Pushing the Limits – Using System Dynamics to Forecast Short-Term-Commodity Price Movements

Florian Kapmeier*

Reutlingen University
ESB Business School
Alteburgstraße 150
72762 Reutlingen
Germany

Phone: +49 7121 271 3000
email: florian.kapmeier@reutlingen-university.de

Alexander Voigt

Greenwood Strategic Advisors AG

Zugerstrasse 40
6314 Unterägeri
Switzerland

Phone: +41 41 754 7447
email: alex.voigt@greenwood-ag.com

* corresponding author

The authors are indebted to Mr. Andreas Harbig (andreas.harbig@greenwood-ag.com) and Mr. Craig Stephens (craig.stephens@greenwood-ag.com) of Greenwood Strategic Advisors, Unterägeri, Switzerland, leaders on the project described in this article, for their support and for their invaluable recollections of the project.

Abstract

Information derived from forecasts forms the basis for companies' strategic decision-making processes. Forecast accuracy can have a significant impact on the bottom line. While scholars discuss the usefulness of System Dynamics for carrying out forecasts in general, it is recognized that it should mainly be applied to provide insight into long-term developments, and not for short-term forecasts. In this paper we challenge both assumptions and propose that when expertly and rigorously wielded it can produce valuable results.

To illustrate our point we provide the example of a short-term price forecasting model developed for a large global petrochemical company. We first portray limitations and benefits of forecasts in general and System Dynamics-based forecasts in particular. We then present the challenge posed by the company, and how a somewhat unorthodox yet still rigorous process based on System Dynamics tools, methods and insights delivered value to the client – even if the eventual forecasting model used to do so contained no significant feedback loops. We conclude by discussing the benefits of using System Dynamics outside of what can be considered its typical field of application, and by stressing the extensive and varied use of data that permeated the whole process.

1 Usefulness and limitations of forecasting

Forecasts are a fundamental part of businesses' planning processes (Lyneis, 2000; Makridakis & Wheelwright, 1977; Makridakis, Wheelwright, & Hyndman, 1998). They are conducted to reduce risk that derives from uncertainty in the business environment (Makridakis, 1996) and thus to shed light on uncertain future developments. As a consequence, companies spend multi-million dollars for forecasting (Sherden, 1998). Policy-designers and decision-makers rely on insights from expectations about future developments of market demand, profit, revenues, costs, supply-chain partners' or competitors' actions, etc., and information derived from forecasts is the basis for policy-design and decision-making that heavily determines the future performance of their organizations.

Forecasts might refer to short-term or longer-term future developments (Lyneis, 2000) in both the economic and business world (Makridakis, Hogarth, & Gaba, 2009). Short-term forecasting, on the one hand, may include estimations about the size of order batches in a supply-chain in the automobile industry (Croson & Donohue, 2005, 2006; Lee, Padmanabhan, & Whang, 1997; Zahn, Hülsmann, Kapmeier, & Tilebein, 2007) or about the development of commodity prices like rubber (Khin, Zainalabidin, & Mad. Nasir, 2011), for example. Long-term or strategic forecasts, on the other hand, may concern earnings forecasts (Libby & Rennekamp, 2012) or decision-makers' desire to increase their companies' competitiveness on their Search for Excellence (Kapmeier, Tilebein, Voigt, & Dillerup, 2011; Micheli & Manzoni, 2010; Peters & Waterman, 1982).

When conducting forecasts, policy-designers and decision-makers mostly count on qualitative and quantitative methods that they are familiar with. Qualitative methods include decision trees, juries of executive opinion, or exploration assessments, among others. Such judgmental approaches have traditionally been used in the social sciences, whereas scientists have traditionally focused more on quantitative forecasting approaches (Fildes & Goodwin, 2007). These quantitative methods include single and multiple regression, econometric models, trend exploration and exponential smoothing (Forrester, 1961, 2007; Makridakis & Wheelwright, 1977; Makridakis et al., 1998).

While forecasts have been carried out for decades, the field of organizational forecasting has mainly evolved since the 1970s. According to MAKRIDAKIS ET AL. (2009; 2010) many social scientists in the 1970s and 1980s hoped that increasing computer power and the use of complex mathematical models would make them as successful in forecasting as their colleagues in the physical sciences. Since then, however, scholars have realized that forecasts in the social sciences have failed to achieve the expected levels of accuracy, for the reasons that we explore below.

1.1 Forecasts and uncertainty

Despite the many pitfalls of forecasting in general, they are still essential for the business and the economic world as they seek to estimate future developments when evaluating different decision alternatives. It is thus important for policy-designers and decision-makers to understand what at least can be modeled for forecasting – and this essentially depends on the concept of uncertainty.

