The Bestseller Phenomenon Revisited

The Diffusion of Books

What makes us buy:
Habits, adverts, or social contagion?
1 State of research on diffusion of books

2 Theory: 2 rather basic and 1 integrated model of diffusion

- External Influence Model
- Internal Influence Model
- Mixed Influence Model

3 Empirics: Do book characteristics determine the type of diffusion?

- Two-step approach, using book-level data
- Test for reliability: multi-level approach, still using macro-level data
- Test for validity: looking at readers, using micro-level data
Current state of research on diffusion of books

In general, successful hedonic goods (e.g. movies, music albums, books) exhibit heavy adoptions in their early life cycles and diminishing sales further on (see Clement et al. 2006 for an overview).


Both studies find two major patterns of diffusion:
- external diffusion (successfull start, strong decay trend)
- internal diffusion (bell-shaped curve, more sustainable growth in popularity)

Both studies offer no explanation of diffusional heterogeneity.

This presentation links varying patterns to readers’ knowledge on writers.
External Influence Model

\[ f(t) = a(1 - F(t)) \] , with \( a > 0 \) and \( N^* = \sum_{t} f(t) = 1. \)

\[ a = 0.1 \quad \text{---} \]
\[ a = 0.04 \quad \text{-----} \]

\( a = \) Parameter of Innovation, External Influence

*Mostly relevant for books by established writers?*
Example for External Diffusion

H. Mankell, Die Brandmauer → associated star power (existing knowledge on writer, high media coverage)
Internal Influence Model

\[ f(t) = bF(t)(1 - F(t)), \text{ with } F(t = 0) > 0 \text{ and } b > 0. \]

\[ b = 0.3 \quad \text{---} \]
\[ b = 0.15 \quad \text{-----} \]

\[ F(t = 0) = 0.01 \]

\[ b = \text{Parameter of Imitation, Internal/Social Influence} \]


 MOSTLY RELEVANT FOR NEWCOMER BOOKS?
Example for Internal Diffusion

L. Swann, Glennkill → no associated star power
Mixed Influence Model

\[ f(t) = (a + bF(t))(1 - F(t)) \]

Examples for MIM’s flexibility:

- Blockbuster Diffusion \( a = 0.04, \ b = 0.01 \)
- Newcomer Diffusion \( a = 0.001, \ b = 0.1 \)
Mixed Influence Model

\[ f(t) = (a + bF(t))(1 - F(t)) \]

*Estimation of both, EI and II, is possible!*

Examples for MIM’s flexibility:

- Blockbuster Diffusion: \( a = 0.04, \ b = 0.01 \)
- Newcomer Diffusion: \( a = 0.001, \ b = 0.1 \)

Hypotheses:

- H1: \( a_{Established} > a_{Newcomer} \)
- H2: \( b_{Established} < b_{Newcomer} \)
Data: Novels ranked Top 50 in Germany (2001 – 2006, \( n = 798 \)).
Only books included, which ranked Top 50 for a duration of \( \geq 10 \) weeks, \( n = 354 \).
Thus, 354 diffusions are examined.

Dep. variable: \( f(t) = y_t \), inverted sales rank in \( t \) (range 1 – 50).
Regressors: \( F(t) = x_t \), cumulated prior inverted sales ranks including \( t-1 \).

Non-linear least squares (NLLS) offers a continous estimation of MIM and thus book-specific \( a, b \) (e.g. Srinivasan/Mason 1989):

\[
y_i = (a + bx_i) (1 - x_i) + u_i
\]

\( a \) and \( b \) are estimated for 354 diffusions each, using 354 NLLS-regressions
OLS Regression, robust standard errors. NLLS estimates used as dep. variables (log \(a\), log \(b\)).

Regressors:  
- established writer  
- publisher’s marketing power  
- book’s quality  
- log run time  
- external influence (log \(a\))

<table>
<thead>
<tr>
<th>Regression Coefficients</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>established writer</td>
<td>0.165*</td>
<td>0.133†</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.056)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>marketing</td>
<td>0.006*</td>
<td>0.007*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.079)</td>
<td></td>
</tr>
<tr>
<td>quality</td>
<td></td>
<td></td>
<td>0.153**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.090)</td>
</tr>
<tr>
<td>log run time</td>
<td>-1.690***</td>
<td>-1.697***</td>
<td>-1.691***</td>
</tr>
<tr>
<td></td>
<td>(-0.818)</td>
<td>(-0.822)</td>
<td>(-0.819)</td>
</tr>
<tr>
<td>genre controlled</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>1.686***</td>
<td>1.592***</td>
<td>0.927**</td>
</tr>
<tr>
<td>(n)</td>
<td>354</td>
<td>354</td>
<td>354</td>
</tr>
<tr>
<td>(R^2) adj.</td>
<td>0.665</td>
<td>0.669</td>
<td>0.674</td>
</tr>
</tbody>
</table>

Unstand. coefficients, standardized coefficients in parentheses, † \(p<0.1\), * \(p<0.05\), ** \(p<0.01\), *** \(p<0.001\).
OLS Regression, robust standard errors. NLLS estimates used as dep. variables (log a, log b).

