PERFORMANCE MARKETING LIBRARY / MODEL HANDBOOK

MMM Model Manual
Inputs, methodology, priors, validation, and interpretation

A practical and technical reference for the browser-side empirical-Bayes MMM. It explains what the model calculates, where uncertainty comes from, how experimental and country evidence is transferred, and where the result must not be interpreted causally.

Edition 1.0 / 2026-07-23
MMM methodology 1.1.0 / Korean and English parity edition
Processing policy: uploaded CSV data stays in browser memory and is not sent to a server.

Contents

How to use this manual

Operators can start with Chapters 1-4 and 10-12. Analysts and reviewers should also read Chapters 5-9 and 14-15. The final checklists are designed for presentation approval.

1. What the model does - and does not do

1.1 One-sentence definition

The in-app MMM decomposes weekly KPI variation into natural demand, trend, seasonality, events, and media contributions. It estimates channel effect direction, uncertainty, and response curves while accounting for carryover and saturation, entirely in the browser.

> Essential distinction
> The model borrows selected ideas from Meridian, but it is not Meridian. It is a conditional Gaussian empirical-Bayes approximation with plug-in residual variance, Normal priors, and weak ridge regularization; it is not a fully hierarchical NUTS/MCMC model.

1.2 Questions it can support

- Which channels show a positive relationship with a KPI after modeled trend, seasonality, 0/1 events, and regime steps, and how stable are the positive probability and 90% interval?
- Where is current spend on the response curve, and how much marginal response appears to remain?
- Does adding experimental or reference-country evidence improve held-out prediction, and how sensitive is the conclusion to that evidence?
- Which modeled drivers account for weekly KPI movement after trend, seasonality, and events are separated?
- How do conclusions change when the same media inputs are evaluated against any of five mapped KPIs?

1.3 Questions it must not claim to answer

- It cannot prove a channel's pure incrementality from observational data alone. Simultaneous movement and omitted confounders can make individual channels unidentified.
- A low positive probability is not proof of zero effect; it can be lack of information.
- The budget response is not a binding plan when auctions, inventory, contracts, or creative quality change.
- A reference-country coefficient is not transferable merely because a column name matches.
- High in-sample R2 does not establish causal validity.

1.4 Decision-use levels

Level	Conditions	Appropriate use	Required wording
Exploratory	Observational only; identification warnings	Prioritize experiments and data fixes	Provisional association-based result
Evidence-adjusted	Prior applied; evidence-specific checks pass	Ranged scenarios and channel comparison	Disclose prior and validation
Decision support	Experiment aligns; base OOS and health are sound	Incremental budget changes with measurement	Supporting evidence, not causal proof

2. Quick start: from upload to presentation

2.1 Ten-step workflow

1. Prepare date, channel spend, KPI, and 0/1 event/regime columns in one CSV.
2. Upload daily data directly; the app groups it into Monday-start weeks.
3. Map whichever of Traffic, Registrations, Reactivations, Revenue, and Purchasers are available.
4. Classify media as Performance or Brand and verify 0/1 event/regime columns.
5. Run the no-prior base model first.
6. Review predictive accuracy, interval coverage, residual autocorrelation, and negative natural demand.
7. If experimental data exists, use Auto detection or explicitly select On/Off or Geo and generate a prior.
8. If country panels exist, review eligible countries and the rolling 12-week recommendation.
9. Toggle priors on and off for the same KPI to assess sensitivity.
10. Present the data scope, model version, prior source, validation, and limitations together.

2.2 Data-length guidance

State	Suggested weeks	Use
Operational recommendation	104+	Best support for two-year seasonality, spend variation, and repeated validation
Analyzable	52-103	Model can run; budget action still requires identification gates
App low-information	Under 52	Exploratory effects only; budget guidance is blocked
Very short	Under 24	Hard to separate time structure, saturation, and media

> More rows are not automatically more informative
> A shorter period with separable channel movements can identify more than a long period in which every channel follows the same budget pattern.

2.3 Sixty-second preflight

- Is the KPI summable by week? Do not mix cumulative or cohort-window metrics into calendar-week outcomes.
- Do spend and KPI share geography, currency, time zone, and conversion window?
- Do target, experiment, and country-prior files map the same channel names?
- Are supported promotions, outages, and holidays encoded as exact 0/1 events or regime steps, and is the inability to map arbitrary continuous price/competition controls disclosed?
- Are any channels always switched on and off together?

3. MMM CSV input specification

3.1 Meaning of one row

A row can be a day or a week. Daily dates are automatically assigned to a Monday-start week and aggregated. A pre-aggregated weekly file can use its existing week identifier. With no time column, row-order t is allowed only after the user explicitly confirms that the file already contains one row per week ordered oldest to newest. Do not mix daily and weekly rows in one file.

3.2 Field dictionary

Role	Example header	Required	Weekly rule	Caution
Date/week	date or week	One, or confirm row order	Daily rows map to Monday	ISO date/week preferred
Country	country	Optional / required for market prior	Split by country	Exactly one target code
Platform/segment	platform	Optional	Split or aggregate Total	Keep Android/iOS definitions fixed
Media spend	meta_spend	One or more	Sum	Non-negative; same currency
Traffic	traffic, sessions	Optional Y	Sum	May be Reg + React if appropriate
Registrations	registrations	Optional Y	Sum	Stable deduplication
Reactivations	reactivations	Optional Y	Sum	Stable reactivation definition
Revenue	revenue	Optional Y	Sum	Currency/refund policy fixed
Purchasers	purchasers	Optional Y	Sum	Weekly unique preferred
Event/regime flag	holiday, promo	Optional	Occurred in week / step	Only numeric or recognized binary 0/1

Numeric inputs accept thousands separators, scientific notation, and explicit magnitude suffixes (K/M/B and Korean 천/만/백만/억). For example, 1.2M becomes 1,200,000 and 3억원 becomes 300,000,000. Ambiguous suffixes such as 1.2MB are rejected rather than silently becoming 1.2. Five-digit Excel date serials use the same 1900-system UTC conversion in the main MMM, experiment, and country-prior paths. A row-segmented platform file must map date or week so rows for the same period can be combined; row-order confirmation alone is not allowed.

