The Economic Value of Integrated Data Models for the Financial Industry

Andreas Glatter[1], Martin Janssen[2] and Paolo Vanini[3]

Keywords: Data modelling, data exploration, value data systems, irreversible investment,

JEL-classification: G 20, G 29, C 61, D 70, D 74

[1]ECOFIN Research and Consulting Ltd., Neumünsterallee 6, CH-8008 Zürich, e-mail: glatter@ecofin.ch

[2]ECOFIN Research and Consulting Ltd., Neumünsterallee 6, CH-8008 Zürich and University of Zurich, Department of Economics. E-mail: janssen@ecofin.ch

[3]ECOFIN Research and Consulting Ltd., Neumünsterallee 6, CH-8008 Zürich and University of Zurich, Department of Economics. E-mail: vanini@ecofin.ch (Corresponding author)

Abstract

A necessary condition for successful decision-making, analysis and processing in the banking industry is the availability of comprehensive and consistent data. Basically, there are two approaches for structuring data: Data exploration and data modeling. These two methods are useful under different circumstances and we first discuss the main criteria to determine which approach should be followed. In a second step a dynamic cost analysis for the two methods of structuring data is performed, where the model captures the criteria discussed such as semantic and syntactic consistency, comprehensiveness, cost of data acquisition and data entry, development costs, project risk and transaction costs. The basic question we analyze is under which conditions should one optimally switch from data exploration to data modelling, what is the optimal time for a firm to switch and how does the model parameters affect this optimal switching time.

1 Introduction

Economic research in the areas of financial instruments and of the banking industry usually takes the availability of data for granted. Little effort is made to assert the value of consistent data for analysis, decision making and processing.

This paper will address exactly this issue.

Data can be physically centralized in a big database or spread across different data bases (often several hundred in large financial institutions). Conceptually data can be structured according to one single concept or to different, unrelated concepts. Concepts for structuring data are called "Data Model". The physical and the conceptual approach are not necessarily correlated. Data can be spread over different databases but stored according to a single, common data model. Also a centralized database may contain information totally unrelated and even highly redundant. Concepts for structuring data for an entire institution are known under the term "Data Warehouse". Data warehousing covers a wide range of approaches to data management, out of which only two following more or less opposite philosophies will be addressed in this paper.

There are two basic approaches to structuring these data. The first is to develop a consistent and comprehensive data model, set up a database according to this model and transfer all data from existing databases into this database which then serves as a reference data base. All new applications can rely on this database as source for their data. This approach is called data modeling (DM) throughout this paper. The second approach is to extract data for each application from existing databases. This approach makes use of the fact, that most data are already collected somewhere within each financial institution. So, only some remodeling may be required for data not already maintained in any of the existing databases. This approach will be called data exploration (DE) throughout this paper. This is not a technical expression, but useful for referring to the method (the article of Frawley et al. (1990) presents an overview of the methods and problems encountered in data mining systems).

A data model is a concept for structuring data. Any meaningful use of data requires a data model, whether this is made explicit or not. Data are modeled for different purposes. Each purpose requires its own data model. If different data models are used, it has to be decided how the data should be permanently stored. Using different models for the same data independently of each other would mean multiple storage of data with the associated costs for collecting and entering the data and a high probability of inconsistencies and faulty data. No institution follows exactly this approach. All financial institutions decide on how they store data within a reference database independently of an individual application. This requires a second layer of data modeling. The data modeling for applications becomes independent of the data modeling the reference database. Developing a data model for a reference database and thus for different applications requires the designers to know the needs of current applications and to anticipate the needs of future applications. Since designers are confined to their knowledge acquired by education and experience, this means that only a limited scope can be handled. Data exploration and data modeling take different approaches defining their scopes of data. Data exploration relies on past decisions to establish different and unrelated databases thereby accepting multiple data entry, inconsistency or inefficiency. The reason lies not so much in mismanagement but in the more or less independent operation of business units having their own, historically grown IT-departments and lacking infrastructure for communication between these entities. Data exploration tries to combine the efforts in the past and to feed new applications from the existing databases on the basis of today's communication technology. Therefore usually no transition process is necessary. In contrast, data modeling breaks with the past storing data according to an entirely new data model. It requires a transition process usually involving an initial load for the new database from the old ones and subsequently feeding the existing databases from the new one.

These different data structures lead to the basic question for a firm: What are the main criteria to determine which approach should be followed? The various criteria and trade-offs are analyzed in Section 2. This analysis assumes that the firm can take this decision without inheritance, i.e. as if no database is actually used. Clearly, this is not true for most firms, since they are on the contrary actually using databases. Suppose they use DE and the basic question for them is then: It is optimal for the firm to maintain DE or it is optimal to switch to a DM system? If this is optimal, what is the optimal time to perform the change? This question will be analyzed formally in Section 3. In Section 4 the same questions as in section 3 are analyzed in a setup where the the DE cost function is embedded in a market. Section 5 discusses the results and some issues not tackled in the paper and in Appendix A a categorization of financial data is provided. Appendix B contains the mathematical details of the investment analysis.

2 Trade-offs between data modeling and data exploration

Data modeling and data exploration are useful under different circumstances. The main criteria to determine, which approach should be followed, are depicted below.

