

# Project in Applied Econometrics

## Report

Lucas Javaudin, Robin Le Huérou-Kérisel, Rémi Moreau

March 2018

### Abstract

This project aims at reproducing a paper by Moretti (2011) on social learning effects in movie sales with R. We then confront his theory and predictions with French data. We find evidence of social learning in French movie sales but our results are less robust than the results of Moretti.

## Contents

<b>1</b>	<b>Intuitions and Detailed Presentation of the Model</b>	<b>2</b>
1.1	Utility Estimation before a Movie Is Released . . . . .	2
1.1.1	Actual Utility . . . . .	2
1.1.2	Signal: Noisy Utility . . . . .	2
1.1.3	Expected Utility . . . . .	3
1.1.4	Cost of Watching . . . . .	3
1.1.5	Probability of Watching . . . . .	4
1.2	Utility Estimation with Social Learning . . . . .	4
1.2.1	Signal with Feedback . . . . .	4
1.2.2	Expected Utility . . . . .	5
1.2.3	Probability of Watching . . . . .	6
1.2.4	Ruling Out Network Externalities . . . . .	6
<b>2</b>	<b>Analysis and Main Results</b>	<b>7</b>
2.1	Data Cleaning . . . . .	7
2.2	Identification of the Surprises . . . . .	7
2.3	Divergence of the Sales . . . . .	8
2.4	Precision of the Prior . . . . .	10
2.5	Size of the Social Network . . . . .	12
2.6	Does learning decline over time? . . . . .	12
<b>3</b>	<b>Conclusion</b>	<b>15</b>

# 1 Intuitions and Detailed Presentation of the Model

Moretti develops a model that defines social learning as the effect of one consumer's opinion on a given good's consumption, on her acquaintances' willingness to consume. Applied to cinema, his idea is that consumers go and see a certain film if they get a sufficient expected utility to compensate the cost of viewing.

## 1.1 Utility Estimation before a Movie Is Released

### 1.1.1 Actual Utility

The utility provided by a given film (called "quality") is estimated by the consumers before the movie is being launched. They use the objective elements about the film, available to them all, to get a *prior*. Consumers also get a personal *signal* that indicates how much they are attracted by the film's concept.

Movie theatres only get a prior, and estimate the appeal for the film from it. They are profit-maximizing firms, so they have incentives to correctly estimate the number of screens for a given film using the prior.

An individual  $i$  gets a utility  $U_{i,j}$  watching the film  $j$  with

$$U_{i,j} = \alpha_j^* + v_{ij}, \quad (1)$$

where  $\alpha_j^*$  is the quality of the movie for the average individual and  $v_{ij} \sim \mathcal{N}(0, \frac{1}{d})$  represents how much individual  $i$ 's appeal for movie  $j$  differs from the average individual's taste for this movie (that is the reason why the variable  $v_{ij}$  is zero-mean).

It is worth noticing that the variance of  $v_{ij}$  is independent of both  $i$  and  $j$ : the distribution of the tastes around the mean is supposed to be constant (in absolute value) no matter the film.

These two parameters are supposed to be unobserved by the consumers. This is consistent for  $\alpha_j^*$  as an individual may find it hard to estimate precisely the advice of the *average consumer*: she does not observe all advices, thus cannot precisely determine the mean of all the tastes. So  $\alpha_j^*$  can be considered as a random variable; to make the model rather simple, this variable is supposed to be normal, i.e. totally characterized by its mean and variance.

This variable's mean (referred to previously as the prior) can be estimated from all the observable features of the film (such as the director, the budget, the casting...), that are aggregated into the variable  $X_j$ . Thus,  $\mathbb{E}(\alpha_j^*) = X_j' \beta$ , like in a simple linear model. This expression embodies one of Moretti's key ideas in his paper, that a film's quality is the only determiner of the utility if we omit interpersonal differences in tastes.

The variance of  $\alpha_j^*$  does not depend on the individual doing the estimation (all individual are supposed to have the same information about the film and to analyse it the same way). Nevertheless, it depends on the movie. Sequels and films shot by famous directors are characterised by a much lower variance since the outcome (success or failure) of the previous movies could be observed.

Thus:  $\alpha_j^* \sim \mathcal{N}(X_j' \beta, \frac{1}{m_j})$ .

### 1.1.2 Signal: Noisy Utility

Moretti supposes that consumers do not measure the utility  $U_{i,j}$  he could get by watching the film, but receives instead a noisy signal

$$s_{ij} = U_{i,j} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \frac{1}{k_j}).$$

Introducing an unbiased signal allows that the averages of  $s$  and  $U$  are the same for a certain movie  $j$ , which is a basic requirement in a social learning model (aggregated signal gives an unbiased estimator). Additionally, for a set of movies that give the same utility to a given consumer and that have the same variance for  $\epsilon_{ij}$ , the consumer makes on average correct predictions of her utility starting from the signal she gets.

The variables  $\epsilon_{ij}$  and  $v_{ij}$  are also supposed independent of each other and of  $\alpha_j^*$ ; the idea is that the objective quality of the film, the subjective appeal for it and the error in measuring these values aren't correlated, since they are of different nature.

It is worth noticing, in that sense, that the three variables introduced above are all considered as random, but in different senses. For  $\alpha_j^*$  it can be interpreted as a lack of sufficient information to estimate  $\beta$  correctly. For  $v_{ij}$  the probability distribution allows to consider a set of different individuals with different tastes.  $\epsilon_{ij}$  is rather an error of interpretation (or, more mathematically, of measurement) during the determination of  $U_{ij}$ .

Additionally, we suppose that  $X_j$ ,  $\beta$ ,  $m_j$ ,  $k_j$  and  $d$  are well-known by all the consumers.

### 1.1.3 Expected Utility

The expected utility in the first week (i.e. before the film is being released) is estimated by the weighted average of the prior and the signal:

$$\mathbb{E}_1[U_{ij}|X_j'\beta, s_{ij}] = \omega_j X_j'\beta + (1 - \omega_j)s_{ij}.$$

The idea is to use two different estimators of  $U_{ij}$ : its estimated average value (which is certain) and its value with some zero-mean noise (which cumulates several uncertainties).

The expression of the variance has to put more weight on the most precise term between the two estimators of  $U_{ij}$ . We want to know whether it is better

- to omit widely the personal part  $v_{ij}$  (of variance  $\frac{1}{d}$ ) and the error in estimating the average value  $\alpha_j^*$  (of variance  $\frac{1}{m_j}$ ), i.e. to put more weight on  $X_j'\beta$ ,
- to cope with noise  $\epsilon_{ij}$  (of variance  $\frac{1}{k_j}$ ), i.e. to put more weight on  $s_{ij}$ .

Like in the WLS approach, we minimize the expected value of the square difference between the right-hand side above, and the sought term  $U_{ij}$ , choosing the optimal  $\omega_j$ , i.e. :

$$\min_{\omega_j} \mathbb{E}[(\omega_j X_j'\beta + (1 - \omega_j)s_{ij} - U_{ij})^2] = \min_{\omega_j} \mathbb{E}[(-\omega_j(\alpha_j^* - X_j'\beta + v_{ij}) + (1 - \omega_j)\epsilon_{ij})^2]. \quad (2)$$

According to Appendix 1, the solution is

$$\begin{aligned} \omega_j \left( \frac{1}{h_j} + \frac{1}{k_j} \right) &= \frac{1}{k_j}, \\ \omega_j &= \frac{h_j}{h_j + k_j}, \end{aligned}$$

with  $h_j = \frac{1}{\frac{1}{d} + \frac{1}{m_j}} = \frac{dm_j}{d+m_j}$  ( $h_j = \text{Var}(U_{ij})$  too).