Regressors:
- established writer
- publisher’s marketing power
- book’s quality
- log run time

<table>
<thead>
<tr>
<th>Regressor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>established writer</td>
<td>0.165*</td>
<td>0.133†</td>
<td>0.082</td>
<td>-0.212*</td>
<td>-0.224*</td>
<td>-0.211*</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.056)</td>
<td>(0.035)</td>
<td>(-0.080)</td>
<td>(-0.084)</td>
<td>(-0.079)</td>
</tr>
<tr>
<td>marketing</td>
<td>0.006*</td>
<td>0.007†</td>
<td>0.153**</td>
<td>0.002</td>
<td>0.023</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.079)</td>
<td>(0.090)</td>
<td>(0.063)</td>
<td>(0.023)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>quality</td>
<td>-1.690***</td>
<td>-1.697***</td>
<td>-1.691***</td>
<td>-1.843***</td>
<td>-1.846***</td>
<td>-1.840***</td>
</tr>
<tr>
<td></td>
<td>(-0.818)</td>
<td>(-0.822)</td>
<td>(-0.819)</td>
<td>(-0.786)</td>
<td>(-0.787)</td>
<td>(-0.785)</td>
</tr>
<tr>
<td>log run time</td>
<td>-1.690***</td>
<td>-1.697***</td>
<td>-1.691***</td>
<td>-1.843***</td>
<td>-1.846***</td>
<td>-1.840***</td>
</tr>
<tr>
<td></td>
<td>(-0.818)</td>
<td>(-0.822)</td>
<td>(-0.819)</td>
<td>(-0.786)</td>
<td>(-0.787)</td>
<td>(-0.785)</td>
</tr>
<tr>
<td>genre</td>
<td>constant</td>
<td>1.686***</td>
<td>1.592***</td>
<td>0.927**</td>
<td>3.066***</td>
<td>3.029***</td>
</tr>
<tr>
<td></td>
<td>controlled</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
</tr>
<tr>
<td></td>
<td>R² adj.</td>
<td>0.665</td>
<td>0.669</td>
<td>0.674</td>
<td>0.631</td>
<td>0.631</td>
</tr>
</tbody>
</table>

Unstand. coefficients, standardized coefficients in parentheses, † p<0.1, * p<0.05, ** p<0.01, *** p<0.001.
Two Level-Analysis
Efficient, integrated estimation of book characteristics’ effects on diffusion curves (see Rossman et al. 2008 on payola in pop music radio).

Dep. variable: \( y_{it} \), book \( i \)'s inverted sales rank in \( t \)

Regressors Level 1: \( x_{it}, x_{it}^2 \), cumulated inverted ranks (Bass’ linearization); within book variation

Level 2: book characteristics as interactions of \( x_{it}, x_{it}^2 \); between books variation
- writer’s star power
- publisher’s marketing power
- quality
- log run time
Effects of book characteristics on external and internal influence.

Random Effects Model.
Estimation by Restricted Maximum-Likelihood (REML), e.g. Snijders/Bosker (1999).

*R² within: 0.55, R² between: 0.87. Results extracted.*

<table>
<thead>
<tr>
<th></th>
<th>external influence</th>
<th>internal influence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>% change</td>
</tr>
<tr>
<td>established writer</td>
<td>+ 0.008*** (0.001)</td>
<td>+ 17.4</td>
</tr>
<tr>
<td>marketing</td>
<td>0.000 (0.000)</td>
<td>0</td>
</tr>
<tr>
<td>quality</td>
<td>− 0.002 (0.001)</td>
<td>− 4.3</td>
</tr>
</tbody>
</table>

Unstand. coefficients, standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001.
Effects of book characteristics on external and internal influence.

Random Effects Model.
Estimation by Restricted Maximum-Likelihood (REML), e.g. Snijders/Bosker (1999).

$R^2$ within: 0.55, $R^2$ between: 0.87. Results extracted.

<table>
<thead>
<tr>
<th></th>
<th>external influence</th>
<th>internal influence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td></td>
<td>% change</td>
<td>% change</td>
</tr>
<tr>
<td>established writer</td>
<td>+ 0.008***</td>
<td>− 0.023***</td>
</tr>
<tr>
<td></td>
<td>17.4</td>
<td>− 30.7</td>
</tr>
<tr>
<td>marketing</td>
<td>0.000 (0.000)</td>
<td>− 0.000* (0.000)</td>
</tr>
<tr>
<td>quality</td>
<td>− 0.002 (0.001)</td>
<td>+ 0.012** (0.004)</td>
</tr>
<tr>
<td></td>
<td>− 4.3</td>
<td>+ 16.0</td>
</tr>
</tbody>
</table>

Unstand. coefficents, standard errors in parentheses, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Results from two-step approach: + 17 / 8 % − 21 %
Which diffusion channel is being used to stimulate individual adoption?

