Institut für Soziologie, LMU
Marc Keuschnigg

The Bestseller Phenomenon Revisited

The Diffusion of Books

What makes us buy:

Habits, adverts, or social contagion?

Diffusion of Books: Contents

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

Beck (2007) offers an exemplary fit of an elaborate diffusion model on book sales (n = 4).

Deschâtres / Sornette (2005) analyze books‘ sales patterns on Amazon.com (n = 138).

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.

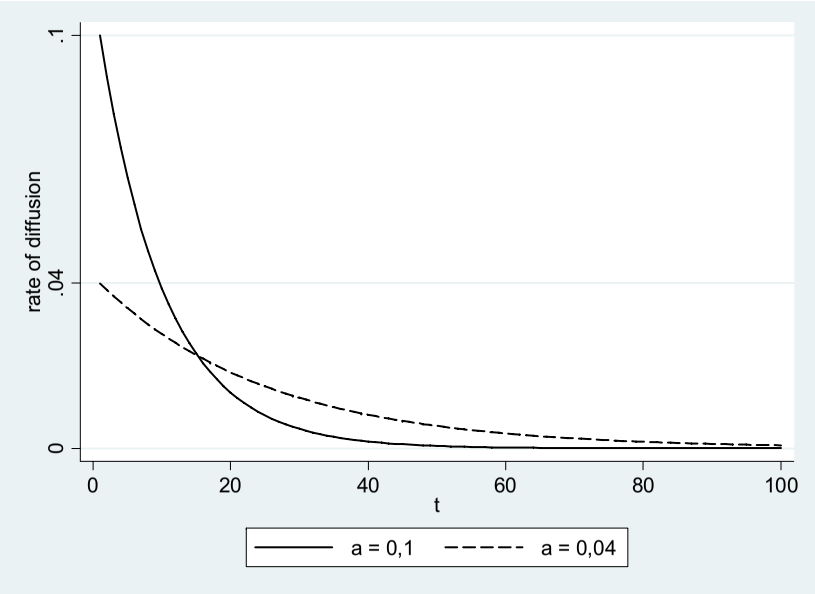
Basic Models of Diffusion: EIM

External Influence Model

$f(t) = a(1 - F(t))$, with $a > 0$ and $N^* = \sum_t f(t) = 1$.

$a = 0.1$ ——

$a = 0.04$ - - - -

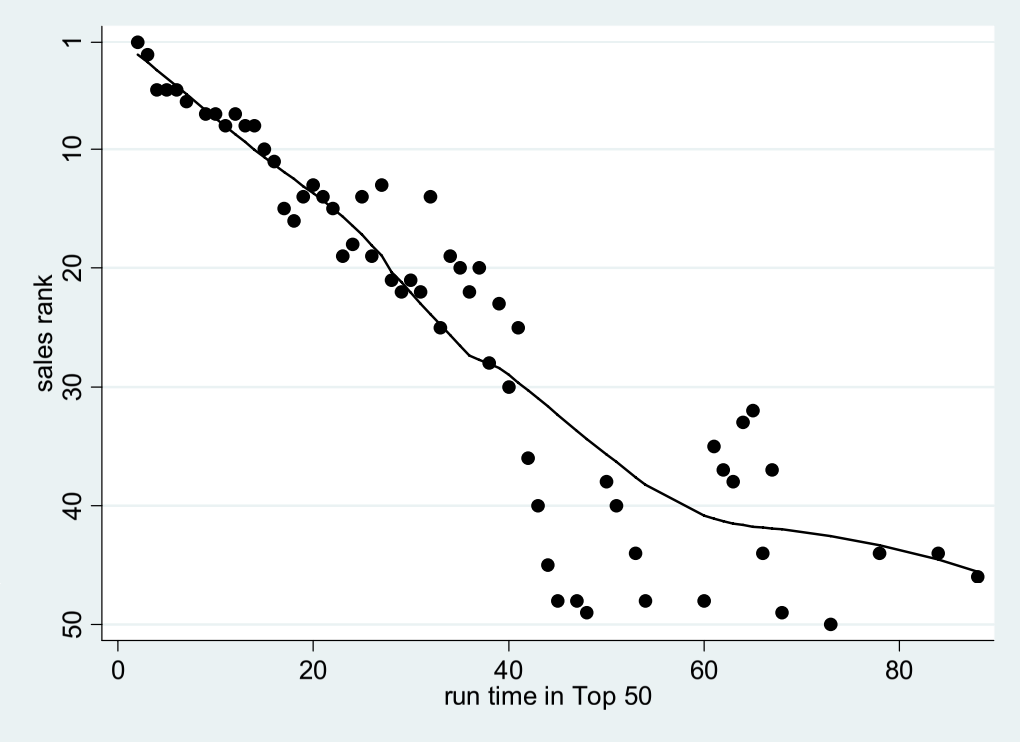


a = Parameter of Innovation, External Influence

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)