Semantics and consistency
Independently developed databases are not consistent because they employ different semantics for their data. In addition, some data needed may be missing. This problem grows with the complexity of data and with the number of databases to be

consolidated. Due to the difficulties of handling the unrelated semantics of different databases the result is often not fully consistent and often may not sufficient to be used for processing as an example. Therefore, the more complex the data in question are and the more data are missing, the more a data model turns out to be the only possible solution.

Costs of Data Acquisition and Data Entry
Using databases developed independently of each other usually leads to great redundancies in data stored. For example a trading department may store the same dividends, as does the back office despite using only slightly different details of information. Nearly all departments collect prices of securities. This redundancy leads to high costs of data acquisition and data entry. It can only be eliminated if all data are collected at a single entry point, which in turn requires an appropriate data model. Operationally, data exploration usually does not incur additional costs for maintaining more data. But, all redundancies are kept. Data modeling leads to elimination of these redundancies.

Development Costs
Development costs of a comprehensive and consistent data model are high, but they can be estimated roughly. Development costs for data exploration are low, as long as the complexity of data and the number of databases to be consolidated are low. The costs of data exploration can outgrow the costs of data modeling considerably if small data exploration activities are carried out in great numbers. Generally it is much harder to predict the costs of data exploration since the combined effort of exploring data is never made explicit.

Project risk
The risk of data exploration lies in possibly lacking data and the impossibility to consolidate data collected and stored in different databases developed independently and according to some unrelated semantics. The risk of data modeling in turn lies in the difficulty to determine the data required and to define the objectives of modeling. Data modeling projects often fail because the necessary expertise can not be made available and the scope of modeling is not clearly enough defined. These risks can be reduced in several ways.

1. Using a stepped approach defining the scope of a detailed model. In this approach the area to be modeled has to be broken down from a top level to a more detailed level. The scope of the detailed model can be limited by this approach while also specifying the interfaces with other parts of the overall model.

2. Using commercially available models entirely or as a starting point for own modeling activities.

The number of projects failed and the respective amounts, that have been spent, indicate that data modeling is not a risk-free venture. The catastrophic results of an unsuccessful attempt to create a comprehensive and consistent data model are unlikely to be matched by unsuccessful data exploration since smaller amounts are spent at each occasion in terms of money, human resources and human goodwill.

Transition Costs
Basing applications on a new data model requires all interfaces to be rewritten. This is a costly and time-consuming task. While data exploration eliminates these costs data modeling requires a transition period during which the old databases are filled from the new one. These will continue to be used as the basis for all existing applications. New applications will then be written on the basis of the new database and allow the old databases to be eliminated one after the other over time. Therefore the transition costs can be kept low even with an entirely new data model provided it contains all information already stored in the old environment.

Development over time
Data modeling always takes place in a certain environment given at any time. Over time new requirements will come up and new technologies will emerge. This will mean additional databases will be created using more modern techniques. Therefore every data modeling approach will turn into a data exploration approach sooner or later. The greater the variety of databases becomes the more difficult it will be to stick to data exploration. So the question is not really whether to use this or that approach but to find the optimal point in time to start a development from scratch.

Stability
Databases are used by many applications. All these applications have to be rewritten, if the data model changes. Therefore keeping data models stable may save considerable amounts of money. On the other hand keeping the old models and filling them from a new one greatly reduces the problem.

The criteria described above suggest that data exploration is the appropriate technique for consolidating a limited number of databases, containing data of limited complexity and exhibiting no or little holes. As soon as the number of databases becomes large, data are complex or a lot of the required data is stored in neither database data modeling becomes the more promising approach. In any institution a mixed strategy will be used: data exploration will be seen as the appropriate solution for minor problems while for bigger steps data modeling has to be considered. Each time a major new model has been introduced it will take some time until a new model will be developed or introduced.

This paper addresses two questions related to data modeling: When is it optimal to take the decision to switch from data exploration to data modeling? If data modeling is the solution, is it better to develop a proprietary model or to buy an existing one on the market?

The first question will be addressed in several steps. A first step in analyzing the decision problem is to take into account the development risk associated with either decision. The second step takes into account, that some data exploration activity is always taking place in financial institutions. The problem in this environment is to find the optimal point in time for switching from an ongoing process of data exploration to a big investment in data modeling. Finally the effects of data modeling and data exploration on the resulting operational costs and benefits in terms of data entry and availability of an increased amount of data will be considered.

The second question will be analyzed within a framework of contract theory. The "make-"decision can usually not be taken independently of the existing resources. Usually some expertise from the front departments is required. But these front departments earn their money by doing their transactions. They would appreciate a better environment but would not do the job for other departments depriving themselves from potential earnings out of their usual business, which they know best. Building a proprietary data model will have the advantage of exactly fitting the needs of the own institution but is clearly more expensive than buying a model on the market. Buying a model in the market in turn reduces project risks and provides some commonality with data formats used in other institutions, which facilitates the exchange of information. Our analysis in Section 4 will take into account these factors.