3.3 Multiple KPI mapping

Only available, mapped targets become active. After analysis, the sticky target selector switches among Traffic, Registrations, Reactivations, Revenue, and Purchasers while retaining the same media/event/regime structure. If Traffic is absent but both Registrations and Reactivations are present, their sum can occupy the Traffic slot when the business definition permits it; the same derived value is not exposed again as a sixth RR/Total target.

> When the number of Y columns differs
> A KPI that exists in the target MMM file but is not mapped in the experiment or country-prior file receives no prior. That target remains a base observational model and must be labeled as such.

3.4 Matching rules

- To calibrate target meta_spend, map the prior file's corresponding channel to meta_spend as well.
- Do not trust automatic linking when case, whitespace, or prefixes differ; verify the mapper. Unicode headers use stable name-based internal keys, but semantic equivalence still requires human review.
- KPI mappings require the same business definition, not merely the same header.
- Align currency and KPI scale before transferring coefficients across countries.

3.5 Platform splits

Map a platform column or Android/iOS-tagged columns to switch between Total and platform-specific results. Total sums KPI and spend for the same week; Android and iOS are fitted as separate Y models. A large platform difference is not automatically pooled or borrowed across operating systems. Verify audience, pricing, and measurement-window differences and report separate conclusions when needed.

> Experiment priors stay at the same platform grain
> An Android or iOS result requires experiment rows for that same platform or an explicitly mapped Y tagged to that platform. Total sums every mapped platform Y, but a row is excluded when any required component Y is missing. The app never substitutes Total or another OS for the selected platform.

3.6 Long-format input

The app also accepts week/date | channel | spend | Y. Channel rows for the same period and segment must repeat the identical total Y and event/regime values; only spend varies by channel. Spend is pivoted into wide media columns, while repeated Y is never summed. Blank time/channel rows and conflicting repeated Y/event values block analysis instead of being silently dropped or resolved by taking the first row. Reversing source-row order therefore produces the same result or the same blocking diagnostic.

4. Daily-to-weekly preprocessing

4.1 Monday-start weeks

Every date is assigned to the Monday that starts its calendar week. Rows through Sunday are grouped and ordered in time. Weekly modeling avoids treating day-level noise as if it were longer advertising carryover.

```
week_start(d) = Monday of the calendar week containing date d
weekly_spend[w] = sum(spend[d]) for d in week w
weekly_kpi[w] = sum(kpi[d]) for d in week w
weekly_event[w] = 1 if any event[d] = 1 in week w, else 0
```

4.2 Sums versus occurrence

Spend, traffic, registrations, reactivations, revenue, and purchasers are flows and are summed. Binary holidays, promotions, and outages represent whether an event occurred in the week. Event and regime columns must be exactly numeric or recognized binary 0/1; a blank or another number blocks analysis rather than silently becoming zero. Three event days do not automatically imply an effect three times larger.

4.3 Missing values and partial weeks

Issue	App behavior	Operator action
Invalid/blank date	Drop row and raise a blocking issue	Fix source date
Duplicate date/week	Block at the same platform grain	Resolve grain and deduplicate
Missing week	Block on a calendar gap	Add actual KPI/spend row
Internal partial week	Block below modal 5/7-day cadence	Recover missing days
First/last partial	Exclude with a warning	Confirm complete-week scope
Blank spend/KPI	Do not silently sum as zero; block selected Y/channel	Use 0 only for verified no-spend
Negative media spend	Block analysis	Resolve refunds/corrections under the source definition
Event/step not binary 0/1	Block analysis	Correct the binary value; never fill a blank silently with zero
Rate KPI	No automatic repair	Redefine from numerator/denominator

What weekly aggregation cannot fix

It does not solve channel collinearity, omitted confounders, or inconsistent measurement. It only aligns time grain and reduces daily noise. The app does not directly map arbitrary continuous price or competition controls. Encode a defensible change as a 0/1 event or regime step when appropriate; otherwise disclose the omitted factor as a limitation.

5. Model architecture

5.1 Outcome equation

Each KPI is fitted separately. Natural-demand level and trend, seasonality, events/regime changes, and transformed media features add up to the modeled weekly KPI.

```
y[t] = intercept + trend[t] + seasonality[t] + events[t]
+ sum_m beta[m] * Hill(Adstock(spend[m,t])) + error[t]
error[t] ~ Normal(0, sigma^2)
```

5.2 Non-media components

Component	Role	Interpretation caution
Trend	Intercept plus long-run direction	Not identical to all organic demand
Seasonality	Fourier cycles for configured periods	Irregular events need explicit columns
Holidays & Events	Known shocks	Future events require scenario values
Regime change	Step changes in product/measurement/policy	Needs a defensible change date
Performance / Brand	Reported media contribution groups	Group totals do not ensure channel identification

The default seasonality grid is annual 52.18 weeks and quarterly 13.04 weeks, with one sine/cosine pair for each period. The app does not automatically create separate monthly, weekday, or country-calendar effects.

> No automatic local calendar or arbitrary steps
> The model does not hard-code Korean holiday weeks or insert arbitrary steps at week 42/55. It uses only event and regime columns mapped by the user. A collinear step is not silently deleted; it is surfaced in diagnostics so the analyst can retain, combine, or remove it based on the real change event.

5.3 Standardization and back-transformation

Features are centered and scaled for numerical stability, then coefficients and contributions are returned to the original transformed-feature and KPI units. Calibrated-prior precision is transformed consistently, so changing revenue units from currency units to thousands does not arbitrarily change relative prior strength.

```
z[j,t] = (x[j,t] - mean(x[j])) / sd(x[j])
beta_raw[j] = beta_standardized[j] / sd(x[j])
Var(beta_standardized prior) = Var(beta_raw prior) * sd(x[j])^2
```

5.4 Collinearity

When two channels move together at nearly a fixed ratio, the model may explain their combined movement but cannot reliably separate them. A prior adds information; it must not disguise non-identification. Aggregate channels, obtain more varied periods, or run a holdout/geo experiment.