Thus:

$$\begin{aligned} \mathbb{E}_1[U_{ij}|X_j'\beta, s_{ij}] &= \omega_j X_j'\beta + (1 - \omega_j)s_{ij}, \\ \omega_j &= \frac{h_j}{h_j + k_j}, \\ h_j &= \frac{dm_j}{d + m_j}. \end{aligned}$$

### 1.1.4 Cost of Watching

In the model, a consumer decides to go and watch a film if her utility is higher than her subjective cost of watching in the week considered (parameter  $t$ ), depending for example on whether she has a lot or a few activities planned during the week:

$$\mathbb{E}_1[U_{ij}|X_j'\beta, s_{ij}] > q_{it}. \quad (3)$$

To model the idea that the cost depends on the individual and on the week, Moretti supposes that  $q_{it} = q + u_{it}$  with  $u_{ij} \sim \mathcal{N}(0, \frac{1}{r})$  with all  $u_{it}$  independent.  $u_{it}$  is zero-mean so the cost of watching has the same average value (between individuals) every week, and the same average for a given individual during a long period of time. As a consequence, the consumers are considered similar; the structural difference of timetable between positions and jobs is not taken into account, or it is supposed not to influence the possibility to go to the movies.

### 1.1.5 Probability of Watching

Let us now compute the probability for individual  $i$  to go to see movie  $j$  in the first week. To do so, we will have to change the viewpoint, from the consumer deciding whether or not to see a movie, to an observer of the market looking at the result of consumer's decision at an aggregate level. The aim is to get a formula that is possible to use. More precisely, all variables with indexes  $i$  won't be observable any more, but the average  $\alpha_j^*$  is now observable (thus precisely determined); this induces a certain form for the result. The sought probability of watching is in this context:

$$P_1 = \mathbb{P}(\mathbb{E}_1[U_{ij}|X_j'\beta, s_{ij}] > q_{it}) = \Phi\left(\frac{(1 - \omega_j)(\alpha_j^* - X_j'\beta) + X_j'\beta - q}{\sigma_{j1}}\right), \quad (4)$$

$$\sigma_{j1}^2 = (1 - \omega_j)^2 \left(\frac{1}{k_j} + \frac{1}{d}\right) + \frac{1}{r},$$

where  $\Phi$  is the cumulative function of a standard normal distribution  $\mathcal{N}(0, 1)$ .

Moretti underlines that the term  $\alpha_j^* - X_j'\beta$  that appears in the previous equation measures the *surprise*. It is indeed the difference between

- the 'true quality' of the movie,  $\alpha_j^*$ , in the sense of average utility that consumers get by viewing the film, that only the observer of the market sees, and
- the prior of quality,  $X_j'\beta$ , which is the only piece of information movie theatres have and one of the only two informations consumers have to estimate the quality of the movie. The observer of the market can also see it.

We note that in the first week a positive difference  $\alpha_j^* - X_j'\beta > 0$  increases the probability of watching and conversely. The presence of this term is justified by the fact that

- movie theatres have less information than consumers,
- both information consumers receive are unbiased,
- consumers are numerous and thus, on average, the utility they expect is close to the true utility (or quality)  $\alpha_j^*$ , which is totally determined by the consumer's appeal for the film.

With neither social learning, nor decrease in utility for viewing a film again and again, the previous formula for  $P_1$  remains valid as long as the movie can be watched.

## 1.2 Utility Estimation with Social Learning

### 1.2.1 Signal with Feedback

Let us add social learning in the model now. In week 2, we consider a consumer  $i$  who has  $N_i$  peers,  $n_i$  of which see the movie in Week 1. They all give their utility  $U_{pj}$  ( $p \in [1, n_i]$ ) as a feedback to  $i$  after watching. Consumer  $i$  receives two informations at the same time:

- The fact that his acquaintances who saw the film had a sufficiently high expected (ex-ante) utility to do so, i.e.  $\omega_j X_j'\beta + (1 - \omega_j)s_{pj} > q_{p1}$ , and that the  $N_i - n_i$  other peers did not verify such inequality.

- Their ex-post utility itself.

By maximum likelihood estimation we obtain an estimate  $S_{ij2}$  such that (see Appendix 3):

$$S_{ij2} = \frac{1}{n_i} \sum_{p=1}^{n_i} U_{pj} - \frac{1 - \omega_j}{d\sigma_V} \frac{N_i - n_i}{n_i} \frac{\phi\left(\frac{q - \omega_j X'_j \beta - (1 - \omega_j) S_{ij2}}{\sigma_V}\right)}{\Phi\left(\frac{q - \omega_j X'_j \beta - (1 - \omega_j) S_{ij2}}{\sigma_V}\right)} \quad (5)$$

The second part of the expression is negative so that  $S_{ij2}$  is lower than the average of the ex-post utilities consumer  $i$  receives; this is the impact of non-viewers.

Moretti underlines that the estimator obtained is unbiased and asymptotically normal (this second property comes from the fact that the estimator is a likelihood maximizer):

$$S_{ij2} \sim \mathcal{N}\left(\alpha_j^*, \frac{1}{b_{i2}}\right), b_{i2} = dn_i + (N_i - n_i) \frac{\phi(c)}{\Phi(c)} \left(c + \frac{\phi(c)}{\Phi(c)}\right) \left(\frac{1 - \omega_j}{\sigma_V}\right)^2 \quad (6)$$

### 1.2.2 Expected Utility

Progressively, the acquaintances' opinion on the film is also included in this weighted average. Consumer  $i$  will do a weighted average of the three informations she has,  $X'_j \beta$ ,  $s_{ij}$  and  $S_{ij2}$ , just like in the first week:

$$\mathbb{E}_2[U_{ij}|X'_j \beta, s_{ij}, S_{ij2}] = \frac{h_j}{h_j + k_j + z_{i2}} X'_j \beta + \frac{k_j}{h_j + k_j + z_{i2}} s_{ij} + \frac{z_{i2}}{h_j + k_j + z_{i2}} S_{ij2}, \quad (7)$$

$$h_j = \frac{dm_j}{d + m_j}, z_{i2} = \frac{b_{i2}d}{b_{i2} + d}.$$

Likewise, in week  $t \geq 2$ , the expected utility is

$$\begin{aligned} \mathbb{E}_t[U_{ij}|X'_j \beta, s_{ij}, S_{ij2}, \dots, S_{ijt}] &= \frac{h_j X'_j \beta}{h_j + k_j + \sum_{s=2}^t z_{is}} + \frac{k_j s_{ij}}{h_j + k_j + \sum_{s=2}^t z_{is}} + \sum_{w=2}^t \frac{z_{iw} S_{ijw}}{h_j + k_j + \sum_{s=2}^t z_{is}}, \\ \mathbb{E}_t[U_{ij}|X'_j \beta, s_{ij}, S_{ij2}, \dots, S_{ijt}] &= \omega_{j1t} X'_j \beta + \omega_{j2t} s_{ij} + \sum_{w=2}^t \omega_{j3w} S_{ijw}, \end{aligned} \quad (8)$$

$$\begin{aligned} \omega_{j1t} &= \frac{h_j}{h_j + k_j + \sum_{s=2}^t z_{is}}, \\ \omega_{j2t} &= \frac{k_j}{h_j + k_j + \sum_{s=2}^t z_{is}}, \\ \omega_{j3w} &= \frac{z_{iw}}{h_j + k_j + \sum_{s=2}^t z_{is}}, \\ h_j &= \frac{dm_j}{d + m_j}, \\ z_{it} &= \frac{b_{it}d}{b_{it} + d}. \end{aligned}$$

This equation shows that a given piece of information has a decreasing importance in the final decision from week to week.

