Original source of publication: Path dependent stochastic models to detect planned and actual technology use: A case study of OpenOffice.
Information & Software Technology 53(11): 1209-1226 (2011),
DOI: http://dx.doi.org/10.1016/j.infsof.2011.05.002
The original publication is available at http://www.sciencedirect.com/science/article/pii/S0950584911001042

Path dependent stochastic models to detect planned and actual technology use: A case study of OpenOffice

Bruno Rossi, Barbara Russo *, Giancarlo Succi

Center for Applied Software Engineering, Faculty of Computer Science, Free University of Bolzano-Bozen, Italy

A B S T R A C T

Context: Adopting IT innovation in organizations is a complex decision process driven by technical, social and economic issues. Thus, those organizations that decide to adopt innovation take a decision of uncertain success of implementation, as the actual use of a new technology might not be the one expected. The misalignment between planned and effective use of innovation is called assimilation gap.
Objective: This research aims at defining a quantitative instrument for measuring the assimilation gap and applying it to the case of the adoption of OSS.
Method: In this paper, we use the theory of path dependence and increasing returns of Arthur. In particular, we model the use of software applications (planned or actual) by stochastic processes defined by the daily amounts of files created with the applications. We quantify the assimilation gap by comparing the resulting models by measures of proximity.
Results: We apply and validate our method to a real case study of introduction of OpenOffice. We have found a gap between the planned and the effective use despite well-defined directives to use the new OS technology. These findings suggest a need of strategy re-calibration that takes into account environmental factors and individual attitudes.
Conclusions: The theory of path dependence is a valid instrument to model the assimilation gap provided information on strategy toward innovation and quantitative data on actual use are available.

1. Introduction

Adopting innovation is a complex process that comprises technical, organizational, economic, and social issues [14,39]. Innovation is about new ideas and getting ideas accepted involves the human sphere of individual choices and attitudes. Researchers in Information Systems have studied the subject for a long time since the initial work of Rogers [42]. Among other works, the Technology Acceptance Model (TAM) [14,50] and the Theory of Innovation Adoption (TIA) [28,43,62] have dominated the research in the field. They investigate what influences adoption: TAM models how users accept a technology and which are the factors that influence the choice whereas TIA focuses on the decision process and its key actors. In particular, TIA structures the adoption process in two major stages:

(1) A primary adoption, the decision made at the strategic-level that includes technology evaluation and selection.
(2) A secondary adoption, the actual adoption and use by individuals throughout the organization.

Examples in literature report that often primary and secondary adoptions are not aligned [8,35,52]. The actual use of technology that has been selected at primary adoption may be inhibited by social and technical issues that slow down or even stop secondary adoption [10,50]. In the specific case of IT, it is common to find workers that are reluctant to use new tools provided by the organization and that continue to use either old tools, which are accustomed to, or tools that they have installed themselves on their machines. Even when users have a positive attitude toward the new technology, the assimilation process might be slowed down by the users' learning pace. The gap between the predicted and the actual use of innovation is called assimilation gap [22]. As such, the assimilation gap indicates the misalignment between primary and secondary adoption.

Since the early works of Rogers [42], there have been several studies on factors that reduce or expand the assimilation gap [10,14,28,34,44,53,57]. There are factors related to the technology itself like the innovation attributes of Rogers [44] or to the user's perception like ease of use and usefulness of TAM [14]. Recently Turner et al. [56] have discussed to which extent this research has been able to estimate the *actual use* of technology – and the assimilation gap – rather than simply the user's *intention to use a* technology. The outcome is not satisfactory. In the majority of the works reviewed by the authors, the key predictors of technology acceptance – like ease of use and usefulness perceived by the users – correlate with the intention of use, but do not correlate with the actual use of a technology. As such, traditional models of technology acceptance are poor predictors of the actual technology use and the corresponding assimilation gap.

On such premises, our work moves away from the research of factors of technology acceptance and perceived use of IT. In this respect, this study aims at addressing the following questions:

RQ: How the assimilation gap can be objectively quantified? Is there any rigorous approach that compares the use of technology foreseen in the strategic plan toward IT innovation and the actual use?

With our method, the actual use is directly modeled from objective use data. The resulting model is then compared to the model of technology use forecasted by the strategy at primary adoption. Specifically, we build two temporal models that estimate the use of competing technologies across the entire adoption process: one predicted by the strategy and one derived from use data of the technologies. The difference between these two models gives us a quantifiable approximation of the assimilation gap. In the ideal case, when such difference gets less and less over time, the adoption process converges to a positive end: at some point the use foreseen by the strategy coincides with the actual use. In reality, the gap may persist and be significant all along the process of adoption. With our tool, the gap is traced seamlessly and reported to managers. When our tool reports a significant gap, a recalibration is needed and the manager can decide either to review the strategy or its implementation with actions that facilitate the actual use of the technology.

Therefore, the challenges of our work become how to build such models. To create the two models, we use one single mathematical approach applied to two different types of data: the information gathered from the strategic plan and the actual use data. The former is collected by interviewing decision makers. The latter is collected automatically from the files' repositories. For the latter, we have developed an ad hoc script that collects basic information on files stored in a server. With this tool, we get historical data at once. This contributes to reduce the time for data collection and limit the intrusion into company's activities, but also limits the type of information at our disposal to the minimum necessary for the present research. Other sophisticated tools for data collection can provide more complete and real-time data (e.g. PROM, [9]). For example, such tools also store information about time of use of an application. As such, it could complement the work of this paper and might be matter of future research.

The mathematical approach we consider derives from the work of Arthur on path-dependent processes and the theory of increasing returns (1994). Path-dependent processes are processes where a new status of a system is based on its past states, that is, its history. Path-dependent processes have been long adopted in describing the fluctuation of the users' choice among competing technologies in the economic and technologic market [1,2,4,5,13,38,47]. In this paper, we use this theory to describe the evolution of technology use during the adoption of a new technology. The user makes a choice every day and even more than once per day: s/he selects a technology to use among a set of technologies that are perceived as equivalent to perform given tasks and then repeats a choice over time. For example, in the case of open source software (OSS), the user may select to stay with the old technology or to move to the new OSS technology, regardless of the recommendation of the strategic levels of his/her organization. The choice is principally made on the user's experience – her/his history. This choice is called non-ergodic.

The principle of returns explains the rationale of the users' choice in the selection of goods in the economic market. In particular, Arthur describes the choice of a technology with the increasing returns, which is based on the claim "the more you sell, the more you sell" (1989, 1994). This indicates that a technology that dominates the selling quotes will dominate in the future as well. Simulations done in the past by Marchesi et al. [36] confirm the validity of this model also for software systems. In our work, we further extend the concept to software use rephrasing this claim with "the more you use, the more you use," meaning that future use is driven by the current one.

The theory of path dependence and the principle of increasing returns aim at identifying equilibrium points in the evolution of the users' choice. In particular, they aim at determining the winning technology or equivalently the final choice of the user. Unfortunately, the mathematical models proposed in the application of these theories [3,4] can just provide a partial solution to the issue. Namely, the mathematical models can have multiple attraction points at which the user's choice converge and any of them can be finally chosen as a consequence of little, unpredictable events happening during the process, as also discussed in the mentioned work of Marchesi et al. [36]. Thus, using the model for future final predictions can be inefficacious. Rather, Arthur [3] hints using this theory to understand the day-to-day process of user's choice. We embrace this view and we identify in the *urn model* – a specific path-dependent model expressing increasing returns – the most suitable mathematical tool to build the two models of technology use. This is because urn models are defined by clear quantitative rules that make their use simple in the case of quantitative data.

Finally, we apply our approach to the specific case of the adoption of OSS as a prototypical case of innovation. According to one of the most used definition of innovation [55], OSS is innovative in terms of its distribution and accessibility policy: users can freely download, modify, and use at home the same OSS application they use at work without any extra charge or break of the license. Unfortunately, this open policy has some drawbacks that hinder the assimilation of these technologies. For example, OSS often lacks of support and training that inevitably cause adverse attitude and resistance to its use [23,29]. Therefore, predicting, quantifying, and reducing the assimilation gap in the adoption of an OSS technology is of foremost importance for modern software companies that want to make their business with OSS.

For these reasons, we choose to validate our approach in the specific case of adoption of OpenOffice as replacement for the Microsoft Office suite. We analyse the transition in a company for a period of 24 months. The company prefers to remain anonymous, thus, for simplicity, in this paper we refer to as Softech.

Following our approach, we build urn models in two ways: (1) analysing the strategy of the company underlying the migration to OSS and (2) extracting automatically data on the files created with OpenOffice and Microsoft Office. In both cases, urn models estimate the user's choice of one application between two competing ones (Microsoft or OpenOffice) given the choice being determined by the proportion of files previously created with one application. In the former case, we identify urn models on the ideal proportion of files created assuming that the strategy is successfully implemented. In the latter case, we build urn models from a sample of 4,000,000 files created over 353 days with spreadsheets and text applications. With Monte Carlo simulations, we define measures of model proximity and rank models by their proximity with the real dataset. Findings validate our novel method showing that models constructed from objective data of use outperform in the ranking. In addition, results show that predicted and actual uses of the OSS technology differ significantly and consistently during the migration. As we have presented our results to the managers of Softech, we got confirmation of what happened. With our analysis, managers realized that there was an assimilation problem that was not resolving with time. With this feedback, managers implemented actions to reduce the gap and train the users to the new technology.

Altogether, we have proved that our method can be accurate in describing the reality of adoption processes beyond what can be perceived from a qualitative or subjective perspective. As such, we have provided managers with an evidence-based instrument for the decision making process toward innovation. In particular, as we have used evolutionary models, with our method we have estimated the assimilation gap over time providing timely feedback on what is effectively happening. We have helped managers to be more prompt and effective in recalibrating the adoption process.

Finally, our work contributes to the debate on the adoption of OSS. In fact, guiding the adoption of such technologies is a crucial and unexplored issue given the scarce documentation and information available. Our work contributes to the discussion with a case study that reports experience and solutions for recalibration of OSS adoption processes in the case of parallel use of proprietary and open competing technologies. Findings provide evidence not only on the adoption of the open technology itself, but also on the use of the competing proprietary one. In organizations, such case studies are rare as often the adoption is implemented in an asynchronous way, when a new technology completely replaces the existing one. To the best of our knowledge, analyzing the choice of use among technologies – miming the choice in the economic market – has not been studied yet.

The paper is organized as follows. In Section 2, we present the related works and their contribution to our study. Then, we discuss the proposed method (Section 3). In Section 4, we detail the organization where we run the case study and the data collected. In Section 5, we describe the application of our method. In Section 6, we analyse the resulting models. In Section 7, we discuss the limitations of this work. In Section 8, we draw some conclusions and we identify lines for future research. Appendix A gives more detailed information about the urn models used in the paper. Appendix B reports the complete ranking of the models by measure of proximity and type of application.

2. Related work

2.1. Models of Innovation

Research in innovation adoption is mainly qualitative [33]. It focuses on those factors that facilitate or inhibit the introduction of innovation in organizations both at primary [44] or secondary level [28] and analyses them with surveys, questionnaires, and interviews. This is the case of Cooper and Zmud [10] that have evaluated the usage of Material Resource Planning software in the context of innovative characteristics (task–technology compatibility, technical task complexity) by questionnaires to employees or of Lefebvre and Harvey [34] that have based their work on the opinion of the customer, the technological suppliers and the strategic motivations. Many others can be cited, but the majority of them have one point in common: they refer to the theory of the Technology Acceptance Model (TAM) [14], later extended in Davis et al. [15], Taylor and Todd [53], Thompson et al. [54], and Moore and Benbasat [37], and culminated in the Unified Theory of Acceptance and Use of Technology (UTAUT) in Venkatesh et al. [57]. TAM is built on four major perceived factors of technology acceptance: perceived ease of use, perceived usefulness, attitude toward use, and behavioural intention to use. Empirical research investigates the correlation of these factors with the actual use. Unfortunately, the implementations of TAM have reported little evidence of correlation of ease of use and usefulness with the actual use of the technology. This has been illustrated in the literature review performed in Turner et al. [56]. In particular, they have found that there is scarce empirical evidence on the correlation between the acceptance factors, like easy of use and usefulness of a technology, and the actual use of the technology.

The empirical works studying OSS adoption again focus on qualitative aspects and/or developer oriented issues. In 2006, a special issue of Management Science has illustrated new frontiers of research on OSS [59]: no research on adoption of OSS by end-users has been included. Few other articles have focused on traditional investigation of qualitative aspects like the strategic choice in primary adoption [16], individual motivations in secondary adoption [21], which compares the usability of StarOffice 5.2 with Microsoft Office 2000 or user acceptance models for OSS [27].

The quantitative approach we propose, aims at complementing the existing qualitative research providing also a quantitative instrument that compares evolutionary models of technology use. Specifically, our approach measures the “distance” between models of use predicted by the strategy and model of use resulting from the data. This would provide a tool for objective measurement that can be integrated with qualitative models in the decision making process to adopt a technology.

2.2. Models of file creation and objective measurement of use

There are various models of file creations. There are models based on static analysis of data collected at one point in time. For example, Doucher and Bolosky [18] perform a static analysis on file types by end-user job task over 140 millions of files in 4.801 client installations. Similarly, Satyanarayanan [48] analyse 86,000 files created by users of an academic network. Other models focus on evolutionary models where files’ occurrences are recorded and analysed over time. Vogels [58] monitors the evolution of file systems on Windows NT platform against file size, age, number of system read/write, and data throughput. Roselli et al. [45] analyse the dynamic evolution of file system modifications, recording type changes at runtime.

