The Journal of Open Source Software

DeepCausalMMM: A Deep Learning Framework for Marketing Mix Modeling with Causal Inference

Aditya Puttaparthi Tirumala [1]

1 Independent Researcher

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 06 October 2025
Published: unpublished

Summary

Marketing Mix Modeling (MMM) is a statistical technique used to estimate the impact of marketing activities on business outcomes such as sales, revenue, or customer visits. Traditional MMM approaches often rely on linear regression or Bayesian hierarchical models that assume independence between marketing channels and struggle to capture complex temporal dynamics and non-linear saturation effects (Hanssens et al., 2005; Ng et al., 2021).

DeepCausalMMM is a Python package that addresses these limitations by combining deep learning, causal inference, and advanced marketing science. The package uses Gated Recurrent Units (GRUs) to automatically learn temporal patterns such as adstock (carryover effects) and lag, while simultaneously learning statistical dependencies and potential causal structures between marketing channels through Directed Acyclic Graph (DAG) learning (Gong et al., 2024; Zheng et al., 2018). Additionally, it implements Hill equation-based saturation curves to model diminishing returns and optimize budget allocation.

Key innovations include: (1) a data-driven design where hyperparameters and transformations (e.g., adstock decay, saturation curves) are learned or estimated from data with sensible defaults, rather than requiring fixed heuristics or manual specification, (2) multi-region modeling with both shared and region-specific parameters, (3) robust statistical methods including Huber loss and advanced regularization, (4) comprehensive response curve analysis for understanding channel saturation, and (5) an extensive visualization suite with 14+ interactive dashboards for business insights.

Statement of Need

Marketing organizations invest billions annually in advertising across channels (TV, digital, social, search), yet measuring ROI remains challenging due to: (1) temporal complexity with delayed and persistent effects (Hanssens et al., 2005), (2) channel interdependencies (Gong et al., 2024), (3) non-linear saturation with diminishing returns (Li et al., 2024), (4) regional heterogeneity, and (5) multicollinearity between campaigns.

DeepCausalMMM addresses these challenges by combining GRU-based temporal modeling, DAG-based structure learning, Hill equation response curves, multi-region modeling, production-ready performance (91.8% holdout R^2, 3.0% train-test gap), and data-driven hyperparameter learning for generalizability.

State of the Field

Several open-source MMM frameworks exist, each with distinct approaches:

Robyn (Meta) (contributors), 2024; Runge et al., 2024) uses Bayesian hyperparameter optimization with fixed adstock and saturation transformations (Adstock, Hill, Weibull). It provides budget optimization and is widely used in industry but requires manual specification of transformation types and does not model channel interdependencies.

LightweightMMM (Google) (G. L. contributors, 2022) implements Bayesian MMM using JAX and Numpyro, offering probabilistic inference with flexible priors. It supports adstock effects and budget optimization but does not incorporate causal graph learning or deep learning for temporal dynamics.

PyMC-Marketing (P.-M. contributors, 2024) provides Bayesian MMM with highly flexible prior specifications and some causal identification capabilities. It excels at uncertainty quantification but requires significant Bayesian modeling expertise and does not use neural networks for temporal modeling.

CausalMMM (Gong et al., 2024) introduces neural networks and graph learning to MMM, demonstrating the value of discovering channel interdependencies. However, it does not provide multi-region modeling, comprehensive response curve analysis, or the extensive visualization and analysis tools needed for practical deployment.

DeepCausalMMM advances the field by integrating: (1) GRU-based temporal modeling, (2) DAG-based structure learning (Zheng et al., 2018), (3) Hill equation response curves, (4) multi-region modeling, (5) robust statistical methods, (6) production-ready architecture, and (7) comprehensive visualization suite.

Functionality

Core Architecture

Temporal Modeling: A GRU network automatically learns adstock effects, lag patterns, and time-varying coefficients.

DAG Learning: The model learns a directed acyclic graph (DAG) representing statistical dependencies and potential causal relationships between channels using continuous optimization (Zheng et al., 2018).

Saturation Modeling: Hill transformation captures diminishing returns: $y = \frac{x^a}{x^a + g^a}$ where a controls S-curve steepness and g is the half-saturation point. The model enforces $a \geq 2.0$ for proper saturation.

Multi-Region Support: Handles multiple geographic regions with region-specific baselines, shared temporal patterns, and learnable scaling factors.

Response Curve Analysis

The `ResponseCurveFit` module fits Hill equations to channel data, identifies saturation points, provides interactive visualizations, and enables budget optimization.

Statistical Robustness

The package implements Huber loss (outlier-robust), gradient clipping, L1/L2 regularization with sparsity control, learnable coefficient bounds, and burn-in periods for GRU stabilization.