### 1.2.3 Probability of Watching

Just like before, the probability of watching at week  $t$  is (see Appendix 4):

$$P_t = \mathbb{P}(\mathbb{E}_t[U_{ij}|X'_j\beta, s_{ij}, S_{ij2}, \dots, S_{ijt}] > q_{it}) = \Phi\left(\frac{(1 - \omega_{j1t})(\alpha_j^* - X'_j\beta) + X'_j\beta - q}{\sigma_{jt}}\right), \quad (9)$$

with:

$$\sigma_{jt}^2 = (\omega_{j2t})^2 \left( \frac{1}{k_j} + \frac{1}{d} \right) + \frac{\sum_{p=2}^t z_{ip}}{(h_j + k_j + \sum_{s=2}^t z_{is})^2} + \frac{1}{r}.$$

To analyze this equation we can consider first the case  $X'_j = \alpha_j^*$  (no surprise). Then  $P_t = \Phi\left(\frac{X'_j\beta - q}{\sigma_{jt}}\right)$ : the higher the difference between the prior and the average cost (average for every week and/or every consumer), the higher the probability for a viewer to watch the film, and thus the higher the attendance to see the movie considered. The probability of viewing is constant equal to one-half when  $X'_j\beta = q$ .

To study the evolution over time we can derivate (discretely)  $P_t$  two times w.r.t  $t$  (see Appendix 5).

When  $X'\beta = q$ , we obtain a consistent model, as announced by Moretti, if some condition is verified by the parameters ( $h_j < k_j$  or  $h_j < \frac{2k_j^2}{d}$  is sufficient). If we have a positive surprise, i.e.  $\alpha_j^* > X'_j\beta$ , the probability of going to watch the film is increasing from one week to the following, whereas a negative surprise  $\alpha_j^* < X'_j\beta$  leads to a decrease in this probability across the weeks. Additionally, without any surprise, i.e.  $\alpha_j^* = X'_j\beta$ , the probability of watching is constant over time.

Moretti underlines that the second derivative has the opposite sign from the first derivative, meaning that:

$$\frac{\partial^2 P_t}{\partial t^2} < 0 \text{ (concavity) when } \alpha_j^* > X'_j\beta \text{ and } \frac{\partial^2 P_t}{\partial t^2} > 0 \text{ (convexity) when } \alpha_j^* < X'_j\beta.$$

The concavity or convexity is the mark of a social learning multiplier: from one week to another, the quality becomes more and more precisely known, additional advices from peers bring marginally less and less informations from one week to another.

If we relax the hypothesis  $X'\beta = q$ , the same trends are observed but with a threshold. Considering again the sufficient conditions above, when  $X'\beta > q$ , the prior on quality of the movie exceeds the cost of viewing, so the impact of a negative surprise on the probability of watching the film is lower. This is due to the fact that people, in that case, are to perform a trade-off between a strongly positive prior and a negative surprise. We have a symmetric situation when  $X'\beta < q$  and people's signal shows a negative surprise.

### 1.2.4 Ruling Out Network Externalities

Moretti wants to show a social learning effect and to rule out a network externality effect. The latter is present when a consumer has a utility that depends only on the number of people who have watched the film (for example, so as to speak with them about the film). The theory he develops models the idea that the evolution in sales is due to the process of estimation of a movie's quality, and not merely on the attendance for a given movie.

## 2 Analysis and Main Results

Moretti's purpose is to provide evidence of social learning in consumption, that is to say that people tend to take into account their peers' experience to get a more precise idea of the value of a good. Economists, Moretti says, have had difficulties showing such social learning effects because of the absence of useful microdata on the matter. Moretti's innovation lies in his use of market-level data to identify social learning. He does so by defining what he calls "surprises" in movie sales: surprises, as their name suggests, consist in the difference between expected and actual sales. Moretti proposes that if we observe a surprise, we should also observe social learning effects: if a film is better or worse than expected, then by gathering experiences through peers, people should reconsider their expectations and we might be able to see it in the data. In particular, Moretti makes five predictions on things we should be able to observe in presence of social learning:

1. in presence of social learning, sales of movies with positive and negative surprises should diverge: sales of better-than-expected movies should decrease at a lower rate than worse ones (see 2.3);
2. we should observe less social learning effects from a movie on which quality we have a precise idea and more social learning effects from movies which have a more uncertain quality (see 2.4);
3. we should observe more social learning effects when people have a greater social network (see 2.5);
4. we should be able to observe that the effects of a surprise decline over time: once the information on the quality of a movie has been shared, what was a surprise should not play a major role in sales (see 2.6);
5. we should not observe social learning effects when a surprise is due to elements other than quality of the film (let say weather).

We have replicated Moretti's work and tried to confront his predictions with French data.

### 2.1 Data Cleaning

The first step before replicating the results of Moretti was to clean the US data in the same way as he does (conversion of variables in 2005 dollars, removal of movies with typos, generation of useful variables...). While doing so, we found some oddness in the code of Moretti. Notably, he removes all the movies for which the growth rate of the number of screens is higher than 5 between the first and the second week. He says in the article that the removed movies are the movies opening only in New York and Los Angeles. This justification appeared strange to us, even though we obtained very close results when computing the regressions without removing those movies.

The French data were already cleaned. We decided to change the structure of the data to bring it closer to the structure of the US data. To replace the initial dataset where there is one variable for tickets sold of week 1 to 13 (*entree\_fr1*, ..., *entree\_fr13*), we built a dataset where there is a single variable for tickets sold (*entree\_fr*) but one observation per week. This manipulation allowed us to use a very similar code for both dataset.

While the US data indicate the sales of the movies, the French data report the tickets sold. Hence, there is no need to correct the data for inflation effects like we did with the US data.

We computed all the regressions with the tickets sold and the number of screens in Paris only and in France as a whole. The results are very similar between these two subsets. Here, we report only the results of the regressions when using the data of France as a whole.

### 2.2 Identification of the Surprises

Surprises consist in the residuals of the regression of the log-number of sales in the first week on the log-number of screens available (opened by theaters). This definition of surprises holds because we suppose that theaters are profit-maximizing agents and make use of all the available information to predict the success

of a movie. If this definition is correct, we should expect log-number of screens opened by theaters first week to be a good indicator of knowledge available on the movie quality before it is released. In the Table 1 we reproduce Moretti’s regression of *log\_sales\_first\_we* on *log\_screens\_first\_week*. Each column is the result of the regression when we control with some variables (film genre, rating available, cost, distributor, weekday, month, week, year). The fact that adding control variables doesn’t change the robustness of the regression proves Moretti’s point which is that theaters take into account these factors when deciding their number of available screens.