6. Empirical-Bayes inference

6.1 Conditional Gaussian analytical approximation

The app does not draw thousands of posterior samples. It plug-in estimates residual variance sigma², then conditions on that value while combining a linear Gaussian likelihood with Normal priors and weak ridge penalties. Coefficient means and covariance are solved by matrix algebra. This is an empirical-Bayes analytical approximation, not a Student-t posterior integrating sigma² or a fully hierarchical joint posterior. Results are deterministic for the same input and fast enough for in-browser use.

```
Likelihood: y | beta ~ Normal(X beta, sigma^2 I)
Calibrated prior: beta[j] ~ Normal(mu[j], v[j])
Posterior precision = X'X / sigma^2 + prior_precision
Posterior mean = inverse(Posterior precision) * (X'y / sigma^2 + prior_precision * mu)
```

6.2 Weak regularization and calibrated priors

Coefficients without external evidence receive weak ridge regularization. Media coefficients with experimental or country evidence receive an actual Normal(mean, variance) precision. Residual scale is plug-in estimated and the prior penalty is fixed-point updated so that KPI unit changes do not distort the prior-data balance. Uncertainty in sigma² itself is not integrated.

6.3 Positive probability and the 90% interval

A prior-locked channel uses its conditional Gaussian coefficient mean and standard deviation for positive probability and its 90% interval. For a no-prior channel, transform-candidate probabilities are BIC-weighted and the 90% effect interval directly inverts the weighted Gaussian-mixture CDF at 5% and 95%, preserving asymmetry or separated candidate conclusions instead of replacing the mixture with a Normal approximation.

```
F_mix(x) = sum_k weight[k] * Phi((x - mean[k]) / sd[k])
CI90 = [inverse_F_mix(0.05), inverse_F_mix(0.95)]
```

No sampling diagnostics

Because there are no MCMC chains, R-hat, effective sample size, and trace plots cannot be computed. The product reports them as not applicable rather than manufacturing substitutes.

6.4 Trade-offs

Benefit	Trade-off
Fast, deterministic browser execution	No fully sampled nonlinear/hierarchical joint posterior
Explicit prior mean and variance	Prior does not repair misspecification or collinearity
Fast switching across KPIs	KPIs are separate models, not a multivariate outcome
No server upload	Large geo panels and huge candidate sets exceed browser comfort

7. Carryover, saturation, and model averaging

7.1 Geometric adstock

A fraction alpha of the previous media state carries into the next week. Higher alpha implies longer carryover. For Geo experiments, state is computed independently within each geo so the last week of one geo never leaks into the first week of another.

```
adstock[t] = spend[t] + alpha * adstock[t-1], 0 <= alpha < 1
```

7.2 Hill saturation

The half-saturation ec is the adstock level at half of the normalized maximum response; slope controls curvature. It represents diminishing marginal response as exposure grows.

```
hill(a; ec, slope) = a^slope / (a^slope + ec^slope)
```

7.3 Representative transform

The engine compares alpha x ec x slope candidates for every channel. The fixed grids are alpha = 0.0, 0.1, ..., 0.8; ec = the 40th, 60th, and 80th percentiles of that channel's positive adstock distribution; and slope = 0.6, 0.8, 1.0, 1.4. It selects a representative transform by the magnitude of its relationship with KPI residuals after trend, seasonality, 0/1 events, and regime steps; coefficient inference determines the sign later. That set drives decomposition and forecast. Profile-fit budget is 48 per channel and 480 total without priors, or 240 total when any external prior exists, divided evenly across no-prior channels that still require profiling.

7.4 One-channel-at-a-time profile averaging

For channels without priors, effect probability, intervals, response curves, and budget guidance do not rely on only the representative transform. That channel's candidates are refitted while other channels stay at their representatives, and BIC differences determine weights. On/Off intensity uses the target representative alpha/ec/slope. A Geo effect is first divided by its linear-adstock DiD effect and then converted to the target coefficient unit with the Hill derivative at the overlapping national target-adstock operating point. Each reference market is refitted with the same target parameters. A channel with external evidence is therefore locked to the target representative transform, and the same prior numbers are never reused under a different transform.

```
weight[k] proportional to exp(-0.5 * (BIC[k] - min(BIC)))
mean_mix = sum_k weight[k] * mean[k]
var_mix = sum_k weight[k] * (sd[k]^2 + mean[k]^2) - mean_mix^2
effective_candidates = 1 / sum_k weight[k]^2
```

Scope of averaging

The main effect estimate is a one-channel-at-a-time profile average. When transform uncertainty is high, the two most uncertain channels are additionally checked across at most four top-candidate combinations. This bounded joint diagnostic does not automatically replace coefficients and is not a full all-channel joint posterior.

7.5 Limited baseline-knot selection

With at least 78 valid weeks, the model compares 0-, 1-, and 2-knot baseline candidates. A hinge knot is applied only when BIC improves by at least 6; otherwise the base trend/Fourier/step structure is retained. Candidate count is intentionally limited because a flexible baseline can absorb genuine media changes.

7.6 Experiment-prior mROI display

Experiment priors remain applied in transformed-feature coefficient units. The UI also multiplies the coefficient prior by the Hill derivative at the current weekly spend to display expected KPI per added spend (mROI). This is a unit-converted view of the existing prior, not a separately sampled ROI posterior.

7.7 Steady-state response curves

The response-curve x-axis is a sustainable weekly spend level, not a single-week pulse. Each candidate therefore uses steady-state adstock = spend/(1-alpha). Changing alpha changes accumulated exposure and marginal response at the same weekly spend.

8. Experimental priors: On/Off and Geo

8.1 Why raw test data is useful

Raw experiment periods allow the app to estimate effect size and uncertainty from the same design. On/Off divides the KPI treatment effect by the treatment effect on exposure transformed with the target alpha, ec, and slope. Geo does not apply nonlinear Hill separately within each geo and then aggregate it. It first divides the KPI DiD by the geo-level linear-adstock DiD, then converts that return into the target coefficient unit with the Hill derivative at the national target-adstock operating point shared by the experiment and target periods. Cross-equation covariance, the delta method, a bounded Fieller ratio interval, and jackknife uncertainty are all propagated.

```
On/Off prior_mean = effect_KPI / effect_Hill(adstock)
Geo r_A = DiD_effect_KPI / DiD_effect_geo_adstock
Geo prior_mean = r_A / Hill_derivative(target_overlap_adstock)
Var(y/x) approximately Var(y)/x^2 + y^2 Var(x)/x^4
- 2*y*Cov(y,x)/x^3
require a bounded two-sided 90% Fieller ratio interval
prior_variance = max(delta_variance * inflation^2, Fieller_equivalent_variance,
jackknife_SE^2 * inflation^2)
prior_precision = 1 / prior_variance
```

8.2 Experiment-type detection and accepted layouts

The upload default is Auto. Auto first uses a type column when present, then looks for a Geo wide or Geo long schema, and otherwise falls back to On/Off. The type column must contain one interpretable file-wide value such as Geo or OnOff. If detection does not match the intended design, the Experiment type selector can force On/Off or Geo; this user choice overrides the type column and schema detection. The screen reports both the resolved type and its source.