MAKRIDAKIS ET AL. (2009, 2010), introduce three knowledge levels to illustrate three uncertainty types (see Figure 1): 'known knowns', 'known unknowns', and 'unknown unknowns'. On the one hand, systems whose behavior is dominated by known knowns do not require any forecast as there is no uncertainty; the future is perfectly known. On the other hand, systems whose behavior is dominated by unkonwn unknowns" are defined by rare events that are fully unexpected and unimagined, where decision-makers and policy-designers can only recognize events after they have taken place. These would be the so-called black swans, following TALEB's (2010) definition – like, for example, the invention of the internet and its impact on the business and the economic world. By definition, these uncertainties cannot be quantified in advance.

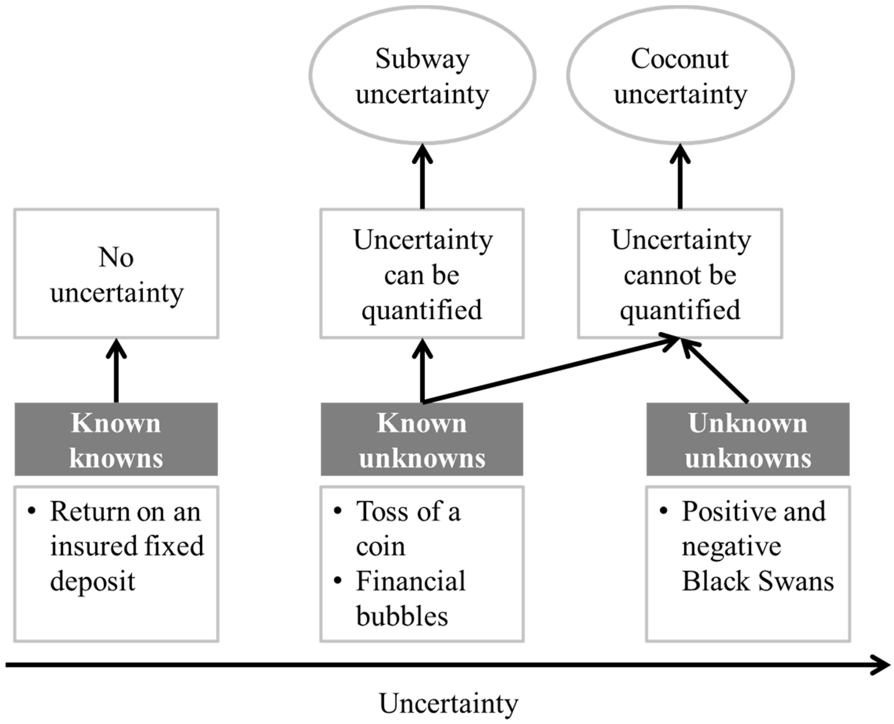

Figure 1: Drivers of 'subway' and 'coconut' uncertainty, adapted from Makridakis et al. (2009)

To summarize, for the very different reasons described above, forecasting cannot be usefully applied to either known knowns or unknown unknowns In between these two extremes are the "known unknowns" – and these consist of uncertainties that can be quantified, and of those that cannot be quantified. For assessing this, MAKRIDAKIS ET AL. (2009, 2010) distinguish between two different types of uncertainty: 'subway' uncertainty and 'coconut' uncertainty. Subway uncertainty describes uncertainties that can be quantified to some degree, like, for example, how long the subway commute will take on any given day. Coconut uncertainties in contrast, refer to uncertainties that cannot be quantified, like, for instance, the likelihood of being hit with a coconut while walking under a palm tree. In the first case, there is continuity between past and future, and quantification of uncertainties can be dealt with via assumptions based on the past performance of the system. Coconut uncertainties, however, refer to events with such small likelihoods of happening that they cannot be predicted based on the past. They constitute a discontinuity between past and future. If these occur, they happen unexpectedly - if they occur they carry with themselves critical consequences.

1.2 Forecasts still fail to consistently predict subway schedules

MAKRIDAKIS ET AL. (2009, 2010) provide a number of examples of failed forecasts in the economic and business environment, incl. the unforeseen demise of the once strong demand for copper which was mainly replaced by fiber optics in the 1970s and 1980s. Also, nobody forecasted the 80.5% decline of the Nikkei 225 stock index from 1989 to 2003 or GE's profit decline from 2007 until 2009 which was accompanied by a share price decline from $60 in October 2000 to $6.66 in March 2009. At the same time, however, failed forecasts also refer to positive events, like the unforeseen growth and business success of Amazon, Google,Apple, or Samsung. Some of these cases cited above are examples of systems dominated by unknown unknowns. In these situations, forecasting, following MAKRIDAKIS ET AL.'s (2009, 2010) definition, does not lead to any additional insight. However, even in these cases, forecasting can be useful again after the unknown unknown has taken place: its consequences become known, or at least estimatable. The uncertainty thus becomes known unknown. In other words, and to use the analogies defined by MAKRIDAKIS ET AL. (2009, 2010), coconut (uncertainties) could typically transform into subway coconut (uncertainties) after they have fallen. Consequently, most systems are inherently amenable to forecasting most of the time based on the uncertainty they are subject to. This leads to the question why so many forecasts still fail.