Basic hypotheses:
- Books by established writers are adopted thru external stimulation.
- Books by newcomers are adopted thru internal, social stimulation.

Link to social network analysis:
- Socially well embedded readers (e.g. city size) adopt thru internal stimulation.
Data: GfK Consumer Panel 2003-04; 20,000 individual participants (Germans only).

Sample: The analysis is confined to hardcover books, which were ranked Top 50. To get reasonable group sizes, only books with a total number of ≥ 30 observed readers are included. This leads to 1,181 observed purchases $i$ of 24 books $j$. 
Data: GfK Consumer Panel 2003-04; 20,000 individual participants (Germans only).

Sample: The analysis is confined to hardcover books, which were ranked Top 50. To get reasonable group sizes, only books with a total number of ≥ 30 observed readers are included. This leads to 1,181 observed purchases $i$ of 24 books $j$.

Dep. var.: „How did you learn about the book you bought?“

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>freq.</th>
<th>%</th>
<th>freq.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 media</td>
<td>147</td>
<td>12,5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 advertisement</td>
<td>92</td>
<td>7,8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 catalogue</td>
<td>391</td>
<td>33,1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 shop decor</td>
<td>296</td>
<td>25,1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 bookseller’s tip</td>
<td>30</td>
<td>2,5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 word of mouth</td>
<td>129</td>
<td>10,9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 bestseller list</td>
<td>96</td>
<td>8,1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>External Influence ($y=0$)</td>
<td>956</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Influence ($y=1$)</td>
<td>225</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$ purchases</td>
<td>1,181</td>
<td></td>
<td>1,181</td>
<td></td>
</tr>
<tr>
<td>$m$ books</td>
<td>24</td>
<td></td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Marc Keuschnigg

30.11.2009 # 19
Hierarchical logistic regression with random intercepts

Dichotomous dep. var.: \( y_{ij} = 1 \), individual adoption thru internal stimulation
\( y_{ij} = 0 \), individual adoption thru external stimulation

Regressors  Level 1: **buyer characteristics; within book variation**
- embeddedness/social capital
  - city size  (5 categories, quasi-metric)
  - education  (dummy, 1 if \( \geq \) Fachhochschulreife)
  - high income  (dummy, 1 if \( \geq \) 3,000 EUR)
- sex  (dummy, 1 if buyer is female)
- week of purchase  (# weeks since release)

Level 2: **book characteristics; between books variation**
- established writer  (dummy, 1 if previous Top 50-titles exist)
- publ’s marketing power  (dummy, 1 if belonging to conglomerate)
- log run time  (log # weeks ranked in Top 50)
Odds for social contagion.
1,181 purchases of 24 books.
Dichotomous dep. variable: 1 if individual adoption by internal stimulus.

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratios</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>established writer</td>
<td>0.550**</td>
<td>-45.0</td>
</tr>
<tr>
<td>marketing</td>
<td>0.668*</td>
<td>-33.2</td>
</tr>
<tr>
<td>log run time</td>
<td>1.538***</td>
<td>+53.2</td>
</tr>
<tr>
<td>city size</td>
<td>1.164*</td>
<td>+16.4</td>
</tr>
<tr>
<td>high education</td>
<td>1.591**</td>
<td>+59.1</td>
</tr>
<tr>
<td>high income</td>
<td>1.523**</td>
<td>+52.3</td>
</tr>
<tr>
<td>female</td>
<td>1.745**</td>
<td>+74.5</td>
</tr>
<tr>
<td>week of purchase</td>
<td>1.008</td>
<td>+0.8</td>
</tr>
</tbody>
</table>

$m$ purchases 1,181
$n$ books 24

*Nagelkerkes $R^2$* 0.079

Odds ratios ($e^\beta$),
% change ($[e^\beta - 1] \cdot 100$),
† $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. 
Diffusion of books differs by writers’ star power, i.e. readers’ knowledge on writers.

- **Book level analysis**
  
  Diffusion for **established writers** follows a **blockbuster-shaped curve** (hints external diffusion).
  
  Diffusion for **newcomers** follows a **hump-shaped curve** (hints internal diffusion).

- **Reader level analysis**
  
  Indeed, **established books** get adopted primarily by **external stimulus**.
  
  Indeed, **newcomer books** get adopted primarily by **social stimulus**.

Also: Type of individual adoption depends on

- reader’s network **embeddedness**: well connected agents make greater use of **social learning**.
- reader’s **sex**: female readers make greater use of **social learning**.
Thank you all for your attention!

References