In this context, our work proposes an evolutionary model of file creations during the migration to OSS. The model both estimates and predicts the file creations of open and close technologies in parallel. Our work compares the actual evolution of software adoption with the predicted one shedding in addition some lights on the use of open data standards [49]. The evolutionary model we used is based on urn models, which are Markov chains defined on specific rules of transition typically used to describe the evolution of items in the economic market or in organizations. Such models have proven to be efficacious in describing the evolution of the choice of goods in the market. As such, we believe that these models well represent the choice of users among competing technologies in the settings of our case study.

Measuring objective use of a technology has been explored recently by few other studies. Objective measures are generally retrieved from logs of file use [11] or application access [19]. In our research, we propose to investigate the history of file creation to describe the process of use as the evolution of the user’s choice. This supplies a continuous model of adoption rather that a snapshot of the use in one point in time. In this way, managers have a thorough understanding of the whole adoption process.

2.3. Path dependence and urn models

Path dependence [12] is a characteristic of evolutionary processes for which “history matters.” The way it matters defines different concepts of dependence [38]. In common sense, path dependence means that current and future states/actions/decisions depend on the path of previous states/actions/decisions.

In stochastic processes that describe/predict the preference of users for competing software technologies – and in general for high-tech products that involves high degree of knowledge and low costs of production – increasing returns applies and can creates contagious effects. As such, increasing returns is equivalent to "the more you use the more you use."

Path dependent processes are typically modeled stochastic process whose future behaviour depends cumulatively on their history [1,13]. Urn models are specific path dependent processes, namely Markov chains, whose future depend only on the present state [24,31,38,40,46].

Literature in technological change claims that path dependence models and urn models play a key role in explaining evolutionary processes in the economic market and in organizations. They have been studied for their application to economics, political and social sciences [13,17,32,47]. For example, they have been used: in economics to characterize the emergence of a single clustering location of industrial companies [5,2], in the technology sector, to evaluate the pattern of selection of technologies [1] like the predominance of a given keyboard standard [12]), in medical sciences to model the assignment of patients to hospitals [30,60], in social sciences to understand the emergence of conformity in specific environmental settings or the network formation [51,61], and in politics to study the evolution of political institutions [40].

There are two major approaches in using path dependence models to describe technology adoption. The traditional approach explores their limit behaviour to predict expected equilibrium points, like a final configuration of the market share or the predominance of a given technology or product [4]. An alternative approach classifies them (e.g. [24]) in terms of their analytic expressions that model the evolution itself. The two approaches have different objectives: the former studies the stable configurations while the latter the various types of evolution of the adoption process. In the present study, we are interested in the evolutionary classification of the adoption and thus, we adopt the second perspective.

Urn models are specific examples of path dependent stochastic models. Urn models are Markov chains [6] that represent the process of selecting and adding balls to an urn. The probability to move to a future state of the Markov chain is given by the proportions of balls in the urn at present state.

In the following, we give an overview of urn models; readers already familiar with this subject can skip this section. Interested readers can get a further complete description of path dependence and urn models in [1,13,24].

Urn models can be defined with an arbitrary number of coloured balls. We describe the case of balls of two colours, as it is the model we use in our study. The extension to the general case of more colours can be found in [13].

Urn models are Markov chains, whose paths are initiated by considering an urn of infinite capacity that contains balls of two colours, say red and white. At the beginning, the urn contains w_1 red balls and w_2 white balls. At each step of the process, a ball is drawn and a certain number of balls are inserted in the urn. If the selected ball is red, then it is reinserted back in the urn together with other α red balls and β white balls. If the selected ball is white, then it is reinserted back in the urn together with other γ red balls and δ white balls. Thus, urn paths are specified by a matrix M that defines the numbers of balls added at each step after a selection (Table 1).

The urn process is then denoted by $U(\alpha, \beta, \gamma, \delta)$ or $U(w_1, w_2, \alpha, \beta, \gamma, \delta)$ depending on whether the initial urn conditions (w_1, w_2) are stressed. By convention, a negative entry value means that the balls of the corresponding colour are taken out of the urn. Thus, for example, if $\alpha = -2$ and a red ball is selected, then the selected red ball is re-inserted and two red balls are taken out of the urn.

Table 1
Entries of the urn matrix *M*.

	Number of red balls inserted	Number of white balls inserted
If red ball is selected	α	β
If white ball is selected	γ	δ

Without losing in generality, we can assume that β or γ in (Table 1) is non-negative; if they are not, we simply multiply the whole matrix by -1 reducing the discussion to the case in which only α or δ may be negative. An urn model with negative diagonal entries stops the process as soon as the number of balls to withdrawn exceeds the total number of balls of a given colour in the urn; an urn is called *tenable*, whenever one of its diagonal entries is less than minus one, the diagonal entry divides its corresponding initial value and the other entry belonging to the same column. In this case, there is a sequence of selections of balls that does not stop the process [24]. For the rest of the paper we consider tenable urn models.

The outcome of the Markov chain at step n is the number of balls of the two colours in the urn. If after n steps, there have been n trials with k selections of red balls and $(n-k)$ selections of white balls., then the frequencies of red balls $(W_{\text{red}}(n))$ and of white balls $(W_{\text{white}}(n))$ in the urn at n are:

$$W_{\text{white}}(n) = \frac{w_2 + \delta(n-k) + \beta k}{w_1 + w_2 + k(\alpha+\beta) + (n-k)(\delta+\gamma)}$$

$$W_{\text{red}}(n) = \frac{w_1 + \alpha k + \gamma(n-k)}{w_1 + w_2 + k(\alpha+\beta) + (n-k)(\delta+\gamma)}$$

Frequencies $W_{\text{red}}(n)$ and $W_{\text{white}}(n)$ belong to the interval $[0, 1]$ and $W_{\text{red}}(n) + W_{\text{white}}(n) = 1$. Tenable urns can reach extreme values in $[0, 1]$. For example, a tenable urn with $\alpha < -1$ has no red balls after n steps with $n = -w_1/\alpha$. Thus, the selection of a red ball is not possible anymore. The process continues with the selection of white balls only. The process is similar if $\delta < -1$.

By definition, the Markov chain is completely described by the transition rules defined by the so-called urn functions:

$f_{\text{red}}(x) = \Pr\{$the selected ball is red at observation $n+1 | W_{\text{red}}(n) = x\}$
$f_{\text{white}}(x) = \Pr\{$the selected ball is white at observation $n+1 | W_{\text{white}}(n) = x\}$

This means that the probability of selecting a ball of a given colour at $n+1$ depends on the frequency of balls of that colour in the urn at n. Therefore, from the expressions of the frequencies, the urn functions depend on the matrix M. This indicates that the user's future choice depends on the frequencies W and the configuration of the urn at time "minus one." The simplest case of urn functions is when $f_{\text{red}}(x) = f_{\text{white}}(x) = x$, the probability that a user selects a ball of a given colour at $n+1$ is the frequency of that colour in the urn at n. In general, the urn function may be non linear [5] and the points x_0 at which an urn function $f(x_0) = x_0$ are called equilibrium points. The equilibrium points are the points at which the urn model settles down – in statistical terms, the ones to which the urn model converges with probability one. Equilibrium points are those values for which the future choice of the user is purely determined by the use of a technology, in that the choice is determined by no other factors but the frequency of use. Among them, there are the attracting points, which are the candidate outcomes of the user's choice in the long run [1]. As such, attracting point are points toward which the user's choice converges.

A major goal of modeling with urn is to determine the expression of M for a given context. In our case, we derive it from data. Our method starts with the analysis of irreducible balanced urn models [24]. Flajolet et al. [24] classify irreducible balanced urn models by their analytic expression. We briefly overview the classification in the following.

Irreducible urn models have the great common divisor of the M entries equal to one. Any non-irreducible matrix is the multiplication of an irreducible matrix by a scalar and the scalar represents the overall magnitude of the balls additions. Irreducible balanced urn models are irreducible models in which the row sums of the matrix M are constant (σ):

$$\alpha + \beta = \delta + \gamma = \sigma$$

Therefore, at any time n, the number of balls in the urn is a deterministic quantity:

$$N_n = w_1 + w_2 + n\sigma$$

This property makes balanced urn models solutions of ordinary differential equation systems [24] and as such, it makes them completely determined. Balanced urn models are classified by the dissimilarity index p:

$$p = \gamma - \alpha = \beta - \delta$$

in altruistic, neutral, or selfish (Table 2).

Flajolet et al., found that balanced urn models are associated to an ordinary differential system with a well-known explicit solution completely described by the dissimilarity index p. Examples of irreducible balanced urn models are illustrated in Table 3.

2.4. Monte Carlo simulations to visualize urn paths

From the definition of urn model, a path of an urn model is a set of selections and additions of balls from the urn in a sequence of steps. Starting from the same initial values (w_1, w_2), we can draw various possible paths depending on the matrix M and the probability of selection of a ball of a given colour. These paths are called random walks. To sample different random walks, we use the method of Monte Carlo and the uniform probability distribution for the simulation of the selection of a ball. Each simulation starts from the same initial number of balls (w_1, w_2). At each step, a ball is selected randomly (using uniform distribution). Then the urn rule is applied according to the matrix M. Fig. 1 illustrates one hundred simulated paths for a Pólya model with U(1100, 1, 0, 0, 1) (Fig. 1a) and a Friedman model with U(1100, 0, 1, 1, 0) (Fig. 1b). In both graphs, the x-axis represents the number of steps, the y-axis the proportion of the frequencies $W_{red}/W_{white+red}$ at each step. Fig. 1 also reveals a typical difference between Pólya and Friedman's urn models: a Pólya's model statistically converges to a random variable [17] whereas a Friedman's model to the constant value ½ [25]. This indicates that the Friedman's model statistically convergence to the equilibrium point of equal number of red and white balls.

Our original dataset represents the path described by the actual additions to the urn that follow the real preference of the user. Simulated random walks can be compared with the actual path by measures of proximity (Fig. 4). The comparison determines a ranking of the urn models in that the more the simulated random walks approach the real data the higher the urn model is ranked (Section 6.2). We use the ranking to validate our constructive method to build the urn model from objective data of technology use as the top ranked models coincide with the models we derive by regression of the data (Section 5.2).

Table 2
Classification of the balanced urns by dissimilarity index p.

Name	Description
Altruistic urn	$p > 0$ the ball chosen determines more new balls of the opposite colour
Neutral urn	$p = 0$ the ball chosen determines the same number of new balls in both colours
Selfish urn	$p < 0$ the ball chosen determines more new balls of its same colour

3. The method

Our method starts identifying the matrices M that model predicted and effective use of the two office suites. For predicted use, we investigate the strategy of the organization (Section 5.1) and we derive the matrix of what we call *abstract models*. For effective use, we apply statistical correlations and linear regression on the distributions of the amount of files created in each format (OpenOffice or Microsoft Office) and we derive the matrix of what we call *empirical models* (Section 5.3). Then, we detect an assimilation gap by comparing the designs of matrix M of abstract and empirical models of the same type of application (Word processor or Spreadsheet application). High discrepancy between the designs indicates high assimilation gap.

Finally, to validate the procedure that determines M for the empirical models, we define two measures of proximity (Section 6). With these two measures, we rank abstract, empirical and known urn models. The resulting rankings also give a quantitative value of the assimilation gap in that they measure the distance from matrices of empirical models. The overall method is illustrated in Fig. 2.

If the matrices M of abstract and empirical models coincide then the assimilation gap is null. The strategy has been successfully implemented. In this case, the ranking given by the proximity measures, identifies the best abstract model that also represents the actual use. If this does not happen, the ranking quantifies the gap between the abstract and empirical models. In particular, if all the empirical models outperform in the ranking, our constructive method (Section 5.3) is validated as the matrices M of the empirical models have the nearest random walks to the actual path.

Table 3
Most known urn models. Source: [24].

Name	Model	P	Meaning and example of use of the model
The contagion model (Pólya)	U(1, 0, 0, 1)	$p < 0$	*Meaning*: evolution of a contagion process *Example*: a population with two kinds of genes that randomly comes into contact with external entities to which their gene is transmitted [20]
The adverse-campaign model (Friedman)	U(0, 1, 1, 0)	$p > 0$	*Meaning*: evolution with negative feedback *Example*: a propaganda campaign in which candidates are so bad that people that listen to them will vote the opposite candidate [25,26]
The records urn	U(1, 0, 1, 0)	$p = 0$	*Meaning:* The number of records in a random permutation *Example:* The "Chinese Restaurant Problem" [41]

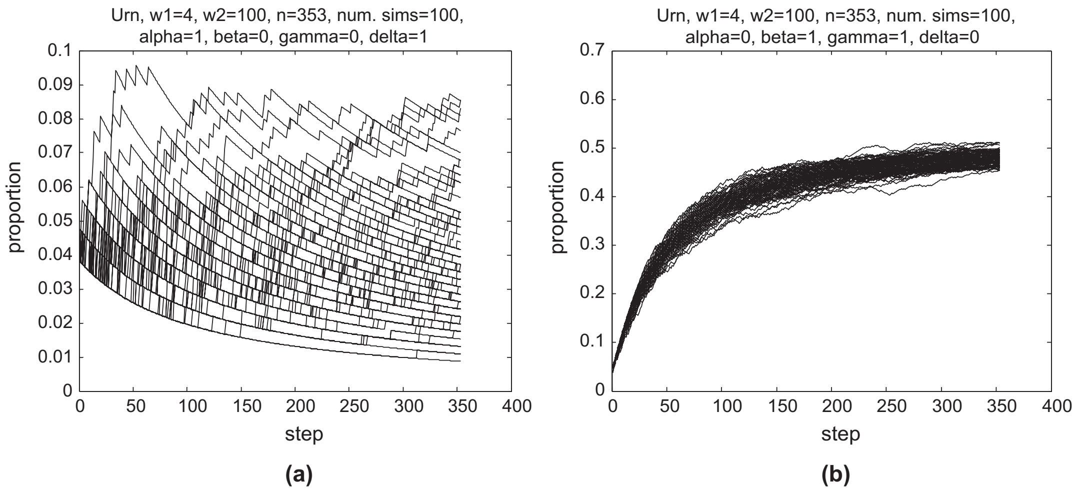

Fig. 1. (a) 100 simulations of a Pólya's model, b) 100 simulations of the a Friedman's model.