Layout	Required columns	Normalization
On/Off wide	type (optional), time, state, KPI, tested channel spend	Used as one row per week
Geo long	type (optional), time, period, geo, arm, KPI, tested channel spend	Used as a geo x week panel
Geo wide	type (optional), time, period, target_geo, control_geo, paired target_/control_ KPI and spend	Each row becomes treatment and control geo rows
Long media	time, channel, spend plus design fields and repeated KPI	Pivots spend into headers matching the main MMM channels

Rename template fields to the actual mappings

The screen provides On/Off, Geo wide, and Geo long example CSV downloads. Replace registrations and meta_spend with the exact Y and channel headers mapped in the main MMM. A long-media channel value must also map to a main-MMM channel. If repeated KPI values disagree within the same time, geo, and platform, the app pauses instead of selecting the first row.

8.3 Building experiment weeks from daily rows

When dates are parseable, daily experiment rows are grouped by geo x Monday-start week, or by Monday-start week when geo is absent. KPI and tested-channel spend are summed. Invalid-date rows are dropped; a boundary week with mixed state/arm/period and a week with any missing KPI or spend source row are excluded in full. The modal cadence classifies five- versus seven-day feeds. First/last partial weeks are removed, while an internal partial week, missing period, or duplicate weekly period pauses the prior so adstock time is never compressed.

Diagnostic	Meaning
source rows	Raw uploaded row count
usable weeks	Geo-weeks/weeks retained for prior calculation
unparseable date rows	Rows excluded because the date could not be parsed
mixed boundary weeks	Weeks excluded because state/arm/period changed within the week
missing weeks	Weeks excluded because any source KPI/spend value was missing
partial/gap diagnostics	Boundary partial exclusions and blocking internal partial, missing-period, or duplicate-period reasons

Conservative exclusion is intentional
A boundary week is not forced into ON or POST by majority vote. Source rows, usable weeks, and every exclusion count are retained so the prior sample is reproducible.

A platform-specific prior uses only same-platform rows or the corresponding Android/iOS-tagged mapped Y. Total sums all mapped platform Y values, and a row is excluded if any component Y is missing. Total or another OS is never borrowed for the selected platform prior.

8.4 On/Off schema

Column	Meaning	Example
time	Date or week	2026-01-05
state	ON/OFF state	on, off
channel spend	Spend for the tested channel	meta_spend
KPI	Same target definition	registrations

Repeated ON-OFF-ON periods are required for useful state replication. The prior needs at least 16 validated weekly observations, at least four ON weeks, at least four OFF weeks, and at least two state transitions. A one-way switch or a 1-versus-15 split does not qualify. A state may be inferred only from verified observed spend zeros; missing spend is never treated as OFF. Time-series uncertainty uses HAC standard errors and a block jackknife. If HAC estimation fails, the app pauses instead of falling back to ordinary OLS standard errors. This converter does not regress on extra promotion or outage controls; disclose concurrent shocks separately.

8.5 Geo schema

Column	Meaning	Example
time	Aligned observation week	2026-W08
geo	Geographic unit	Seoul, Busan
arm	Target/control	treatment, control
period	Pre/post	pre, post
channel spend	Geo-level tested spend	search_spend
KPI	Geo-level outcome	revenue

A Geo prior requires a comparable date/week column, at least 16 validated geo-week observations, at least four geos, and at least two geos in each treatment and control arm. Each geo's arm must remain fixed; every week must have one common pre/post state across geos; every geo must appear in both pre and post; and all four treatment/control x pre/post cells must exist. The app estimates treatment x post DiD with geo and week fixed effects against geo-level linear adstock. It does not fall back to OLS standard errors when cluster covariance fails, and LOCO is accepted only if every leave-one-geo-out fit succeeds. The two-sided 90% Fieller interval for the KPI/intensity ratio must be bounded. At least four experiment weeks must overlap target MMM weeks to define the national target-adstock Hill derivative; no overlap or a near-zero derivative pauses unit alignment.

```
inflation = max(1, t_critical(0.90, df) / 1.645)
prior_variance = max(cluster_delta_variance * inflation^2, Fieller_equivalent_variance,
LOCO_SE^2 * inflation^2)
```

When multiple spend columns move in one holdout

When the file has several spend columns, the app first checks which ones move materially with state/arm/period. Exactly one clear channel can proceed; two or more moving together remain a bundle that cannot identify channel-specific effects. Even the selected channel is paused when transformed-exposure contrast is not distinguishable from zero, preventing an unstable ratio prior.

8.6 How the confidence interval controls strength

The UI does not accept only a summarized CI. It estimates the effect/intensity ratio from period-level rows, including cross-equation covariance, robust delta variance, a two-sided 90% Fieller ratio interval, and block or LOCO jackknife uncertainty. An unbounded or disjoint Fieller interval does not produce a prior. For a bounded interval, the farther endpoint from the mean is divided by 1.645 to form a conservative symmetric Normal-variance candidate. The widest of robust delta, Fieller-equivalent, and jackknife uncertainty becomes the prior variance. Wider intervals therefore reduce precision and let MMM data update the prior more easily; there is no arbitrary strength slider.

8.7 Same-KPI/time reuse and transfer checks

Do not put the same outcome into both prior and likelihood

Calendar overlap alone is not an automatic failure, but copying the same population's KPI values into both the experiment file and main MMM does not create independent evidence. For national On/Off, the app treats at least four comparable overlapping weeks with 95% or more numeric matches to the same-week main-MMM Y as outcome reuse and blocks the prior. If the two clocks themselves cannot be compared, such as calendar dates versus an unlabeled numeric index, the On/Off prior is also paused because reuse cannot be ruled out. Partial overlap and Geo evidence report the overlapping and exact-match week counts; the analyst must confirm that the effect came from separately identified treatment/control aggregation.

- Do experiment and MMM share KPI definition, conversion window, currency, and population?
- Is test intensity within the MMM's observed range?
- Did price, promotion, product, or creative change at the same time?
- Do tested channel and target channel map to the same key?
- Has uncertainty from timing/population transfer been acknowledged?