3 Cost Analysis

We formally analyze in this section the cost evolutions of DM and DE, respectively. This model allows us to determine when does it is optimal to switch from one system to the other. The methodology of our approach is beautifully presented in the book of Dixit and Pindyck (1994). That for,

1. we need reasonable models which are based on the economic analysis of the last Section and

2. we have to introduce an objective function which is a good measure for a potential switching between the two kinds of data systems under consideration.

We first present the model and then the relation to the economic discussion of the last Section is provided. We assume that the benefits of both technologies are the same but that the costs of implementation, maintainance and risk in data exploration are different.

3.1 Model

We consider first the situation where a firm is actually running a DE. The maintainance costs $V(t)$ of the DE are determined by

$$dV_t = \alpha(N(t))V_t dt + \sigma V_t dB_t \ , V(0) = V_0 \tag{1}$$

with B_t a standard Brownian motion, σ the constant volatility and α the drift component of the cost process. $N(t)$ describes the number of different stations in DE and the drift is an increasing and concave function of N. The systematic risk term $\sigma V_t dB_t$ models risk in a data exploration system due to unforeseen consequences of the inconsistency or the lack of completeness of the data system for example. If we want to connect the costs of DE to the competitive market costs, we would chose a mean-reverting dynamics

$$dV_t = \eta(t)(\tilde{V} - V_t)dt + \sigma V_t dB_t \ , V(0) = V_0 \tag{2}$$

with $\tilde{V}$ the long run costs of DE in the competitive market. Since (1) is easier to handle we consider in this section the dynamics (1) and discuss the competitive market costs case in the next section.

The DM-cost function is modelled as follows:

$$dI_t = \mu I_t dt \ , I(0) = I_0, \tag{3}$$

i.e. there is no systematic risk component since the system is consistent and spanning (i.e. any contingent claim contract can be hedged).

After the introduction of the cost functions the following question are basic for the firm:

1. What are the optimality conditions for the trade-off between large sunk costs and systematic risk costs?

2. If such optimal conditions exist, what is the optimal time to switch from one system to the other one?

The objective function of the firm is

$$F(V) = -\max_{\tau} E[e^{-rt}(V_\tau - I_\tau)] \tag{4}$$

with r the risk free interest rate. Hence, the firm chooses the optimal time so that the net cost gains by changing from DE to DM are a maximum. This optimal stopping problem is carried out under the conditions that the dynamics (1) and (3) hold.

The model captures the following intuitions. We assume that $I(0) >> V(0)$, i.e. the sunk costs are much larger for the DM than for the DE. The sunk costs for DM are to be understood as an aggregate quantity. That is, transition costs (rewriting of all interfaces for example) and the development costs are two main objects hidden in the sunk costs. The fact that independently developed databases are not consistent due to different semantics employed is modeled by the systematic risk component in (1). A comparable term is missing for DM in equation (3). The systematic risk modeled by the Brownian motion is also used to model project risks due to lacking data and the impossibility to consolidate data collected and stored in different databases. Finally, the model is flexible enough to model different cost evolution scenarios for the DE.

We have not modeled the important tasks related to the *stability* of the data systems. Hence, the preferences for stable data models do not enter the analysis below.

3.2 Results

We consider the solution of the optimization problem (1), (3) and (4) for two different cost evolution scenarios.

3.2.1 Sunk Cost Case

This is the simplest form of the model described in the last section where the maintainance cost rate μ for DM is zero. I.e. the choice of the firm is to keep going on with DE and the cost evolution (1) or to make an irreversible, sunk cost investment I_0 in DM. We assume further that the deterministic cost rate α in (1) is a constant. Although this situation overfavorizes DM it reveals some major issues - under suitable modifications these also hold in the less biased cost analysis presented in the next section. Figure 1 illustrates the content of Proposition 1.

Insert Figure 1 around here

The figure shows that for $r \leq \alpha$ a finite stopping time exists, i.e. a time where it is optimal to invest in DM. If the last inequality is reversed the stopping time is infinite, i.e. the DE is maintained for ever. The intuition is that for a risk free interest rate which exceeds the cost rate of the DE, the opportunity choice to invest in DM is worth nothing. The first main Proposition is:

Proposition 1 *Suppose that $r \leq \alpha$. Then, it is always optimal to switch from DE technology to DM technology before the costs of DE equal the sunk costs of DM. I.e. the critical value $\tilde{V}$, which defines the optimal stopping time (see the Appendix B), is strictly smaller than the sunk cost I_0.*

The Proposition is proved in Appendix B. The intuition for this fact follows the same logic as in the case where V describes the value of a project and not its costs.

In this latter case, the opportunity to postpone an investment decision possess a value which implies that an optimal decision is taken later than in the case where a simple but wrong net present value calculation is carried out without considering the option value of postponing an investment decision. Since we investigate costs in this paper, the argument is mutis mutandis the same but with a reversed conclusion. This result shows that postponing investment decisions to switch from DE to DM is not an optimal strategy for a firm.

The next result considers the impact of increasing average costs on the optimal stopping time.

Proposition 2 *If the cost rate α increases, the optimal stopping time τ moves to the left, i.e. the value $F(V)$ of the option to invest in DM decreases with decreasing running costs α whereas the opposite comparative statics holds if the noise σ increases.*

The intuition for this result is straightforward. What causes α to increase? An increasing drift α for the diffusion $V(t)$ can be due to increasing complexity of the data, the lack of containing all necessary information and if the number of number of databases increases. Roughly, the intuition is catched that for larger firms the expected switching time to DM is shorter than for smaller firms.

