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

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 acquaitancies's willingness to consume.
Applied to cinema, his idea is that comsumers go and see a certain film if they get a sufficient expected utiliy 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 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 are incentivized to choose the number of screens for a given film following the prior.

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

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

where α_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 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 v_{ij}'s variance is independant of both i and j: the distribution of the tastes aroung the mean is supposed to be constant (in absolute value) no matter the film.

These two parameters are supposed to be unobserved. This is consistent for α_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 α_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 precedently 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.
α_j^*'s variance 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 precedent movies'outcome (success or failure) could be observed.
Thus: $\alpha_j^* \sim \mathcal{N}(X_j'\beta, \frac{1}{m_j})$

1.1.2 Signal: Noisy Utility

Moretti suposes that consumers don't measure the utility $U_{i,j}$ he could get by watching the film, but recieves instead a noisy signal

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

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). Additionnaly, for a set of movies that give the same utility to a given consumer and that have the same variance for ϵ_{ij}, the consumer makes on average correct predictions of her utility starting from the signal she gets.
ϵ_{ij} and v_{ij} are also supposed independant of each other and of α_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 α_j^* it can be interpreted as a lack of sufficient information to estimate β correctly. For v_{ij} the probability distribution allows to consider a set of different individuals with different tastes. ϵ_{ij} is rather an error of interpretation (or, more mathematically, of measurement) during the determining of U_{ij}

Additionaly we suppose that X_j, β, m_j, k_j and d are well-known by all the consumers.

1.1.3 Expected Utiliy

The exprected utiliy 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 to 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 α_j^* (of variance $\frac{1}{m_j}$)), i.e. to put more weight on $X_j'\beta$,
- or to cope with noise ϵ_{ij} (of variance $\frac{1}{k_j}$), i.e. to put more weight on s_{ij}.

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

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

The solution is, according to Appendix 1:

$$\omega_j(\tfrac{1}{h_j} + \tfrac{1}{k_j}) = \tfrac{1}{k_j} \ \omega_j = \tfrac{h_j}{h_j+k_j}.$$

with $h_j = \dfrac{1}{\frac{1}{d} + \frac{1}{m_j}} = \frac{dm_j}{d+m_j}$ ($h_j = Var(U_{ij})$ too) Thus:

$$\mathbb{E}_1[U_{ij}|X_j'\beta, s_{ij}] = \omega_j X_j'\beta + (1-\omega_j)s_{ij}, \omega_j = \frac{h_j}{h_j+k_j}, h_j = \frac{dm_j}{d+m_j} \tag{2}$$

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:

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

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} independant. 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 stuctural 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

To finish with, let us 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 α_j^* is now obervable (thus precisely determined); this induces a certain form for the result. The sought probability of watching is in this context:

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

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

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

We note that 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 recieve are unbiased
- consumers are numerous and thus, on average, the utility they expect is close to the true utility (or quality) α_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 utiliy U_{pj}, $p \in [1, n_i]$ as a feedback to i after watching. Consumer i recieves two information in the same time:
- the fact that his acquaintancies 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.
- and their ex-post utility itself.

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

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

The second part of the expression is negative so that S_{ij2} is lower than the average of the ex-post utilities consumer i recieves; 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):

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

1.2.2 Expected Utility

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

$$\mathbb{E}_2[U_{ij}|X_j'\beta, s_{ij}, S_{ij2}] = \frac{h_j}{h_j+k_j+z_{i2}}X_j'\beta + \frac{k_j}{h_j+k_j+z_{i2}}s_{ij} + \frac{z_{i2}}{h_j+k_j+z_{i2}}S_{ij2}, \tag{7}$$