An experiment is not automatically a perfect prior

Google's Meridian calibration guidance also notes that estimand, time, and population can differ between an experiment and an MMM. A narrow test interval should not be forced onto the full history without a transfer assessment.

9. Reference-country priors

9.1 Purpose and schema

Upload multiple reference countries in the same format as the target MMM to discover transferable channel evidence. The main MMM CSV must contain exactly one COUNTRY value, the reference CSV needs a COUNTRY column, and every reference country requires at least 24 valid weekly rows. If the target cannot be identified, all market priors are paused to prevent self-reference. Matching headers only establish a technical mapping; the user must verify business definition, units, and conversion windows.

9.2 Two-stage selection

- Eligibility fits at most 12 reference countries after checking data sufficiency, overlapping media/KPI, timeline compatibility, and fit feasibility. Eligible countries remain individually visible.
- All eligible countries and later combinations use the same non-overlapping rolling-origin folds. This is more stable than one holdout, but reusing the folds for shortlisting and set selection does not remove selection bias and is not independent OOS.
- Only the top four countries on those common folds enter combinations of up to three countries and at most 24 sets.
- Selection considers fold instability, at least 2% improvement, fold win rate, whether paired-fold mean improvement exceeds one SE of its fold noise, country-count complexity, and a one-standard-error preference for a simpler set.
- No prior is the correct recommendation if the baseline wins or improvements are unstable.

9.3 Repeated 12-week validation

The default moves a 12-week window backward in 12-week steps, producing up to three non-overlapping folds with at least 24 training weeks. Country-prior selection requires at least two folds, which typically needs 48 valid target weeks; three folds are preferred. A 36-47 week target yields only one fold, so the app retains the base even if that single score looks favorable. Actual target-country spend and supported 0/1 event/regime features are supplied while KPI values are hidden, and every candidate uses identical folds.

Candidate validation locks target Y scale and representative alpha/ec/slope once at the earliest training cutoff across all folds. Reference rows are also restricted to what existed by that cutoff, so later reference data and hidden target Y cannot enter candidate priors. Because market spend scales differ, each reference channel is multiplied so its positive-adstock median matches the earliest-cut target operating scale before fitting under the same target ec/Hill unit. A market is held if date, transform, or spend-scale alignment fails. Only after selection are the chosen references refit using reference history through the target's final period and full-target transforms/Y scale; reference rows after the target ends are excluded. That final refit has no separate independent OOS score.

```
spend_scale[m,c] = median_positive_adstock_target[m] / median_positive_adstock_reference[m,c]
eligibility requires finite RMSE on every common fold
mean_RMSE = mean(RMSE across all common folds)
score = (mean_RMSE + 0.25 * sd_RMSE)
* (1 + 0.025 * max(0, number_of_countries - 1))
```

9.4 Between-country transfer heterogeneity

With multiple references, a random-effects tau^2 represents between-country disagreement and the result is a predictive prior for a new target market. With one source, heterogeneity cannot be estimated, so posterior SD has a 50% relative floor. With multiple sources, the relative SD floor is 25%; tau^2 and a within-country precision floor are also retained.

```
relative_floor = 0.50 if K = 1, else 0.25
relative_var_floor = (relative_floor * max(|pooled_mean|, sqrt(median(within_var))))^2
precision_floor_var = median(within_var) / max_effective_countries (max = 3)
predictive_var = tau^2 + max(pooled_mean_var, precision_floor_var, relative_var_floor)
transfer_precision = 1 / predictive_var
```

Uploading 47 countries does not mean using 47

Only a small, stable set that improves target prediction should survive. More countries do not automatically score better; complexity and cross-country heterogeneity suppress overconfident pooling.

9.5 Interpretation boundary

Country rolling folds ask whether a country-only prior repeatedly improves hidden 12-week target prediction over the base. Target transforms, Y scale, and reference evidence are locked at the earliest cutoff, preventing later-information leakage. However, the same folds are reused to shortlist markets and select combinations, so they remain candidate-tuning data rather than final independent OOS. After selection, chosen references are refit on full history and combined only in the final model with experiment evidence estimated under a separate design. If experiment and MMM share outcome records, the same-KPI reuse boundary reported by experiment diagnostics still applies. Predictive gain is not proof of causal transportability.

10. Validation and model health

10.1 Three layers

Layer	Question	Metrics
Fit	How well does the model describe observed history?	R2, RMSE, wMAPE
Prediction	How well does it predict hidden future periods?	Last-20% OOS for the no-external-prior base; earliest-cut as-of 12-week country tuning
Identification/health	Could it be right for the wrong reason?	Coverage, residual ACF1, negative baseline, prior shift

10.2 General OOS backtest

With at least 48 weeks, the final 20% base OOS is shown only when no external prior is active. Transform parameters are selected within training data and actual held-out media spend is used for forward prediction. With external evidence active, this base-only score is disabled so it is not mistaken for an independent score of the prior model. Experiment priors use their own effect CI, Fieller, HAC/cluster, and jackknife diagnostics; country priors use country-only versus base rolling 12-week candidate tuning locked to earliest-cut target transforms/Y/reference evidence. Both sources are precision-combined only in the final all-history fit.

10.3 Health metrics

Metric	Meaning	Warning	Action
wMAPE	Absolute errors / absolute actuals	Worse than baseline	Review events/regimes/transforms/prior
90% coverage	Actual inside predicted 90% interval	Too low or nearly 100%	Check under/over-dispersion
Residual ACF1	Adjacent residual correlation	Absolute value 〉= 0.3	Check time structure
Negative baseline	Weeks with natural demand below zero	Any positive share	Suspect misspecification
Prior shift	Posterior-prior gap in prior SDs	Absolute value 〉 2	Review transportability
Top weight	Weight on the leading transform	Low with many effective candidates	Treat shape as uncertain

10.4 Prediction is not causality

A high R2 can coexist with confounded channel coefficients when a price change or another driver is omitted. The app cannot directly map a continuous price control; encode a defensible change point as a 0/1 event/step when appropriate and disclose the rest as a limitation. Conversely, a market with little media variation can yield wide channel intervals even if media truly works. Combine predictive validation, identification review, and external experiments.

