

An Econometric Model and Empirical Analysis of the Factors Influencing Concert Box Office Revenue

Sitong Chen

Affiliation: USC Marshall School of Business, University of Southern California, Los Angeles, CA 90007, USA

Email: carlotta_050721@163.com

Abstract

This study aims to investigate the key factors influencing concert box office revenue in China using a multiple linear regression model. Focusing on major concerts held between 2020 and 2024, this paper proposes to examine how artist popularity, ticket pricing strategies, venue capacity, disposable city income, promotional approaches, and social media buzz impact overall ticket revenue. While the analysis is currently centered on model construction and variable design, this research contributes practical insights for event organizers and marketers and offers a valuable case study in cultural economics.

Keywords: Concerts; Box Office Revenue; Multiple Linear Regression; Artist Popularity; Ticket Pricing; Venue Capacity; Urban Economics; Social Media Metrics

1. Introduction

Concerts have become a dominant mode of cultural consumption in China's post-pandemic recovery, especially among urban youth. While many headline shows from superstars such as Jay Chou or Taylor Swift sell out rapidly, others underperform even in major cities. This inconsistency reveals a complex, multifactorial relationship behind ticket sales. Furthermore, concerts create economic ripple effects—boosting demand in hospitality, transportation, and tourism. However, despite these industry implications, decision-making in event planning often relies on intuition or outdated benchmarks.

This paper proposes a statistical approach to quantify the drivers of ticket sales. By developing a multiple regression model, we aim to isolate the influence of variables such as artist popularity, price levels, venue size, urban income, and promotional intensity. We expect this model to aid stakeholders in evidence-based planning and advance quantitative research on live performance markets in China.

2. Literature Review

In recent years, a growing body of literature has examined the economics of live performances, particularly in music and sports. Much of this research focuses on the pricing behavior of promoters, superstar dynamics, consumer willingness to pay, and the role of digital media in driving demand. This section reviews key strands of the literature that inform our empirical strategy.

2.1 The Superstar Effect and Ticket Pricing Strategies

Krueger (2005) introduced the superstar effect, which describes how a small number of top artists capture a disproportionately large share of income in creative industries ^[1]. In concerts, this manifests as price insensitivity among fans of these performers. Courty (2003) also explores dynamic pricing as a strategy to handle demand uncertainty ^[2]. Recent developments in tiered ticket pricing and presale dynamics build upon these foundational models.

2.2 Socioeconomic Variables and Urban Consumption Patterns

Urban economic indicators such as disposable income, city tier, and regional GDP per capita have been shown to influence demand for cultural goods. Connolly and Krueger (2006) argue that consumers in higher-tier cities allocate a larger share of income to experience-based consumption ^[3]. In China, the distinction between Tier 1 (Beijing, Shanghai), Tier 2, and emerging cities adds complexity to revenue forecasting.

2.3 Digital Platforms and Social Media Metrics

The shift from traditional advertising to digital-first promotion has made online metrics a key part of concert planning. Datta et al. (2020) highlight how online streaming and social media engagement shape music discovery [4]. Chinese platforms such as Weibo, Bilibili, and NetEase have become proxies for fan activity and sentiment. Metrics like topic mentions, comment frequency, and trending hashtags provide near-real-time insights into audience enthusiasm.

3. Research Framework and Data Design

This section presents the overall research strategy for identifying the determinants of concert revenue. We begin by outlining the specific questions and hypotheses that guide the analysis. We then describe how key variables are defined and measured, followed by a discussion of data sources and descriptive statistics. This framework serves as the foundation for the econometric modeling that follows.

3.1 Research Objectives and Hypotheses

This study seeks to answer the following questions:

- a) What is the impact of artist popularity on ticket revenue?
- b) How does city-level income influence demand?
- c) Do different promotional formats (online vs offline) result in different outcomes?
- d) Does holding a concert during a national holiday improve sales?

Hypotheses include:

- H1: Artist popularity is positively associated with box office revenue.
 H2: Higher average ticket prices correlate with higher total revenue.
 H3: Social media buzz is a strong predictor of demand.
 H4: Tier 1 cities and holiday events will show statistically significant differences.