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

Likewise, in week $t \geqslant 2$, the exprected utility is

$$\mathbb{E}_t[U_{ij}|X_j'\beta, s_{ij}, S_{ij2}, ..., S_{ijt}] = \frac{h_j X_j'\beta}{h_j+k_j+\sum_{s=2}^t z_{is}} + \frac{k_j s_{ij}}{h_j+k_j+\sum_{s=2}^t z_{is}} + \sum_{w=2}^t \frac{z_{iw}S_{ij2}}{h_j+k_j+\sum_{s=2}^t z_{is}},$$

$$\mathbb{E}_t[U_{ij}|X_j'\beta, s_{ij}, S_{ij2}, ..., S_{ijt}] = \omega_{j1t}X_j'\beta + \omega_{j2t}s_{ij} + \sum_{w=2}^t \omega_{j3w}S_{ijw}, \tag{8}$$

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

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):

$$\begin{aligned} P_t &= \mathbb{P}(\mathbb{E}_t[U_{ij}|X_j'\beta, s_{ij}, S_{ij2}, ..., S_{ijt}] > q_{it}) \\ &= \Phi\left(\frac{(1-\omega_{j1t})(\alpha_j^* - X_j'\beta) + X_j'\beta - q}{\sigma_{jt}}\right), \end{aligned}$$

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. Additionaly, without any surprise, i.e. $\alpha_j^* = X_j'\beta$, the probability of watching in 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 advice from peers bring marginally less and less information from one week to another.

If we relax the hypothesis $X'\beta = q$, the same tendancies 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 symetric situation when $X'\beta < q$ and people's signal shows a negative surprise.

1.2.4 Ruling Out Nework 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 experience 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 2.2);
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 2.3);
3. we should observe more social learning effects when people have a greater social network (see 2.4);
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 2.5);
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 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

	Dependent variable:						
	log_sales_first_we						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log_screens_first_week	0.893^{***}	0.896^{***}	0.883^{***}	0.871^{***}	0.803^{***}	0.806^{***}	0.813^{***}
	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
R^2	0.907	0.909	0.910	0.912	0.932	0.936	0.938
Adjusted R^2	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 France data from 2004 to 2008 and find quite similar results (see table 2).

Table 2: Regression of first-week entries on number of screens for France

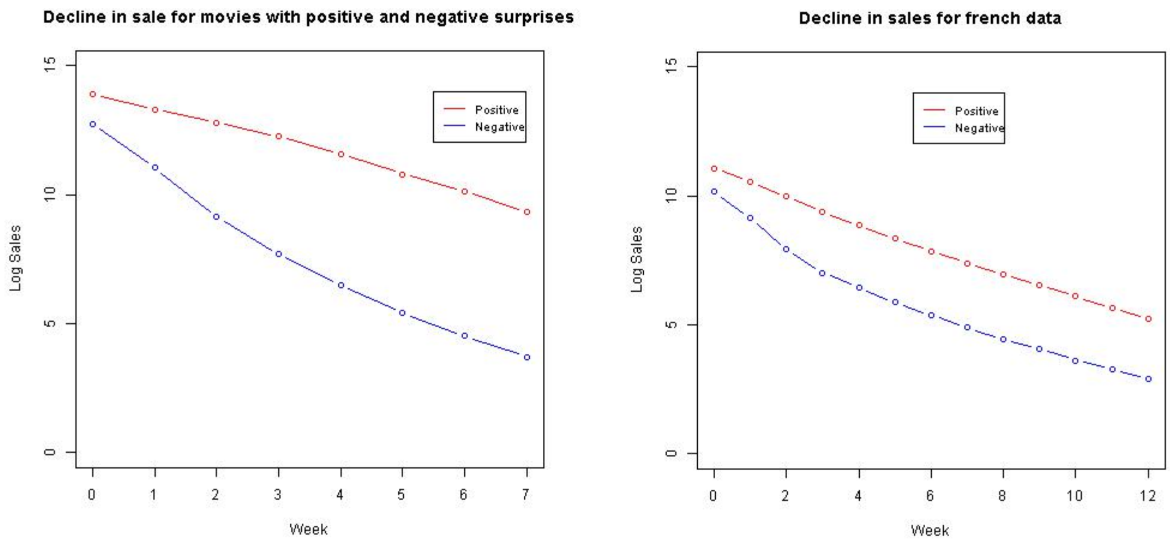
	Dependent variable:						
	log_entree_fr						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log_seance_fr	1.208^{***}	1.237^{***}	1.237^{***}	1.279^{***}	1.282^{***}	1.287^{***}	1.196^{***}
	(0.009)	(0.010)	(0.010)	(0.014)	(0.014)	(0.014)	(0.014)
R^2	0.893	0.899	0.900	0.917	0.924	0.925	0.943
Adjusted R^2	0.893	0.898	0.898	0.910	0.915	0.916	0.935
Note:							*p<0.1; **p<0.05; ***p<0.01

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).