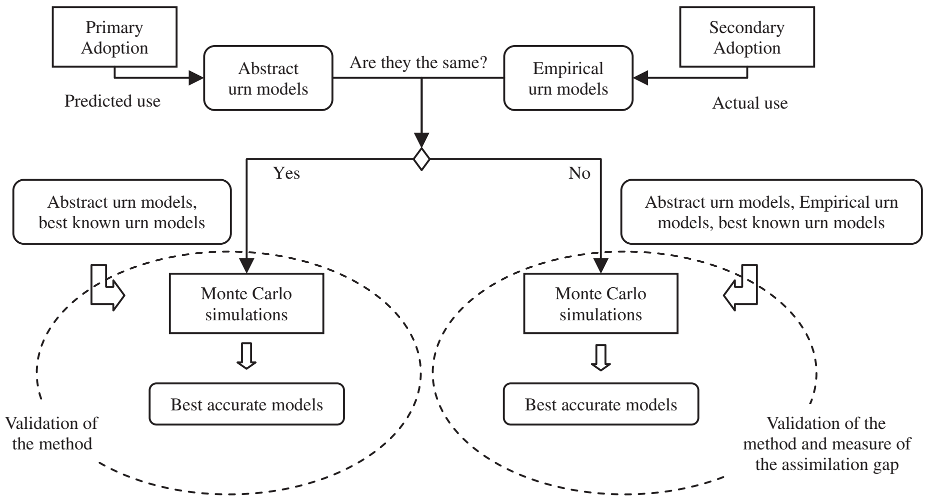

Fig. 2. The method for measuring the assimilation gap.

3.1. Modeling the use of Microsoft Office and OpenOffice with urn models

As mentioned, the first objective of this work is modeling predicted and the actual use of Microsoft Office and OpenOffice during the migration to OpenOffice. The selection of one ball corresponds to choice to use one of the two office suites, the newly adopted OpenOffice or the existing Microsoft Office. In our case, each ball colour corresponds to one office suite and each step corresponds to a day in the migration period. The preference toward one office suite is determined by the probability of the function urns at day n:

$f_{\text{OpenOffice.org}}(x) = \Pr\{\text{to use OpenOffice at } n+1 \mid W_{\text{OpenOffice.org}}(n) = x\}$
$f_{\text{Microsoft Office}}(x) = \Pr\{\text{to use Microsoft Office at } n+1 \mid W_{\text{Microsoft Office}}(n) = x\}$.

To quantify the frequencies $W_{\text{OpenOffice.org}}(n)$ and $W_{\text{Microsoft Office}}(n)$, we determine the number of files created in the native format of that application at day *n*. Therefore, our urn contains files in two data standards, the Microsoft and the OpenOffice one. By definition of the urn functions, the future preference toward a specific office suite depends on its percentage of files present in the urn. We need to anticipate though that in general settings, the urn functions overestimate the probability to use Microsoft Office, as OpenOffice can create and edit files in the native format of Microsoft Office. On the other hand, they perfectly represent the probability of use of data standards – Pr(to use OpenOffice data standard) or Pr(to use Office data standard). We need to say, though, that creation of files well represented application use in our case study. We will discuss this in Section 7.

Table 4 illustrates the urn matrix M in our case:

- The first column of M indicates the OpenOffice files added to the urn respectively after selecting OpenOffice (α) or Microsoft Office (γ);

Table 4
The urn matrix *M* for the generation of files.

	OpenOffice.org	Microsoft Office
OpenOffice.org	α	β
Microsoft Office	γ	δ

- The second column indicates the Microsoft Office files added respectively after selecting OpenOffice (β) or Microsoft Office (δ);
- The first row defines the urn rule once the OpenOffice has been selected;
- The second row defines the urn rule once the Microsoft Office has been selected.

In our settings, the values of β and δ indicate the levels of maturity of secondary adoption as they express the addition of Microsoft Office files. We define mature an adoption process whose matrix M has zero or negative values for β and δ. For instance, converting an existing document from Microsoft Office to OpenOffice.org is a sign of mature process. In terms of M, converting files removes balls of Microsoft Office format decreasing δ – and add balls of OpenOffice format increasing γ. On the contrary, β or δ may be positive if the adoption has just started or it is not (yet) completely mature. In a not-yet mature migration to OpenOffice, users keep on creating files with Microsoft Office and new balls of Microsoft Office format are inserted in the urn. As such, positive values of β or δ can explain a learning curve, an initial resistance to change, or the need of exchanging data with (external) users. In particular, not-skilled end-users might feel safer to create files in a format they already know and they prefer to create Microsoft Office files anytime they fill to be under time pressure.

The entries of M can also tell the success of a secondary adoption. A successful migration to OpenOffice requires $\alpha > 0$. Thus, a migration is a failure if $\alpha < 0$. It is a mild success if $\alpha = \delta$. This is the case of the Pólya urn ($\alpha = \delta = 1$), according to which the end-users split in exactly two groups, each preferring only one office suite. In this case, files are added to the urn only in the format of the selected application and the probability of one application to dominate the other tends to be perfectly even. A migration is a complete success if $\beta = 0$, $\delta \leqslant 0$ and $\gamma \geqslant 0$, which means that the end-users and the surrounding environment (if $\delta = 0$) is ready to jump to, and use only OpenOffice.org. Note that having $\delta < 0$ might mean that there is some activity to delete Microsoft Office files – as we mentioned – and the adoption although successful is not completely mature.

In organizations where the daily activity has a constant pace, the matrix M can be also balanced, as it does not matter which office application is selected to determine the number of total daily additions to the urn (the dissimilarity index p).

4. The case study

Our case study concerns Softech, a public software company that provides and maintains the IT infrastructure of a large number of public councils (hundred sixteen municipalities, eight social service offices and about thirty offices for social support) that are its associates. It has about sixty employees that assist about 2500 employees working in the associates. Ten Softech employees are developers, forty are IT experts coordinating the municipalities' IT departments and maintaining about 3500 PCs, 180 servers, routers and switches. This large and complex structure of the municipalities needs a solid strategy that defines standardization of procedures, coordination with councils, and alignment of underlying IT infrastructure, all in respect of the local autonomy of each associate. Thus, the major mission of Softech is creating and maintaining standards and quality of the services across its associates. The management of Softech had four strategic reasons for experimenting with OSS and migrating from proprietary products:

- To limit public expenditure. No financial constraints were involved and there was no urgency to reduce the costs of software. However, the money saved by using OSS could be used for other purposes. The change also had an ethical dimension: sparing public money represented efficient management of the "res publica".
- To be independent from the external vendors. Most of the consortium's municipalities are unwilling to spend much on updates for their desktop computers and servers. Softech has chosen solutions that do not require frequent updates and has adapted tools to run on old hardware.
- To give citizens the possibility of using open formats and software.
- To adhere to the government's promotion of open source software and standards.

To achieve these strategic objectives, the IT management allocated time to searching and analyzing alternative OSS solutions. The migration to OpenOffice initially caused strong resistance from the councils' Mayors. This opposition was exacerbated by the first pilot migration that failed. Softech learned from this experience. It revised its deployment plan, adding an initial training course and hiring a specialist.

The approach taken for the migration started with an incremental parallel adoption of OpenOffice, to minimize resistance. Thus, OpenOffice was installed in the PCs of the users without uninstalling Microsoft Office. A training of about 20 h was administered to all the employees. In total, about 2500 employees were given training over a 2 months period. Moreover, the migration was further facilitated by the fact that all the existing official document templates written in Microsoft Word were translated into OpenOffice ones by Softech; in particular, additional support was provided to translate macros of Microsoft Office to OpenOffice.org. The transition started at the beginnings of 2003. As of February 2005, OpenOffice was installed on more than 2500 PC and it ran in parallel with Microsoft Office; end-users could freely choose between the Microsoft Office and OpenOffice (parallel adoption). It is worth noticing that the choice of the employees was different from the typical choice that users in the free economic market have. Although employees had the choice to use the application they preferred, they were not in fact so free to choose it. They had to follow the directives defined by the Mayor of the council and exchange files in open format with Softech and other councils already migrated to OpenOffice. File exchange in proprietary format (intranet) became harder and harder with the increase of the number of councils that joined the initiative.

When we interviewed Softech's managers, we discovered that

(1) The migration was performed council by council and was preceded by a long negotiation with the Mayor of the council. The major reason of Mayor's resistance was the image of the council in the association. For this reason, Softech's managers met personally with each Mayor to illustrate the adoption plan, strategy, and directives.
(2) Employees had ad hoc training on office features implemented in the open technology. This type of training was never given with Microsoft Office and employees perceived it as a special reward. In addition, Softech converted but also restyled all the templates in the open format. Again, employees perceived it beneficial.
(3) Many of the councils have no employee expert in IT. Skills and knowledge of employees were limited to some basic office features. The majority of the employees was not even aware of the possibility to open and save Microsoft Office documents with OpenOffice. Intentionally, Softech did not instruct employees on this feature to avoid the creation of files in proprietary format. As employees could still use Microsoft Office suite to read files in the closed format (parallel adoption), managers believed that this information was of no critical

importance. Of course, with time and more stable versions of OpenOffice, some of the employees knew about this feature, but at the end of our analysis in January 2005, we found that only 30.3% of the employees (447 out of 1475) used this feature at least once in twenty working days.

(4) The use of Internet browsers was very limited. This implied that employees rarely downloaded files from external sources.

4.1. Data collection

We collected date of creation and daily amount of files of Word, Excel, Writer and Calc applications with an automated tool we developed[1]. We did not consider PowerPoint and Impress, as we collected too scarce data, not enough to identify a significant sample. We neither consider files that were only modified as we already collected a large amount of data. Modifications would have given a further understanding of the adoption process, but analysing them would have been too time-consuming for this initial exploration of the effectiveness of our method.

As we monitor the use of the office suites for a period of over 2 years, we could select about 1 year (353 observations) in which the transition to OpenOffice was enough mature and the use of the application enough spread among councils. This period spans from 30th September 2003 to 16th September 2004. Given the complex structure of the councils' association, the migration to OpenOffice did not complete in a short time. Actually, in 2011, the migration is still on-going in some of the councils and not all the councils have decided to migrate yet. Anytime, a council can decide to migrate to OpenOffice and new independent information enters the dataset changing its configuration. For this reason, we cannot really tell upfront whether the period we selected can significantly describe the whole evolution of the assimilation process in the councils' association. At any rate, we do not intend to provide an overall model of assimilation gap, but rather illustrate and validate a method to measure the assimilation gap. In particular, we have compared our results with the analysis of the time spent using each application (Section 7).

For the analysis, we used network shares where more than 2000 users have created and modified files related to their daily activities. We gathered information about 5,409,689 files created by different applications. We considered 4,350,239 files created by text processors and spreadsheet applications. Files have been created by the employees of the councils that had migrated or were migrating to OpenOffice at the time of our analysis (Table 5). The assimilation can be completed in some of the authorities but not in all. As we measured the file creations as a whole over all the authorities that migrated to OpenOffice the analysis in subsequent years might not guarantee a finer/different result.

Table 6 details the descriptive statistics of the datasets. The last row indicates the initial amounts of files in the database. These will be used pair wise to define the initial values (w_1, w_2) of the urn model.

5. Derived urn models

5.1. Urn models of actual use forecasted in primary adoption

In this section, we discuss the matrix of the urn models derived from primary adoption as if no assimilation gap existed. These models are model of use of an application in which the strategic view is perfectly implemented. As we said, we call them *abstract*. In our case study, the strategy supports the adoption of OpenOffice. As such, abstract models need to represent a successful adoption (at least $\alpha > 0$). The maturity of the adoption may vary instead. We briefly discuss this in the following.

Rogers present two major strategies in adopting innovation, the *Authority Innovation-Decision and the Optional Innovation Decision* [43]. An Authority Innovation Decision is made by few individuals in position of influence and has the goal to push the use of the new technology by providing precise directives. Opposite to the Authority Innovation Decision, Rogers defines the optional innovation decision, in which the use of a technology is an individual choice. In this case, the management provides the new technology – and even can support it – but it does not implement a set of directives to impose it. As no directives are implemented, the process of adoption can slow down by the effects of the adverse attitude of the end-users and the time pressure [10].

In both strategies, specific directives, personal characteristics, and environmental settings may have direct effects on file creation with the old application and, as such, on the level of maturity of the adoption process. In our case, we have isolated four major causes: exchange of documents with external organizations, conversion of data standards, adverse attitude of the end-users, and time pressure for task delivery. Table 7 illustrates the impact of these factors on the value of β and δ. The sign refers to the positive (+) or the negative (−) file additions.

Exchange of documents with third parties constraints a company to create files in proprietary format. This increases the number of Microsoft Office files (values of β or δ). This can happen both in the authoritative or optional innovation decision. The massive conversion to the new data standard is mainly an authoritative decision that decreases the value of δ. The adverse attitude toward the new technology OSS has a significant effect in the optional innovation decision as users decide to use a technology on their experience.

Table 5
Summary of the analyzed data standards.

	Application	Data standard	Files creations
Microsoft	Word	.doc	3,700,208
OpenOffice.org	Writer	.sxw	130,904
Microsoft	Excel	.xls	510,524
OpenOffice.org	Calc	.sxc	8603

Table 6
Descriptive statistics of daily files creations.