Summarizing, the analysis revealed that the optimal acquisition decision of a firm depends on (i) the number of databases of the firm, (ii) the magnitude of systematic risk due to inconsistency and in-comprehensiveness of the data, (iii) the relative relation of the expected cost growth rate of the two systems, (iv) the sunk costs induced by the investment decision to switch from one technology to the other, (v) the complexity of the data under consideration and (vi) the history of the cost process.

3.2.2 Increasing DM costs

We now assume that (3) holds with a non-vanishing running cost rate for DM, but that the rate μ is positive. Then basically the results of the last section are distorted in favor of DE but the major results that a firm should not wait too long to switch to DM. Basically, the switching region $\tilde{V}$ is no longer a constant but also growing exponentially. Since the expected costs of DE, $E[V_t] = V_0 e^{\alpha t}$, are also exponentially growing, it follows that the fine tuning between the rates μ and α determines whether in the long run it is profitable to switch from DE to DM.

Proposition 3 *If $\mu \geq \alpha$ and $I_0 > V_0$, it is never optimal to switch from a data exploration system to a data model.*

The proof is straightforward and therefore omitted. Clearly, the lower the sunk costs for DE are and the lower the growth rate of the costs is, the smaller is the probability (given a fixed σ) that the DE-costs hit the critical curve $\tilde{V}(t)$ such that a switch to DM pays for the firm. This last result has the interpretation that for a small firm - i.e. a firm with a low drift and small volatility in their maintainance costs for DE - replacing the data exploration system by a data model is unlikely to be optimal. This fact is reinforced if the sunk costs for the data model are increasing.

4 Cost Analysis in a Competitive Market

We consider the same model than we did in the last section but the dynamics (1) is replaced by the dynamics (2). In this model the DE costs of a firm are connected to the competitive market costs of DE system. Although the mathematical modification seems to be a minor one, it turns out that the analysis is much harder to do and the results are not simple modifications of those with an underlying geometric Brownian motion.

The intuition behind (2) is due to the following facts. In the long run, the expected DE costs $E[V_t]$ converge towards $\hat{V}$, the long run market equilibrium. The parameter η measures the speed of convergence towards the equilibrium (if $\eta > 0$). The derivation of these results is standard and omitted (see Dixit and Pindyck (1994) for the derivations and further properties of the dynamics (2)).

Proposition 4 *Suppose that the model (1), (2) and (4) are given and that $\eta + r > 0$ holds. Then the region in the (t, V)-plane where it is never optimal to switch to the DM is increasing if:*

1. *The sunk costs I_0 increase.*
2. *The market equilibrium costs $\hat{V}$ increase.*

Contrary to the model studied in the last section, not solely the sunk costs matter for the region where no switching occurs (see equation (6) in Appendix B) but also the market equilibrium.

What are further implications of the model? More specifically, how do the statements of the Propositions 1 to 3 change? At this point mathematical complexity forces us to apply numerical methods since closed form analytical solutions are no longer feasible to obtain. The next Proposition states the reason for this fact.

Proposition 5 *The optimal stopping costs $\tilde{V}$, which result from the solution of a boundary value problem with a Smooth Pasting and a Value Matching condition, is the solution of the following transcendental equation:*

$$V^2(1 - k_2) + V(\rho(V) + k_2 I_0) - I_0 \rho(V) = 0$$

with $\rho = (4+b)\frac{a_2}{b_2}\frac{M(a_2+1,b_2+1,-\frac{4+b}{V})}{M(a_2,b_2,-\frac{4+b}{V})}$ *and* $M(a,b,x)$ *the Kummer function. The parameters of the Kummer function are:* $a_2 = \frac{4k_2+4-2a+bk_2+b(2-a)}{4+b}$, $b_2 = 2k_2+2-a$, k_2 *is the negative solution of* $k^2+(1-a)k+c=0$ *and* $a = -\frac{2\eta}{\sigma^2}, b = -a\hat{V}, c = -\frac{2r}{\sigma^2}$.

Nevertheless interesting insights can be deduced from Proposition 5. That for we choosed numerical values for some parameter[4] and we expanded the transcendental equation in Proposition 5 around the market equilirium value up to the second order in the maintainance costs V. The resulting approximation is shown in the following figures under different circumstances.

In Figure 2 the parameters of the optimal stopping costs condition where the maintainance costs V and the convergence speed η.