2.2 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 durations we computed, though it is clear that trajectories diverge in the first four weeks of projections.

Figure 1: Comparing decline in sales between Moretti's and French data



To test for differences in trajectories, Moretti estimates models of the form:

$$\ln(y_{jt}) = \beta_0 + \beta_1 * t + \beta_2(t * S_j) + d_j + u_{jt} \tag{9}$$

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 β_2 because we want to identify an effect of the surprise on the dynamic of sales over time. Table 3 estimates equation 9 and then differentiates between positive, top, middle and bottom surprises. Even though our value of β_2 is much smaller than Moretti's (0.076<0.463), it statistically different from zero. This means that we have a significant difference in trajectories between positive and negative surprises, but less important than in Moretti's paper.

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

	Dependent variable:			
	log_entree_fr			
	(1)	(2)	(3)	(4)
t	-0.526^{***}	-0.526^{***}	-0.571^{***}	
	(0.002)	(0.002)	(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^2	0.851	0.853	0.853	0.854
Adjusted R^2	0.838	0.841	0.841	0.841
Note:				*p<0.1; **p<0.05; ***p<0.01

2.3 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, the information they would learn less 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 genre variances of the surprises in the first week as a proxy for the precision of their prior.

Table 4: Precision of the prior

	Dependent variable:			
	log_entree_fr			
	(1)	(2)	(3)	(4)
t	-0.570^{***}	-0.698^{***}	-0.678^{***}	-0.578^{***}
	(0.003)	(0.013)	(0.004)	(0.003)
t×positive_surprise	0.105^{***}	0.109^{***}	0.009	0.078^{***}
	(0.005)	(0.018)	(0.006)	(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^2	0.855	0.854	0.880	0.855
Adjusted R^2	0.843	0.842	0.870	0.843
Note:				*p<0.1; **p<0.05; ***p<0.01

Moretti estimates models of the form:

$$\ln(y_{jt}) = \beta_0 + \beta_1 * t + \beta_2(t * S_j) + \beta_3(t * precision_j) + \beta_4(t * S_j * precision_j) + d_j + u_{jt} \qquad (10)$$

where precision j is a measure of the precision of the prior for movie j. The coefficient of interest is the coefficient on the triple interaction between the time trend, the surprise and the precision of the prior, β_4. We make three hypothesis:

1. as in Moretti (2011), we test that sequels ("*saga*") had higher priors and where subject to less social learning effect;

2. we also suppose that art-house cinema ("*art et essai*") have a more unpredictable quality;

3. and we wonder if 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. 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. Our results support those of Moretti.

2.4 Size of the Social Network

Consumers with a larger social network receive more feedbacks from their peers and thus they 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:

$$\ln(y_{jt}) = \beta_0 + \beta_1 t + \beta_2(t \times S_j) + \beta_3(t \times NS_j) + \beta_4(t \times S_j \times NS_j) + d_j + u_{jt} \tag{11}$$

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 β_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 11 with a dummy for teen movies. He finds that the estimate of the coefficient β_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 the way Moretti build 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 used to defined 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 β_4 is significantly negative. Using *teen3*, we find that the estimate of β_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 β_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 11 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.