Table 1: Regression of first-weekend sales on number of screens (US Data)

	<i>Dependent variable:</i>						
	<i>log_sales_first_we</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>log_screens_first_week</i>	0.893*** (0.004)	0.896*** (0.005)	0.883*** (0.005)	0.871*** (0.005)	0.803*** (0.006)	0.806*** (0.006)	0.813*** (0.006)
Genre		✓	✓	✓	✓	✓	✓
Ratings			✓	✓	✓	✓	✓
Production cost				✓	✓	✓	✓
Distributor					✓	✓	✓
Weekday, month, week						✓	✓
Year							✓
R <sup>2</sup>	0.907	0.909	0.910	0.912	0.932	0.936	0.938
Adjusted R <sup>2</sup>	0.907	0.908	0.910	0.912	0.928	0.931	0.933

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

We have performed the same kind of regression on French data from 2004 to 2008 and find quite similar results (see table 2). We can see that the number of sales in first week is highly explained by the number of screens opened. This result holds even when adding controls: each column corresponds to a regression in which we added a control variable (genre, ratings, distributors, month and week, year, and some other variables).

### 2.3 Divergence of the Sales

The first prediction of Moretti is that if there are social learning effects in movie sales, we should observe diverging trajectories between movies with positive and negative surprises. The idea is simple: without social learning, sales of movies with positive and negative surprises should decrease at the same rate; in other words, surprises would not have any effect on sales. Indeed, people would not take surprises as a new information on the movie quality. In the figure 1, we have reproduced Moretti’s graph and plotted the graph for French data. In Moretti’s graph we clearly see the diverging trajectories of the sales. Our graph shows less clear diverging trajectories on the whole duration we computed, though it is clear that trajectories diverge in the first four weeks of projections. To test for differences in trajectories, Moretti estimates models of the form:

$$\ln(y_{jt}) = \beta_0 + \beta_1 * t + \beta_2(t * S_j) + d_j + u_{jt} \quad (10)$$

where  $\ln(y_{jt})$  is the log of box-office sales in week  $t$ ;  $S_j$  is surprise;  $d_j$  is a movie fixed effect. The variable of interest is  $\beta_2$  because we want to identify an effect of the surprise on the dynamic of sales over time.

Table 3 estimates equation 10 and then differentiates between positive, top, middle and bottom surprises.



Table 2: Regression of first-week tickets sold on number of screens (French Data)

	<i>Dependent variable:</i>						
	log_entree_fr						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log_seance_fr	1.208*** (0.009)	1.237*** (0.010)	1.237*** (0.010)	1.279*** (0.014)	1.282*** (0.014)	1.287*** (0.014)	1.196*** (0.014)
Genre		✓	✓	✓	✓	✓	✓
Interdiction			✓	✓	✓	✓	✓
Distributeur				✓	✓	✓	✓
Mois, semaine					✓	✓	✓
Année						✓	✓
MoyennePresse							✓
SigmaNotePresse							✓
PoidsCasting							✓
Pub							✓
R <sup>2</sup>	0.893	0.899	0.900	0.917	0.924	0.925	0.943
Adjusted R <sup>2</sup>	0.893	0.898	0.898	0.910	0.915	0.916	0.935

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Figure 1: Comparing decline in sales between US Data and French data

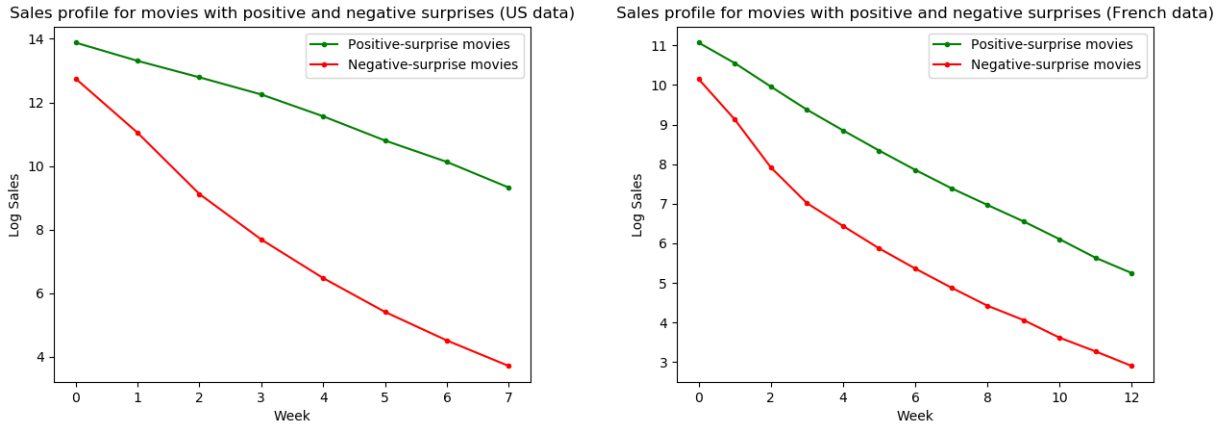


Table 3: Decline in box-office sales by opening week surprise

	<i>Dependent variable:</i>			
	log_entree_fr			
	(1)	(2)	(3)	(4)
t	−0.526*** (0.002)	−0.526*** (0.002)	−0.571*** (0.003)	
t×surprise		0.076*** (0.004)		
t×positive_surprise			0.087*** (0.004)	
t× top surprise				−0.459*** (0.004)
t× middle surprise				−0.546*** (0.004)
t× bottom surprise				−0.574*** (0.004)
Observations	26,598	26,598	26,598	26,598
R <sup>2</sup>	0.851	0.853	0.853	0.854
Adjusted R <sup>2</sup>	0.838	0.841	0.841	0.841

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

## 2.4 Precision of the Prior

Another prediction of Moretti is that the effect of surprises should vary with the precision of the prior people have on movies. Indeed, if people were to have a precise idea of the quality of the film, they would learn less informations from their peers' experience. To empirically identify which movies are likely to have precise priors, Moretti proposes to add dummies for sequels, and to use variances of the surprises by genre in the first week as a proxy for the precision of their prior.

Moretti estimates models of the form:

$$\ln(y_{jt}) = \beta_0 + \beta_1 * t + \beta_2(t * S_j) + \beta_3(t * \text{precision}_j) + \beta_4(t * S_j * \text{precision}_j) + d_j + u_{jt} \quad (11)$$

where  $\text{precision}_j$  is a measure of the precision of the prior for movie  $j$ . The coefficient of interest is the coefficient  $\beta_4$  on the triple interaction between the time trend, the surprise and the precision of the prior. We make three hypothesis:

1. as in Moretti (2011), we test that sequels (“*saga*”) had more precise priors and thus were subject to less social learning effect;
2. we suppose that art-house cinema (“*art et essai*”) have a more unpredictable quality;
3. we suppose that film produced outside the occidental world (“*Reste Monde*”) have more unpredictable quality.