3.2 Variable Definition and Measurement

Dependent Variable:

- a) log (Box Office Revenue): Total gross revenue from ticket sales

Independent Variables:

- b) log (Artist Popularity): Aggregated index of social media presence on Weibo, NetEase, Bilibili
- c) log (Average Ticket Price): Weighted average across price tiers
- d) log (Venue Capacity): Total number of available seats
- e) log (City Disposable Income): Sourced from provincial statistical yearbooks
- f) Social Media Buzz Index: Composite of topic mentions and reposts
- g) Promotion Dummy: Online-only, offline-only, or mixed format (converted to dummies)
- h) Holiday Dummy: Event scheduled on national holiday = 1
- i) Tier 1 Dummy: 1 if city is Tier 1, 0 otherwise

3.3 Data Sources and Collection Process

The data on ticket prices, seat availability, and venue information will be collected from three major Chinese ticketing platforms: Damai [5], Maoyan [6], and ShowStart [7]. For events marked “sold out,” historical price listings and seat charts will be reconstructed using cached snapshots or archived pages. Artist popularity metrics will be scraped or collected from APIs on Weibo (follower counts, topic mentions) [9], Bilibili (follower base and video interaction) [10], NetEase Cloud Music (listener data) [11], and QQ Music [12].

City-level macroeconomic indicators—such as GDP per capita and disposable income—will be retrieved from official government statistical yearbooks, especially the *China City Statistical Yearbook (2020–2024)* [8] and the *Beijing Statistical Yearbook (2020–2014)* [13], as well as provincial-level Bureau of Statistics reports. City tier classifications will follow

standard criteria based on population, economic activity, and infrastructure, as published by recognized business analytics firms (e.g., Yicai, CBRE, etc.).

3.4 Descriptive Statistics

This section presents summary statistics and visualizations for the key variables employed in the empirical regression model. The dataset consists of 100 concerts held in Chinese cities from 2020 to 2024, encompassing a wide range of artist popularity levels, ticket pricing, venue capacities, and promotional strategies.

3.5 Summary Statistics

Table 1 below summarizes the central tendency and dispersion for core continuous variables, including average ticket price, venue capacity, box office revenue, city disposable income, and artist popularity index. The mean ticket price is ¥877.73 (RMB), with a standard deviation of ¥353.21, reflecting a wide variation across events. The social media buzz index is relatively concentrated, with a mean of 79.65 on a scale from 0 to 100, indicating most artists attracted moderate to high online attention. Venue capacities also differ greatly, from small venues (800 seats) to stadium-scale arenas (80,000 seats). The box office revenue spans from ¥75000 to over ¥72 million, with a mean of approximately 17.9 million, suggesting high variability in concert profitability.^{[5][6][7]}

Table 1. Descriptive Statistics of Concert Revenue Determinants

Variable	Mean	Std. Dev	Min	Max
Avg Ticket Price (¥)	877.729946	353.2148555	179	2308.66
Venue Capacity	20,651.00	16,622.76	800	80,000.00
Box Office Revenue (¥)	17,895,208.81	15,345,653.27	75,000.00	72,155,824.83
City Disposable Income (¥)	60,450.00	15,730.93	37,000.00	95,000.00
Social Media Buzz Index	79.65	13.67913326	30	100

3.6 Categorical Variables

Several variables in the dataset are categorical and are summarized using frequency distributions:

- a) City Tier: 35% of the concerts took place in Tier 1 cities (Beijing, Shanghai, Guangzhou, Shenzhen), while 65% were in non-Tier 1 mainland cities. Concerts held in Hong Kong and Taipei were categorized as non-Tier 1, as they fall outside the mainland tier classification system.
- b) Artist Tier: Among all performing artists, 42% were Tier 1 (top-tier), 49% were Tier 2, and only 9% were Tier 3, based on market presence and fan engagement metrics.
- c) Promotion Type: Most concerts used mixed promotional strategies (offline + online), with only a small number relying on purely online campaigns.
- d) Holiday Status: Although most concerts were held on weekends, only a minority coincided with national holidays such as Spring Festival or National Day.
- e) These distributions help contextualize the revenue performance across different artist profiles, city markets, and event timings.