2.5 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:

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

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

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

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

	Dependent variable:			
	log_sales			
	(1)	(2)	(3)	(4)
t	-1.200^{***}	-1.289^{***}	-1.286^{***}	-1.221^{***}
	(0.012)	(0.009)	(0.009)	(0.012)
$t\times$positive_surprise	0.610^{***}	0.645^{***}	0.638^{***}	0.606^{***}
	(0.017)	(0.013)	(0.013)	(0.017)
$t\times$teen	-0.213^{***}			
	(0.019)			
$t\times$positive_surprise$\times$teen	0.072^{***}			
	(0.026)			
$t\times$teen2		-0.025		
		(0.089)		
$t\times$positive_surprise$\times$teen2		-0.602^{***}		
		(0.137)		
$t\times$teen3			-0.304^{***}	
			(0.097)	
$t\times$positive_surprise$\times$teen3			0.353^{**}	
			(0.163)	
$t\times$nb_screens_first_week				-0.121^{***}
				(0.013)
$t\times$positive_surprise$\times$nb_screens_first_week				0.077^{***}
				(0.016)
Observations	39,936	39,936	39,936	39,936
R^2	0.788	0.787	0.787	0.787
Adjusted R^2	0.758	0.757	0.756	0.757
Note:		*p<0.1; **p<0.05; ***p<0.01		

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

	Dependent variable:	
	log_entree_fr	
	(1)	(2)
t	-0.663^{***}	-0.451^{***}
	(0.007)	(0.005)
t×positive_surprise	0.061^{***}	0.076^{***}
	(0.010)	(0.006)
t×tout_public	0.115^{***}	
	(0.008)	
t×positive_surprise×tout_public	0.031^{***}	
	(0.011)	
t×seance_fr_first_week		-0.033^{***}
		(0.001)
t×positive_surprise×seance_fr_first_week		0.011^{***}
		(0.001)
Observations	26,598	26,598
R^2	0.856	0.867
Adjusted R^2	0.844	0.856
Note:	*p<0.1; **p<0.05; ***p<0.01	

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

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

where $\mathrm{se}(2(\hat{\beta}_2 + \hat{\beta}_4)) = 2[\mathrm{Var}(\hat{\beta}_2) + \mathrm{Var}(\hat{\beta}_4) + 2 \cdot \mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_4)]^{1/2}$ 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 β_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)

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

3 Conclusion: some comments

A R codes

Figure 2: R code used to clean French data

```
###################
#  Data Cleaning  #
###################

# In this part, we change the dataset to make it closer to the dataset of Moretti.

# Remove the movies without any screen in France during the first week (667 movies).
fr_df <- fr_df[!is.na(fr_df$seance_fr1),]
# Remove the movies without any id_distributeur (4 movies).
fr_df <- fr_df[!is.na(fr_df$id_distributeur),]

# Set MoyennePresse and MoyenneSpectateur to the mean if no value is specified.
mean_moy <- mean(fr_df[!is.na(fr_df$MoyennePresse), 'MoyennePresse'])
fr_df[is.na(fr_df$MoyennePresse), 'MoyennePresse'] <- mean_moy
mean_moy <- mean(fr_df[!is.na(fr_df$MoyenneSpectateur), 'MoyenneSpectateur'])
fr_df[is.na(fr_df$MoyenneSpectateur), 'MoyenneSpectateur'] <- mean_moy

# Repeat each columns 13 times.
n <- nrow(fr_df)
df <- fr_df[rep(1:n, each=13),]

# Add a column to indicate the week.
df$t <- rep(0:12, n)

# Replace the variables for each week (e.g. 'entree_paris1') with a global variable (e.g. 'entree_paris')
for (i in 0:12) {
for (variable in c('entree_paris', 'seance_paris', 'entree_fr', 'seance_fr')) {
# Concatenate the variable name with and indicator for the week (e.g. 'entree_paris1').
variable_t <- paste(c(variable, toString(i+1)), collapse='')
# For each week, the variable in the new df (e.g. 'entree_paris') is taken from the old df (e.g. 'entree_paris1').
df[df$t==i, variable] <- fr_df[,variable_t]
}
}

# Keep only the useful variables.
df <- df[,c(1:6, 33:43, 70:85)]

# Replace the NAs in seance_fr with zeros.
df[is.na(df$seance_fr), 'seance_fr'] <- 0

# Generate logarithm of sales and screens.
df$log_entree_paris <- log(df$entree_paris + 1)
df$log_seance_paris <- log(df$seance_paris + 1)
df$log_entree_fr <- log(df$entree_fr + 1)
df$log_seance_fr <- log(df$seance_fr + 1)

# Variable id_distributeur is a factor.
df$id_distributeur <- as.factor(df$id_distributeur)

# Variable id is a factor (this is used for movie dummies with the package lfe).
df$X <- as.factor(df$X)
df$X.eff <- rnorm(nlevels(df$X))
```