Table 4 summarizes our regressions. We included dummies for sequels, art-house cinema and “outside occident” movies (movies not produced in Europe nor in the US). Results show that *saga* has a negative effect on the impact of a surprise over time, meaning that *sagas* have indeed a weaker social learning effect. On the opposite, and as expected, *art et essai* and *Reste Monde* both have significant positive effect on the impact of a surprise over time. We also include the variance of the first-week surprises computed by genre

Table 4: Precision of the prior

	<i>Dependent variable:</i>			
	log_entree_fr			
	(1)	(2)	(3)	(4)
t	−0.570*** (0.003)	−0.698*** (0.013)	−0.678*** (0.004)	−0.578*** (0.003)
t×positive_surprise	0.105*** (0.005)	0.109*** (0.018)	0.009 (0.006)	0.078*** (0.005)
t×saga	−0.027 (0.016)			
t×positive_surprise×saga	−0.145*** (0.019)			
t×var_surprise		0.370*** (0.035)		
t×positive_surprise×var_surprise		−0.062 (0.050)		
t×art_essai			0.259*** (0.006)	
t×positive_surprise×art_essai			0.066*** (0.008)	
t×ResteMonde				0.106*** (0.013)
t×positive_surprise×ResteMonde				0.066*** (0.016)
Observations	26,598	26,546	26,598	26,598
R <sup>2</sup>	0.855	0.854	0.880	0.855
Adjusted R <sup>2</sup>	0.843	0.842	0.870	0.843

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

in the regression (reg 2), we get an insignificant and negative result<sup>1</sup>. The use of dummies tend to support the results of Moretti but using first-week variance as a proxy for prior does not produce significant results.

## 2.5 Size of the Social Network

Consumers with a larger social network receive more feedbacks from their peers and thus are able to evaluate more precisely the quality of the movie. Hence, social learning should be stronger for consumers with a larger social network. More formally, this prediction can be tested by estimating models of the form:

$$\ln(y_{jt}) = \beta_0 + \beta_1 t + \beta_2(t \times S_j) + \beta_3(t \times NS_j) + \beta_4(t \times S_j \times NS_j) + d_j + u_{jt} \quad (12)$$

where  $S_j$  is a dummy for positive surprise and  $NS_j$  is a variable representing the network size of the audience of movie  $j$ . If social learning is stronger with higher values of  $NS_j$ , then the coefficient  $\beta_4$  of the triple interaction between the time trend, the surprise and the network size should be positive.

In his article, Moretti uses two different measurements of network size. First, he makes the assumption that teenagers have a more developed social network than adults and he estimates the model of equation 12 with a dummy for teen movies. He finds that the estimate of the coefficient  $\beta_4$  is indeed positive but with a very weak significance level. We also wanted to point out that there is no indicator for teen movies in the data and that the way Moretti builds a dummy for teen movies is quite surprising. He uses *genre1* (one of the three variables indicating the genre of the movies) and he considers that teen movies are movies of the genre action, adventure, comedy, fantasy, horror, sci-fi or suspense. We would have appreciated more justification for the assumption that teenagers have a larger social network and for the way teen movies are defined. To investigate further these issues, we used the two other variables indicating the genre of the movies: *genre2* and *genre3*. Both variables have a category *Children* and a category *Youth* that we use to define two new dummies. The dummy *teen2* (respectively *teen3*) indicates that the movies is in the category *Children* or *Youth* of the variable *genre2* (respectively *genre3*). Using *teen2*, we find that the estimate of  $\beta_4$  is significantly negative. Using *teen3*, we find that the estimate of  $\beta_4$  is significantly positive. The results of the regressions are reported in table 5. We conclude that using teen movies is not a good way to test this prediction.

Moretti uses the number of theaters broadcasting the movie during the opening week as a second measurement of the size of the social network. If a movie opens in lots of theater, the consumers should receive more feedbacks from their peers. As expected, he estimates that the coefficient of  $\beta_4$  is significantly positive. We estimated the same model with the French data. The results are reported in column (2) of table 6. We find that the coefficient of the triple interaction is indeed significantly positive.

In the French data, the variable *tout\_public* is a dummy indicating movies which are suitable for any kind of audience. We can assume that consumers have more feedbacks from their peers for movies opened to anyone. Hence, we estimated the model of equation 12 using the *tout\_public* dummy to measure network size. The results of the estimated are reported in column (1) of table 6. Consistently with our assumption, the coefficient of the triple interaction is significantly positive.

## 2.6 Does learning decline over time?

The model predicts that the effects of positive and negative surprises should decline over time. More precisely, sales profile should be a concave function of time for positive-surprise movies and a convex function of time for negative-surprise movies. To test this prediction, we need to estimate the sales profile which is assumed to be a quadratic function of time. Therefore, we estimate the following model:

$$\ln(y_{jt}) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3(t \times S_j) + \beta_4(t^2 \times S_j) + d_j + u_{jt}$$

---

<sup>1</sup>The value should be positive in order to confirm the idea that a large variance increases social learning effect: we interpret a high variance as a lack of precision for the prior.

Table 5: Precision of peers' signal (US Data)

	<i>Dependent variable:</i>			
	log_sales			
	(1)	(2)	(3)	(4)
$t$	-1.200*** (0.012)	-1.289*** (0.009)	-1.286*** (0.009)	-1.221*** (0.012)
$t \times \text{positive\_surprise}$	0.610*** (0.017)	0.645*** (0.013)	0.638*** (0.013)	0.606*** (0.017)
$t \times \text{teen}$	-0.213*** (0.019)			
$t \times \text{positive\_surprise} \times \text{teen}$	0.072*** (0.026)			
$t \times \text{teen2}$		-0.025 (0.089)		
$t \times \text{positive\_surprise} \times \text{teen2}$		-0.602*** (0.137)		
$t \times \text{teen3}$			-0.304*** (0.097)	
$t \times \text{positive\_surprise} \times \text{teen3}$			0.353** (0.163)	
$t \times \text{nb\_screens\_first\_week}$				-0.121*** (0.013)
$t \times \text{positive\_surprise} \times \text{nb\_screens\_first\_week}$				0.077*** (0.016)
Observations	39,936	39,936	39,936	39,936
R <sup>2</sup>	0.788	0.787	0.787	0.787
Adjusted R <sup>2</sup>	0.758	0.757	0.756	0.757

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 6: Precision of peers' signal (French Data)

	<i>Dependent variable:</i>	
	log_entree_fr	
	(1)	(2)
$t$	-0.663*** (0.007)	-0.451*** (0.005)
$t \times \text{positive\_surprise}$	0.061*** (0.010)	0.076*** (0.006)
$t \times \text{tout\_public}$	0.115*** (0.008)	
$t \times \text{positive\_surprise} \times \text{tout\_public}$	0.031*** (0.011)	
$t \times \text{seance\_fr\_first\_week}$		-0.033*** (0.001)
$t \times \text{positive\_surprise} \times \text{seance\_fr\_first\_week}$		0.011*** (0.001)
Observations	26,598	26,598
R <sup>2</sup>	0.856	0.867
Adjusted R <sup>2</sup>	0.844	0.856
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

where  $S_j$  is a dummy for positive surprise. The results are reported in table 7. The second derivative of  $y_{jt}$  for negative-surprise movies is

$$\left. \frac{\partial^2 y_{jt}}{\partial t^2} \right|_{S_j=0} = 2\beta_2.$$

The second derivative of  $y_{jt}$  for positive-surprise movies is

$$\left. \frac{\partial^2 y_{jt}}{\partial t^2} \right|_{S_j=1} = 2(\beta_2 + \beta_4).$$

We can test the hypothesis of convexity ( $2\beta_2 > 0$ ) and the hypothesis of concavity ( $2(\beta_2 + \beta_4) < 0$ ) with Student tests. For instance, to test  $H_0 : 2(\beta_2 + \beta_4) < 0$  against  $H_1 : 2(\beta_2 + \beta_4) > 0$ , the  $t$  statistic is

$$t = \frac{2(\hat{\beta}_2 + \hat{\beta}_4)}{\text{se}(2(\hat{\beta}_2 + \hat{\beta}_4))}$$

where  $\text{se}(2(\hat{\beta}_2 + \hat{\beta}_4)) = 2[\text{Var}(\hat{\beta}_2) + \text{Var}(\hat{\beta}_4) + 2 \cdot \text{Cov}(\hat{\beta}_2, \hat{\beta}_4)]^{1/2}$  is the standard error of  $2(\hat{\beta}_2 + \hat{\beta}_4)$ .