Basic Models of Diffusion: IIM

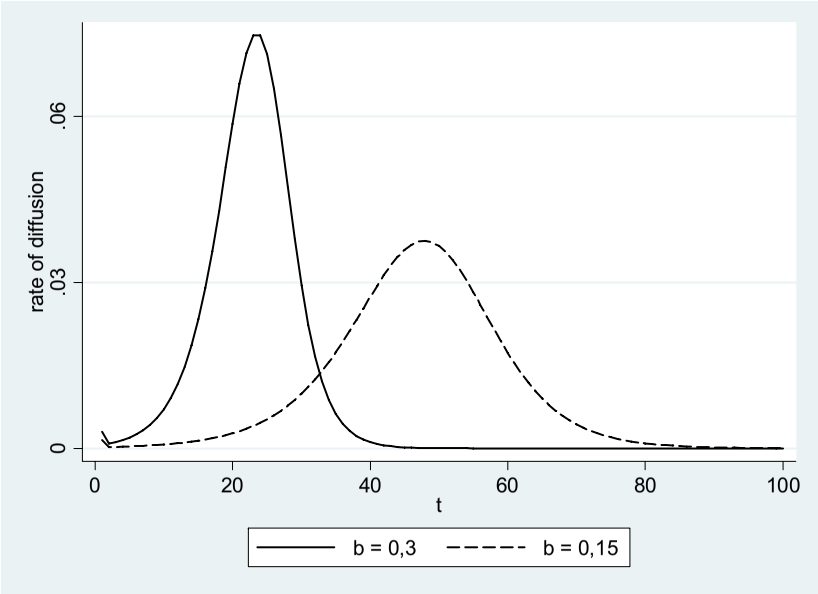
Internal Influence Model

$f(t) = bF(t)\,(1 - F(t))$, with $F(t=0) > 0$ and $b > 0$.

b = 0.3 ——

b = 0.15 - - - -

F(*t* = 0) = 0,01

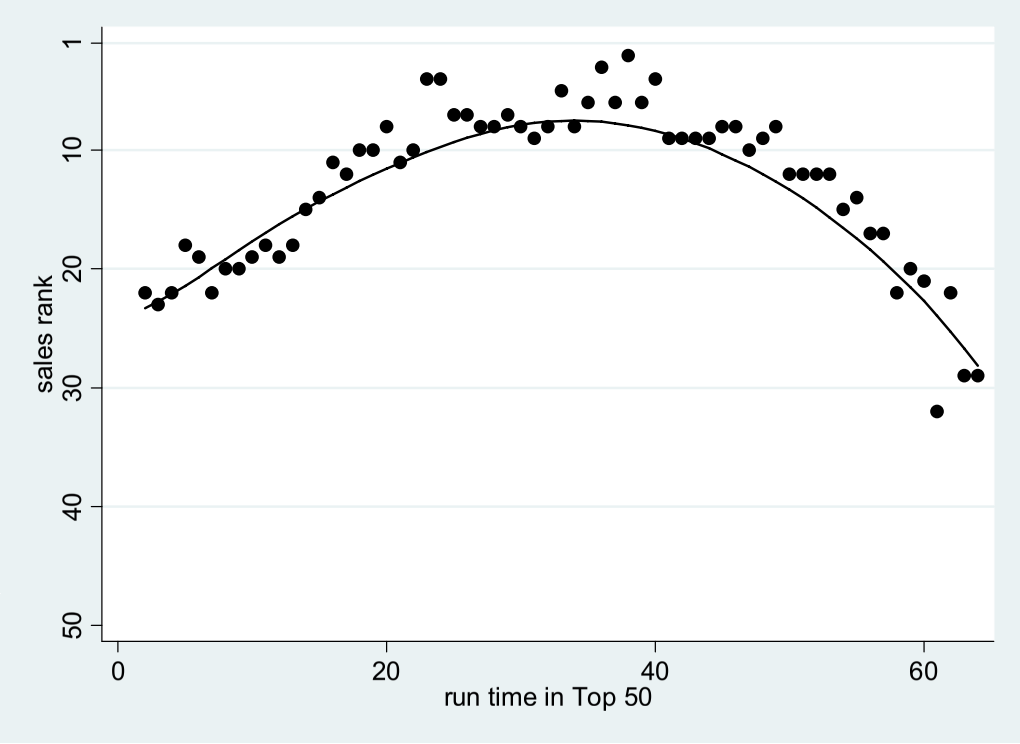


b = Parameter of Imitation, Internal/Social Influence

Mostly relevant for newcomer books?