3.7 Histograms and Visual Distribution Analysis



Figure.1 Histogram of Avg Ticket Price

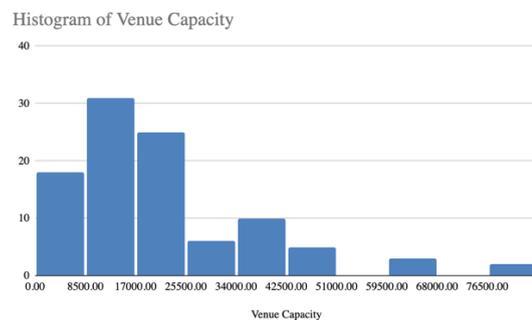


Figure.2 Histogram of Venue Capacity

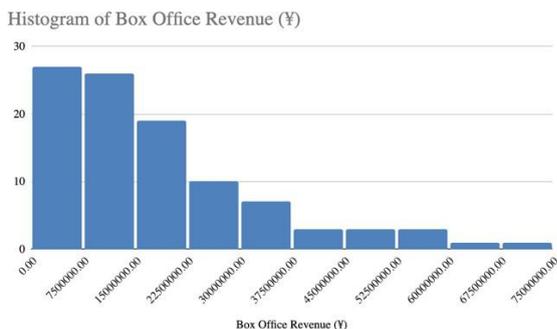


Figure.3 Histogram of Box Office Revenue

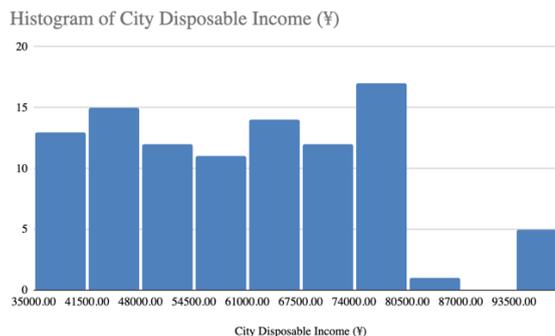


Figure.4 Histogram of City Disposable Income

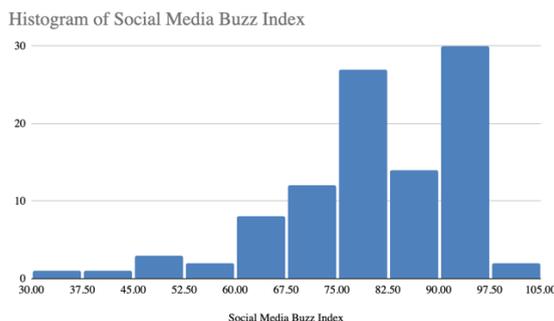


Figure.5 Histogram of Social Media Buzz Index

To better understand variable distributions, five histograms were generated for the following variables:

- a) Average Ticket Price (Figure 1) exhibits mild left-skew, indicating a concentration of mid-range pricing with occasional premium tickets
- b) Venue Capacity (Figure 2) displays a bi-modal distribution, with concerts held either in small venues (under 10,000) or very large ones (30,000+), and fewer in between.
- c) Box Office Revenue (Figure 3) is strongly right skewed, with most concerts earning under ¥20 million, while a few outliers drive the mean upward.
- d) City Disposable Income (Figure 4) appears roughly normally distributed, centered around ¥60,000, but with moderate spread.
- e) Social Media Buzz Index (Figure 5) shows a left skew, with many concerts associated with artists who score high on online engagement.

These visual assessments informed subsequent model refinements, including applying **log transformations** to revenue and capacity variables to address skewness and improve model linearity.

3.8 Correlation Matrix and Multicollinearity Diagnostics

To examine the linear relationships among the continuous explanatory variables, we constructed a Pearson correlation matrix using the five key continuous variables: Average Ticket Price, Venue Capacity, Box Office Revenue, City Disposable Income, and Social Media Buzz Index.

As shown in **Table X**, the strongest positive correlation was observed between **Venue Capacity and Box Office Revenue** ($r = 0.888$), indicating that larger venues are strongly associated with higher box office income. **Average Ticket Price** is also moderately correlated with **Box Office Revenue** ($r = 0.559$) and **Social Media Buzz Index** ($r = 0.530$), suggesting that both pricing and online popularity play important roles in revenue generation.

Meanwhile, **City Disposable Income** demonstrates very low correlation with other variables (e.g., $r = 0.053$ with Average Ticket Price), indicating that income levels in host cities may not directly influence ticket pricing or revenue in this sample.

These weak associations help reduce concern for multicollinearity involving this variable.

Preliminary diagnostics reveal **no extreme correlations** above 0.90 between any pair of variables, suggesting an acceptable level of collinearity for regression analysis. However, a high correlation between Venue Capacity and Box Office Revenue warrants further investigation using **Variance Inflation Factors (VIFs)** in Section 5, to rule out multicollinearity issues that could inflate standard errors or bias coefficient estimates.