Figure 3: R code used to obtain French surprises

```
###############
#  Surprises  #
###############

# In this part, we estimate the surprises of the movies.

# Regression of first week sales on number of screens.
regSurprise1 <- lm(log_entree_fr ~ log_seance_fr, data = df, subset = (t==0))
# Including dummies for genre
regSurprise2 <- lm(log_entree_fr ~ log_seance_fr + genre, data = df, subset = (t==0))
# Including dummies for ratings
regSurprise3 <- lm(log_entree_fr ~ log_seance_fr + genre + interdiction, data = df, subset = (t==0))
# Including dummies for distributor
regSurprise4 <- lm(log_entree_fr ~ log_seance_fr + genre + interdiction + id_distributeur, data = df,
     subset = (t==0))
# Including dummies for month and week
regSurprise5 <- lm(log_entree_fr ~ log_seance_fr + genre + interdiction + id_distributeur + factor(
    mois) + factor(semaine), data = df, subset = (t==0))
# Including dummies for year
regSurprise6 <- lm(log_entree_fr ~ log_seance_fr + genre + interdiction + id_distributeur + factor(
    mois) + factor(semaine) + factor(annee), data = df, subset = (t==0))
# Including other variables
regSurprise7 <- lm(log_entree_fr ~ log_seance_fr + genre + interdiction + id_distributeur + factor(
    mois) + factor(semaine) + factor(annee) + MoyennePresse + sigma_note_presse + PoidsCasting + pub
    , data = df, subset = (t==0))

# Print a table with the results of the last regressions.
stargazer(regSurprise1, regSurprise2, regSurprise3, regSurprise4, regSurprise5, regSurprise6,
    regSurprise7, type='text', keep=c('log_seance_fr'), omit.stat=c("f", "ser"), title='Regression
    of first-week entries on number of screens')

# Surprises are defined as the residuals of the last regression.
surprise <- residuals(regSurprise7)
df$surprise <- rep(residuals(regSurprise3), each = 13)
quantile(df$surprise, probs = c(0, .05, .1, .25, .5, .75, .9, .95, 1))

# Generate additional variables for surprises.
df$positive_surprise <- df$surprise >= 0
q_surprise <- quantile(df$surprise, probs = c(1/3, 2/3))
df$bottom_surprise <- df$surprise < q_surprise[1]
df$middle_surprise <- df$surprise >= q_surprise[1] & df$surprise < q_surprise[2]
df$top_surprise <- df$surprise >= q_surprise[2]
```

Figure 4: R code used to obtain French sales dynamics

```
##############################################
#  Prediction 1: Surprises and Sale Dynamics  #
##############################################

# In this part, we study the difference in rate of decline between movies with a positive surprise
    and movies with a negative surprise.

# Regression of sales on the interaction between time and surprises.
# We use the command felm of the package lfe to compute linear regressions with thousands of dummies.
regSaleDynamics1 <- felm(log_entree_fr ~ t | X, data = df)
regSaleDynamics2 <- felm(log_entree_fr ~ t + t : surprise | X, data = df)
regSaleDynamics3 <- felm(log_entree_fr ~ t + t : positive_surprise | X, data = df)
regSaleDynamics4 <- felm(log_entree_fr ~ I(t * top_surprise)+ I(t * middle_surprise) + I(t * bottom_
    surprise) | X, data = df)

# Print a table with the results of the regressions.
stargazer(regSaleDynamics1, regSaleDynamics2, regSaleDynamics3, regSaleDynamics4, omit.stat=c("f", "
    ser"), title='Decline in box-office sales by opening week surprise')
```