Insert Figure 2 around here

It follows that for a positive, increasing speed, i.e. the DE maintainance costs of the firm are actually converging towards the market value, the optimal stopping costs decrease. Therefore, the quicker the firm converges towards the equilirium, the sooner it is optimal to switch to a DM. The situation changes dramatically if the convergence speed becomes negative, i.e. the DE costs are not converging toward the equilirium. Then there will in general no longer exist a value V such that for $\eta < 0$ the point (η, V) lies on the curve which separates the black and white regions (i.e. the set which defines the optimal value $\tilde{V}$). This is intuitively clear: If your data exploration costs are exploding relative to the market costs, you should switch immediately to a DM. Contrary to the case of a geometric Brownian motion dynamics, the optimal decision of switching to a DM from a DE is not only determined by their relative cost and expected cost evolution but also by the relative positioning of the actual DE in the DE-market. I_0 discussion

Insert Figure 3 around here

σ

Insert Figure 4 around here

5 Discussion of Results

We have shown that there is no straightforward answer to the question: How should our firm structure the financial data? Should we use a data model system or a data exploration system? The analysis revealed that different factors matter for the

[4] The parameters are $r = 0.05, \sigma^2 = 0.2, \hat{V} = 2, I_0 = 1, \eta = 1$. This are the default values, i.e. if we consider for example the case where η is fixed the corresponding value is attributed.

optimal decision to be chosen which have to be considered carefully. Nevertheless, there are some rough rules which set the border lines for the decision process. First, large firms should optimally use a data model or it will be optimal for them to switch from a data exploration system to a data model sooner than for smaller firms. These results are reinforced if the data to be structured are of a high complexity. Contrary, for small firms which a low expected maintainance cost growth rate the sunk costs for a data model may prevent them to implement such a model. This outcome is reinforced if the project risk is small for the firm or if comprehensiveness is not an important factor.

Although the answer to the questions mentioned above are intricate, there is a clear answer to the time resolution of the trade-off between accepting the exposure to the systematic risk of DE or the large sunk costs for a data model. If there is an optimal finite switching time, it is optimal to switch before the costs of DE reach the level of the sunk costs of DM due to the systematic risk component in the model.

Contrary, the need for smaller firms to switch from DE technology to DM is less urgent. If the optimal switching time is reached it may pay to develop the data model in-house if the model does not need to many individuals involved in its development.

6 Appendix A: Categorization of Financial Data

The analysis will be confined to data on financial instruments including the availability of prices. Macro economic research, research on the financial structure of a company, or on management re-muneration schemes is not subject to this paper. For the analysis of portfolios data can be categorized into various classes according to complexity, amount and availability:

- Market Data (e.g. quotes on financial instruments, interest rates)
  - Low complexity
  - Huge amounts
  - Easily available
- Publicly available information on financial instruments and market participants
  - Very high complexity
  - Large amounts
  - Limited availability
- Privately available information on financial instruments and market participants

  - High complexity
  - Large amounts
  - Only private availability

  The availability of each of these data categories requires different handling facilities:

- Market data are usually easily structured. They are often available from multiple sources. One problem with handling these data is usually to cope with the huge amount. The other problem is that for many financial instruments no data can be obtained from any publicly available source.

- Publicly available information on financial instruments and market participants can usually be obtained in written form. For storing the data in a database these have to be structured and entered. This requires a appropriately structured database - and therefore a data model - as well as a costly process to enter the data. This last requirement can be reduced if an agent collects and enters the data centrally and sells them to the market participants.

- Privately available information on financial instruments and market participants consists of all data not of enough interest for most market participants as well as proprietary information, that should not leave the area of a single financial institution, e.g. private credit relationships and positions. This type of information requires the same well-structured database as publicly available information but there is no chance of having it entered by a central agent. It has to be collected and entered by each institution individually.

7 Appendix B

The mathematical methods used in this Appendix are intuitively explained in chapter II in the book of Dixit and Pindyck (1994). A more formal treatment Oksendal (1995), chapters IX and X. We assume that the drift of the value process is smaller than r, else the value of the option to invest will be infinite.

Proof of Proposition 1:

To start with, we determine first the set U where it is never optimal to stop the process

$$U = \{(s, V) | Ag(s, V) > 0\} \subset R^{2,+} \ , \ A = \partial_s + \alpha V \partial_V + \frac{1}{2}\sigma^2 V^2 \partial^2_{VV} \ . \tag{5}$$

We get

$$Ag = e^{-rs}(-r(I_0 - V) - \alpha V) > 0 \iff V < \frac{rI_0}{r - \alpha} =: V_0 \ . \tag{6}$$

Hence,

$$U = \{(s,V)|0 \le s < \infty, 0 \le V \le V_0\} \subset R^{2,+} \ . \tag{7}$$

We now consider the set D, the conitinuation region, where we know that $U \subset D$ and it is optimal to stop at τ_D, the exit time of the process from the region D (see Oksendal, chapter X, (1995) for the proofs). The theory implies, that finally we will have to solve a boundary value problem. As in the preceding example, D is an infinite strip of the form

$$D = \{(s,V)|0 \le s < \infty, 0 \le V < \tilde{V}, V_0 \le \tilde{V}\} \subset R^{2,+} \ . \tag{8}$$

