (K^S + x^S)$ ) captures diminishing returns. This formulation is standard in the MMM literature (Jin et al., Chan & Perry).

### Parameter Estimation

**Phase 1: Joint K/S/lambda optimization.** Nonlinear least squares via L-BFGS-B with 12 multi-start initializations. K prior centered at raw spend median (not adstocked median, which biases upward for high-lambda channels). S prior near 1.5. Lambda prior at industry benchmarks (TV ~0.80, search ~0.25).

**Phase 2: Ridge regression for betas.** Information-scaled ridge (70% OLS-proportional + 30% equal-share blend). Addresses multicollinearity that makes OLS unstable in small-sample marketing data.

### Confidence Intervals

Deliberate design choice: **ridge point estimates + OLS standard errors**. Ridge produces stable betas but underestimates SEs. OLS SEs naturally include variance inflation from correlated regressors (VIF). The result: CIs wider than pure ridge would report, but more honest about the identification problem. In illustrative demos with known ground truth, these CIs have shown strong coverage properties, though real-world coverage depends on the degree to which model assumptions hold.

## Structured Incorporation of Domain Knowledge

An optional LLM-conducted interview extracts domain knowledge from the marketing team and translates it into model priors. Every extracted prior passes through a 5-check validation gate before incorporation:

| Knowledge Type | Example | Model Use |
|----------------|---------|-----------|
|----------------|---------|-----------|

|                     |                                                          |                                       |
|---------------------|----------------------------------------------------------|---------------------------------------|
| Natural experiments | Paused Instagram for 2 weeks; branded search dropped 30% | Prior on Instagram→Search interaction |
| Lift studies        | Meta CLS showed 1.8x incremental ROAS                    | Calibration constraint on Meta beta   |
| Confounders         | Outerwear brand — 45% of revenue is Oct-Dec              | Baseline seasonality control          |
| Saturation beliefs  | Diminishing returns on Display above \$50K/week          | Informative K prior for Display       |

Validation checks: impossible values, direction consistency, magnitude bounds, lag plausibility, evidence strength. Priors that fail are flagged or auto-corrected, not silently incorporated. The model runs without priors if the interview is skipped.

## Cross-Channel Interaction Detection

A 3-method cascade designed to minimize false positives:

| Method               | Approach                                             | Strength                        |
|----------------------|------------------------------------------------------|---------------------------------|
| Partial F-test       | Multiplicative Hill features of lagged channel pairs | Direct, strong effects          |
| Residual correlation | Fisher z-test with Bartlett effective N              | Subtle, persistent correlations |
| Block bootstrap      | 500 resamples, 4-week blocks, ridge-stabilized       | Robust to autocorrelation       |

Interaction detected only when **2 of 3 methods agree**. Client priors can contribute an additional vote. With N=52 weeks, the minimum detectable effect (MDE) at 80% power is approximately 0.40. Interactions below this threshold may exist but will not be reported as significant.

## Uncertainty and Identification Risks

- **Multicollinearity:** Ridge provides stability; OLS SEs capture VIF. Channels with VIF > 5 show wider CIs, appropriately reflecting identification difficulty.
- **CI interpretation:** 90% CIs mean “given the Hill + adstock + ridge specification, this range is consistent with the data.” They do not account for model misspecification, omitted variables, or structural breaks.
- **Weak identification:** High channel correlation (>0.8) makes individual betas poorly identified. Total marketing effect remains estimable. Fewer than 30 weeks results in very wide CIs and minimal interaction detection power.

## Web Application Architecture

Galileo is delivered as a FastAPI web application with a React frontend (no build step — JSX transpiled via Babel standalone). The pipeline runs as a subprocess, and all state is stored on disk.

| Layer    | Component                       | Detail                                                  |
|----------|---------------------------------|---------------------------------------------------------|
| Frontend | React 18 + Babel + Tailwind CDN | No build step; JSX transpiled in browser                |
| API      | FastAPI (uvicorn)               | Upload, interview (WebSocket), pipeline, report routers |
| Auth     | Session-based + bcrypt          | Time-limited accounts, job-level access control         |
| Storage  | On-disk (jobs/{id}/)            | No external database; atomic writes via tmp+replace     |
| Pipeline | Subprocess                      | 9-stage audit pipeline runs in isolated process         |