Example for Internal Diffusion

L. Swann, Glennkill → no associated star power



Integrated Model of Diffusion: MIM

Mixed Influence Model

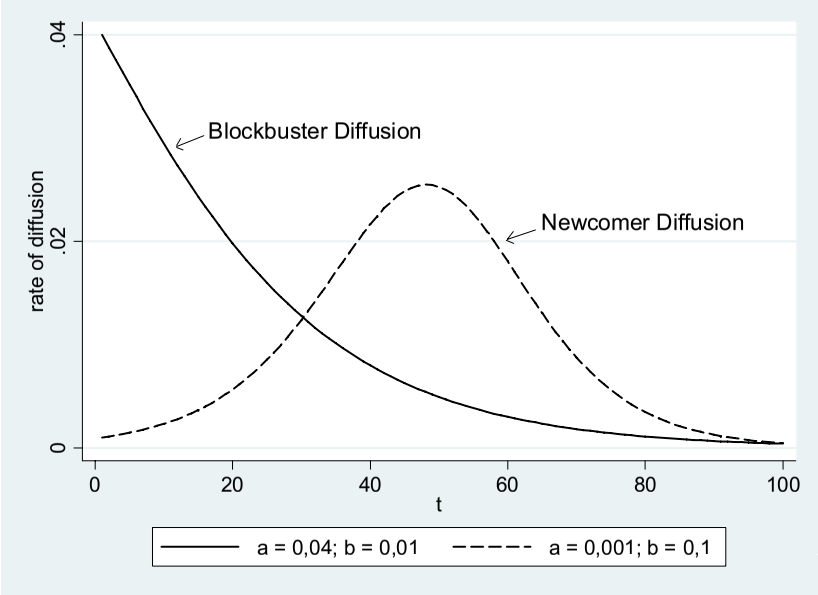
Estimation of both, EI and II, is possible!

$$f(t) = (a + bF(t))\,(1 - F(t))$$

Examples for MIM's flexibility:

Blockbuster Diffusion —— $a = 0.04$, $b = 0.01$

Newcomer Diffusion - - - - $a = 0.001$, $b = 0.1$



Integrated Model of Diffusion: MIM

Mixed Influence Model

Estimation of both, EI and II, is possible!

$$f(t) = (a + bF(t))\,(1 - F(t))$$

Examples for MIM's flexibility:

Blockbuster Diffusion —— $a = 0.04$, $b = 0.01$

Newcomer Diffusion - - - - $a = 0.001$, $b = 0.1$

Hypotheses:

H1 $a_{Established} > a_{Newcomer}$

H2 $b_{Established} < b_{Newcomer}$

Estimation of *a* and *b*

Data: Novels ranked Top 50 in Germany (2001 – 2006, *n* = 798).
Only books included, which ranked Top 50 for a duration of ≥ 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):

$$y_t = (a + bx_t)(1 - x_t) + u_t$$

a and *b* are estimated for 354 diffusions each, using 354 NLLS-regressions

Explanation of *a* and *b*

OLS Regression, robust standard errors. NLLS estimates used as dep. variables (log *a*, log *b*).

Regressors:
- established writer (dummy, 1 if previous Top 50-titles exist)
- publisher's marketing power (# of published Top 50-titles in 2001-06)
- book's quality (# of Amazon stars)
- log run time (log # weeks ranked in Top 50)

	external influence (log *a*)		
	1	2	3
established writer	0.165*	0.133†	0.082
	(0.070)	(0.056)	(0.035)
marketing		0.006*	0.007*
		(0.066)	(0.079)
quality			0.153**
			(0.090)
log run time	-1.690***	-1.697***	-1.691***
	(-0.818)	(-0.822)	(-0.819)
genre			controlled
constant	1.686***	1.592***	0.927**
n	354	354	354
R^2 *adj.*	0.665	0.669	0.674

Unstand. coefficents, standardized coefficients in parentheses, † $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Explanation of *a* and *b*

OLS Regression, robust standard errors. NLLS estimates used as dep. variables (log *a*, log *b*).