It follows, that D istime-invariant in the following sense

$$D + (t_0, V) = D \ . \tag{9}$$

We set

$$g^*(s,V) = g_{V_{\max}}(s,V) = E^{(s,V)}[g(Y_{\tau_{V_{\max}}}] \ . \tag{10}$$

Setting $f(V,s) = g_{V_0}$, we know that $f(V,s)$ is the solution of the boundary value problem

$$Af(s,V) = 0 \ , \ 0 < V < \tilde{V} \ , \ \text{(PDE)} \tag{11}$$

and the boundary conditions

$$f(s,\tilde{V}) = -e^{-rs}(\tilde{V} - I_0) \ , \quad \text{(Value Matching)} \tag{12}$$

$$f(s,0) = 0 \ , \quad \text{(0 is an absorbing boundary)} \ . \tag{13}$$

The solution of this problem leads to $f(V,s) = g_{V_0}$. In order to determine the maximum $V_{\max}$, we use a third condition

$$\frac{\partial f(V,s)}{\partial V}|_{V=\tilde{V}} = \frac{\partial g(V,s)}{\partial V}|_{V=\tilde{V}} \ , \quad \text{(Smooth Pasting)} \ . \tag{14}$$

In the mathematical literature this condition is also called the "high contact principle" or the "smooth fit" condition. The solution obtained by the method of separation of variables is

$$f(s,V) = h(V)m(s) \ , m(s) = e^{-rs} \tag{15}$$

implies

$$h(V) = c_1 V^{\beta_1} + c_2 V^{\beta_2} \tag{16}$$

where $\beta_2 < 0 < 1 < \beta_1$ solve the fundamental quadratic equation $\frac{1}{2}\beta(\beta-1)+\alpha\beta-r = 0$. The constants c_i are determined by the Value Matching condition and that 0 is an absorbing boundary: (24) implies $c_2 = 0$ and the Value Matching condition implies

$$c_1 = -\frac{\tilde{V} - I_0}{\tilde{V}^{\beta_1}} \ . \tag{17}$$

Smooth Pasting finally leads to

$$0 < \tilde{V} = \frac{\beta_1 I_0}{\beta_1 + 1} < I_0 \ , \text{ since } \beta_1 > 1. \tag{18}$$

From $\tilde{V} < I_0$ follows, that it is optimal to stop the data exploration system before the costs augmented to the sunk costs of the data model. This proves Proposition 1.

Since all parameters of the solution $c_1, c_1, \tilde{V}$ of the model were determined using the three conditions Smooth Pasting, Value Matching and that 0 is an absorbing boundary , we get for the value of the option to invest

$$F(V) = \frac{I_0}{\beta_1^{\beta_1}}(\beta_1 + 1)^{\beta_1+1} \ . \tag{19}$$

Why is it optimal not to invest before $\tilde{V}$? That for, assume that $V < \tilde{V}$ holds. Then, the value of the option to invest $F(V)$ is strictly larger than $V - I_0$, i.e.

$$F(V) > V - I_0 \Longrightarrow V < I_0 + F(V) \ . \tag{20}$$

Hence, the value of the project is strictly smaller than the full costs, which is the sum of the direct costs I and the opportunity costs $F(V)$ to invest now rather than later.

Proof of Proposition 2: We calculate

$$\frac{\partial \tilde{V}}{\partial \alpha} \ = \ \frac{\partial \tilde{V}}{\partial \beta_1}\frac{\partial \beta_1}{\partial \alpha} = \tilde{V}\beta_1 \frac{1}{\beta_1(\beta_1+1)}\frac{\partial \beta_1}{\partial \alpha} \ . \tag{21}$$

Since $\frac{\partial \beta_1}{\partial \alpha} < 0$ and $\tilde{V}\beta_1\frac{1}{\beta_1(\beta_1+1)} > 0$ hold, it follows $\frac{\partial \tilde{V}}{\partial \alpha} < 0$. But this proves the Proposition 2 that the optimal stopping time, which defines optimal switching between DE and DM, decreases if the cost rate α increases.

Proof of Proposition 4:

The infenitesimal generator of the Ornstein-Uhlenbeck process is

$$A = \partial_s + \eta(\hat{V} - V)\partial_V + \frac{1}{2}\sigma^2 V^2 \partial^2_{VV} \ .$$

This operator is applies to $g = -e^{-rs}(V - I_0)$. If $\eta + r \geq 0$, it follows

$$Ag > 0 \iff V < \frac{\eta \hat{V} + rI_0}{\eta + r} =: V_0 \ .$$

Hence, the set U, where it is never optimal to stop, is equal to

$$U = \{(s, V) |\ 0 \leq s < \infty, 0 \leq V \leq V_0\} \ .$$

The area of the strip U is therefore increasing if I_0 and/or $\hat{V}$ are increasing. Furthermore, if the speed η increases in absolute values, V_0 decreases and it converges towards the equilibrium value $\hat{V}$. If $\eta + r < 0$, i.e. the expected maintainance costs of the firm's DE system explodes it follows that it is optimal to switch immediately since the U is in this case a strip in the third quadrant. This proves Proposition 4.