	.sxw	.doc	.sxc	.xls
N cases	353	353	353	353
Minimum	0	49	0	1
Maximum	3748	89,630	195	3175
Initial values (w)	8733	21,7470	27	732

Table 7
Factors influencing the creation of files with the old application in abstract models.

	β	δ
Authority innovation-decision	+Exchange of documents	+Exchange of documents −Conversion of data standards
Optional innovation decision	+Exchange of documents +Adverse attitude	+Exchange of documents +Adverse attitude +Time pressure

[1] FLEA. FiLe Extension Analyser It is a Windows application which scans disk drives and catalogues all the files found according to their extension, e.g. .doc, .xls, .sxw, etc. FLEA records in a log file extension, size, and date of creation of each file found.

The confidence with the previous technology and the freedom of choice induces some resistance of the users toward the new technology increasing the value of β or δ. Finally, the pressure due to time constraints induces the users to keep using the old application with which they feel to be faster. This increases the value of δ.

In our case study, Softech have adopted a hybrid approach. The management followed an Authority Innovation Decision strategy with the definition of clear directives pushing the use of OSS and implemented an optional innovation decision letting users (apparently) free to choose the office application they prefer. This approach was crucial to mitigate any resistance of the users, as Softech does not have a direct authority on some of the councils, which keep their autonomy in the choice of their IT infrastructure. Managers have reported that adopting only one of the two strategies would have resulted in a complete failure of the migration to OpenOffice. At the end of the migration, Microsoft Office was uninstalled and no other option was possible anymore.

Softech's directives during the migration were the following ones:

(1) Substitution of the proprietary templates with the open source ones (massive conversion of data standards).
(2) Creation of documents in proprietary format only for exchange with external parties or on specific request of other councils.
(3) Creation of new documents only with OpenOffice.

Softech's directives imply that no new Microsoft Office document can be created except for exchange purposes ($\beta = 0$). In other words, if the choice is OpenOffice no new Microsoft Office files can be created and if the choice is Microsoft Office the files created in this format are only for the purpose of exchange with third parties. In addition, the effect of the document exchange (positive δ) – that was kept limited – is absorbed by the conversion into the new format (negative δ) – as existing documents in Microsoft Office will be massively substituted with OpenOffice ones. Thus, we may consider that the predicted use of the two office suites has M with β zero and δ negative. Consequently, no balanced selfish urn model can describe the predicted use (as $p = \beta - \delta > 0$). Managers did not implement any further action to limit time pressure or adverse attitude. As such the simplest irreducible models that can represent the Softech strategy in the migration toward OpenOffice are:

The balanced urn models in (a) and (b) of Table 8 represent completely successful adoptions. In (a), the effect of "Conversion of data standards" is visible: the matrix is altruistic ($p > 0$) and has negative entry δ. Therefore, the process of secondary adoption is not completely mature. In (b) (the "records urn"), these effects disappear and the process is mature. The urn model in (c) is an example of non-balanced but completely successful secondary adoption. Again, in this case the process is not completely mature as δ is negative.

5.2. The opponent model

To complete the set theoretical models, we consider the new urn model $U(-1, 1, 0, 1)$ that expresses the adverse attitude toward OpenOffice.org. We call it the *opponent model*. The opponent model represents the failed migration to OpenOffice ($\alpha < 0$). In principle, it is not derived from the management's strategy.

Table 8
Abstract urn model predicted from primary adoption in our case study.

Balanced altruistic	Balanced neutral	Non-balanced
$\begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$
(a)	(b)	(c)

5.3. Urn models of actual use in secondary adoption

In this section, we propose a new method to derive the matrix M from the data of use of two similar technologies. The method is constructive and produces the matrix of urn from the data. We further confirm the validity of our method ranking all the urn models we have introduced by measures of proximity with our dataset (Section 6). If the urn matrices are representative of our data, they will also outperform in this ranking.

χ and Φ are the two time series describing the daily number of files that have been created respectively with the Microsoft Office and OpenOffice data format and Ω is $\chi + \Phi$. We set $\chi_{-1}(t) = \chi(t-1)$ and $\Phi_{-1}(t) = \Phi(t-1)$ and we define the values of the matrix $M = (\alpha, \beta, \delta, \gamma)$ by the round up of the angular coefficient m of the regression line between χ_{-1} and Φ or χ and between Φ_{-1} and Φ or χ (Table 9).

To generalize the definition of M to the population of the two time series, we test the entries for significance with the False Discovery Rate (FDR) correction for multiple tests [7]. In this way, the test guarantees that the four values are significant simultaneously.

In the following sections, we apply this method to two different data sets of word processors and spreadsheet applications.

5.3.1. Microsoft word and writer

χ and Φ are the two random variables of the files created with Microsoft (.doc) and OpenOffice documents (.sxw). A log transformation of χ, and Φ evidences that the two distributions are both bimodal and they both split into two sub-distributions, one of low frequencies and one of high frequencies (Fig. 3, Tables 10 and 11). Thus, we apply our method to the whole dataset and to the two sub-distributions. As, the matrix derived from the dataset of low frequencies has not passed the significance test, we exclude it from our analysis.

Table 12–14 illustrate the result of our method applied to the two datasets.

Following our definition, the matrices in Table 14 represent irreducible and non-balanced urn models of a mild successful and not mature secondary adoption. In particular, $\gamma = 0$, and $\alpha = \delta > 0$ indicate parallel independent use of the two applications. As such, the use of one single application is not predominant and there is no conversion of files into the open standard. Furthermore, the positive values of β indicate the creations of Microsoft files while selecting OpenOffice. This might reflect the need of exchanging files with other organizations or use of the old technology for time pressure. In this case, OpenOffice might have been used to save files in the Microsoft format. In our case study, at the end of our observation period, we found that only about 30% of the users have ever used this feature of OpenOffice. As such, we are enough confident that a significant part of the files we analysed were created in the native application.

There is an interesting difference between the two matrices in Table 14. When the job activity is more intense (sub-distribution

Table 9
Definition of the urn matrix M via linear regression and round up ([.]).

$\alpha = [m(\Phi_{-1}, \Phi)]$
$\beta = [m(\Phi_{-1}, \chi)]$
$\delta = [m(\chi_{-1}, \chi)]$
$\gamma = [m(\chi_{-1}, \Phi)]$

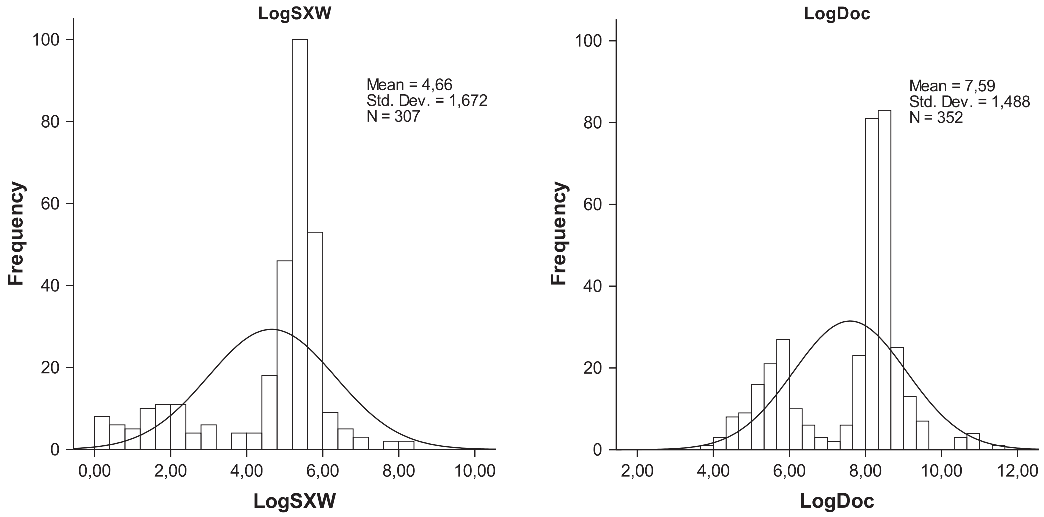

Fig. 3. Log transformation of the data for word processors.

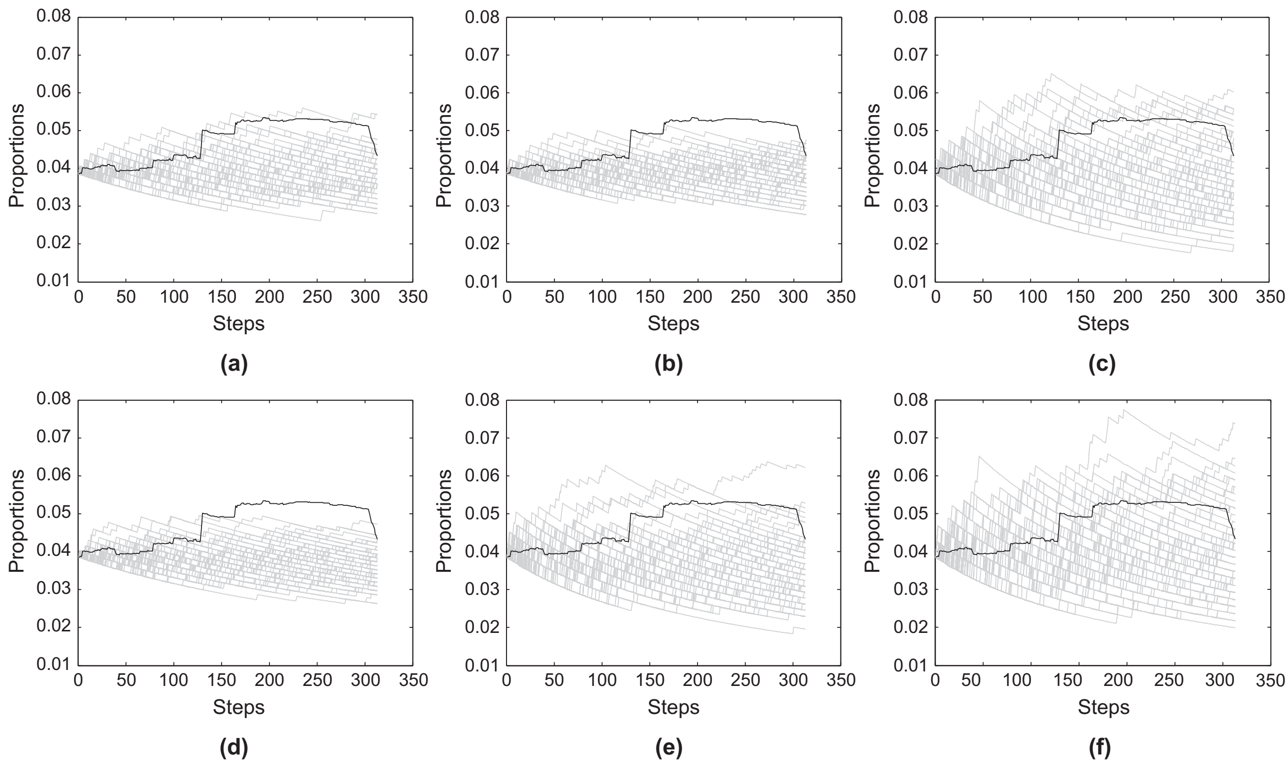

Fig. 4. 100 simulations of different urn models: (a) U(500, 0, 0, 500), (b) U(500, 1000, 0, 500), (c) U(1000, 2000, 0, 1000), (d) U(500, 2000, 0, 500), (e) U(1000, 4000, 0, 1000), (f) U(1000, 0, 0, 1000).

with high number of file creations) the value of β reduces to two (right matrix in Table 14) and the creation of Microsoft Office files decreases. In other words, the more the office activity is intense the less the Microsoft Office files are created.

Table 10
Descriptive analysis of sub-distribution with high number of creations.

Data standard	Lower bound	Upper bound	*N*
.sxw	41	1200	242
.doc	1556	15207	238
Total	1647	15214	238

Table 11
Descriptive analysis of sub-distribution with low number of creations.

Data standard	Lower bound	Upper bound	*N*
.sxw	0	40	107
.doc	49	1555	107
Total	49	1647	107

Table 12
Spearman correlations and entries of M – the whole distribution.

$\rho(\Phi_{-1}, \Phi) = 0.419^{**}$	$\alpha = m(\Phi_{-1}, \Phi) = [0.25] \approx \mathbf{1}$
$\rho(\Phi_{-1}, \chi) = 0.418^{**}$	$\beta = m(\Phi_{-1}, \chi) = [3.76] \approx \mathbf{4}$
$\rho(\chi_{-1}, \chi) = 0.425^{**}$	$\delta = m(\chi_{-1}, \chi) = [0.51] \approx \mathbf{1}$
$\rho(\chi_{-1}, \Phi) = 0.82^{**}$	$\gamma = m(\chi_{-1}, \Phi) = [0.00] \approx \mathbf{0}$

** Two tails, $\alpha = 0.05$ and FDR = 0.031.

Table 13
Spearman correlations and entries of M – sub-distribution of high number of creations.

$\rho(\Phi_{-1}, \chi) = 0.313^{**}$	$\alpha = m(\Phi_{-1}, \Phi) = [0.48] \approx \mathbf{1}$
$\rho(\Phi_{-1}, \Phi) = 0.507^{**}$	$\beta = m(\Phi_{-1}, \chi) = [1.71] \approx \mathbf{2}$
$\rho(\chi_{-1}, \chi) = 0.192^{**}$	$\delta = m(\chi_{-1}, \chi) = [0.27] \approx \mathbf{1}$
$\rho(\chi_{-1}, \Phi) = 0.263^{**}$	$\gamma = m(\chi_{-1}, \Phi) = [0.0] \approx \mathbf{0}$

** Two tails, $\alpha = 0.05$ and FDR = 0.031.

Table 14
Empirical urn rules for word processors.