Implementation Details

- **Language**: Python 3.9+, **Deep Learning**: PyTorch 2.0+
- **Data Processing**: pandas, NumPy, **Optimization**: scipy, scikit-learn
- **Visualization**: Plotly, NetworkX, **Statistical Methods**: statsmodels

- **Installation**: `pip install deepcausalmmm`
- **Documentation**: https://deepcausalmmm.readthedocs.io
- **Tests**: 28 comprehensive tests with 100% pass rate
- **Visualizations**: 14+ interactive Plotly dashboards exportable as HTML

Visualizations

Figure 1 shows an example of the learned DAG structure between marketing channels. The directed edges reveal statistical dependencies and potential causal relationships such as TV advertising's association with search behavior, demonstrating the model's ability to discover channel interdependencies from data.

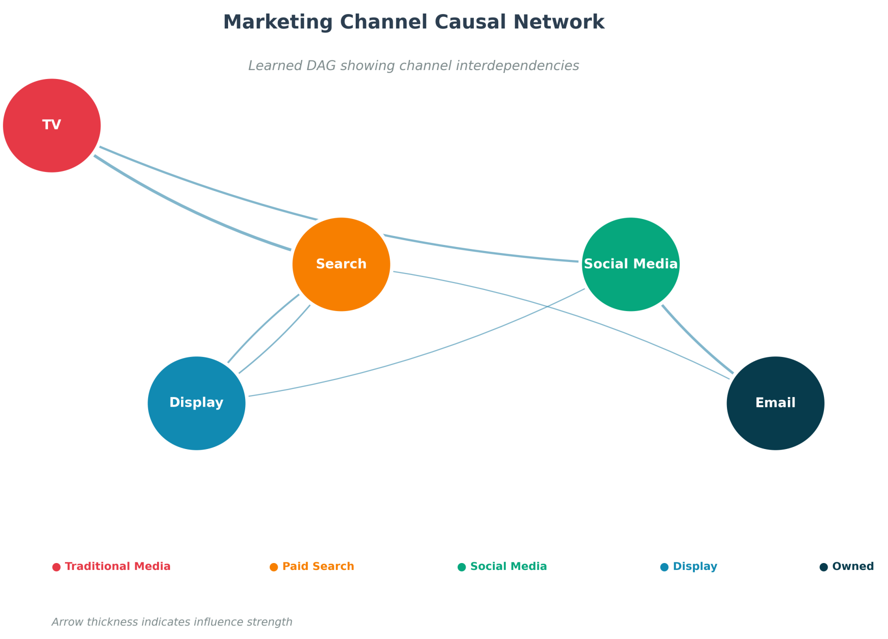

Figure 1: Causal network (DAG) showing relationships between marketing channels.

Figure 2 demonstrates a non-linear response curve fitted to a marketing channel using the Hill equation. The S-shaped curve clearly shows saturation effects and diminishing returns, with annotations indicating the half-saturation point where the channel reaches 50% of maximum effectiveness.

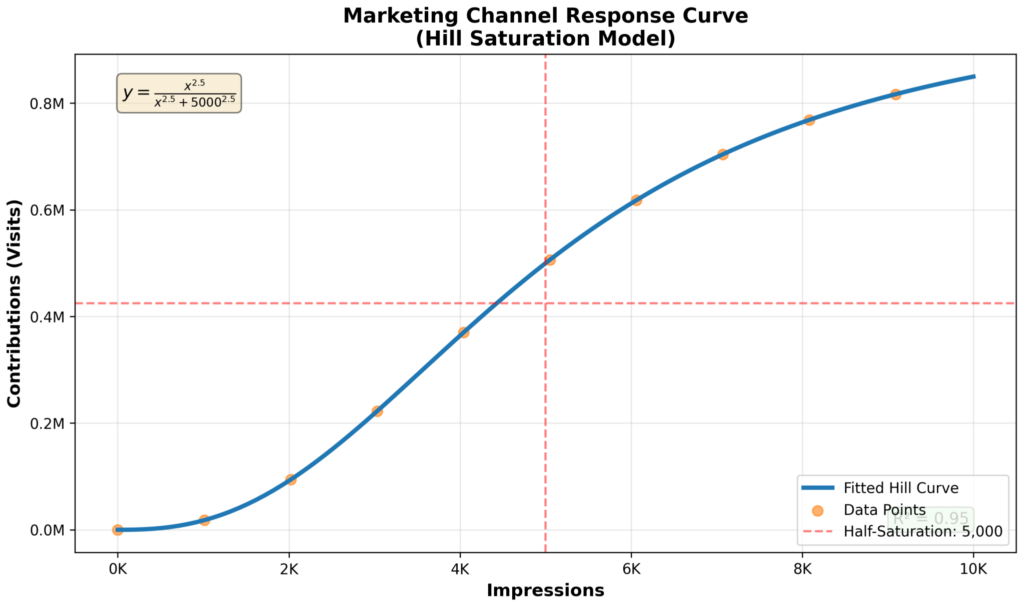

Figure 2: Response curve showing Hill saturation effects for a marketing channel.