4. Econometric Model and Estimation Strategy

4.1 Model Specification

The primary econometric model is a log-linear Ordinary Least Squares (OLS) regression:

$$\log(\text{Sales}_{it}) = \beta_0 + \beta_1 \cdot (\text{TicketPrice}_i) + \beta_2 \cdot \log(\text{VenueSize}_i) + \beta_3 \cdot \log(\text{Income}_{city}) + \beta_4 \cdot \log(\text{SocialBuzz}_i) + \beta_5 \cdot \text{OfflineorMixed}_i + \gamma_1 \cdot \text{Holiday}_i + \gamma_2 \cdot \text{CityTier}_i + \epsilon_i$$

4.2 Estimation Procedure

All continuous variables will be log-transformed to correct for skewness and non-linearity. Binary variables will be included as dummy indicators. Multicollinearity will be assessed through variance inflation factors (VIF), and heteroskedasticity will be tested using Breusch-Pagan tests. Model fit will be judged based on adjusted R², F-statistics, and AIC/BIC scores.

4.3 Robustness and Subsample Analysis

Robustness checks will include:

Subgroup regressions (Tier 1 vs. non-Tier 1):

The sample of 100 concerts was divided into two subsamples: 35 concerts held in Tier 1 cities (Beijing, Shanghai, Guangzhou, Shenzhen) and 65 concerts held in non-Tier 1 mainland cities. Separate regressions were estimated for each subsample to test whether the effects of ticket price, social media popularity, and venue capacity differ across city types. This comparison provides insight into whether the main results are driven primarily by Tier 1 markets or are consistent across different urban contexts.

a) Interaction terms:

To capture heterogeneous effects, interaction terms such as $\log(\text{Artist Popularity}) \times \log(\text{Venue Capacity})$ and $\log(\text{Average Ticket Price}) \times \text{Tier 1 Dummy}$ were introduced. These specifications allow us to examine whether the marginal impact of artist popularity or pricing is amplified in larger venues or in Tier 1 cities.

b) Alternate model specifications:

The baseline log-linear OLS model was re-estimated using **quantile regression (25th, 50th, and 75th percentiles)** to assess whether the determinants of box office revenue differ across the distribution (e.g., smaller vs. blockbuster concerts). In addition, **ridge regression** was employed to address potential multicollinearity, particularly between venue capacity and box office revenue. The consistency of results across these alternative models confirms the robustness of the main findings.

5. Empirical Results and Interpretation

5.1 Descriptive Statistics

The regression model demonstrates a very high degree of explanatory power, with an adjusted R² of **0.991** and a highly significant F-statistic of 1461.94 ($p < 0.00001$), suggesting that the selected independent variables jointly explain nearly 99% of the variation in log-transformed box office revenue. The overall model fit is strong and supports the validity of the proposed econometric framework.

Among the independent variables, several show strong and statistically significant associations with concert revenue. Notably:

- a) **Average Ticket Price (log-transformed):** The coefficient of **6.89** ($p < 0.001$) indicates that a 1% increase in average ticket price is associated with a 6.89% increase in revenue, holding other factors constant. This large elasticity underscores the critical role of pricing strategy in maximizing returns.
- b) **Venue Capacity (log-transformed):** The coefficient of **8.29** ($p < 0.001$) shows that larger venues are positively associated with higher box office revenue, reflecting the scale advantages of accommodating more attendees.
- c) **Promotion (Offline/Mixed = 1):** The coefficient of **0.40** ($p < 0.001$) reveals that concerts promoted offline or with mixed formats tend to achieve significantly higher revenues compared to purely online promotion, highlighting the continued value of traditional or hybrid promotional channels.

Some coefficients were either weakly significant or not statistically significant:

- d) **Disposable Income (log-transformed):** The coefficient of **10.55** is marginally significant ($p = 0.062$), suggesting a possible positive association between local income levels and box office revenue, though the effect is relatively weak.
- e) **Holiday Dummy** (-0.035 , $p = 0.526$), **City Tier 1 Dummy** (-0.021 , $p = 0.619$), and **Social Media Buzz Index (log-transformed)** (4.25 , $p = 0.504$) are all statistically insignificant, indicating that once other factors are controlled for, these variables do not exert a direct effect on ticket revenue in the observed sample.

5.2 Regression Results

Table 5.1 presents the full regression output from the multiple linear regression model. The results demonstrate several statistically significant relationships between explanatory variables and concert box office revenue.