The whole dataset	High number of file creations
$M \approx \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}$	$M \approx \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$

5.3.2. Microsoft excel and calc

In this section, we replicate the application of our method to Microsoft Excel and OpenOffice Calc files to excluding the bias related to the specific type of application (i.e. word processors).

Now, χ and Φ are respectively the distributions of files created with Microsoft excel (.xls) and OpenOffice Calc (.sxc). The relatively limited amount of data does not indicate any sub-distribution. Thus, we consider only the distribution of the whole dataset. As before, we derive the matrix M by computing the regression slope (Table 15).

Thus, the urn matrix M is:

$$M \approx \begin{pmatrix} 1 & 10 \\ 0 & 1 \end{pmatrix}$$

Again, M is an irreducible upper triangular matrix. Its urn model shows a mild increment of the OpenOffice Calc files while increasing the creation of Microsoft Excel files. The urn model is irreducible and non-balanced. As for the word processors, it represents a mild successful and not mature secondary adoption.

6. Analysis of the resulting models

In the previous sections, we have introduced various types of urn models *abstract*, *empirical, opponent* and three well-known *balanced models* for each choice of the sign of the dissimilarity index *p*. In this section, we discuss which of the models best fits our data. By construction, we expect empirical models to perform better. If this is true, our constructive method of secondary adoption is validated.

Table 15
Spearman correlations and regression slopes.

$\rho(\Phi_{-1}, \chi) = 0.325^{**}$	$\alpha = m(\Phi_{-1}, \Phi) = [0.41] \approx 1$
$\rho(\Phi_{-1}, \Phi) = 0.414^{**}$	$\beta = m(\Phi_{-1}, \chi) = [9.65] \approx 10$
$\rho(\chi_{-1}, \chi) = 0.348^{**}$	$\delta = m(\chi_{-1}, \chi) = [0.45] \approx 1$
$\rho(\chi_{-1}, \Phi) = 0.390^{**}$	$\gamma = m(\chi_{-1}, \Phi) = [0.01] \approx 0$

** Two tails, $\alpha = 0.05$ and FDR = 0.031.

With the same analysis, we can also determine the magnitude of the additions (Section 1). All the above urn models are irreducible, that is Greatest Common Divisor (GCD) of α, β, γ, δ equals to one, and they represent the type of rule, but not the magnitude of the additions to the urn. To capture the average amount of files added to the urn we also consider non-irreducible models derived from the irreducible ones by multiplication of a scalar a (low magnitude) or b (high magnitude) with $b > a$. The values of a and b are the same for all the models. For models with $\alpha = -1$ we request that a and b divide w_1 for the ones with $\delta = -1$ we request that a and b divide w_2. We have also considered a and b with values 500 and 1000 to understand extreme behaviours. All the models are summarized in Tables 21 and 22 of Appendix A.

To compare the various models, we introduce two measures of proximity: the *nearest simulation distance* and the *coverage of fit*. With a Monte Carlo simulation, we generate one hundred paths for each urn model. For each simulation, we identify the maximal point distance with the dataset:

$$\max(j) = \max_i\{|s_j^i - o_j|\}$$

where s_j^i is the jth simulation and o_i is the observation at time i.

We then take the minimum value of max(j):

$$m = \min_j\{\max(j)\}$$

A band of width μ around the time series defined by the dataset includes the nearest simulations among the one hundred considered per urn model. We compare the urn models by the values of μ. The best urn model will have at least one simulation, which is the nearest curve among all the simulations of all the urn models. The second measure defines the proximity of the urn models for coverage of fit. We define the discrete area for one hundred simulations $\{s_i^j\}$ of a given urn model:

$$A = \{(i, y) : i\text{th observation},\ \min_j\{s_i^j\} \leqslant y \leqslant \max_j\{s_i^j\}\}$$

We define the coverage of fit as the percentage, %, of original data values that fall in the area A.

6.1. Proximity of the urn models for Microsoft word and Openoffice writer

Tables 16 and 17 show the top six urn models of the ranking according to the nearest simulation distance and the coverage of

Table 16
Ranking by nearest simulation distance. Top 6 models.

Type	Urn model (w_1 = 8733, w_2 = 217470)	μ
P	U(500, 0, 0, 500)	0.006605
E1	U(500, 1000, 0, 500)	0.007652
E1	U(1000, 2000, 0, 1000)	0.007823
E2	U(500, 2000, 0, 500)	0.008032
E2	U(1000, 4000, 0, 1000)	0.008179
P	U(1000, 0, 0, 1000)	0.008193

Table 17
Ranking by coverage of fit. Top 6 models.

Type	Urn model (w_1 = 8733, w_2 = 217470)	%
E1	U(1000, 2000, 0, 1000)	100
P	U(1000, 0, 0, 1000)	99.68
P	U(500, 0, 0, 500)	97.12
E2	U(1000, 4000, 0, 1000)	65.17
E1	U(500, 1000, 0, 500)	43.13
E2	U(500, 2000, 0, 500)	42.17

fit; the complete rankings are included in Appendix B (Tables 23 and 24).

Tables 16 and 17 display type, design, and measure value of the urn models. Types (abstract, empirical, opponent, and well-known balanced models) are defined in Appendix A. In particular, E stands for empirical model, P for Pólya model. The index indicates different matrix designs of the same type. For example, E1 is the empirical model defined by high number of file creations (Table 16). The top six models in Tables 16 and 17 have only three designs, P, E1, and E2 and have GCD greater than 1. In particular, they include all the designs of the empirical models (E1 and E2). Looking at the complete list in Appendix A, we can see that there is a significant jump in the value of μ and% for the remaining designs and types, O, F, C1, C2, and C3 of GCD 500 or 1000. This is a strong validation of our method, considering that the two measures capture two complementary characteristic of proximity and that data has been tested on a set of 28 models. This result confirms the difference between empirical and abstract models and suggests a magnitude of addition greater than one (GCD greater than one). In particular, all the empirical models are not mature and they are mild successful ($\alpha > 0$). Among the best models of Tables 16 and 17, only the Pólya model is not empirical. Nevertheless, it is the most successful and mature, but also deterministic [24]. This indicates that, in the most optimistic case, users create files of the same format for the whole period; in particular, users of OpenOffice Writer do not switch to Microsoft Word nor create files of proprietary format. Fig. 4 displays the one hundred simulations of the top six models for μ. The black line represents the original data. As one can see, the original data of the top three models for coverage of fit (c, f, and a) is completely embedded in the area defined by the simulations' plots. Unfortunately, it is more difficult to visualize the nearest simulation random walk that determines the nearest distance of the best three models (a, b, and c).

Likewise, Tables 18 and 19 show the top six urn models of the ranking according to nearest simulation distance and the coverage of fit in the case of the spreadsheet applications; the complete lists are reported in Appendix B (Tables 30 and 26).

Again, the empirical models are the predominant models and the Pólya urn model is one of the best models for both measures. Unlike the Word/Writer case, Table 18 also includes the opponent model. This might indicate some resistance to the use of Calc and a less mature and successful process of secondary adoption than the one with text processors. This is not a surprise as users reported that spreadsheet applications have sometimes been used with macros that cannot be easily converted into open format. Nevertheless, our results show that the transition has some effects on these applications too.

6.1.1. Variation analysis

To show that our results do not depend from the chosen paths, we perform a variation analysis. We first compute mean and standard deviation of the original data (Table 20). Then we compute mean and standard deviation ranges of all the one hundred paths we have used for the top six models (Tables 28 and 29). As result, we get one hundred means and standard deviations that define two variation ranges (min, max) of mean and standard deviation for every top six model. For Word processors, we found that all the two ranges of all top six models but model d) of Fig. 4 include respectively original mean and standard deviation (Table 20). For spreadsheet applications, we found that all ranges include respectively original mean and standard deviation for all the top six models but two models. Tables 27–30 report in boldface non-correct ranges.

This result is in line with the results of Section 6.1 and the patterns in Fig. 4. Nevertheless, it is still not enough to ensure that our full rankings do not depend on the one hundred paths we used. To investigate this, we perform one hundred simulations with fifty, one hundred, one thousand paths for each model of Tables 21 and 22. Then we compute how many times (over one hundred) the original mean and standard deviation belongs to the respective range of a given model. Figs. 5–8 of Appendix B show the results. Each three consecutive values on the x-axis correspond to fifty, one hundred, one thousand paths of the same urn model. Urn models are displayed on the x-axis as they are in Tables 24 and 26. For example, for Word processors the first eighteen x values refer to the top six models. All the non-zero percentages reported in Figs. 5 and 7 refer to the top six models respectively for Word processors and Spreadsheet applications. In particular, the original path is never embedded in the area defined by the fifty, one hundred, and one thousand simulation paths of the remaining urn models. This confirms our rankings. In addition, the simulations area of the two models b) and d) of Word processors in Fig. 4 embeds the original path for less than 100% of times also with one thousand paths. For Spreadsheet applications, the same is true. The two models that were poorest with one hundred paths remain poorest with one thousand paths. This supports the choice of one hundred paths.

6.2. Brief summary of the findings

From the application of our method, we derived some facts:

- Empirical and abstract models are significantly different for both word processors and spreadsheet applications. Therefore, abstract models do not represent the actual use of the office suites.
- Empirical models are not balanced and always add Microsoft files (β and $\delta > 0$). They all express a mild successful but not mature secondary adoption. This has been confirmed by the results on time of application use we found in January 2005. The average time spent per day with Microsoft Office doubled that with OpenOffice (Section 4.1).

Table 18
Ranking by the nearest simulation distance. Top 6 models.

Type	Urn model ($w_1 = 27$, $w_2 = 732$)	μ
P	U(26, 0, 0, 26)	0.009282
E1	U(26, 260, 0, 26)	0.009282
P	U(13, 0, 0, 13)	0.010385
E1	U(13, 130, 0, 13)	0.010385
O	U(−9, 9, 0, 9)	0.0142782
O	U(−1, 1, 0, 1)	0.014724

Table 19
Ranking by coverage of fit. Top 6 models.

Type	Urn model ($w_1 = 27$, $w_2 = 732$)	%
E1	U(500, 5000, 0, 500)	100.00
E1	U(1000, 10000, 0, 1000)	100.00
P	U(1000, 0, 0, 1000)	100.00
P	U(500, 0, 0, 500)	100.00
E1	U(26, 260, 0, 26)	90.08
E1	U(13, 130, 0, 13)	79.04

Table 20
Mean and st. dev. for the original data.

	Mean	st. dev.
Text processors	0.0473	0.0055
Spreadsheets	0.0176	0.0032

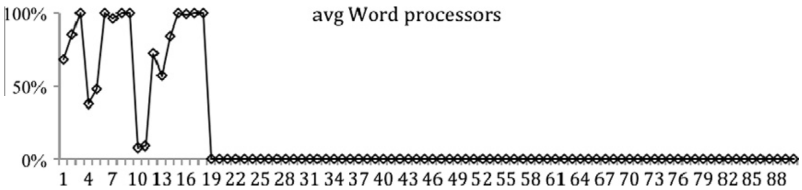

Fig. 5. Percentage of times original mean belongs to mean ranges for 50, 100, 1000 paths over urn models. Triples of consecutive values corresponds of percentage for 50, 100, 1000 paths for a single model.

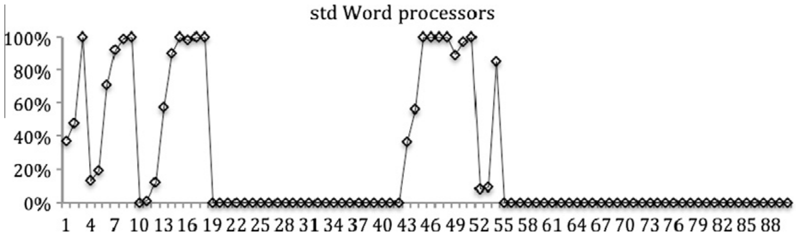

Fig. 6. Percentage of times original standard deviation belongs to standard deviation ranges for 50, 100, 1000 paths over urn models. Triples of consecutive values corresponds of percentage for 50, 100, 1000 paths for a single model.

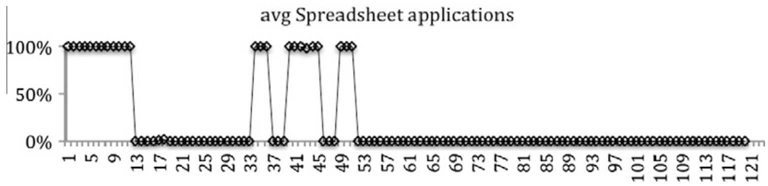

Fig. 7. Percentage of times original mean belongs to mean ranges for 50, 100, 1000 paths over urn models. Triples of consecutive values corresponds of percentage for 50, 100, 1000 paths for a single model.

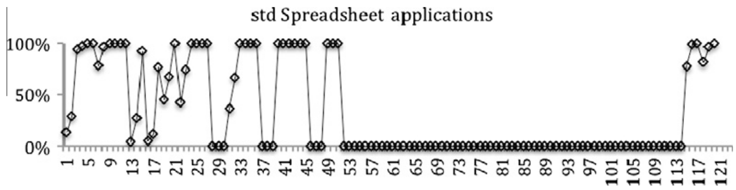

Fig. 8. Percentage of times original standard deviation belongs to standard deviation ranges for 50, 100, 1000 paths over urn models. Triples of consecutive values corresponds of percentage for 50, 100, 1000 paths for a single model.

- The difference between abstract models and empirical models reduces with the increase of file creations of OpenOffice (design E2 outperforms design E1).
- Secondary adoption is more successful and mature with word processors than with spreadsheets.
- The Pólya urn model is the most mature and successful model that accurately fits our data. According to this result, the most optimistic configuration of the adoption process depicts users, who tend to create files in the format they have been using since ever. As it is deterministic [24], its performance might tell the absence of exceptional environmental factors that disturb the adoption process.