FORRESTER (2007) claims that the main reason for forecasts to fail is inherent in the underlying system structure. This is explained in **Fehler! Verweisquelle konnte nicht gefunden werden.**. The figure shows how the variable follows a path that is decreasingly increasing at e, until it levels off at the vertical line which represents 'today', the time of decision-making. The graph then shows two different possible future developments. Firstly, following Scenario A, the variable reaches a plateau and then further increases to reach a higher plateau. Secondly, following Scenario B, it reaches a peak comparatively early and then decreases at an increasing rate, which then slows down. The spread between the two paths defines the future uncertainty range. Despite the large spread in the far future, the system's momentum and inertia will prevent it from developing far away from an extrapolation of its past path during the forecast time horizon. Therefore, a valuable forecast can only be made within the forecast time horizon, during which the past momentum will continue (Forrester, 2007). FORRESTER (2007) points out that such short-term forecasts do not add much value as during this short amount of time, present decisions will not have any effect on the system. After the forecast time horizon, arbitrary occurrences that happen over time will increase uncertainty on the development path. Thus, regarding longer time forecasts, decisions made today may only affect the system during a time when forecasts are influenced by high uncertainty. In our opinion, FORRESTER here espouses a rather fatalistic view on forecasting, in which forecasts can be either accurate or valuable, but not both.

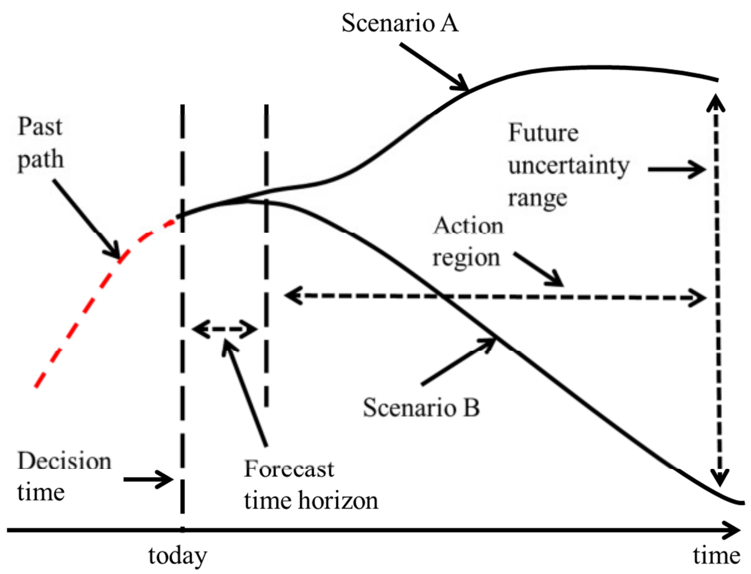

Figure 2: Forecast failure is embedded in the underlying system structure (Forrester, 2007)

Another explanation for the failure of many forecasts is that quantitative models are based on a limited amount of data that describes the past of the system – which inevitably leads to the following limitations (Makridakis et al., 2009). First, the future is not always like the past, extrapolation of past behavior does not provide accurate predictions. Thus, even sophisticated econometric models that show good fit to past data do not necessarily predict future developments. FOR-

RESTER (2007) supports this argument by emphasizing that econometric models seldom outperform naïve extrapolation of past trends. Simpler models, on the other hand, that do not fit past data, do oftentimes predict the future better than more complex models. Then, neither statistical models nor judgmental approaches have been able to capture the full extent of future uncertainty. Users of these forecasts might have been surprised by large forecasting errors. Moreover, the authors found out that expert judgment is inferior to simple statistical models. At the same time, forecasts done by experts are not necessarily more precise than forecasts by knowledgeable people. Instead, if predictions of several individuals are averaged, forecast accuracy is increased. Finally, averaging forecasts of two or more models also increases forecast accuracy. Summarizing, no forecasting approach has yet been able to capture the full complexity of reality, and thus none can claim to have eliminated the full extent of future uncertainty.

1.3 System Dynamics has been successfully applied to forecasting

According to MAKRIDAKIS ET AL. (2009, 2010) the ability to quantify subway uncertainties is beneficial for many applications in the economic and the business world. In spite of the inherent limitations of quantitative forecasting methods described above, MAKRIDAKIS AND TALEB. state that "the field of forecasting has done excellent work with models of subway uncertainty that can be assessed precisely and incorporated into all sorts of analysis to determine optimal decisions" (2009: 806).