### Authentication and Data Lifecycle

User accounts are provisioned with explicit expiry dates. After expiry, authentication is rejected. Job data persists until an administrator runs the cleanup command, which enforces six safety invariants:

| Invariant                    | Purpose                                                         |
|------------------------------|-----------------------------------------------------------------|
| Path containment             | Resolved path must be a direct child of jobs/                   |
| Hex-8 directory names        | Only valid job directories (UUID prefixes) are candidates       |
| Owner verification at delete | Re-reads meta.json immediately before deletion (TOCTOU defense) |
| Owner-less protection        | Jobs with empty owner field are never deleted                   |
| Expiry filtering             | Only users past the configured grace period are targeted        |
| Symlink rejection            | Path.resolve() + explicit symlink check prevents traversal      |

Cleanup supports three modes: dry-run (preview), interactive (per-user confirmation), and force (for automated cron scheduling). All filesystem writes use atomic patterns (write to .tmp, then os.replace).

## What Galileo Does Not Do

- **Not a replacement for experiments.** Lift tests and geo-holdout studies provide causal evidence that observational models cannot match.
- **Not individual-level attribution.** This is an aggregate time-series model. It answers “how much revenue did TV drive?” not “which users converted because of a TV ad.”
- **Not a forecasting tool.** It estimates historical contributions. Forecast extensions exist in prototype but are not part of the standard audit.
- **Not creative optimization.** Operates at channel-budget level, not ad-set or creative level.
- **Does not guarantee correctness.** All observational models are subject to omitted variable bias. The credibility certificate flags known risk factors but cannot rule out unknown confounders.

## Limitations and Assumptions

| Limitation             | Impact                                          | Mitigation                                                     |
|------------------------|-------------------------------------------------|----------------------------------------------------------------|
| Small N (52 weeks)     | Wide CIs, low interaction power                 | Ridge stabilization, prior incorporation                       |
| Channel correlation    | Individual betas poorly identified              | Ridge, VIF-aware SEs, total effect reliable                    |
| Model misspecification | Biased betas if DGP differs from Hill+adstock   | Credibility gates, DW diagnostics, residual checks             |
| No competitor data     | Competitor spend absorbed into channel effects  | Flagged in caveats when detected                               |
| Demo validation only   | Real-world accuracy not independently validated | Ground truth tests with synthetic data; client pilots underway |

Demo results use synthetic data with known ground truth. They demonstrate that the procedure recovers planted parameters under favorable conditions. They should be interpreted as illustrative examples, not guarantees of accuracy on real data.

## Data Requirements

- Weekly granularity, 26+ weeks (52+ preferred for interaction detection)
- At least 2 spend channels + 1 revenue column
- Non-trivial variation in spend (constant-spend channels cannot be identified)
- Assumes stationarity, linearity in betas (post-transform), and no contemporaneous confounding beyond seasonality

## Deliverables

Channel contribution estimates with 90% CIs • Saturation curves with waste estimates • Cross-channel interaction map • Budget-neutral optimization • Counterfactual scenarios • Pattern discovery • Credibility certificate (40+ gates) • Verification checklist (8 checks)

Positioning vs. Alternatives

|                  | Galileo                    | Consultancy MMM           | Open-Source MMM          |
|------------------|----------------------------|---------------------------|--------------------------|
| Time             | Minutes (automated)        | 8–16 weeks                | Days–weeks               |
| Domain knowledge | Structured LLM interview   | Ad hoc meetings           | Manual configuration     |
| Interactions     | 3-method cascade           | Manual specification      | Typically not included   |
| Uncertainty      | Ridge + OLS SE (VIF-aware) | Bayesian posterior (MCMC) | Varies by implementation |
| Audit            | 40+ automated gates        | Consultant judgment       | Basic diagnostics        |
| Flexibility      | Fixed model class          | Fully custom              | Fully custom             |

# Getting Started

1. Prepare a weekly CSV with date, channel spend columns, and revenue
2. Upload through the web interface or API
3. Optionally complete a short conversational interview (5–10 minutes)
4. Receive interactive report + downloadable PDF + credibility certificate

For a pilot audit with your data: [anjan@smartinfer.ai](mailto:anjan@smartinfer.ai)