In the end, the application of our method to the adoption of OpenOffice reveals the existence of an assimilation gap in the organization under study. Although the management has put in place a well-defined strategy to migrate to OpenOffice.org, still the creation of Microsoft Office files is significant. File exchange with third parties was the major reason. These findings support the fact that predictions based on company's strategy and directives can be far from the reality as the actual use of a technology can be assimilated with a different pace.

7. Discussion and limitations

There are some limitations in our work. The first concerns data collection. Due to instrumentation, we did not infer file type from file header information, but from file extension. This fact can have increased the false positives/false negatives for certain standards, although we suspect this phenomenon to be limited. A second limitation regards measuring the application use. We model it with the daily creation of files and urn models. This describes the assimilation process as a user's preference. Other measures, like time of application use, can be combined or used to give a different insight. Namely, we have collected the time of application use for 1475 employees in twenty consecutive days of January 2005 and we have found that, in average, employees spend 48.08 min on OpenOffice documents and 92.93 min on Microsoft Office documents confirming the existence of a gap at the end of our observation period. We exploit this measure further in future work.

Modeling the use of an application with file creations might have another limitation. As OpenOffice can create files in the Microsoft Office format, the number of files created in the proprietary format can overestimate the actual use of Microsoft Office. This might mean that the actual assimilation gap is smaller. For this, we have proved that at the end of our observation period, the use of the OpenOffice feature was limited as well as its effects on the assimilation gap.

Finally, creation of files might not be an accurate measure of application use. For example, files might be created just downloading and saving them into the network server without opening any application. In this case, the creation of files into the server does not indicate application use. There are few options for which a file is not created from the native application: downloading it from 1. Internet, 2. intranet, or 3. emails or 4. creating it from another office application if it is possible. In our case study,

(1) The use of Internet was very limited. This prevented file downloads from the web.
(2) As intranet exchange ran through the network server, there was little interest to email documents among the Softech's employees. As such, the dispatch of files via emails was limited to the exchange with external parties.
(3) Files received from external parties can have made the real difference between application use and files creation. Files can be downloaded from the email application and directly saved or opened and saved from the office application into the server. In the former case, we do not have control on the use of the office application. We must admit though that saving files without opening them is not a common practice, in particular when files are saved in a public share.
(4) Employees created files with the native application. As we mentioned, the majority of the employees were not in general IT expert and aware of the fact that OpenOffice could open and save Microsoft Office files. As such, they created files in the native application they prefer.

In conclusions, we are confident that creation of files and application use are pretty much interchangeable to measure users' preferences in our case study. As such, as the values of β have more than doubled the values of α, we can claim that the assimilation gap exists and is significant in Softech.

8. Conclusions

The gap between predicted and actual use of innovation is a critical issue in any organization as it reveals misalignment between primary and secondary adoption. In IT adoption, this discontinuity can create strong resistance and adverse attitude toward any future adoption of innovation. It is of foremost importance for an organization to keep this gap to minimum. Literature on technology acceptance focuses on factors that drive the use of technology. Those factors have proven to be poor predictors of the actual use of technology, though [56].

In economics, politics, or medical sciences, adoption is typically studied as the dynamic problem of choice. For example, the market fluctuation is driven by the choices of the customers [1,2,4,40], the evolution of political institutions is affected by choices in social contingent events [40] or the hospital allocations of patients is determined by the choice of medical treatments [30,60]. Typically, research in adoption studies two characteristics of the evolution of the choice: whether it is driven by learning from history (non-ergodicity) or is locked in the future (inflexibility) [4].

In our work, we study the non-ergodic evolution of software adoption as the choice of the employees of an organization to use specific software applications in their daily work. In our case study, we assume that employees follow a parallel adoption, for which employees can freely choose among software applications with the same functionalities. As such, their choice is characterized only by positive feedback (increasing returns) as it is purely conditioned by the previous use of the applications [4].

Following the traditional literature in the field, we have chosen to model the users' choice with path dependent processes, specifically urn models. We have then identified the mathematical expression of the urn models predicted by the strategy and by the actual use of the software applications.

In our case, urn models describe the parallel use of OpenOffice and Microsoft Office suites via the evolution of files creation. We have used urn models to compare predicted and actual use of a technology and to determine successful and mature adoptions. With this research, we have provided a constructive and objective method to assess and quantify the assimilation gap (RQ).

The application of our method to a real case study has first reported a misalignment between use predicted from the strategy and actual one. Namely, the actual use of the office suites does not correspond to the one predicted by the strategy. This shows that the adoption of OpenOffice – as new technology – has been only mild successful and not mature at the time of our analysis despite well-defined strategies and directives taken by the managers. By the analysis of the context, we suspected that this was due to the hybrid decision that let the users freely chose between the two office suites and the need to exchange files with external organizations in the proprietary format. Aware of this, the management re-calibrated the strategy of adoption of OSS including actions toward external parties to sensitize them on the OS phenomenon and activities to further training users of OSS.

Overall, this research suggests the need of a seamless monitoring of the adoption process and of objective measurement of the actual use of a technology.

We expect that future research will test our method in different environments and with different OSS as our method can be applied to any technology whose use can be measured and quantified periodically.

9. Future work

This work had the goal to measure the alignment between primary and secondary adoption. To do this, we used path-dependent Markov chains that describe the process of technology use paraphrasing the traditional approaches of modeling user's choices in the socio/economic market with the choice of use of a software application. Markov chains are built using just one-step in history, though. Adding more history to the definition of the future states of a path dependent process can give a finer representation of the process. This will be matter of future work.

As we mentioned, file creation is just one measure of application use. Time of application use can be combined with file creation to provide a more complete understanding of the assimilation of a technology. It is worth noticing that using time of application use alone have its limitations. For example, if time is collected automatically one cannot detect and exclude periods in which the application is open but is not used in fact. Anyhow, time of application use can help clarifying the relation between use of data standards and application use. As we have collected data on time of use since January 2005, we will explore this in future work.

The experience with case study proved that direct interviews with end-users can provide a reference model for our method. In future replications of the work, we will include this practice in the design of the study.

Finally, the dataset, although significant, represents one case study only. We are collecting further data from Softech to compare the current use of OpenOffice with the one we have studied and we are intentioned to replicate the study in other organizations.

10. Lessons learned

Overall, we wanted to provide managers with a rigorous instrument for measuring the assimilation gap. In the interpretation of our results and interviewing managers, we have increased our knowledge on the practical aspects of the assimilation of a new technology. At the end of this experience, we are able to draft some initial recommendations to researchers and managers interested in the adoption process of a new technology:

Recommendation 1. Monitor the actual use of a technology at different angles seamlessly. This will help to draw the overall view of the adoption process. In our case study, we focus on creation of files, but time to use an application as well as periodic reports from users might have been useful to complement our study.

Recommendation 2. Define clearly strategy and directives supporting the adoption process. This will serve as reference for the actual adoption.

Recommendation 3. Act and recalibrate the adoption process as soon as any gap is reported. We learned that the adoption process is evolutionary and affected by the surrounding environment. Actions need to limit the influence of non-controlled external factors.

Recommendation 4. Interview stakeholders to get perception on the adoption. This will augment the knowledge of the adoption process.

Recommendation 5. Perform ad hoc training at the early stages of the adoption process. This definitely avoids users' resistance and help increasing the learning pace.

Recommendation 6. Use parallel adoption. In our study, company did not replace the old technology with the new one. This avoided some resistance from those users that were accustomed with the old technology. On the hand, this has slowed the learning pace of the new technology. We believe that a mixture approach, parallel and then single adoption, is a good compromise.

Acknowledgments

The authors thank Ing. Hugo Leiter, head of the IT department of Softech for the collaboration in collecting data and information about his organization. This work was supported in part by the FP6 European project COSPA, Contract No.: IST-2002-2164. We thank the reviewer for the valuable comments.

Appendix A

Tables 21 and 22.

Table 21
Types of urn models considered in the analysis of Word and Writer files.

ID	Matrix Abstract	ID	Matrix Opponent
C1	U(w_1 = 8733, w_2 = 217470, 1000, 0, 2000, −1000)	O	U(w_1 = 8733, w_2 = 217470, −1000, 1000, 0, 1000)
	U(w_1 = 8733, w_2 = 217470, 500, 0, 1000, −500)		U(w_1 = 8733, w_2 = 217470, −500, 500, 0, 500)
	U(w_1 = 8733, w_2 = 217470, 1, 0, 2, −1)		U(w_1 = 8733, w_2 = 217470, −1, 1, 0, 1)
	U(w_1 = 8733, w_2 = 217470659, 0, 1318, −659)		**Pólya**
C2	U(w_1 = 8733, w_2 = 217470, 1000, 0, 1000, −1000)	P	U(w_1 = 8733, w_2 = 217470, 1000, 0, 0, 1000)
	U(w_1 = 8733, w_2 = 217470, 500, 0, 500, −500)		U(w_1 = 8733, w_2 = 217470, 500, 0, 0, 500)
	U(w_1 = 8733, w_2 = 217470, 659, 0, 659, −659)		U(w_1 = 8733, w_2 = 217470, 1, 0, 0, 1)
	U(w_1 = 8733, w_2 = 217470330, 0, 330, −330)		**Friedman**
	U(w_1 = 8733, w_2 = 217470, 1, 0, 1, −1)	F	U(w_1 = 8733, w_2 = 217470, 0, 1000, 1000, 0)
C3	U(w_1 = 8733, w_2 = 217470, 1000, 0, 1000, 0)		U(w_1 = 8733, w_2 = 217470, 0, 500, 500, 0)
	U(w_1 = 8733, w_2 = 217470, 500, 0, 500, 0)		U(w_1 = 8733, w_2 = 217470, 0, 1, 1, 0)
	U(w_1 = 8733, w_2 = 217470, 1, 0, 1, 0)		
	Empirical		
E1	U(w_1 = 8733, w_2 = 217470, 1000, 2000, 0, 1000)		
	U(w_1 = 8733, w_2 = 217470, 500, 1000, 0, 500)		
	U(w_1 = 8733, w_2 = 217470, 1, 2, 0, 1)		
E2	U(w_1 = 8733, w_2 = 217470, 1000, 4000, 0, 1000)		
	U(w_1 = 8733, w_2 = 217470,V500, 2000, 0, 500)		
	U(w_1 = 8733, w_2 = 217470, 1, 4, 0, 1)		

Table 22
Type of urn models considered in the analysis of Excel and Calc files.

ID	Matrix Abstract	ID	Matrix Opponent
C1	U(w_1 = 27, w_2 = 732, 1, 0, 2, −1)	O	U(w_1 = 27, w_2 = 732, −9, 9, 0, 9)
	U (w_1 = 27, w_2 = 732, 12, 0, 24, −12)		U(w_1 = 27, w_2 = 732, −1, 1, 0, 1)
	U(w_1 = 27, w_2 = 732, 13, 0, 26, −13)		U(w_1 = 27, w_2 = 732, −12, 12, 0, 12)
	U(w_1 = 27, w_2 = 732, 26, 0, 52, −26)		U(w_1 = 27, w_2 = 732, −13, 13, 0, 13)
	U(w_1 = 27, w_2 = 732, 61, 0122, −61)		U(w_1 = 27, w_2 = 732, −26, 26, 0, 26)
	U(w_1 = 27, w_2 = 732, 500, 0, 1000, −500)		U(w_1 = 27, w_2 = 732, −500, 500, 0, 500)
	U(w_1 = 27, w_2 = 732, 1000, 0, 2000, −1000)		U(w_1 = 27, w_2 = 732, −1000, 1000, 0, 1000)
C2	U(w_1 = 27, w_2 = 732, 1, 0, 1, −1)		**Pólya**
	U (w_1 = 27, w_2 = 732, 13, 0, 13, −13)	P	U(w_1 = 27, w_2 = 732, 26, 0, 0, 26)
	U(w_1 = 27, w_2 = 732, 26, 0, 26, −26)		U(w_1 = 27, w_2 = 732, 13, 0, 0, 13)
	U(w_1 = 27, w_2 = 732, 61, 0, 61, −61)		U(w_1 = 27, w_2 = 732, 1, 0, 0, 1)
	U(w_1 = 27, w_2 = 732, 61, 0122, −61)		U(w_1 = 27, w_2 = 732, 500, 0, 0, 500)

(*continued on next page*)

Table 22 (*continued*)

ID	Matrix Abstract	ID	Matrix Opponent
	U(w_1 = 27, w_2 = 732, 500, 0, 500, −500)		U(w_1 = 27, w_2 = 732, 1000, 0, 0, 1000)
	U(w_1 = 27, w_2 = 732, 1000, 0, 1000, −1000)		**Friedman**
C3	U(w_1 = 27, w_2 = 732, 1, 0, 1, 0)	F	U(w_1 = 27, w_2 = 732, 0, 1, 1, 0)
	U(w_1 = 27, w_2 = 732, 13, 0, 13, 0)		U(w_1 = 27, w_2 = 732, 0, 13, 13, 0)
	U(w_1 = 27, w_2 = 732, 26, 0, 26, 0)		U(w_1 = 27, w_2 = 732, 0, 26, 26, 0)
	U(w_1 = 27, w_2 = 732, 500, 0, 500, 0)		U(w_1 = 27, w_2 = 732, 0, 500, 500, 0)
	U(w_1 = 27, w_2 = 732, 1000, 0, 1000, 0)		U(w_1 = 27, w_2 = 732, 0, 1000, 1000, 0)
	Empirical		
E1	U(w_1 = 27, w_2 = 732, 26, 260, 0, 26)		
	U(w_1 = 27, w_2 = 732, 13, 130, 0, 13)		
	U(w_1 = 27, w_2 = 732, 1, 10, 0, 1)		
	U(w_1 = 27, w_2 = 732, 500, 500, 0, 0, 500)		
	U(w_1 = 27, w_2 = 732, 1000, 10000, 0, 1000)		

Appendix B

Tables 23–30.