> Why there is no R-hat
> Meridian diagnoses MCMC chain convergence with R-hat and traces. This analytical model has no chains, so it uses predictive, residual, baseline, and prior-conflict checks instead.

11. Interpreting the results

11.1 Weekly contribution decomposition

The stacked decomposition uses one representative transform set. Trend includes the intercept/natural-demand level and its long-run direction. Seasonality, Events, Regime change, Performance, and Brand add to the fitted KPI; residual is actual minus fitted.

```
actual[t] = sum(group_contribution[g,t]) + residual[t]
```

11.2 Contribution-magnitude share

The driver share is each group's mean squared weekly contribution normalized across groups. It is a diagnostic for which groups moved the model most. It is not Shapley R2, incremental revenue share, or total contribution share.

```
magnitude[g] = mean(contribution[g,t]^2)
share[g] = magnitude[g] / sum_h magnitude[h]
```

11.3 Channel-effect card

Field	Meaning
Mean effect	KPI coefficient per unit of transformed media feature
90% interval	Conditional Gaussian for prior-locked channels; 5%/95% mixture-CDF quantiles otherwise
Positive probability	BIC-weighted candidate Gaussian probabilities
Candidate count	One for a unit-locked prior channel; fitted profile count for a no-prior channel
Effective candidates	How broadly weights are distributed; near 1 means one dominates
Top candidate weight	Largest BIC weight; low values signal transform ambiguity
Current marginal	Response change and 90% interval for a fixed source-currency increment (KRW 1M or USD 1k); the display-currency toggle changes formatting only

11.4 Prior toggle

Compare with-prior and without-prior for the same KPI. Look beyond a mean shift: inspect the 90% interval, positive probability, and prior shift, and check that the interval does not become implausibly narrow. Base OOS is available only without external priors, not as an equivalent independent score for the prior model. Experiment Fieller/CI identification and earliest-cut as-of country tuning are labeled separately. If a conclusion depends completely on one prior, that sensitivity belongs in the presentation.

> Use ranges, not forced rankings
> If a channel's current-marginal 90% interval overlaps the leading channel's interval, the app returns a candidate group rather than a single winner. Treat both as candidates and use an experiment to distinguish them.

12. Response curves, budget guidance, and forecasts

12.1 Response curves

A no-prior channel curve is the BIC-weighted response across profile candidates; a prior channel uses its unit-locked representative transform. The x-axis is sustainable weekly spend and the y-axis is modeled media response. Compare recent spend with the curve to assess saturation and remaining marginal response.

12.2 Budget use

- Prefer measurement over a large increase when the effect interval is broad or positive probability is weak.
- The marginal increment is fixed in source currency at KRW 1 million or USD 1 thousand; changing display currency does not change the calculation.
- Convert sustained post-increment spend to steady-state adstock for each profile candidate. If it exceeds the historical adstock maximum for any candidate in the leading cumulative 95% BIC weight, label extrapolation and exclude it from budget guidance.
- Group channels whose marginal 90% intervals overlap the leader; do not force a single rank.
- Add auction learning, inventory, contracts, creative, and Brand/Performance lag as external constraints.
- Refit after the budget change and confirm incrementality with a holdout where possible.

12.3 Budget-action identification gates

If any gate below fails, the app can still show exploratory direction but blocks direct increase/decrease guidance. Budget action requires both overall model identification and channel-level spend history, effect evidence, and a positive current marginal interval.

Gate	Block when	Meaning
History	Fewer than 52 weeks	App low-information region
Observations/parameters	weeks / estimated parameters 〈 8	Too little information per degree of freedom
VIF	Any media-variable VIF 〉= 10	Individual effects unstable under multicollinearity
Media-related correlation	Any media-media or media-nonmedia explanatory pair |corr| 〉= 0.90	Channel budget actions cannot be separated
Active weeks	Fewer than 20 weeks with spend 〉 0	Insufficient channel-specific variation
Spend variation	Spend is constant across history	Response and marginal effect are unidentified
Recent activity	Spend is zero in all of the last 8 weeks	Not connected to a current budget action
Effect evidence	Channel P(effect 〉 0) 〈 0.80	Insufficient probability support for an increase
Marginal interval	90% lower bound for the fixed increment 〈= 0	Positive response to the next increment is uncertain
Transformed observed range	Post-increment steady-state adstock exceeds the historical adstock maximum for any profile in the leading cumulative 95% BIC weight	Catch flight-channel extrapolation and exclude from guidance

12.4 Forecast mechanics

Future spend scenarios advance adstock state sequentially. Trend and Fourier features advance to the future week; structural steps retain the last known state. Future holidays/events require explicit scenario inputs, otherwise the baseline scenario treats unknown events as zero.

12.5 Predictive interval

The predictive interval includes residual variance and coefficient uncertainty for the future design row, x′Cov(beta)x. Long horizons and out-of-range scenarios require an additional model-risk warning.

```
predictive_var(x*) = sigma^2 * (1 + x*' (X'X + penalty)^-1 x*)
```

Forecasts are conditional scenarios

They assume planned spend occurs and that pricing, product, competition, and measurement remain consistent with the model. They are not guaranteed revenue commitments.

13. Working across KPIs

13.1 Sticky target selector

Mapped KPIs are active in the sticky selector. Switching the target loads that Y's fit, priors, validation, and response curves under the same input structure. Coefficient scale and identification differ by KPI; never copy a conclusion from one target to another.

13.2 Funnel interpretation

Target	Primary question	Caution
Traffic	Which channels move top-funnel volume?	Check overlap if built from Reg + React
Registrations	Which channels move new acquisition?	Stable deduplication/window
Reactivations	Which channels move returning users?	Retargeting vs natural return
Purchasers	Which channels move buyer count?	Weekly unique definition
Revenue	Which channels move value?	Fix price/refund definitions; promotion enters only as an exact 0/1 event

13.3 KPI-prior coverage

Review the channel x KPI prior map. If an experiment contains Registrations only while the MMM also has Reactivations and Revenue, only the Registrations model receives that experiment prior. Country priors follow the same rule; unmatched targets show 'no prior'.

