

Analysis of Information Revelation Impact on the Popularity of English-Language Fiction Books

In this analysis, the factors influencing the popularity of books, as measured by the number of downloads, are explored. The focus is on the Kullback-Leibler Divergence (KLD) values to capture various aspects of narrative structure and content. Specifically, the study examines the skewness, kurtosis, cumulative sum, and rolling mean of KLD values to understand their asymmetry, peakedness, cumulative effect, and local trends throughout the narrative. Additionally, other important book attributes, including narrative pacing (speed), average sentiment score (sentiment_avg), and sentiment volatility (sentiment_vol), are considered. Dummy variables for different genres (genre_*) are also included to capture genre-specific effects. The methodology involves three key steps: data processing to derive book-level measures from KLD scores, regression analysis using Ordinary Least Squares (OLS) to relate these measures and book attributes to log-transformed download counts, and LASSO regression to identify the most significant predictors independently influencing the log of downloads. This comprehensive approach aims to uncover the critical factors driving book popularity and provide insights into the interplay between narrative characteristics and reader engagement.

Detailed Tables and Variable Descriptions

Regression Tables

OLS Regression Results						
Dep. Variable:	log_downloads	R-squared:	0.084			
Model:	OLS	Adj. R-squared:	0.083			
Method:	Least Squares	F-statistic:	111.4			
Date:	Mon, 01 Jul 2024	Prob (F-statistic):	6.17e-157			
Time:	22:26:39	Log-Likelihood:	-12741.			
No. Observations:	8534	AIC:	2.550e+04			
Df Residuals:	8526	BIC:	2.555e+04			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	4.2951	0.244	17.630	0.000	3.818	4.773
skewness_kld	-0.1056	0.044	-2.403	0.016	-0.192	-0.019
kurtosis_kld	0.0108	0.007	1.570	0.116	-0.003	0.024
cumulative_kld	-1.0532	0.169	-6.214	0.000	-1.385	-0.721
rolling_mean_kld	53.9194	8.401	6.418	0.000	37.451	70.387
speed	-4.3381	0.792	-5.480	0.000	-5.890	-2.786
sentiment_avg	-9.1505	0.557	-16.442	0.000	-10.241	-8.060
sentiment_vol	102.3627	8.407	12.175	0.000	85.882	118.843
Omnibus:	1653.220	Durbin-Watson:	1.491			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3892.383			
Skew:	1.091	Prob(JB):	0.00			
Kurtosis:	5.488	Cond. No.	1.07e+04			

LASSO Regression Table

```
LASSO Coefficients:
const          0.000000
skewness_kld   -0.114902
kurtosis_kld   -0.000000
cumulative_kld -0.000000
rolling_mean_kld 0.007440
speed          -0.101239
sentiment_avg  -0.203645
sentiment_vol   0.202774
dtype: float64
Important Predictors identified by LASSO:
['skewness_kld', 'rolling_mean_kld', 'speed', 'sentiment_avg', 'sentiment_vol']
```

Overview of Results

The analysis examines the relationship between narrative information revelation, measured by Kullback-Leibler divergence (KLD), and the popularity of English-language fiction books. Book-level measures, including skewness, kurtosis, cumulative KLD, and rolling mean KLD, were calculated. Our regression model indicates that higher skewness in KLD negatively impacts book downloads, suggesting that books with uneven information distribution are less popular. Conversely, higher rolling mean KLD positively influences popularity, indicating that consistent

information revelation is favored by readers. Cumulative KLD shows a negative relationship, implying that books diverging significantly from the expected narrative are less downloaded. While kurtosis KLD was not statistically significant, it provides additional insight into the shape of the information distribution.

Heterogeneity and Predictors

Incorporating additional variables such as speed, sentiment average, and sentiment volatility into our model reveals that these factors also significantly affect book popularity. The LASSO regression identified skewness KLD, rolling mean KLD, speed, sentiment average, and sentiment volatility as the most important predictors of book popularity. These findings suggest that both the manner and consistency of information revelation, as well as the pacing and emotional tone of the narrative, are crucial for book popularity. Authors and publishers may benefit from focusing on these aspects to enhance the appeal of their works.