Regressors:
- established writer (dummy, 1 if previous Top 50-titles exist)
- publisher's marketing power (# of published Top 50-titles in 2001-06)
- book's quality (# of Amazon stars)
- log run time (log # weeks ranked in Top 50)

	external influence (log *a*)			**internal influence (log *b*)**		
	1	2	3	1	2	3
established writer	0.165*	0.133†	0.082	-0.212*	-0.224*	-0.211*
	(0.070)	(0.056)	(0.035)	(-0.080)	(-0.084)	(-0.079)
marketing		0.006*	0.007*		0.002	0.002
		(0.066)	(0.079)		(0.023)	(0.021)
quality			0.153**			0.139*
			(0.090)			(0.073)
log run time	-1.690***	-1.697***	-1.691***	-1.843***	-1.846***	-1.840***
	(-0.818)	(-0.822)	(-0.819)	(-0.786)	(-0.787)	(-0.785)
genre			controlled			controlled
constant	1.686***	1.592***	0.927**	3.066***	3.029***	2.523***
n	354	354	354	354	354	354
R^2 *adj.*	0.665	0.669	0.674	0.631	0.631	0.639

Unstand. coefficents, 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).

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

	external influence		**internal influence**	
	(*a*)	% change	**(*b*)**	% change
established writer	+ 0.008***	+ 17.4	– 0.023***	– 30.7
	(0.001)		(0.005)	
marketing	0.000	0	– 0.000*	0
	(0.000)		(0.000)	
quality	– 0.002	– 4.3	+ 0.012**	+ 16.0
	(0.001)		(0.004)	

Unstand. coefficents, 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.

	external influence		**internal influence**	
	(*a*)	% change	**(*b*)**	% change
established writer	+ 0.008***	+ 17.4	– 0.023***	– 30.7
	(0.001)		(0.005)	
marketing	0.000	0	– 0.000*	0
	(0.000)		(0.000)	
quality	– 0.002	– 4.3	+ 0.012**	+ 16.0
	(0.001)		(0.004)	

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

Results from two-step approach: + 17 / 8 % – 21 %

Test for Validity: Looking at Readers

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.

Micro Level-Analysis

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*.

Micro Level-Analysis

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?“

Stimulus		freq.	%		freq.	%
1	media	147	12,5			
2	advertisement	92	7,8			
3	catalogue	391	33,1	External Influence (*y*=0)	956	81
4	shop decor	296	25,1			
5	bookseller's tip	30	2,5			
6	word of mouth	129	10,9			
7	bestseller list	96	8,1	Internal Influence (*y*=1)	225	19
	n purchases	1.181			1.181	
	m books	24			24	

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 ≥ Fachhochschulreife)
  - high income (dummy, 1 if ≥ 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.

	Odds Ratios	% change
established writer	0.550**	– 45.0
marketing	0.668*	– 33.2
log run time	1.538***	+ 53.2
city size	1.164*	+ 16.4
high education	1.591**	+ 59.1
high income	1.523**	+ 52.3
female	1.745**	+ 74.5
week of purchase	1.008	+ 0.8
m purchases	1,181	
n books	24	
Nagelkerkes R^2	0.079	

Odds ratios (e^{β}),
% change $([e^{\beta} - 1] \cdot 100)$,
† $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Summary

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

Bass F 1969: A New Product Growth Model for Consumer Durables. Management Science 15: 215-227.

Beck J 2007: The Sales Effect of Word of Mouth: A Model for Creative Goods and Estimates for Novels. Journal of Cultural Economics 31: 5-23.

Clement M, et al. 2006: Diffusion of Hedonic Goods: A Literature Review. International Journal on Media Management 8: 155-163.

Deschâtres F, Sornette D 2005: The Dynamics of Book Sales: Endogenous versus Exogenous Shocks in Complex Networks. Physical Review E 72.

Mahajan V, Peterson R 1985: Models for Innovation Diffusion. Beverly Hills: Sage Publications .

Rossman G, et al. 2008: Modelling Diffusion of Multiple Innovations via Multilevel Diffusion Curves: Payola in Pop Music Radio. Sociological Methodology 38: 201-230.

Snijders T, Bosker R 1999: Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks: Sage Publications.

Srinivasan V, Mason C 1986: Non-Linear Least Squares Estimation of New Product Diffusion Models. Marketing Science 5: 169-178.

Van den Bulte C, Joshi Y 2005: New Product Diffusion with Influentials and Imitators. Working Paper, University of Pennsylvania.

Appendix