14. Comparison with Google Meridian

14.1 Side-by-side

Area	This app	Google Meridian
Execution	Browser memory; no CSV server upload	Python package; local/cloud compute
Inference	Plug-in sigma² empirical-Bayes conditional Gaussian approximation	NUTS MCMC posterior sampling
Spatial structure	One target + optional country coefficient priors	Hierarchical geo random coefficients or national
Media data	Weekly spend-based features	Media execution + spend; reach/frequency support
Carryover/saturation	Geometric adstock + Hill; no-prior profile averaging, prior transform lock	HillAdstock with configurable lag/adstock
Prior types	Normal coefficient prior from experiments/countries	ROI, mROI, contribution, coefficient priors
Time baseline	Trend, Fourier seasonality, events, steps	Knot-based time-varying intercept + controls
Uncertainty	Sigma²-conditional Gaussian + per-channel profile-mixture CDF	Joint posterior samples across parameters
Validation	Base-only 20% OOS; earliest-cut as-of country tuning; experiment Fieller/CI; fit/coverage/ACF/baseline/prior shift	holdout_id, model fit, posterior predictive and MCMC diagnostics
Sampling health	R-hat/ESS/trace not applicable	R-hat, trace, ESS
Repeatability	Deterministic for identical input	Sampling variation managed by chains/seeds
Scale	Fast marketer-facing multi-KPI UX	Compute-intensive; GPU can be recommended

14.2 Difference from the broader fully Bayesian MMM family

Area	This app	Fully Bayesian MMM family
Residual variance	Plug-in fixed-point sigma²	Can assign priors to and integrate sigma²/error distribution
Nonlinear parameters	Per-channel profile candidates + BIC weights	Can jointly infer adstock, saturation, and coefficients
Hierarchy	Transfers country effects as external priors	Can partially pool geo/channel coefficients inside one model
Compute	Deterministic matrix algebra	MCMC, VI, SMC, or related probabilistic inference
Diagnostics	Prediction, residual, identification, prior shift	Those plus chain convergence, ESS, posterior predictive checks

Calling this app simply non-Bayesian would be misleading, but calling it equivalent to Meridian or a fully Bayesian hierarchical MMM would also overstate it. The precise description is a conditional Gaussian empirical-Bayes MMM that applies Normal priors with explicit variances.

14.3 Meridian principles adopted

- Separate carryover from saturation.
- Treat a prior as a probability distribution with mean and variance, not a manual weight.

- Use experimental point estimates and standard errors as calibration evidence.
- Separate in-sample fit from holdout prediction; inspect prior-posterior conflict and baseline pathologies.
- State that predictive accuracy does not establish causal validity.

14.4 Not currently implemented

Capability	Why/impact	Use when
NUTS/MCMC	Browser compute and latency; no fully sampled joint posterior	Run Meridian in Python/local compute
Hierarchical geo effects	Country priors transfer coefficients; they are not partial pooling	Rich geo panels and population data
Reach/frequency	Kept as a follow-up diagnostic, not a coefficient input	Stable reach/frequency definitions and overlap data
ROI/mROI priors	Displayed and unit-converted from the experiment coefficient prior at current spend	Cost, revenue, and estimands are rigorously aligned
Knot baseline	Limited 0/1/2-knot BIC selection when history is at least 78 weeks	Long nonlinear baselines with enough history
Joint transform average	Diagnostic over top two uncertain channels and at most four combinations	Full joint posterior requires offline MCMC/SMC
Posterior predictive sampling	Uses a conditional Normal predictive reference interval	Asymmetry/heavy tails are material

Which is better?

This app favors fast, private, browser-based, multi-KPI analysis. Meridian is more appropriate when geo partial pooling, reach/frequency, a joint posterior, and standard MCMC diagnostics are essential. The choice is about architecture and purpose, not a universal accuracy rank.

15. Accuracy, causality, and limitations

15.1 Major failure modes

Cause	Symptom	Response
Channel collinearity	Overlapping intervals, unstable signs	Aggregate, experiment, add varied periods
Omitted confounder	Inflated media effect, residual ACF	Use supported promo/outage 0/1 events or steps. Continuous price/competition controls are not directly mapped; split the period, use external analysis, or disclose the limitation
Measurement change	Residual regime break	Add step or split period
Failed prior transfer	Large prior shift, unstable country tuning	Widen or remove prior
Short history	Transform disagreement	Add time or reduce candidate grid
Out-of-range budget	Curve extrapolation	Use incremental in-range changes

15.2 Language for positive probability

State	Recommended	Do not say
High; interval positive	Strong evidence supporting a positive effect	Causal effect is certain
Middle; interval crosses 0	Positive direction but uncertain; test	No effect
Low/negative	Limited evidence of positive effect under these data/assumptions	Advertising is certainly harmful
Collinearity warning	Individual channel is unidentified	Definitive worst channel

15.3 Prior-selection bias

Trying many countries and selecting the best creates a winner's-curse risk. Multiple folds, an instability surcharge, and a complexity penalty reduce but do not eliminate it. Confirm the recommendation in a later period or independent experiment.

15.4 Privacy

CSV analysis stays in browser memory and is not transmitted or stored for modeling. Still, upload only aggregated week/country/channel data and exclude personally identifiable information.

16. Troubleshooting

16.1 Upload and mapping

Symptom	Likely cause	Fix
Dates do not aggregate	Mixed or invalid date formats	Normalize to YYYY-MM-DD
KPI inactive	Y not mapped or values nonnumeric	Map one of the five targets
No prior for one Y	That Y is absent in the prior file	Expected; base model remains
Channel prior missing	Channel keys differ	Align target and prior mapping keys
Countries mixed	Country missing/blank	Populate a consistent code on every row

16.2 Model

Symptom	Meaning	Action
Very wide interval	Low variation, collinearity, transform uncertainty	Add time/test; aggregate channels
Prior makes interval tiny	Test CI/transfer variance too small	Verify definitions; widen/remove
High R2, poor base OOS	Overfit or regime change	Review events and steps
High residual ACF	Missing time structure	Review trend, seasonality, adstock, events
Negative natural demand	Baseline misspecification/extrapolation	Stop decision use; redesign
Low top candidate weight	Carryover/saturation weakly identified	Use curve as a range

16.3 Experimental prior

Symptom	Check
On/Off not generated	At least 16 valid weeks, four weeks per state, two transitions, bounded Fieller, HAC, and block-jackknife diagnostics
Geo not generated	At least 16 geo-weeks, four geos, two per arm, common pre/post, all four cells, cluster/all-LOCO, bounded Fieller, and target Hill-derivative alignment
Same KPI reuse blocked	Check whether national On/Off KPI exactly matches main-MMM Y in the same weeks; use only separately identified experiment aggregation
Very large SE	Duration, geo count, intensity, and concurrent events; a weak prior is valid
Unexpected sign	KPI direction, state/arm/period coding, spend unit, and time order
Geo estimates unstable	Geo sample sizes, pre-trends, and enough clusters

16.4 Country priors

- Zero eligible countries: inspect overlapping channels, KPI, date range, and country mapping.
- Confirm at least 24 valid weekly rows for every reference country.
- Every candidate loses to baseline: no prior is the correct result.