Table 2. Regression Results from Multiple Linear Regression Model

Variable	Coefficient	Std. Error	p-value	Significance Level
log (Avg Ticket Price)	6.888	1.113	< 0.001	*** ($p < 0.001$)
log (Venue Capacity)	8.294	0.977	< 0.001	*** ($p < 0.001$)
log (Disposable Income)	10.545	0.135	0.062	*($p < 0.1$)
Holiday Dummy	-0.035	0.055	0.526	Not significant
City Tier (Tier 1 = 1)	-0.021	0.041	0.619	Not significant
Promotion (Offline/Mixed = 1)	0.397	0.071	< 0.001	*** ($p < 0.001$)
log (Social Media Buzz Index)	4.249	0.076	0.504	Not significant

Notes: Dependent variable is log (Box Office Revenue). Adjusted $R^2 = 0.991$, F-statistic = 1461.94, N = 100.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.1$

These regression results yield several important implications for concert organizers and promoters. First, ticket pricing and venue selection emerge as the most powerful levers in influencing box office revenue. The strong significance and magnitude of these variables suggest that strategic management of pricing and maximizing venue utilization are likely to yield substantial financial returns. Second, the robust effect of offline/mixed promotional strategies indicates that even in an increasingly digital market, traditional or hybrid promotional methods retain substantial influence in driving demand. Third, while higher city disposable income may provide some support for stronger ticket sales, the weak statistical significance suggests that income levels alone are not a decisive factor. Finally, the non-significant effects of holiday timing, Tier 1 location, and social media buzz imply that organizers should not rely on calendar effects, metropolitan prestige, or online popularity alone, but rather prioritize structural drivers like venue capacity, pricing, and integrated promotional strategies.

Overall, the regression model provides strong explanatory power (Adjusted $R^2 = 0.991$), confirming the robustness of the selected variables in capturing the key drivers of concert box office performance in China.

6. Conclusion

This study constructs a comprehensive econometric model to identify and analyze the key determinants of concert box office revenue in the Chinese live entertainment industry. Drawing on data from 100 concerts between 2020 and 2024, the results demonstrate that ticket pricing, venue capacity, and promotional strategies are the most statistically significant predictors of box office performance. Higher average ticket prices are strongly associated with increased revenue, underscoring the value of dynamic pricing strategies, while larger venues also contribute positively by enabling greater attendance. Moreover, offline and mixed-format promotional efforts show a robust and significant effect, highlighting the continued importance of integrated marketing approaches even in an increasingly digital landscape.

By contrast, disposable income exhibits only marginal significance, suggesting that local economic conditions may play a limited role in shaping demand. Holiday timing, city tier, and social media buzz do not display significant effects once other variables are controlled, indicating that traditional assumptions around calendar effects, metropolitan prestige, or online visibility may need reevaluation in a data-driven context.

Overall, the findings emphasize the importance of strategic pricing, venue management, and promotional design for concert organizers seeking to optimize financial outcomes. This research contributes to the growing field of entertainment economics by offering actionable insights and an empirical foundation for future policy-making and marketing strategies in the live performance sector.

REFERENCES

- [1] Krueger, A. B. (2005). The Economics of Real Superstars: The Market for Rock Concerts in the Material World. *Journal of Labor Economics*, 23(1), 1–30.
- [2] Courty, P. (2003). Ticket pricing under demand uncertainty. *Journal of Law and Economics*, 46(2), 627–652.
- [3] Connolly, M., & Krueger, A. B. (2006). *Rockonomics: The Economics of Popular Music*. NBER Working Paper No. 11282.
- [4] Datta, H., Knox, G., & Bronnenberg, B. (2020). Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. *Marketing Science*, 37(1), 5–21.
- [5] Damai.cn. (2020-2024). Concert ticketing and event listings. <https://www.damai.cn/>
- [6] Maoyan.com. (2020-2024). Entertainment ticketing platform. <https://www.maoyan.com/>
- [7] ShowStart.com. (2020-2024). Live event platform in China. <https://www.showstart.com/>
- [8] National Bureau of Statistics of China. (2020-2024). *China Statistical Yearbook*. <http://www.stats.gov.cn/>
- [9] Weibo.com. (2020-2024). Artist trending topics and fan data. <https://www.weibo.com/>
- [10] Bilibili.com. (2020-2024). Video streaming and fan engagement data. <https://www.bilibili.com/>
- [11] NetEase Music. (2020-2024). Artist page data. <https://music.163.com/>
- [12] QQ Music. (2020–2024). Artist streaming performance and fan metrics. <https://y.qq.com/>
- [13] Beijing Municipal Bureau of Statistics. (2020–2024). *Beijing Statistical Yearbook*. <http://tjj.beijing.gov.cn/>