Table 23
.doc and .sxw files. Ranking of urn models by the score nearest distance. The Simulation# refers to nearest simulation.

Type	Urn model	Simulation #	μ
P	U(w_1 = 8733, w_2 = 217470, 500, 0, 0, 500)	50	0.006605
E1	U(w_1 = 8733, w_2 = 217470, 500, 1000, 0, 500)	11	0.007652
E1	U(w_1 = 8733, w_2 = 217470, 1000, 2000, 0, 1000)	39	0.007823
E2	U(w_1 = 8733, w_2 = 217470, 500, 2000, 0, 500)	58	0.008032
E2	U(w_1 = 8733, w_2 = 217470, 1000, 4000, 0, 1000)	57	0.008179
P	U(w_1 = 8733, w_2 = 217470, 1000, 0, 0, 1000)	36	0.008193
C1	U(w_1 = 8733, w_2 = 217470, 1, 0, 2, −1)	20	0.013151
C2	U(w_1 = 8733, w_2 = 217470, 1, 0, 1, −1)	3	0.013966
C3	U(w_1 = 8733, w_2 = 217470, 1, 0, 1, 0)	1	0.013999
F	U(w_1 = 8733, w_2 = 217470, 0, 1, 1, 0)	65	0.013999
E1	U(w_1 = 8733, w_2 = 217470, 1, 2, 0, 1)	66	0.014783
P	U(w_1 = 8733, w_2 = 217470, 1, 0, 0, 1)	73	0.014785
E2	U(w_1 = 8733, w_2 = 217470, 1, 4, 0, 1)	7	0.014808
O	U(w_1 = 8733, w_2 = 217470, −1, 1, 0, 1)	30	0.014865
O	U(w_1 = 8733, w_2 = 217470, −639, 639, 0639)	17	0.015988
O	U(w_1 = 8733, w_2 = 217470, −369, 369, 0369)	47	0.028724
O	U(w_1 = 8733, w_2 = 217470, −500, 500, 0, 500)	67	0.029799
O	U(w_1 = 8733, w_2 = 217470, −1000, 1000, 0, 1000)	18	0.038952
F	U(w_1 = 8733, w_2 = 217470, 0, 500, 500, 0).	72	0.26831
F	U(w_1 = 8733, w_2 = 217470, 0, 1000, 1000, 0)	100	0.34758
C3	U(w_1 = 8733, w_2 = 217470, 500, 0, 500, 0)	1	0.38851
C2	U(w_1 = 8733, w_2 = 217470330, 0, 330, −330)	100	0.41153
C1	U(w_1 = 8733, w_2 = 217470, 330, 0660, −330)	16	0.50818
C2	U(w_1 = 8733, w_2 = 217470, 500, 0, 500, −500)	65	0.5281
C3	U(w_1 = 8733, w_2 = 217470, 1000, 0, 1000, 0)	1	0.55344
C1	U(w_1 = 8733, w_2 = 217470, 500, 0, 1000, −500)	26	0.60278
C2	U(w_1 = 8733, w_2 = 2174706591, 0, 659, −659)	16	0.62496
C1	U(w_1 = 8733, w_2 = 217470, 659, 0, 1318, −659)	75	0.69495
C2	U(w_1 = 8733, w_2 = 217470, 1000, 0, 1000, −1000)	73	0.73918
C1	U(w_1 = 8733, w_2 = 217470, 1000, 0, 2000, −1000)	79	0.7593

Table 24
.doc and .sxw files. Ranking of the urn models by the coverage of fit.

Type	Urn model	%
E1	U(w_1 = 8733, w_2 = 217470, 1000, 2000, 0, 1000)	100%
P	U(w_1 = 8733, w_2 = 217470, 1000, 0, 0, 1000)	99.68%
P	U(w_1 = 8733, w_2 = 217470, 500, 0, 0, 500)	97.12%
E2	U(w_1 = 8733, w_2 = 217470, 1000, 4000, 0, 1000)	65.17%
E1	U(w_1 = 8733, w_2 = 217470, 500, 1000, 0, 500)	43.13%
E2	U(w_1 = 8733, w_2 = 217470, 500, 2000, 0, 500)	42.17%
P	U(w_1 = 8733, w_2 = 217470, 1, 0, 0, 1)	0.32%
F	U(w_1 = 8733, w_2 = 217470, 0, 1000, 1000, 0)	0.32%
F	U(w_1 = 8733, w_2 = 217470, 0, 500, 500, 0)	0.32%
E2	U(w_1 = 8733, w_2 = 217470, 1, 4, 0, 1)	0.32%
E1	U(w_1 = 8733, w_2 = 217470, 1, 2, 0, 1)	0.32%
O	U(w_1 = 8733, w_2 = 217470, −500, 500, 0, 500)	0%
O	U(w_1 = 8733, w_2 = 217470, −1, 1, 0, 1)	0%
F	U(w_1 = 8733, w_2 = 217470, 0, 1, 1, 0)	0%

Table 24 (*continued*)

Type	Urn model	%
O	U(w_1 = 8733, w_2 = 217470, −1000, 1000, 0, 1000)	0%
C1	U(w_1 = 8733, w_2 = 217470, 1000, 0, 2000, −1000)	0%
C2	U(w_1 = 8733, w_2 = 217470, 1000, 0, 1000, −1000)	0%
C3	U(w_1 = 8733, w_2 = 217470, 1000, 0, 1000, 0)	0%
C1	U(w_1 = 8733, w_2 = 217470, 500, 0, 1000, −500)	0%
C2	U(w_1 = 8733, w_2 = 217470, 500, 0, 500, −500)	0%
C3	U(w_1 = 8733, w_2 = 217470, 500, 0, 500, 0)	0%
C1	U(w_1 = 8733, w_2 = 217470, 1, 0, 2, −1)	0%
C2	U(w_1 = 8733, w_2 = 217470, 1, 0, 1, −1)	0%
C3	U(w_1 = 8733, w_2 = 217470, 1, 0, 1, 0)	0%
C2	U(w_1 = 8733, w_2 = 217470330, 0, 330, −330)	0%
O	U(w_1 = 8733, w_2 = 217470, −369, 369, 0369)	0%
C1	U(w_1 = 8733, w_2 = 217470659, 0, 1318, −659)	0%
C2	U(w_1 = 8733, w_2 = 217470, 659, 0, 659, −659)	0%
C1	U(w_1 = 8733, w_2 = 217470, 330, 0660, −330)	0%

Table 25
.xls and .sxc files. Ranking of the urn models by the score m. The Simulation# refers to nearest simulation.

Type	Urn model	Simulation #	μ
P	U(w_1 = 27, w_2 = 732, 26, 0, 0, 26)	61	0.009282
E1	U(w_1 = 27, w_2 = 732, 26, 260, 0, 26)	100	0.009282
P	U(w_1 = 27, w_2 = 732, 13, 0, 0, 13)	3	0.010385
E1	U(w_1 = 27, w_2 = 732, 13, 130, 0, 13)	2	0.010385
O	U(w_1 = 27, w_2 = 732, −9, 9, 0, 9)	88	0.0142782
O	U(w_1 = 27, w_2 = 732, −1, 1, 0, 1)	25	0.014724
O	U(w_1 = 27, w_2 = 732, −12, 12, 0, 12)	73	0.0157326
O	U(w_1 = 27, w_2 = 732, −13, 13, 0, 13)	1	0.0160894
O	U(w_1 = 27, w_2 = 732, −26, 26, 0, 26)	3	0.0184209
E1	U(w_1 = 27, w_2 = 732, 1, 10, 0, 1)	9	0.019633
P	U(w_1 = 27, w_2 = 732, 1, 0, 0, 1)	24	0.0196330
E1	U(w_1 = 27, w_2 = 732, 500, 500, 0, 0, 500)	1	0.0290102
O	U(w_1 = 27, w_2 = 732, −500, 500, 0, 500)	1	0.0290102
P	U(w_1 = 27, w_2 = 732, 500, 0, 0, 500)	1	0.0290102
E1	U(w_1 = 27, w_2 = 732, 1000, 10000, 0, 1000)	86	0.0313635
O	U(w_1 = 27, w_2 = 732, −1000, 1000, 0, 1000)	1	0.0321983
P	U(w_1 = 27, w_2 = 732, 1000, 0, 0, 1000)	1	0.0321983
F	U(w_1 = 27, w_2 = 732, 0, 1, 1, 0)	99	0.2475490
C3	U(w_1 = 27, w_2 = 732, 1, 0, 1, 0)	1	0.3205920
C2	U(w_1 = 27, w_2 = 732, 1, 0, 1, −1)	1	0.4151420
F	U(w_1 = 27, w_2 = 732, 0, 13, 13, 0)	63	0.440326
F	U(w_1 = 27, w_2 = 732, 0, 26, 26, 0)	40	0.461956
F	U(w_1 = 27, w_2 = 732, 0, 500, 500, 0)	56	0.4861010
C1	U(w_1 = 27, w_2 = 732, 1, 0, 2, −1)	63	0.5139320
F	U(w_1 = 27, w_2 = 732, 0, 1000, 1000, 0)	31	0.5464470
C3	U(w_1 = 27, w_2 = 732, 13, 0, 13, 0)	1	0.841992
C3	U(w_1 = 27, w_2 = 732, 26, 0, 26, 0)	1	0.905202
C1	U(w_1 = 27, w_2 = 732, 12, 0, 24, −12)	14	0.942822
C1	U(w_1 = 27, w_2 = 732, 13, 0, 26, −13)	68	0.944266
C2	U(w_1 = 27, w_2 = 732, 13, 0, 13, −13)	22	0.945516
C1	U(w_1 = 27, w_2 = 732, 26, 0, 52, −26)	88	0.963138
C2	U(w_1 = 27, w_2 = 732, 26, 0, 26, −26)	47	0.964918
C2	U(w_1 = 27, w_2 = 732, 61, 0, 61, −61)	33	0.974156
C1	U(w_1 = 27, w_2 = 732, 61, 0122, −61)	13	0.974456
C2	U(w_1 = 27, w_2 = 732, 61, 0122, −61)	13	0.974456
C3	U(w_1 = 27, w_2 = 732, 500, 0, 500, 0)	1	0.976335
C3	U(w_1 = 27, w_2 = 732, 1000, 0, 1000, 0)	1	0.979915
C2	U(w_1 = 27, w_2 = 732, 500, 0, 500, −500)	35	0.981244
C1	U(w_1 = 27, w_2 = 732, 500, 0, 1000, −500)	3	0.981254
C1	U(w_1 = 27, w_2 = 732, 1000, 0, 2000, −1000)	1	0.986344
C2	U(w_1 = 27, w_2 = 732, 1000, 0, 1000, −1000)	1	0.986344

Table 26
.xls and .sxc files. Ranking of the urn models by the coverage of fit.

Type	Urn model	%
E1	U(w_1 = 27, w_2 = 732, 500, 500, 0, 0, 500)	100.00
E1	U(w_1 = 27, w_2 = 732, 1000, 10000, 0, 1000)	100.00
P	U(w_1 = 27, w_2 = 732, 1000, 0, 0, 1000)	100.00
P	U(w_1 = 27, w_2 = 732, 500, 0, 0, 500)	100.00
E1	U(w_1 = 27, w_2 = 732, 26, 260, 0, 26)	90.08
E1	U(w_1 = 27, w_2 = 732, 13, 130, 0, 13)	79.04
O	U(w_1 = 27, w_2 = 732, −1, 1, 0, 1)	41.64
O	U(w_1 = 27, w_2 = 732, −9, 9, 0, 9)	31.16
O	U(w_1 = 27, w_2 = 732, −12, 12, 0, 12)	22.66
O	U(w_1 = 27, w_2 = 732, −13, 13, 0, 13)	20.96
C2	U(w_1 = 27, w_2 = 732, 13, 0, 13, −13)	9.92
O	U(w_1 = 27, w_2 = 732, −26, 26, 0, 26)	9.92
P	U(w_1 = 27, w_2 = 732, 1, 0, 0, 1)	1.13
E1	U(w_1 = 27, w_2 = 732, 1, 10, 0, 1)	0.57
F	U(w_1 = 27, w_2 = 732, 0, 1000, 1000, 0)	0.28
F	U(w_1 = 27, w_2 = 732, 0, 500, 500, 0)	0.28
F	U(w_1 = 27, w_2 = 732, 0, 1, 1, 0)	0.28
F	U(w_1 = 27, w_2 = 732, 0, 13, 13, 0)	0.28
P	U(w_1 = 27, w_2 = 732, 13, 0, 0, 13)	0.28
F	U(w_1 = 27, w_2 = 732, 0, 26, 26, 0)	0.28
P	U(w_1 = 27, w_2 = 732, 26, 0, 0, 26)	0.28

Table 26 (*continued*)

Type	Urn model	%
O	U(w_1 = 27, w_2 = 732, −1000, 1000, 0, 1000)	0.00
O	U(w_1 = 27, w_2 = 732, −500, 500, 0, 500)	0.00
C1	U(w_1 = 27, w_2 = 732, 1000, 0, 2000, −1000)	0.00
C2	U(w_1 = 27, w_2 = 732, 1000, 0, 1000, −1000)	0.00
C3	U(w_1 = 27, w_2 = 732, 1000, 0, 1000, 0)	00.00
C1	U(w_1 = 27, w_2 = 732, 500, 0, 1000, −500)	0.00
C2	U(w_1 = 27, w_2 = 732, 500, 0, 500, −500)	0.00
C3	U(w_1 = 27, w_2 = 732, 500, 0, 500, 0)	0.00
C1	U(w_1 = 27, w_2 = 732, 1, 0, 2, −1)	0.00
C2	U(w_1 = 27, w_2 = 732, 1, 0, 1, −1)	0.00
C3	U(w_1 = 27, w_2 = 732, 1, 0, 1, 0)	0.00
C3	U(w_1 = 27, w_2 = 732, 13, 0, 13, 0)	0.00
C1	U(w_1 = 27, w_2 = 732, 13, 0, 26, −13)	0.00
C3	U(w_1 = 27, w_2 = 732, 26, 0, 26, 0)	0.00
C2	U(w_1 = 27, w_2 = 732, 26, 0, 26, −26)	0.00
C1	U(w_1 = 27, w_2 = 732, 26, 0, 52, −26)	0.00
C1	U(w_1 = 27, w_2 = 732, 12, 0, 24, −12)	0.00
C1	U(w_1 = 27, w_2 = 732, 61, 0122, −61)	0.00
C2	U(w_1 = 27, w_2 = 732, 61, 0, 61, −61)	0.00

Table 27
Mean range of the top 6 models over 100 simulation paths. Word Processors. In boldface models with inaccurate range.

