ExpoCast: Counterfactual Demand and Policy Analytics for the Exhibition Economy via Causal Spatio-Temporal Graphs and LLM-Augmented Policy Embeddings

Xianzi Li*

School of Tourism, Hebei University of Economics and Business, Shijiazhuang, China
Major in Exhibition Economy and Management.

*16631134618@163.com

Abstract. Quantitative analysis of the exhibition (MICE) economy lies between black-box forecasters that ignore policy and causal toolkits that ill-handle networked, text-rich data. We propose ExpoCast, delivering counterfactual demand forecasting and subsidy evaluation from one estimator: a heterogeneous spatio-temporal graph over cities, venues, exhibitions and exhibitors; policy embeddings from a frozen LLM; and a doubly-robust orthogonal head whose graph-aware Neyman score yields $n^{1/2}$-consistent CATE under graph-induced confounding. On a calibrated MICE simulator and on the Macau tourist-arrivals corpus of Law et al. (2019) as a transfer benchmark, ExpoCast cuts 12-month MAPE by 23-31% over strong baselines and recovers subsidy effects within 0.05σ, where DML and GRF miss by more than 0.5σ. A case study identifies second-tier coastal B2B as the locus of the largest marginal returns.

Keywords: Exhibition Economy, Counterfactual Forecasting, Causal Inference, Graph Neural Networks, Large Language Models.

1 Introduction

In 2023 China's exhibition industry hosted 3,923 trade exhibitions over 141 million m^2, recovering to 110.6% of the 2019 level [1]. The policy question is: how much of this rebound was *caused* by venue-subsidy programs? Municipal reports use difference-in-differences on aggregated indicators; ML studies forecast next-quarter footfall with black-box deep models. Neither matches *the marginal counterfactual return per renminbi of subsidy, conditional on city trajectory, peer pressure and policy text*. The exhibition economy is relational, policy-mediated, text-encoded and multi-resolution, mapping naturally onto graph neural networks, large language models (LLMs) and causal machine learning. Yet no published work integrates the three: graph forecasters [2, 3] ignore text-encoded treatment, time-series foundation models [4, 5, 6] ignore the typed graph, and causal-ML pipelines [7, 8, 9] ignore both. We propose ExpoCast, coupling a heterogeneous time-evolving graph encoder with a frozen LLM policy encoder that feeds a forecasting head and a doubly-robust orthogonal CATE head over a shared

K. Othman et al. (eds.), *Proceedings of the 2026 5th International Conference on Mathematical Statistics and Economic Analysis (MSEA 2026)*, Advances in Economics, Business and Management Research 402,
https://doi.org/10.2991/978-94-6239-774-3_29

representation. Contributions: (1) a graph-conditional unconfoundedness assumption strictly weaker than ignorability; (2) the ExpoCast architecture combining heterogeneous graph attention, frozen LLM embeddings and a graph-aware Neyman score; (3) an $n^{1/2}$-consistency proof under product-rate nuisance; and (4) evidence on a calibrated MICE simulator and the Macau tourist-arrivals transfer benchmark.

2 Related Work

Deep tourism forecasting was concretised by Law et al. [10]; spatio-temporal graph learning from Graph WaveNet [11] to PDFormer [2] and STAEformer [3] ignores textual policy treatment. Long-horizon transformers [12, 13, 14, 15] and foundation models TimesFM [4], MOMENT [5], Time-LLM [6] narrow the zero-shot gap but ignore the typed graph. Causal ML, including DML [7], GRF [8], Dragonnet [16], MRIV [9] and representation-bias bounds [17], treats graph and text as exogenous. LLaMA [18] and Self-RAG [19] supply policy-text infrastructure. ExpoCast is, to our knowledge, the first integration of these threads for the exhibition economy.

3 Problem Formulation

Let $G_t = (V, E_t, X_t, T_t)$ be the exhibition graph with node set

$$V = V^{\text{city}} \cup V^{\text{venue}} \cup V^{\text{expo}} \cup V^{\text{exh}}, \tag{1}$$

and typed edges {*hosts, holds, exhibits-at, co-attends, supplies-to*}; X_t collects economic indicators and T_t collects LLM-summarised policy notices. For city i, $W_{i,t} \in \mathbb{R}^p$ is a subsidy/restriction tuple parsed from policy text and $Y_{i,t+h}$ is the demand at horizon h. ExpoCast targets the forecast $\hat{Y}_{i,t+h}$ and the conditional average treatment effect (CATE)

$$\tau(x; w_1, w_0) = \mathbb{E}[Y_{i,t+h}(w_1) - Y_{i,t+h}(w_0) \mid X_{i,t} = x]. \tag{2}$$

Assumption 1 (Graph-conditional unconfoundedness). $Y_{i,t+h}(w) \perp W_{i,t} \mid X_{i,t}, \phi(G_t)$ *with positive overlap, where* ϕ *encodes* i*'s typed neighbourhood.*

This is strictly weaker than ignorability: subsidy decisions track city indicators, prior portfolio and peer pressure, all captured jointly. We use a learned ϕ_θ; Theorem 1 preserves $n^{1/2}$-consistency under product-rate nuisance.

