Ivitskiy Igor
PhD, Associate Professor, Founder,
Doctor Ads LTD

DOI: https://doi.org/10.36059/978-966-397-543-6-24

FROM REGRESSION TO CAUSATION: WHEN AND HOW TO EMBED CONTROLS, INSTRUMENTS, AND PRIORS IN MMM

Modern marketing mix models face an identification problem. Budgets, bids, and creative refreshes react to the same demand shocks that drive sales; platform optimizers also chase conversion spikes. A plain regression that maps spend to revenue therefore misattributes correlation to impact. The remedy is not a single trick but a toolkit: embed explicit statistical controls when selection on observables is plausible, deploy instrumental variables when endogeneity is structural, and encode defensible priors when experiments or business constraints carry information. The decision is empirical and can be formalized.

Start with a baseline structural regression that respects media physics. Let the outcome at time t be y_t, and let $x_{k,t}$ be raw exposures for channel k. Adstock transforms carryover, and response saturates. A minimally identified MMM writes as $y_t = \alpha + \sum_{k=1}^{K} \beta_k \, g(s_{k,t}) + \delta' z_t + u_t$, where $s_{k,t}$ are adstocked exposures, $g(\cdot)$ is a concave saturation function, z_t are controls, and u_t is the error. The canonical transforms are

$$s_{k,t} = \lambda_k s_{k,t-1} + x_{k,t}$$

where $0 \leq \lambda_k < 1$ governs carryover, and

$$g(s) = \frac{s^\eta}{\theta^\eta + s^\eta}$$

where $\eta > 0$ controls steepness and $\theta > 0$ is the half-saturation. These forms are standard in Bayesian MMM stacks and are now first-class citizens in open implementations [3; 7].

When are controls enough. If the remaining endogeneity can be captured by observable confounders, controls suffice. In practice this means adding price and promo indicators, macroeconomic covariates, competitor intensity, and seasonality decompositions; it also means building synthetic controls that track the target when media is off. Bayesian structural time series (BSTS) formalizes this by estimating the counterfactual path and attributing lift to the intervention while letting empirical priors shrink noisy coefficients [1]. Concretely, the state-space layer absorbs trend and seasonality, and the regression layer selects contemporaneous covariates; the result is a model that attributes only what remains after the counterfactual is explained [1].

When instruments are required. If spend is chosen in response to unobservables that also move sales, controls will not rescue identification. Consider weekly bid changes that react to latent demand or creative swaps triggered by performance slumps; both contaminate u_t. Two-stage least squares targets the exogenous variation isolated by instruments Z. The estimator is:

$$\widehat{\beta^{2SLS}} = (X'P_Z X)^{-1} X' P_Z y$$

with projection matrix $P_Z = Z(Z'Z)^{-1}Z'$. Valid advertising instruments include budget caps imposed by finance cycles, auction discontinuities at threshold bids, supply-side shocks to reach or inventory, and staggered geo rollouts that predate observed demand shocks. The identification claim is Local Average Treatment Effect, so interpretability hinges on who is compliers. Applied guidance from modern orthogonalization shows how to pair IV with high-dimensional controls while keeping valid inference [2].

Where priors carry the signal. In low-sample or highly collinear settings, experiments and domain knowledge are too valuable to ignore. A Bayesian MMM encodes them directly:

$$\beta_k \sim Normal(m_k, v_k)$$

with m_k set by calibrated lift studies and v_k tuned to reflect uncertainty. Posterior inference combines likelihood and prior,

$$p(\beta \mid y) \propto p(y \mid \beta)\, p(\beta)$$

so credible intervals and budget recommendations propagate experimental evidence instead of treating it as an afterthought. Recent open frameworks make this workflow practical at geo scale, including hierarchical pooling across regions and explicit modules for prior calibration and causal estimands [3; 5]. In production, this materially changes the optimizer: the objective is no longer a point estimate of response but a posterior over response, and budget allocation should target expected utility under that distribution.