Proof of Proposition 5:

The continuation region D is time-invariant as in the proof of Proposition 1. The boundary value problem which is to be solved possess also three boundary conditions: The Value Matching condition, the Smooth Pasting condition and that 0 is an absorbing boundary. Applying the same separation of variable approach as in (15) for the new PDE $Af(s, V) = 0$ implies the ordinary differential equation

$$V^2 h'' + (aV + b)h' + ch = 0 \ , \ a = -\frac{2\eta}{\sigma^2}, b = -a\hat{V}, c = -\frac{2r}{\sigma^2} \ . \qquad (22)$$

The non-homogenity of equation (22) makes its solution more cumbersome than the solution of the corresponding equation in Proposition 1. First we substitute

$$V = \xi^{-1} \ , h(V) = \xi^k e^{\xi} w(\xi)$$

in (22), where k solve the quadratic $k^2 + (1 - a)k + c = 0$. (22) is then equivalent to

$$\xi w'' + ((2 - b)\xi + 2k + 2 - a)w' + ((1 - b)\xi + 2k + 2 - a - bk)w = 0$$

with primes denoting derivatives w.r.t. to ξ. This hypergeometric differential quation has the solution (after all resubstitutions)

$$h(V) = c_2 V^{-k_2} e^{-b} M(a_2, b_2, -\frac{4 + b}{V}) \ . \qquad (23)$$

The function M is the so-called Kummer function or the hypergeometric function $F_{1|1}$, c_2 is a constant to be determined, k_2 is the negative solution of the quadratic $k^2 + (1 - a)k + c = 0$ and the constants a_2, b_2 are defined by:

$$a_2 = \frac{4k_2 + 4 - 2a + bk_2 + b(2 - a)}{4 + b} \ , \ b_2 = 2k_2 + 2 - a \ .$$

Although (22) is a second order differential equation, one solution has already been neglected using that 0 is an absorbing boundary (i.e. the constant c_1 of this solution is equal to 0 to satisfy the mentioned boundary condition). On the solution (23) we apply the Value Matching condition (which determines c_2) and then the Smooth Pasting condition, which fixes the optimal stopping boundary value $\tilde{V}$. Contrary to the case of a geometric Brownian motion, the resultin equation from the Smooth Pasting condition is transcendent. Explicetly, the equation reads

$$V^2(1-k_2) + V(\rho(V) + k_2 I_0) - I_0 \rho(V) = 0 \; , \tag{24}$$

with $\rho(V) = (4+b)\frac{a_2}{b_2}\frac{M(a_2+1, b_2+1, -\frac{4+b}{V})}{M(a_2, b_2, -\frac{4+b}{V})}$. To derive the last result we used the following relation for the Kummer function:

$$\frac{d^n M(a_2, b_2, x)}{dx^n} = \frac{(a_2)_n}{(b_2)_n} M(a_2+n, b_2+n, x)$$

with the Pochhammer symbols $(a)_n = a(a+1)\ldots(a+n-1), (a)_0 = 1$. The numerical analysis is based on equation (24) and for the used asymptotic expansions we refer to the work of Gradshteyn and Ryzik for example.

8 References

Dixit, A.K. and Pindyck, R.S, (1994), Investment under Uncertainty, Princeton University Press, New Jersey

Frawley, I.F., Piatetsky-Shapiro, G. and Matheus, C.I., (1990), Knowledge Discovery in Databases: An Overview, AI Magazine 13, 3

Oksendal, B., (1995), Stochastic Differential Equations: An Introduction with Applications, Universitext Springer-Verlag Berlin, Fourth Edition

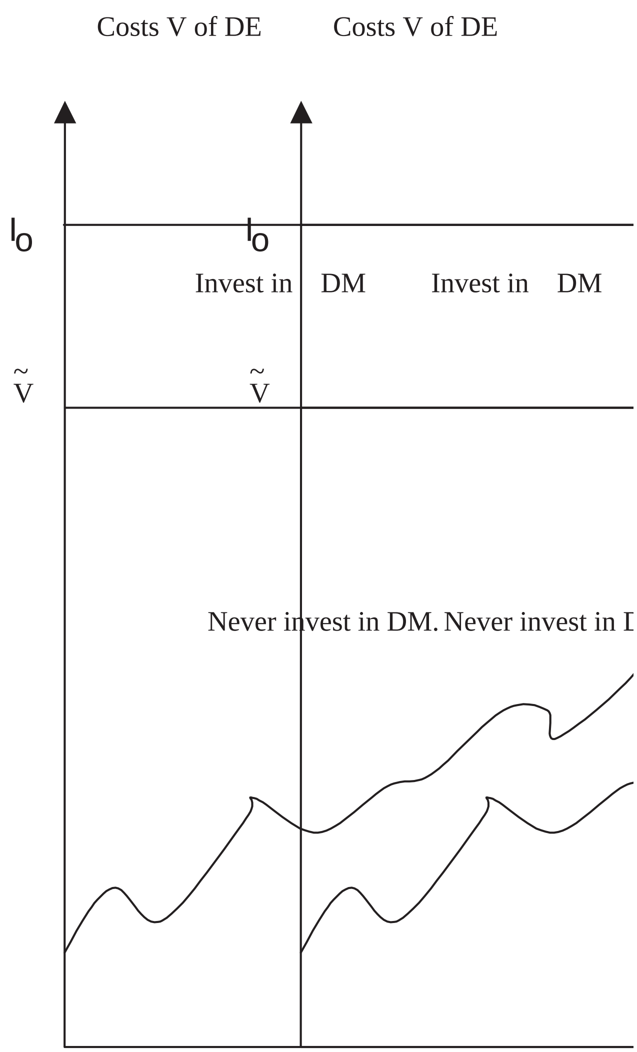
Costs V of DE
Costs V of DE
I0
I0
Invest in DM
Invest in DM
Ṽ
Ṽ
Never invest in DM.
Never invest in D