- With only 36-47 target weeks, one fold is insufficient and the base is retained. Require at least two folds (typically 48 weeks), preferably three.
- Many countries visible: distinguish the eligible-country list from the one recommended combination.
- Fold winners change: use score including fold SD and complexity, not mean alone.
- Country coefficients disagree: tau^2 should reduce precision.

17. Review and presentation checklist

17.1 Data approval

- Document geography, period, week definition, currency, KPI definition, and conversion window.
- Verify daily-to-weekly and event-occurrence rules.
- Review missingness, duplicates, partial weeks, and measurement changes.
- Encode supported promotions, holidays, and outages as exact 0/1 events/steps. Disclose price-definition changes and the inability to map arbitrary continuous price/competition controls.

17.2 Model approval

- Compare the no-prior and with-prior models.
- Review OOS wMAPE only for the no-external-prior base, plus coverage, residual ACF, negative baseline, and prior shift.
- Record candidate count, effective candidates, and top weight.
- Do not force rankings where intervals overlap or channels are collinear.
- Do not label contribution-magnitude share as Shapley R2.

17.3 Prior approval

- Disclose experiment design, period, KPI, intensity, and uncertainty calculation.
- Review overlap and exact-match diagnostics between experiment KPI and main-MMM Y, and confirm that the same outcome was not used twice.
- Human-verify country definitions, currency, and conversion windows.
- Disclose country rolling folds and complexity penalty.
- Confirm that unmatched Y targets received no prior.

17.4 Suggested disclosure

> Presentation language
> This result is a conditional Gaussian empirical-Bayes MMM using weekly observational data, plug-in residual variance, carryover/saturation candidates, and Normal priors where available. Transform uncertainty is incorporated through per-channel profile BIC averaging; this is not a fully hierarchical joint MCMC model. Channel estimates are ranged decision evidence to be used with experiments and holdouts, not causal point truth.

18. Glossary

Term	Definition
Adstock	Media state combining current execution and prior carryover
alpha	Weekly persistence fraction in geometric adstock
Hill	Saturating response function
ec / half-saturation	Adstock at half normalized response
slope	Curvature/steepness of the Hill function
Prior	Parameter distribution before fitting target observations
Posterior	Parameter distribution after combining data and prior
Precision	Inverse variance; larger means stronger evidence
90% interval	Range describing coefficient uncertainty under the model/approximation
Profile model	Refit that varies one channel's transform candidate
BIC weight	Relative candidate weight balancing fit and complexity
Effective candidates	Inverse concentration of candidate weights
OOS	Out-of-sample period excluded from training
wMAPE	Sum absolute error divided by sum absolute actual
Coverage	Share of actuals inside a predictive interval
Residual ACF1	Correlation of adjacent weekly residuals
DiD	Difference in pre/post changes between treatment and control
HAC SE	Standard error robust to time autocorrelation/heteroskedasticity
Cluster-robust SE	Standard error allowing within-geo error correlation
Jackknife	Instability estimate from repeatedly omitting blocks
tau^2	Between-country transfer heterogeneity variance
Ridge	L2 regularization reducing extreme coefficients
R-hat	MCMC convergence metric; not applicable here
Shapley R2	Permutation-based allocation of explanatory power; not the current driver-share table
Incrementality	Effect relative to the counterfactual without advertising

Appendix A. Calculation-to-screen map

Screen element	Calculation source	Review question
Weekly contribution	Representative transform + posterior mean	Is representative-candidate dependence disclosed?
Effect 90% interval	Conditional Gaussian for prior channels / profile-mixture CDF quantiles otherwise	Does it cross zero; are candidates diffuse?
Positive probability	BIC-weighted candidate Gaussian probabilities	Is it being overstated as causal probability?
Response curve	Weighted steady-state adstock responses	Is spend inside the observed range?
Magnitude share	Normalized mean(contribution^2)	Is it incorrectly called Shapley R2?
Experiment prior	Effect/exposure ratio + conservative variance	Is transportability defensible?
Country prior	Country coefficient + tau^2 + rolling selection	Is selection bias disclosed?
Forecast	Future features + posterior predictive variance	Are event/regime assumptions explicit?

Appendix B. Official references

These official Google pages and repository are the comparison basis. Meridian changes over time; re-check current documentation before an implementation decision.

Google Meridian - Model specification

https://developers.google.com/meridian/docs/advanced-modeling/model-spec

Google Meridian - ModelSpec API

https://developers.google.com/meridian/reference/api/meridian/model/spec/ModelSpec

Google Meridian - ROI priors and calibration

https://developers.google.com/meridian/docs/advanced-modeling/roi-priors-and-calibration

Google Meridian - Choosing treatment prior types

https://developers.google.com/meridian/docs/advanced-modeling/how-to-choose-treatment-prior-types

Google Meridian - Collect data

https://developers.google.com/meridian/docs/pre-modeling/collect-data

Google Meridian - Geo selection and national data

https://developers.google.com/meridian/docs/pre-modeling/geo-selection-national-data

Google Meridian - Model health checks

https://developers.google.com/meridian/docs/post-modeling/health-checks

Google Meridian - Model fit

https://developers.google.com/meridian/docs/post-modeling/model-fit

Google Meridian - Model diagnostics

https://developers.google.com/meridian/docs/user-guide/model-diagnostics

Google Meridian - Interpret visualizations

https://developers.google.com/meridian/docs/post-modeling/interpret-visualizations

Google Meridian source repository

https://github.com/google/meridian

> Version note
> This handbook covers MMM methodology 1.1.0: the conditional Gaussian empirical-Bayes MMM, per-channel profile averaging, experiment/country priors, repeated validation, and health checks. Archive the handbook date with the model version used in a presentation.