Type	Urn model (w_1 = 8733, w_2 = 217470)	Min (μ)	Max (μ)
P	U(500, 0, 0, 500)	0.030912	0.048836
E1	U(500, 1000, 0, 500)	0.032251	0.047403
E1	U(1000, 2000, 0, 1000)	0.024474	0.055895
E2	**U(500, 2000, 0, 500)**	0.030293	**0.045707**
E2	U(1000, 4000, 0, 1000)	0.025684	0.056412
P	U(1000, 0, 0, 1000)	0.025748	0.058952

Table 28
St. dev. range of the top 6 models over 100 simulation paths. Word Processors. In boldface models with inaccurate range.

Type	Urn model (w_1 = 8733, w_2 = 217470)	Min (st. dev)	Max (st. dev)
P	U(500, 0, 0, 500)	0.001128	0.006305
E1	**U(500, 1000, 0, 500)**	0.000826	**0.004127**
E1	U(1000, 2000, 0, 1000)	0.001346	0.006844
E2	**U(500, 2000, 0, 500)**	0.000875	**0.003787**
E2	U(1000, 4000, 0, 1000)	0.001149	0.006747
P	U(1000, 0, 0, 1000)	0.001537	0.013279

Table 29
Mean range of the top 6 models over 100 simulation paths. Spreadsheets. In boldface models with inaccurate range.

Type	Urn model (w_1 = 27, w_2 = 732)	Min (μ)	Max (μ)
P	U(26, 0, 0, 26)	0.007520	0.186231
E1	U(26, 260, 0, 26)	0.007520	0.042635
P	U(13, 0, 0, 13)	0.011445	0.141909
E1	U(13, 130, 0, 13)	0.011445	0.047802
O	**U(−9, 9, 0, 9)**	0.000920	**0.013947**
O	**U(−1, 1, 0, 1)**	**0.019954**	0.027429
E1	U(500, 500, 0, 0, 500)	0.000789	0.046013
E1	U(1000, 10000, 0, 1000)	0.000430	0.044645
P	U(1000, 0, 0, 1000)	0.000430	0.955835
P	U(500, 0, 0, 500)	0.000789	0.880317

Table 30
St. dev. range of the top 6 models over 100 simulation paths. Spreadsheets. In boldface models with inaccurate range.

ype	Urn model (w27 = 1, w_2 = 732)	Min (st. dev)	Max (st. dev)
P	U(26, 0, 0, 26)	0.002993	0.032835
E1	U(26, 260, 0, 26)	0.002399	0.008461
P	U(13, 0, 0, 13)	0.002411	0.033457
E1	U(13, 130, 0, 13)	0.002352	0.008934
O	**U(−9, 9, 0, 9)**	**0.003906**	0.010548
O	**U(−1, 1, 0, 1)**	**0.003830**	0.007521
E1	U(500, 500, 0, 0, 500)	0.001838	0.013372
E1	U(1000, 10000, 0, 1000)	0.001175	0.009677
P	U(1000, 0, 0, 1000)	0.001175	0.074328
P	U(500, 0, 0, 500)	0.001838	0.055508

References

[1] W.B. Arthur, Competing technologies, increasing returns, and lock-in by historical events, The Economic Journal 99 (1989) 116–131.
[2] W.B. Arthur, Silicon valley' locational clusters: when do increasing returns imply monopoly?, Mathematical Social Sciences 19 (3) (1990) 235–251
[3] W.B. Arthur, Increasing Returns and path Dependency in Economy, University of Michigan Press, 1994, pp. 224.
[4] W.B. Arthur, Increasing returns and the new world of business, Harvard Business Review, 1996 July–August.
[5] W.B. Arthur, Y.M. Ermoliev, Y.M. Kaniovski. Path-dependent processes and the emergence of macrostructure', in: W.B. Arthur (Ed.), Increasing Returns and Path Dependence in the Economy, The University of Michigan Press, 1987 (Chapter 3).
[6] M. Baron, Probability and Statistics for Computer Scientists, Chapman and Hall/CRC, 2006.
[7] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal Statistical Society Series B (Methodological) 57 (1) (1995) 289.
[8] M. Bradford, J. Florin, Examining the role of innovation diffusion factors on the implementation success of enterprise resource planning systems, Information Systems 4 (2003) 205–225.
[9] I. Coman, A. Sillitti, G. Succi, A case-study on using an Automated In-process Software Engineering Measurement and Analysis system in an industrial environment, Proceedings of the 31st conference on Software Engineering, ICSE 2009, 2009.
[10] R.B. Cooper, R.W. Zmud, Information technology implementation research: a technological diffusion approach, Management Science 36 (2) (1990) 123–139.
[11] S. Dasgupta, M. Granger, N. Mcgarry, User acceptance of e-collaboration technology: an extension of technology acceptance model, Group Decision and Negotiation 11 (2) (2002) 87–100.
[12] P.A. David, Clio and the economics of QWERTY, American Economic Review 75 (2) (1985) 332–337.
[13] P.A. David, Path dependence, its critics and the quest for historical economics, Evolution and Path Dependence in Economic Ideas: Past and Present 15 (2001) 40.
[14] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly 13 (3) (1989) 319–340.
[15] F.D. Davis, R.P. Bagozzi, P.R. Warshaw, Extrinsic and intrinsic motivation to use computers in the workplace, Journal of Applied Social Psychology 22 (14) (1992) 1111–1132.
[16] J. Dedrick, J. West, An exploratory study into open source platform adoption: system sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on System Sciences, HICSS 04, 2004.
[17] G. Dosi, Y. Kaniovski, On 'badly behaved' dynamics: some applications of generalized urn schemes to technological and economic change, Journal of Evolutionary Economics 4 (2) (1994) 93–123.
[18] J. Doucher, W. Bolosky, A large-scale study of file-system contents, in: Proceedings of the 1999 ACM Sigmetrics Conference, 1999, 59-70.
[19] T. Dyba, N.B. Moe, E. Arisholm, Measuring software methodology use: challenges of conceptualization and operationalization, in: Proceedings of the International Symposium of Empirical Software Engineering (ISESE 05), 2005, 447-458.
[20] F. Eggenberger, G. Pólya, Ueber die statistik verketteter vorgaenge, Zeitschrift fuer Angewandte Mathematik und Mechanik 3 (1923) 279–289.
[21] K. Everitt, S. Lederer, A usability comparison of Sun StarOffice 5.2 vs. Microsoft Word 2000, 2001. <www.sims.berkeley.edu/courses/is271/f01/projects/WordStar/>.
[22] R.G. Fichman, C.F. Kemerer, The illusory diffusion of innovation: an examination of assimilation gaps, Information Systems Research 10 (3) (1999) 255–275.
[23] B. Fitzgerald, Open source software adoption: anatomy of success and failure, International Journal of Open Source Software & Processes 1 (1) (2009) 1–23.
[24] P. Flajolet, P. Dumas, V. Puyhaubert, Some exactly solvable models of urn process theory, Discrete Mathematics & Theoretical Computer Science 59 (2006) 118.
[25] D.A. Freedman, Bernard friedman's urn, The Annals of Mathematical Statistics 36 (1965) 956–970.
[26] B. Friedman, A simple urn model, Comm. Pure Appl. Math. 2 (1949) 59–70.
[27] M.D. Gallego, P. Luna, S. Bueno, User acceptance model of open source software, Computers in Human Behavior 24 (5) (2008) 2199–2216.
[28] M.J. Gallivan, Organizational adoption and assimilation of complex technological innovations: development and application of a new framework, ACM SIGMIS Database 32 (3) (2001) 51–85.
[29] E. Glynn, B. Fitzgerald, C. Exton, Commercial adoption of open source software: an empirical study, 2005, in: Proceedings of the International Symposium on Empirical Software Engineering (ISESE 2005), 17–18 November 2005, Noosa Heads, Australia, 2005, 225–234.
[30] N.E. Heckman, A local limit theorem for a biased coin design foe sequential test, The Annals of Statistics 13 (2) (1985) 785–788.
[31] J.P. Jouanolou, Èquations de Pfaff algèbriques, Lecture Notes in Mathematics, 708, 1979, Springer, Berlin.
[32] S.A. Kauffman, The evolution of economics webs, in: P.W. Anderson, K.J. Arrow, D. Pines (Eds.), The Economy as a Complex Evolving System, Addison-Wesley, New York, 1988, pp. 125–146.
[33] W.R. King, J. He, A meta-analysis of the technology acceptance model, Information and Management 6 (2006) 740–755.
[34] L. Lefebvre, J. Harvey, Intangible assets as determinants of advanced manufacturing technology adoption in SME's: toward an evolutionary model, IEEE Transactions on Engineering Management 43 (3) (1996) 307–322.
[35] P. Legris, J. Ingham, P. Collerette, Why do people use information technology? A critical review of technology acceptance model, Information and Management 40 (3) (2003) 191–204.
[36] M. Marchesi, G. Succi, B. Russo, A model of the dynamics of the market of COTS software, in the absence of new entrants, Information Systems Frontiers 9 (2–3) (2007) 257–265.
[37] G.C. Moore, I. Benbasat, Development of an instrument to measure the perceptions of adopting an information technology innovation, Information Systems Research 2 (3) (1991) 192–222.
[38] P. Page, Path dependence, Quarterly Journal of Political Science 1 (2006) 87–115.
[39] R. Pemantle, A survey of random processes with reinforcement, Probability Surveys, 4, 2007, 1–79. <http://www.i-journals.org/ps/viewarticle.php?id=94>.
[40] P. Pierson, Increasing returns, path dependence, and the study of politics, The American Political Science Review 94 (2000) 251–267.
[41] J. Pitman, Exchangeable and partially exchangeable random partitions, Probability Theory and Related Fields 102 (1995) 145–158.
[42] E.M. Rogers, Diffusion of Innovations, Free Press, Glencoe, 1962. 327.
[43] E.M. Rogers, Diffusion of Innovations, fourth ed., Free Press, 1995.
[44] E.M. Rogers, Diffusion of Innovations, fourth ed., Free Press, 2003.
[45] D. Roselli, R.L. Jacob, T.E. Anderson, A comparison of file system workloads, in: Proceedings of the USENIX Annual Technical Conference, San Diego, CA, 2000.
[46] N. Rosenberg, Learning by using: technologies and economics, Cambridge University Press, 1982.
[47] K. Safarzyńska, J.C.J.M. Van Den Bergh, Evolutionary models in economics: a survey of methods and building blocks, Journal of Evolutionary Economics 20 (3) (2009) 329–373.
[48] M. Satyanarayanan, A Study of File Sizes and Functional Lifetimes. 8th ACM SOSP, 1981, 96–108.
[49] R. Shah, J.P. Kesan, An Empirical Examination of Open Standards Development, Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS 08), 2008.
[50] J.H. Sharp, Development, extension, and application: a review of the technology acceptance model, Information Systems Education Journal 9 (2007) 1–9.
[51] B. Skyrms, R. Pemantle, A dynamic model of social network formation, Proceedings of National Academy of Science 97 (2000) 9340–9346.
[52] D. Straub, M. Limayem, E. Karahanna-Evaristo, Measuring system usage – implication for is theory testing, Management Science 41 (8) (1995) 1328–2134.
[53] S. Taylor, P. Todd, Assessing IT usage: the role of prior experience, MIS Quarterly 19 (4) (1995) 561–570.
[54] R.L. Thompson, C.A. Higgins, J.M. Howell, Personal computing: toward a conceptual model of utilization, MIS Quarterly 15 (1) (1991) 125–143.
[55] P. Trott, Innovation Management and New Product Development, Trans, Atlantic Publications, 1998.
[56] M. Turner, B. Kitchenam, P. Brereton, S. Charters, D. Budgen, Does the technology acceptance model predict actual use? A systematic literature review, Information and Software Technology 52 (2010) 463–479.
[57] V. Venkatesh, M. Morris, G. Davis, F. Davis, User acceptance of information technology: toward a unified view, MIS Quarterly 27 (3) (2003) 425–478.
[58] W. Vogels, File system usage in windows NT 4.0', in: Proceedings of the Seventeenth Symposium on Operating Systems Principles, 1999, 93–109.
[59] G. Von Krogh, E. Von Hippel, The promise of research on open source software, Management Science 52 (7) (2006) 975–983.
[60] L.J. Wei, An application of an urn model to the design of sequential controlled clinical trials, Journal of the American Statistical Association 73 (363) (1978) 559–563.
[61] J. Xue, A Pólya Urn Model of Conformity, Cambridge Working Papers in Economics, February 2006.
[62] G. Zaltman, R. Duncan, J. Holbeck, Innovations & Organizations, Wiley & Sons, New York, 1973.