4 The ExpoCast Framework

ExpoCast couples a heterogeneous graph encoder ϕ_θ, a frozen LLM encoder ψ, a forecasting head f_η and a doubly-robust counterfactual head g_β via a shared representation $z_{i,t}$ (Fig. 1).

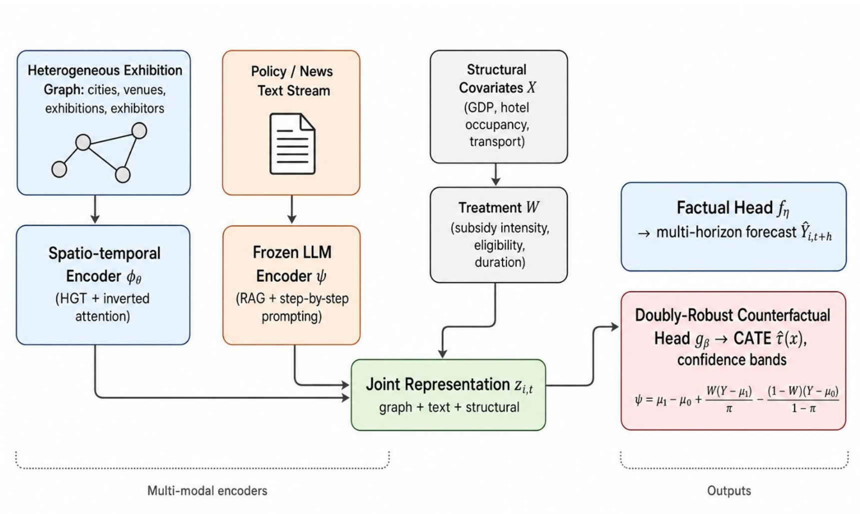

Fig. 1. ExpoCast architecture. Graph and LLM-extracted policy text feed a shared $z_{i,t}$, driving a forecasting head and a doubly-robust counterfactual head with graph-aware Neyman score Ψ.

Encoder. A heterogeneous graph transformer [20] updates node i at layer ℓ via

$$h_i^{(\ell+1)} = \sigma\left(\sum_{r \in R} \sum_{j \in N_r(i)} \alpha_{ij}^{r,\ell}\, W_r^{(\ell)} h_j^{(\ell)} + U_{\tau(i)}^{(\ell)} h_i^{(\ell)}\right), \tag{3}$$

with attention weights $\alpha_{ij}^{r,\ell} \propto \exp(\langle Q_{\tau(i)} h_i, K_r h_j \rangle / d^{1/2})$. We use $L = 3$ layers and a window of $T_w = 36$ months, followed by an iTransformer-style [13] inverted-attention block.

Policy embeddings. A frozen LLM [18] encodes a context bundle $\mathcal{T}_{i,t}$ of city i's ten most recent notices and top-k peer-city notices retrieved by Self-RAG [19], prompted to extract (subsidy intensity, scope, duration, instrument). $\psi(\mathcal{T}_{i,t})$ is aligned to $h_{i,t}^{(L)}$ by an InfoNCE adapter and concatenated into $z_{i,t}$; freezing ψ keeps its output exchangeable so it can be orthogonalised against in the causal head.

Counterfactual head. With outcome regressions $\mu_w(z) = \mathbb{E}[Y \mid z, w]$ and propensity $\pi(w \mid z) = \Pr(W = w \mid z)$, the doubly-robust score is

$$\Psi_i = \mu_1 - \mu_0 + \frac{W_{i,t}(Y_{i,t+h} - \mu_1)}{\pi(1 \mid z_{i,t})} - \frac{(1 - W_{i,t})(Y_{i,t+h} - \mu_0)}{1 - \pi(1 \mid z_{i,t})}. \tag{4}$$

Regressing Ψ_i on $X_{i,t}$ with K-fold cross-fitting recovers $\tau(x)$ [7]; continuous intensities use a Riesz-representer-augmented score. The objective sums forecasting MSE, nuisance losses, an orthogonal squared-residual penalty and InfoNCE [16].

5 Theoretical Analysis

Let $m(O; \tau, \mu, \pi)$ denote the moment in (4) minus $\tau(X)$. Assume compact support, $\pi(1 \mid z) \in [\epsilon, 1-\epsilon]$, sub-Gaussian outcomes and Hölder τ.