With the US data, both hypotheses cannot be rejected with a good confidence. With French data, the  $p$ -value for the test of convexity of negative-surprise movies is really close to 0 ( $t \approx 36.72$  and  $p \approx 0$ ). However, the hypothesis of concavity of positive-surprise movies must be rejected ( $t \approx 9.41$  and  $p \approx 1$ ). What we can say however is that the sales profile of positive-surprise movies is *more concave* than the sales profile of negative-surprise movies because the estimates show that the coefficient  $\beta_4$  is significantly negative. These statements are confirmed by the graphs of the sales profile of figure 1 where the sales profile of negative-surprise movies is clearly convex and the sales profile of positive-surprise movies seems linear.

Table 7: Convexity of the sales profile (French Data)

	<i>Dependent variable:</i>
	log_entree_fr
$t$	$-0.978^{***}$ (0.011)
$t^2$	$0.034^{***}$ (0.001)
$t \times \text{positive\_surprise}$	$0.393^{***}$ (0.016)
$t^2 \times \text{positive\_surprise}$	$-0.026^{***}$ (0.001)
Observations	26,598
$R^2$	0.861
Adjusted $R^2$	0.850
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

### 3 Conclusion

## Appendix 1: Derivating the Least-Square Estimator for $U_{ij}$

Like in the WLS approach, we minimize the expected value of the square difference between the right-hand side above, and the sought term  $U_{ij}$ , choosing the optimal  $\omega_j$ , i.e. :

$$\min_{\omega_j} \mathbb{E}((\omega_j X'_j \beta + (1 - \omega_j) s_{ij} - U_{ij})^2) = \min_{\omega_j} \mathbb{E}((-\omega_j(\alpha_j^* - X'_j \beta + v_{ij}) + (1 - \omega_j)\epsilon_{ij})^2).$$

By independance of the different random variable studied, and linearity of the expected value:

$$\begin{aligned} \mathbb{E}((\omega_j X'_j \beta + (1 - \omega_j) s_{ij} - U_{ij})^2) &= \omega_j^2 \mathbb{E}((\alpha_j^* - X'_j \beta + v_{ij})^2) + (1 - \omega_j)^2 \mathbb{E}((\epsilon_{ij})^2) \\ &= \omega_j^2 \text{Var}(\alpha_j^* - X'_j \beta + v_{ij}) + (1 - \omega_j)^2 \text{Var}(\epsilon_{ij}), \end{aligned}$$

since all the variables considered here are zero-mean (the expectation of their square value is thus their variance). Hence:

$$\begin{aligned} \mathbb{E}((\omega_j X'_j \beta + (1 - \omega_j) s_{ij} - U_{ij})^2) &= \omega_j^2 (\text{Var}(\alpha_j^* - X'_j \beta) + \text{Var}(v_{ij})) + (1 - \omega_j)^2 \text{Var}(\epsilon_{ij}) \\ &= \omega_j^2 \left( \frac{1}{m_j} + \frac{1}{d} \right) + (1 - \omega_j)^2 \frac{1}{k_j}. \end{aligned}$$

The problem to solve is thus

$$\min_{\omega_j} \omega_j^2 \left( \frac{1}{m_j} + \frac{1}{d} \right) + (1 - \omega_j)^2 \frac{1}{k_j}.$$

The first-order condition w.r.t  $\omega_j$  is

$$2\omega_j \left( \frac{1}{m_j} + \frac{1}{d} \right) - 2(1 - \omega_j) \frac{1}{k_j} = 0.$$

so, with  $h_j = \frac{1}{\frac{1}{d} + \frac{1}{m_j}} = \frac{dm_j}{d+m_j}$ , we can rewrite:

$$\omega_j \left( \frac{1}{h_j} + \frac{1}{k_j} \right) = \frac{1}{k_j} \quad \omega_j = \frac{h_j}{h_j + k_j}.$$

Thus:

$$\mathbb{E}_1[U_{ij}|X'_j \beta, s_{ij}] = \omega_j X'_j \beta + (1 - \omega_j) s_{ij}, \omega_j = \frac{h_j}{h_j + k_j}, h_j = \frac{dm_j}{d + m_j}. \quad (13)$$

## Appendix 2: Probability of Watching in the First Week

The sought probability of watching is:

$$\begin{aligned} P_1 &= \mathbb{P}(\mathbb{E}_1[U_{ij}|X'_j \beta, s_{ij}] > q_{it}) \\ &= \mathbb{P}(\omega_j X'_j \beta + (1 - \omega_j) s_{ij} > q + u_{it}) \\ &= \mathbb{P}(X'_j \beta + (1 - \omega_j)(\epsilon_{ij} + v_{ij} + \alpha_j^* - X'_j \beta) - q - u_{it} > 0) \\ &= \mathbb{P}((1 - \omega_j)(\alpha_j^* - X'_j \beta) + X'_j \beta - q > -(1 - \omega_j)(\epsilon_{ij} + v_{ij}) + u_{it}) \\ &= \Phi \left( \frac{(1 - \omega_j)(\alpha_j^* - X'_j \beta) + X'_j \beta - q}{\sigma_{j1}} \right), \end{aligned}$$

where  $\sigma_{j1}$  is the variance of the expression just above, and  $\Phi$  is the cumulative function of a standard normal distribution  $\mathcal{N}(0, 1)$ . All remaining random variables (which are all zero-mean) have been eliminated in the expression with the normal distribution  $\Phi$  ( $\alpha_j^*$  is no longer random according to the previous remark). Since the variance of a sum of independant normal variables is the sum of the variances:

$$\begin{aligned} \sigma_{j1}^2 &= \text{Var}(-(1 - \omega_j)(\epsilon_{ij} + v_{ij}) + u_{it}) \\ &= (1 - \omega_j)^2 (\text{Var}(\epsilon_{ij}) + \text{Var}(v_{ij})) + \text{Var}(u_{it}) \\ &= (1 - \omega_j)^2 \left( \frac{1}{k_j} + \frac{1}{d} \right) + \frac{1}{r} \end{aligned}$$



### Appendix 3: Signal with Feedback

From the information consumer  $i$  has, he can estimate the real quality  $\alpha_j^*$  by maximizing the likelihood function:

$$\begin{aligned} L_{ij2} &= L[U_{1j}, \dots, U_{n_{ij}}, n_i | \alpha_j^*] \\ &= \prod_{p=1}^{n_i} \mathbb{P}_{\alpha_j^*}(V_{pj} > q \wedge U_{pj} = \alpha_j^* + v_{pj}) \prod_{p=n_i+1}^{N_i} \mathbb{P}_{\alpha_j^*}(V_{pj} < q), \end{aligned}$$

with  $V_{pj} = \omega_j X_j' \beta + (1 - \omega_j) s_{pj} - u_{p2}$  the adjusted expected utility of peer  $p$ . We use the cumulative distribution functions in the maximum likelihood function instead of the densities because one of the elements we observe is the satisfaction (or not) of inequalities.