Figure 1

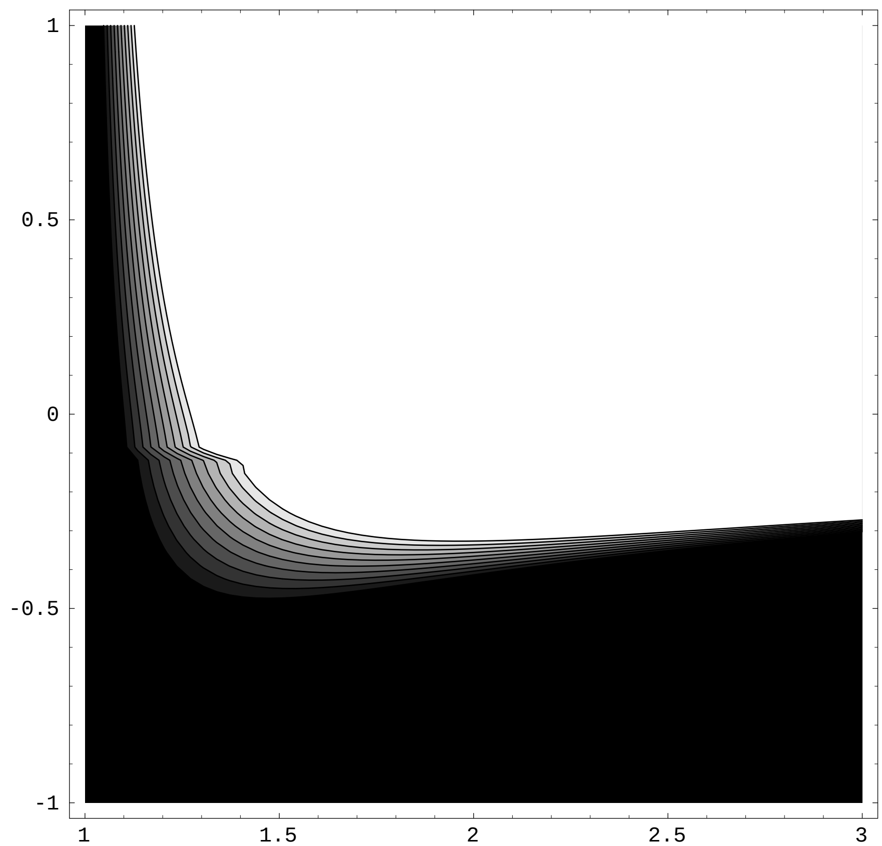

Figure 2: .

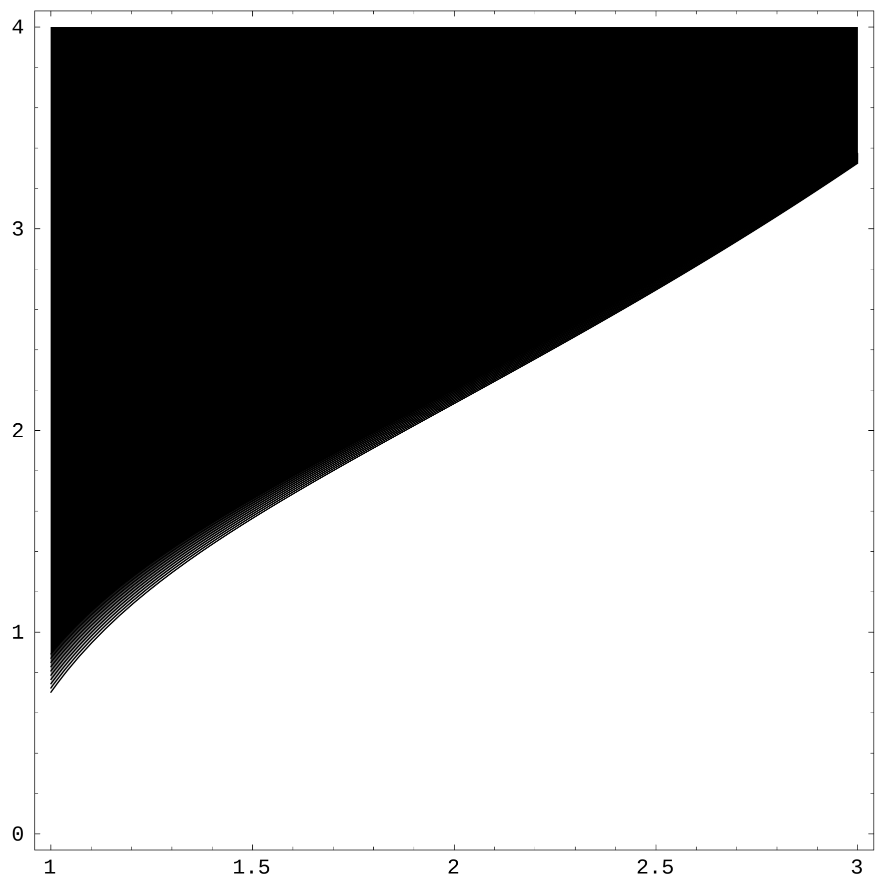

Figure 3: .