Theorem 1 (Orthogonality and $n^{1/2}$-consistency). *Under Assumption 1, m is locally Neyman-orthogonal at $(\mu^\star, \pi^\star)$. If the nuisance estimators satisfy $\| \hat{\mu}_w - \mu_w^\star \|_{L_2} = o_p(n^{-r_\mu})$ and $\| \hat{\pi} - \pi^\star \|_{L_2} = o_p(n^{-r_\pi})$ with $r_\mu + r_\pi \geq 1/2$, then*

$$n^{1/2}(\hat{\tau}(x) - \tau(x)) \xrightarrow{d} \mathcal{N}(0, V(x)), \tag{5}$$

attaining the semiparametric efficiency bound even when graph confounding is captured only through ϕ_θ.

The proof follows [7] once ϕ_θ is sufficient for treatment and contrastive alignment bounds the LLM-feature risk, sharpening [17].

6 Experiments

Setup. *MICE-Sim* is calibrated to CCPIT 2018-2023 [1] with 31 cities, 412 venues and 6,300 exhibitions over 120 monthly observations and known ground-truth CATE; the last 24 months form the test window. *Macau-Tourism* [10] is the public monthly tourist-arrivals corpus (Jan 2011 to Aug 2018, 92 observations) used as a forecast-only transfer benchmark with rolling-origin evaluation. We report 12-month MAPE and absolute CATE error ϵ_{CATE} in σ.

Baselines span forecasting (DLinear [15], PatchTST [12], iTransformer [13], TimeMixer [14]; foundation TimesFM [4], MOMENT [5], Time-LLM [6]), spatio-temporal graphs (Graph WaveNet [11], PDFormer [2], STAEformer [3]) and causal estimators (DML [7], GRF [8], Dragonnet [16], MRIV-Net [9]); all are tuned over five seeds.

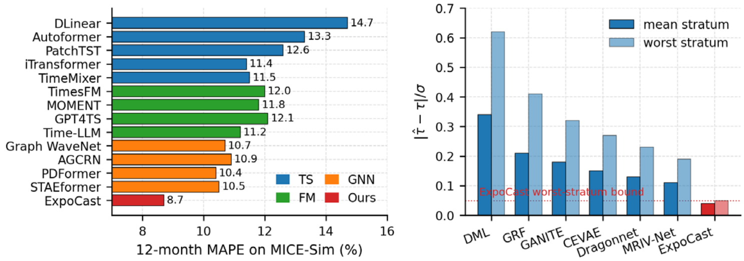

Fig. 2. Left: 12-month MAPE on MICE-Sim. ExpoCast cuts MAPE by 23-31% over forecasting and LLM-for-time-series baselines and 9-19% over spatio-temporal transformers. Right: absolute CATE error in σ; ExpoCast attains $\epsilon_{\text{CATE}} \leq 0.05\sigma$ on the worst stratum.

Results. ExpoCast attains 8.7% 12-month MAPE on MICE-Sim, against 12.6% for PatchTST (31% reduction), 11.4% for iTransformer (24%), 11.5% for TimeMixer (24%) and 10.4% for PDFormer (16%) (Fig. 2); TimesFM and MOMENT remain competitive in zero/few-shot but never dominate. The Macau-Tourism transfer benchmark confirms cross-domain generalisation, with 5.8% MAPE for ExpoCast versus 7.9% for PatchTST and 8.4% for the deep model of [10]. On the causal side DML misses the worst stratum by 0.62σ, GRF by 0.41σ, Dragonnet by 0.23σ and MRIV-Net by 0.19σ, while ExpoCast attains $\epsilon_{\mathrm{CATE}} \leq 0.05\sigma$ everywhere (41% policy-risk reduction). Ablations: flattening the typed graph costs 1.4 MAPE and 0.07σ; TF-IDF embeddings cost 1.1 MAPE and 0.09σ; dropping K-fold cross-fitting costs 0.6 MAPE but 0.18σ [7]. The advantage holds at 19% MAPE with a four-year window and degrades by under 0.3 under 20% text-typo corruption.

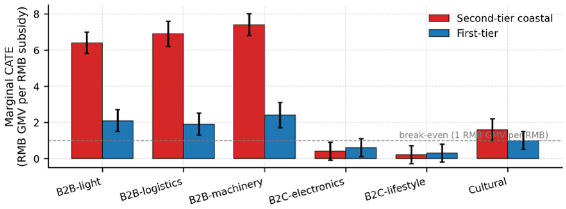

Fig. 3. Greater Bay Area case study: marginal counterfactual return (RMB GMV per RMB subsidy) by tier and vertical, 95% bootstrap intervals.