Let us consider  $f(U, V)$  the joint density of  $U_{pj}$  and  $V_{pj}$ . We need to consider this joint density so as to include all the information we have on individuals who went to see the film. We can then keep  $U_{pj}$  equal to its value at  $\alpha_j^*$  and modify the value of  $V_{pj}$  between  $q$  and  $\infty$ . We will also consider  $\phi$ , the standard normal density. Then:

$$L_{ij2} = \prod_{p=1}^{n_i} \int_q^\infty f(U_{pj}(\alpha_j^*), V) dV \prod_{p=n_i+1}^{N_i} \mathbb{P}(V_{pj} < q).$$

Furthermore:

$$V_{pj} = \omega_j X_j' \beta + (1 - \omega_j) s_{pj} - u_{p2} = \omega_j X_j' \beta + (1 - \omega_j)(U_{pj} + \epsilon_{pj}) - u_{p2},$$

so:

$$\begin{aligned} f(U_{pj}(\alpha_j^*), V_{pj}) &= f(\alpha_j^* + v_{pj}, \omega_j X_j' \beta + (1 - \omega_j)(U_{pj} + \epsilon_{pj}) - u_{p2}) \\ &= f_U(\alpha_j^* + v_{pj}) f_{V|U=U_{pj}}(\omega_j X_j' \beta + (1 - \omega_j)(U + \epsilon_{pj}) - u_{p2}), \end{aligned}$$

where  $f_U$  is the density of variable  $U$ , which distribution is  $\mathcal{N}(\alpha_j^*, \frac{1}{d})$  and  $f_{V|U=U_{pj}}(V_{p,j}) = f(\omega_j X_j' \beta + (1 - \omega_j)(U + \epsilon_{pj}) - u_{p2})$  is the distribution of a  $\mathcal{N}(\omega_j X_j' \beta + (1 - \omega_j)U, (1 - \omega_j)^2 \text{Var}(\epsilon_{pj}) + \text{Var}(u_{p2}))$  variable i.e.  $\mathcal{N}(\omega_j X_j' \beta + (1 - \omega_j)U, (1 - \omega_j)^2 \frac{1}{k_j} + \frac{1}{r})$ . So,

$$\begin{aligned} f_U(U_{pj}) &= \frac{1}{\sqrt{\frac{1}{d}}} \phi\left(\frac{U_{pj} - \alpha_j}{\sqrt{\frac{1}{d}}}\right) = \sqrt{d} \phi(\sqrt{d}(U_{pj} - \alpha_j)) \\ f_{V|U=U_{pj}}(V) &= \frac{1}{\sigma_{V|U_{pj}}} \phi\left(\frac{V - \omega_j X_j' \beta - (1 - \omega_j)U_{pj}}{\sigma_{V|U_{pj}}}\right), \\ f_V(V) &= \frac{1}{\sigma_V} \phi\left(\frac{V - \omega_j X_j' \beta - (1 - \omega_j)\alpha_j^*}{\sigma_V}\right), \end{aligned}$$

with  $\sigma_{V|U_{pj}}^2 = (1 - \omega_j)^2 \frac{1}{k_j} + \frac{1}{r}$  and  $\sigma_V^2 = \sigma_{V|U_{pj}}^2 + (1 - \omega_j)^2 \text{Var}(U_{pj} - \alpha_j^*) = (1 - \omega_j)^2 (\frac{1}{k_j} + \frac{1}{d}) + \frac{1}{r}$ . Thus:

$$L_{ij2} = \prod_{p=1}^{n_i} \sqrt{d} \phi(\sqrt{d}(U_{pj} - \alpha_j^*)) \left(1 - \Phi\left(\frac{q - \omega_j X_j' \beta - (1 - \omega_j)U_{pj}}{\sigma_{V|U_{pj}}}\right)\right) \prod_{p=n_i+1}^{N_i} \Phi\left(\frac{q - \omega_j X_j' \beta - (1 - \omega_j)\alpha_j^*}{\sigma_V}\right).$$

We can derive w.r.t  $\alpha_j^*$ :

$$-\sqrt{d} \sum_{p=1}^{n_i} L_{ij2} \frac{\phi'(\sqrt{d}(U_{pj} - \alpha_j^*))}{\phi(\sqrt{d}(U_{pj} - \alpha_j^*))} + \frac{-(1 - \omega_j)}{\sigma_V} \sum_{p=n_i+1}^{N_i} \frac{\Phi'\left(\frac{q - \omega_j X_j' \beta - (1 - \omega_j)\alpha_j^*}{\sigma_V}\right)}{\Phi\left(\frac{q - \omega_j X_j' \beta - (1 - \omega_j)\alpha_j^*}{\sigma_V}\right)} L_{ij2} = 0.$$

Since  $\Phi'(x) = \phi(x)$  and  $\phi'(x) = -x\phi(x)$ ,

$$\begin{aligned} \sqrt{d} \sum_{p=1}^{n_i} (\sqrt{d}(U_{pj} - \alpha_j^*)) &= \frac{1 - \omega_j}{\sigma_V} (N_i - n_i) \frac{\phi\left(\frac{q - \omega_j X'_j \beta - (1 - \omega_j) \alpha_j^*}{\sigma_V}\right)}{\Phi\left(\frac{q - \omega_j X'_j \beta - (1 - \omega_j) \alpha_j^*}{\sigma_V}\right)} \\ \alpha_j^* &= \frac{1}{n_i} \sum_{p=1}^{n_i} U_{pj} - \frac{1 - \omega_j}{d\sigma_V} \frac{N_i - n_i}{n_i} \frac{\phi\left(\frac{q - \omega_j X'_j \beta - (1 - \omega_j) \alpha_j^*}{\sigma_V}\right)}{\Phi\left(\frac{q - \omega_j X'_j \beta - (1 - \omega_j) \alpha_j^*}{\sigma_V}\right)} \end{aligned} \quad (14)$$

We denote the associated maximum likelihood estimator  $S_{ij2}$ . The second part of the expression is negative so that  $S_{ij2}$  is lower than the average of the ex-post utilities consumer  $i$  receives; this is the impact of non-viewers.

Moretti underlines that the estimator obtained is unbiased and asymptotically normal (this second property comes from the fact that the estimator is a likelihood maximizer):

$$S_{ij2} \sim \mathcal{N}(\alpha_j^*, \frac{1}{b_{i2}}), b_{i2} = dn_i + (N_i - n_i) \frac{\phi(c)}{\Phi(c)} \left( c + \frac{\phi(c)}{\Phi(c)} \right) \left( \frac{1 - \omega_j}{\sigma_V} \right)^2 \quad (15)$$

At the following dates  $t \geq 2$ , using the equation (8) from the next part, the estimator for  $\alpha_j^*$  maximizes the likelihood function:

$$\begin{aligned} L_{ijt} &= L[U_{1j}, \dots, U_{n_{ij}}, n_i | \alpha_j^*, S_{ij2}, \dots, S_{ijt-1}] \\ &= \prod_{p=1}^{n_i} \mathbb{P}_{\alpha_j^*}(V_{pj} > q \wedge U_{pj} = \alpha_j^* + v_{pj}) \prod_{p=n_i+1}^{N_i} \mathbb{P}_{\alpha_j^*}(V_{pj} < q), \end{aligned}$$

with  $V_{pj} = \omega_{j1t} X'_j \beta + \omega_{j2t} s_{ij} + \sum_{w=2}^t \omega_{j3w} S_{ijw} - u_{pj}$