A decision recipe. A workable rule set is:

1. If media is set on exogenous schedules and audit logs confirm no feedback to demand within the modeling window, prefer controls plus a BSTS layer that builds a strong counterfactual; validate with holdouts and placebo interventions [1].
2. If planning reacts to latent demand or algorithmic pacing, design instruments or natural experiments; run weak-IV tests and report partial R^2 for first stages [2; 6].
3. If experiments exist or expert constraints are credible, encode them as priors in a fully Bayesian MMM; perform prior-to-posterior sensitivity and report how budget recommendations move as functions of m_k and v_k [3; 7].

Failure modes and diagnostics. Weak instruments bias estimates toward OLS; treat Stock–Yogo style diagnostics and first-stage F as gating checks. Prior misspecification can over-regularize and understate uncertainty; compare posteriors under reference priors to demonstrate stability. Media physics parameters often absorb unmodeled seasonality; inspect residual autocorrelation and re-estimate adstock on deseasonalized series. Across paradigms, posterior predictive checks or out-of-time fits must be table-stakes. BSTS-style causal impact evaluation is useful as a falsification device around media pauses and shocks [1].

Implementation notes. Modern open stacks operationalize these ideas. Google's Meridian documents causal estimands, hierarchical geo modeling, default and custom priors, and pathways to incorporate experiment lifts directly; it is open and built for GPU-accelerated posterior sampling, which shortens iteration cycles for sensitivity analysis [3; 5].

The Robyn package remains a strong ridge-based baseline with diminishing returns and carryover, plus an explicit calibration step to inject incrementality results; treat it as a frequentist control model or as a prior-setting aid when moving to Bayesian setups [7]. Finally, causal-structure learning for MMM is active research; algorithmic discovery of mediators and interactions is promising yet still requires human judgment on identification [8].

A brief budget implication. Once causality is explicit, the optimizer changes character. Instead of maximizing a deterministic hill curve, the task is to allocate budget to channels with highest expected incremental profit subject to uncertainty. One implementable approach is to sample response curves from the posterior, compute profit surfaces $\Pi(b)$ over candidate budgets b, and choose allocations that maximize expected Π while controlling downside risk through a quantile objective. This makes the plan robust when sales volatility is high or when experiments are sparse. Practical guidance on portfolio-style tradeoffs for bid strategies can complement the causal stack in day-to-day operations [4].

References:

1. Brodersen K. H., Gallusser F., Koehler J., Remy N., Scott S. L. (2015) Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, vol. 9, pp. 247–274. DOI: https://doi.org/10.1214/14-AOAS788. Available at: https://research.google/pubs/pub41854 (accessed 30.09.2025).

2. Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., Newey W., Robins J. (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, vol. 21, no. 1, pp. C1–C68. DOI: https://doi.org/ 0.1111/ectj.12097.

3. Google Developers. About Meridian: Bayesian MMM framework and documentation. 2025. Available at: https://developers.google.com/meridian/docs/basics/about-the-project (accessed 30.09.2025).

4. Ivitskiy I. Mastering Google Ads Portfolio Bid Strategies. *The Doctor Ads Blog*. Available at: https://blog.thedoctorads.com/mastering-google-ads-portfolio-bid-strategies/ (accessed 30.09.2025).

5. Nair H. Meridian is now available to everyone. *Google Ads Blog*. 29.01.2025. Available at: https://blog.google/products/ads-commerce/meridian-marketing-mix-model-open-to-everyone/ (accessed 30.09.2025).

6. Angrist J. D., Pischke J.-S. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press, 2009. 392 p. Available at: https://press.princeton.edu/books/paperback/9780691120355/mostly-harmless-econometrics (accessed 30.09.2025).

7. Robyn Team (Meta). Robyn: Semi-Automated Marketing Mix Modeling. CRAN package page, v3.11.1, 27.06.2024. DOI: https://doi.org/10.32614/CRAN.package.Robyn.

8. Gong C., Yao D., Zhang L., Chen S., Li W., Su Y., Bi J. CausalMMM: Learning causal structure for marketing mix modeling. *arXiv preprint*. 24.06.2024. arXiv:2406.16728. Available at: https://arxiv.org/abs/2406.16728 (accessed 30.09.2025).