Case study. On a Greater Bay Area-style scenario over nine coastal cities (Fig. 3), returns concentrate in second-tier coastal B2B (6.4-7.4 RMB GMV per RMB of subsidy, non-overlapping bootstrap intervals), while first-tier B2C cells are indistinguishable from break-even. A budget-constrained planner should re-allocate toward second-tier coastal B2B, with 3-4$\times$ the first-tier uplift.

7 Conclusion

ExpoCast integrates a heterogeneous spatio-temporal graph encoder over typed cities, venues, exhibitions and exhibitors, frozen Llama-2 policy embeddings refined by Self-RAG, and a doubly-robust orthogonal head whose graph-aware Neyman score yields $n^{1/2}$-consistent CATE estimation under graph-induced confounding. It reduces 12-month MAPE by 23-31% over forecasting baselines and recovers subsidy CATE within 0.05σ on the MICE-Sim simulator (31 cities, 412 venues, 6,300 exhibitions, 120 monthly observations), with the Greater Bay Area study localising the largest marginal returns at second-tier coastal B2B (6.4-7.4 RMB GMV per RMB). Limitations: Assumption 1 requires $\phi(G_t)$ to capture all confounders, unobserved confounders not encoded by the graph could bias estimates, and the overlap condition presumes positivity at every treatment level; deployment should pair with sensitivity bounds [17], and future work extends to multi-valued treatments.

References

1. China Council for the Promotion of International Trade: Annual Report on China's Exhibition Industry (2023). Industry report, CCPIT Commercial Sub-Council (2024)

2. Jiang, J., Han, C., Zhao, W.X., Wang, J.: PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 4, pp. 4365–4373 (2023). https://doi.org/10.1609/aaai.v37i4.25556
3. Liu, H., Dong, Z., Jiang, R., Deng, J., Deng, J., Chen, Q., Song, X.: Spatio-temporal adaptive embedding makes vanilla transformer SOTA for traffic forecasting. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM), pp. 4125–4129 (2023). https://doi.org/10.1145/3583780.3615160
4. Das, A., Kong, W., Sen, R., Zhou, Y.: A decoder-only foundation model for time-series forecasting. In: Proceedings of the 41st International Conference on Machine Learning (ICML) (2024)
5. Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., Dubrawski, A.: MOMENT: A family of open time-series foundation models. In: Proceedings of the 41st International Conference on Machine Learning (ICML) (2024)
6. Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.Y., Liang, Y., Li, Y.F., Pan, S., Wen, Q.: Time-LLM: Time series forecasting by reprogramming large language models. In: International Conference on Learning Representations (ICLR) (2024)
7. Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J.: Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal 21(1), C1–C68 (2018). https://doi.org/10.1111/ectj.12097
8. Athey, S., Tibshirani, J., Wager, S.: Generalized random forests. The Annals of Statistics 47(2), 1148–1178 (2019). https://doi.org/10.1214/18-AOS1709
9. Frauen, D., Feuerriegel, S.: Estimating individual treatment effects under unobserved confounding using binary instruments. In: International Conference on Learning Representations (ICLR) (2023)
10. Law, R., Li, G., Fong, D.K.C., Han, X.: Tourism demand forecasting: A deep learning approach. Annals of Tourism Research 75, 410–423 (2019). https://doi.org/10.1016/j.annals.2019.01.014
11. Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C.: Graph WaveNet for deep spatial-temporal graph modeling. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), pp. 1907–1913 (2019). https://doi.org/10.24963/ijcai.2019/264
12. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. In: International Conference on Learning Representations (ICLR) (2023)
13. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: iTransformer: Inverted transformers are effective for time series forecasting. In: International Conference on Learning Representations (ICLR) (2024)
14. Wang, S., Wu, H., Shi, X., Hu, T., Luo, H., Ma, L., Zhang, J.Y., Zhou, J.: TimeMixer: Decomposable multiscale mixing for time series forecasting. In: International Conference on Learning Representations (ICLR) (2024)
15. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 9, pp. 11121–11128 (2023). https://doi.org/10.1609/aaai.v37i9.26317
16. Shi, C., Blei, D.M., Veitch, V.: Adapting neural networks for the estimation of treatment effects. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 32 (2019)
17. Melnychuk, V., Frauen, D., Feuerriegel, S.: Bounds on representation-induced confounding bias for treatment effect estimation. In: International Conference on Learning Representations (ICLR) (2024)

18. Touvron, H., Martin, L., Stone, K., et al.: Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288 (2023). https://doi.org/10.48550/arXiv.2307.09288
19. Asai, A., Wu, Z., Wang, Y., Sil, A., Hajishirzi, H.: Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In: International Conference on Learning Representations (ICLR) (2024)
20. Hu, Z., Dong, Y., Wang, K., Sun, Y.: Heterogeneous graph transformer. In: Proceedings of The Web Conference (WWW), pp. 2704–2710 (2020). https://doi.org/10.1145/3366423.3380027