## Appendix 4: Probability of Watching after Week 2

Just like before, the probability of watching at week  $t$  is:

$$\begin{aligned} P_t &= \mathbb{P}(\mathbb{E}_t[U_{ij} | X'_j \beta, s_{ij}, S_{ij2}, \dots, S_{ijt}] > q_{it}) \\ &= \mathbb{P}\left(\omega_{j1t} X'_j \beta + \omega_{j2t} (\epsilon_{ij} + v_{ij} + \alpha_j^*) + \sum_{w=2}^t \omega_{j3w} (\alpha_j^* + (S_{ijw} - \alpha_j^*)) - q - u_{it} > 0\right) \\ &= \mathbb{P}\left(\omega_{j1t} X'_j \beta + \omega_{j2t} \alpha_j^* + \sum_{w=2}^t \omega_{j3w} \alpha_j^* - q > u_{it} - \omega_{j2t} (\epsilon_{ij} + v_{ij}) - \sum_{w=2}^t \omega_{j3w} (S_{ijw} - \alpha_j^*)\right) \\ &= \mathbb{P}\left(\omega_{j1t} X'_j \beta + (1 - \omega_{j1t}) \alpha_j^* - q > -\omega_{j2t} (\epsilon_{ij} + v_{ij}) - \sum_{w=2}^t \omega_{j3w} (S_{ijw} - \alpha_j^*) + u_{it}\right). \end{aligned}$$

Thus:

$$P_t = \Phi\left(\frac{(1 - \omega_{j1t})(\alpha_j^* - X'_j \beta) + X'_j \beta - q}{\sigma_{jt}}\right), \quad (16)$$

with (admitted):

$$\sigma_{jt}^2 = (\omega_{j2t})^2 \left( \frac{1}{k_j} + \frac{1}{d} \right) + \frac{\sum_{p=2}^t z_{ip}}{(h_j + k_j + \sum_{s=2}^t z_{is})^2} + \frac{1}{r}$$

## Appendix 5: Temporal Evolution of the Probability of Watching

To study the evolution over time we can derivate (discretely)  $P_t$  two times w.r.t  $t$ . We will denote  $S_t = h_j + k_j + Z_{it}$ ,  $Z_{it} = \sum_{s=2}^t z_{is}$ :

$$\Delta P_t = P_{t+1} - P_t \approx P'_t$$

$$\begin{aligned} \Delta P_t &\approx -\frac{1}{\sigma_{jt}} \left( (\alpha_j^* - X'_j \beta) \frac{d}{dt} \omega_{j1t} + ((1 - \omega_{j1t})(\alpha_j^* - X'_j \beta) + X'_j \beta - q) \frac{d\sigma_{jt}/dt}{\sigma_{jt}} \right) \phi(\Phi^{-1}(P_t)) \\ \Delta P_t &\approx -\frac{1}{\sigma_{jt}} \left( (\alpha_j^* - X'_j \beta)(\omega_{j1t+1} - \omega_{j1t}) + ((1 - \omega_{j1t})(\alpha_j^* - X'_j \beta) + X'_j \beta - q) \frac{\sigma_{jt+1} - \sigma_{jt}}{\sigma_{jt}} \right) \phi(\Phi^{-1}(P_t)) \end{aligned}$$

We can compute the components obtained:

$$\begin{aligned} \omega_{j1t+1} - \omega_{j1t} &= -\frac{h_j z_{it+1}}{S_t S_{t+1}} \\ \sigma_{jt+1} - \sigma_{jt} &= \frac{\sigma_{jt+1}^2 - \sigma_{jt}^2}{\sigma_{jt+1} + \sigma_{jt}} \\ &= \frac{1}{\sigma_{jt+1} + \sigma_{jt}} \left( \left( \frac{1}{k_j} + \frac{1}{d} \right) (\omega_{j2t+1}^2 - \omega_{j2t}^2) + \frac{Z_{it+1} S_t^2 - Z_{it} S_{t+1}^2}{(S_t S_{t+1})^2} \right) \\ &= \frac{1}{(\sigma_{jt+1} + \sigma_{jt})(S_t S_{t+1})^2} \left( k_j^2 \left( \frac{1}{k_j} + \frac{1}{d} \right) (S_t^2 - S_{t+1}^2) + Z_{it+1} S_t^2 - Z_{it} S_{t+1}^2 \right) \\ &= \frac{1}{(\sigma_{jt+1} + \sigma_{jt})(S_t S_{t+1})^2} \left( - \left( Z_{it} + k_j^2 \left( \frac{1}{k_j} + \frac{1}{d} \right) \right) z_{it+1} (2S_t + z_{it+1}) + z_{it+1} S_t^2 \right) \\ &= \frac{z_{it+1}}{(\sigma_{jt+1} + \sigma_{jt})(S_t S_{t+1})^2} \left( \left( \frac{1}{r} - S_t^2 \sigma_{jt}^2 \right) (2S_t + z_{it+1}) + S_t^2 \right) \end{aligned}$$

Thus, altogether, denoting  $\Delta_1 = \frac{\Delta P_t}{\phi(\Phi^{-1}(P_t))}$ :

$$\Delta_1 = (\alpha_j^* - X'_j \beta) \frac{h_j z_{it+1}}{S_t S_{t+1} \sigma_{jt}} + \frac{((1 - \omega_{j1t})(\alpha_j^* - X'_j \beta) + X'_j \beta - q) z_{it+1}}{(\sigma_{jt+1} + \sigma_{jt})(S_t S_{t+1} \sigma_{jt})^2} \left( \left( S_t^2 \sigma_{jt}^2 - \frac{1}{r} \right) (2S_t + z_{it+1}) - S_t^2 \right)$$

If we suppose  $X' \beta = q$ ,

$$\begin{aligned} \Delta_1 &= (\alpha_j^* - X'_j \beta) \left[ \frac{h_j z_{it+1}}{S_t S_{t+1} \sigma_{jt}} + \frac{(1 - \omega_{j1t}) z_{it+1}}{(\sigma_{jt+1} + \sigma_{jt})(S_t S_{t+1} \sigma_{jt})^2} \left( \left( S_t^2 \sigma_{jt}^2 - \frac{1}{r} \right) (2S_t + z_{it+1}) - S_t^2 \right) \right] \\ \Delta_1 &= \frac{(\alpha_j^* - X'_j \beta) z_{it+1}}{(\sigma_{jt+1} + \sigma_{jt}) S_t^2 S_{t+1}^2 \sigma_{jt}^2} \left[ h_j (\sigma_{jt+1} + \sigma_{jt}) \sigma_{jt} S_t S_{t+1} + (1 - \omega_{j1t}) \left( \left( S_t^2 \sigma_{jt}^2 - \frac{1}{r} \right) (2S_t + z_{it+1}) - S_t^2 \right) \right] \end{aligned}$$

The term in brackets seems to be positive as all its components but one are positive, but is not obvious to show in the general case. We can admit this property as a requirement of consistency of the model (supposing the parameters are adapted). Since

$$\left( S_t^2 \sigma_{jt}^2 - \frac{1}{r} \right) (2S_t + z_{it+1}) - S_t^2 = (Z_{it} + k_j)^2 + 2(h_j + k_j + Z_{it}) \frac{k_j^2}{d} + (Z_{it} + k_j + \frac{k_j^2}{d}) z_{it+1} - h_j^2,$$

sufficient conditions for positivity at all dates are  $h_j < k_j$  or  $h_j < \frac{2k_j^2}{d}$ . In these cases we obtain a consistent model, as announced by Moretti.