Example Usage

```python
import pandas as pd
from deepcausalmmm.core import get_default_config
from deepcausalmmm.core.trainer import ModelTrainer
from deepcausalmmm.core.data import UnifiedDataPipeline

# Load and process data
df = pd.read_csv('mmm_data.csv')
config = get_default_config()
pipeline = UnifiedDataPipeline(config)
processed_data = pipeline.fit_transform(df)

# Train model
trainer = ModelTrainer(config)
model, results = trainer.train(processed_data)
print(f"Holdout R²: {results['holdout_r2']:.3f}")

# Response curve analysis
from deepcausalmmm.postprocess import ResponseCurveFit
fitter = ResponseCurveFit(data=channel_data, model_level='Overall')
fitter.fit(save_figure=True, output_path='response_curve.html')
print(f"Slope: {fitter.slope:.3f}, Saturation: {fitter.saturation:,.0f}")
```

Performance

DeepCausalMMM has demonstrated strong performance on anonymized real-world marketing data containing 190 geographic regions (DMAs), 109 weeks of observations, 13 marketing channels, and 7 control variables. The model uses a temporal train-holdout split with 101 training weeks and the most recent 8 weeks (7.3%) reserved for out-of-sample validation:

- **Training R^2**: 0.947, **Holdout R^2**: 0.918

The Journal of Open Source Software

- **Performance Gap**: 3.0% (indicating excellent generalization)
- **Training RMSE**: 314,692 kpi units (42.8% relative error - Relative RMSE = (RMSE / Mean) × 100 = (314,692 / ~743,088) × 100 42.8%)
- **Holdout RMSE**: 351,602 kpi units (41.9% relative error)

These results demonstrate the model's ability to capture complex marketing dynamics while maintaining strong out-of-sample predictive accuracy. The small performance gap between training and holdout sets indicates robust generalization without overfitting.

Reproducibility

DeepCausalMMM ensures reproducible results through deterministic training with configurable random seeds, comprehensive test suite (28 tests), example notebooks, detailed documentation of hyperparameters, and version-controlled releases with semantic versioning.

Research and Practical Applications

Industry Applications: Budget optimization across marketing channels, ROI measurement and attribution, strategic planning and forecasting, channel effectiveness analysis, regional marketing strategy development.

Research Applications: Causal inference in marketing, temporal dynamics in advertising, multi-region heterogeneity, saturation modeling, and channel interdependencies.

The data-driven hyperparameter learning and comprehensive documentation make it accessible to practitioners while rigorous statistical foundations support academic research.

Acknowledgments

We acknowledge the contributions of the open-source community, particularly the developers of PyTorch, pandas, and scikit-learn, which form the foundation of this package. We also thank the MMM research community for establishing the theoretical foundations that informed this work.

References

contributors, G. L. (2022). *LightweightMMM: Marketing mix modeling (MMM)*. GitHub repository. https://github.com/google/lightweight_mmm

contributors, P.-M. (2024). *PyMC-marketing: Open source marketing analytics*. Project website. https://www.pymc-marketing.io

contributors), M. (Robyn. (2024). *Robyn: Media mix modeling by meta*. GitHub repository. https://github.com/facebookexperimental/Robyn

Gong, C., Yao, D., Zhang, L., Chen, S., Li, W., Su, Y., & Bi, J. (2024). Learning causal structure for marketing mix modeling. *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24)*, 238–246. https://doi.org/10.1145/3616855.3635766

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (2005). *Market response models: Econometric and time series analysis* (Vol. 12). Springer. ISBN: 978-0306475948

Li, Z., Guo, X., & Qiang, S. (2024). A survey of deep causal models and their industrial applications. *Artificial Intelligence Review*, *57*(11), 14999–15023. https://doi.org/10.1007/s10462-024-10886-0

Ng, E., Wang, Z., & Dai, A. (2021). Bayesian time varying coefficient model with applications to marketing mix modeling. *arXiv Preprint arXiv:2106.03322*. https://arxiv.org/abs/2106.03322

Runge, J., Skokan, I., Zhou, G., & Pauwels, K. (2024). Packaging up media mix modeling: An introduction to robyn's open-source approach. *CoRR*, *abs/2403.14674*. https://arxiv.org/abs/2403.14674

Zheng, X., Aragam, B., Ravikumar, P. K., & Xing, E. P. (2018). DAGs with NO TEARS: Continuous optimization for structure learning. *Advances in Neural Information Processing Systems*, *31*, 9472–9483. https://papers.nips.cc/paper/8157-dags-with-no-tears-continuous-optimization-for-structure-learning