FORRESTER (2007) claims that System Dynamics models, for example, provide value when used for forecasting, especially in situations of policy change and its effect on the system behavior in situations of continuing effects. In other words, System Dynamics-based forecasts should lead to insights into how different policies might impact future system behavior in different ways. Hence, they may provide the means of understanding the causes of certain system behavior by revealing the underlying system structure. LYNEIS (1999, 2000) describes how System Dynamics was effectively used to forecast the highly non-linear behavior of the airline market, and another successful client application of a System Dynamics forecaster is the shipping market model described by RANDERS AND GÖLUKE (2007), for example.

2 Research propositions

In the end of the previous section we described how System Dynamics has been successfully applied to forecasting, despite the illustrated failures. In the following section we describe how a System Dynamics modeling team extended the field of application of the methodology's tools, methods and insights to apply them to forecast short-term prices for a petrochemical commodity.

On the basis of the context explained above, we postulate the following research propositions. As laid out above, forecasting is controversially discussed in the literature. Even though practitioners ask for forecasts to run their everyday business, forecasts often fail. Within the System Dynamics community, it is stated that, if System Dynamics is applied for forecasting, it should be used for decision-makers to "change policies that will guide future decisions" (Forrester, 2007: 364). As this concerns future decisions, this is restricted to long-term forecasting, leading to:

> *Research proposition 1: The System Dynamics methodology should not be applied for short-term forecasting.*

According to FORRESTER (2007) it is because of the underlying system structure's inertia. Inertia is path dependent and results from the underlying feedback-loop structure. Thus, it is one of the System Dynamics methodology's objectives to identify the system's underlying feedback loop structure. It generates the observed behavior (Sterman, 2000). Consequently, System Dynamics should not be applied for systems without inertia, leading to:

> *Research proposition 2: The System Dynamics methodology should only be used to analyze systems whose behavior is dominated by feedback loops.*

Since understanding the past behavior is crucial for analyzing the nature of inertia and thus to estimate future performance, data are important. As described above, only systems with subway uncertainties where future events follow in the footsteps of the past are amenable to forecasting. Therefore, forecasting models will need to make careful use of the past performance data of the system – which leads us to our third research proposition:

Research Proposition 3: Carefully analyzing data about the past performance of the system, and carefully comparing model output to it, are crucial steps in ensuring forecasting accuracy.

Finally, there is agreement amongst practitioners and scholars (Makridakis et al. 2000) that if predictions of several methods and individuals are averaged, forecast accuracy is increased. We propose to evaluate this claim.

Research Proposition 4: Combining the predictions of different methods and/or individuals increases forecasting accuracy.

We will challenge the research propositions laid out above by introducing a client project in the petrochemical industry. Objective of the project was to forecast short-term price movements of a commodity.

3 Case study: Forecasting short-term price movements of a petrochemical commodity

A global petrochemical company, hereafter referred to as 'Alpha', approached us to help it to improve forecasting accuracy of its price for a certain commodity, hereafter referred to as 'Delta'. As Alpha was a major buyer of Delta it was economically critical for it to estimate possible future short-term and long-term price developments as accurately as possible.

Alpha had installed an internal team to manage a thorough forecasting process. The team consisted of company employees from different geographies and different departments, including purchasing, procurement, production, logistics, product management, key account management, and trading. The team met virtually on a monthly basis to discuss recent market developments and to estimate future price developments. The main focus for the latter was to produce four-week forecasts of both the price for Delta.

Alpha chose to focus on improving the forecasting accuracy for Delta because this commodity had become particularly difficult to forecast in recent years. Delta was produced in four different types of plants, and in most of them it was a byproduct and thus not the main product being produced. Consequently, the supply of Delta was neither only determined by its demand nor by its price, but also by the demand and price of several other petrochemical commodities. Furthermore, the demand of Delta depended on at least four major types of downstream products – in sum, this made Delta's price apparently erratic and highly volatile.

The commodities involved were traded both on a spot market and a contract market. In the latter, companies agreed to buy a certain amount over a full year, at a given price. This 'contract price' was linked to the spot price, varying on a monthly basis. Generally, the contract price equaled the spot price plus a small margin. It was fixed to the spot price at the beginning of each month. Thus, forecasting the contract price was essentially the same as forecasting the spot price. Alpha's objective was to be able to forecast the contract price one month into the future.

In the past, the forecasting team had been fairly successful in estimating the price development within a certain range. Yet, as mentioned the situation had changed significantly in recent years. Alpha's forecasting team was overwhelmed by the increased forecasting uncertainty over the previous two years, as the questions of when to buy, and whether to buy on the spot or on the forecast market became more difficult to answer. This resulted in significant consequences for the client's business performance – a major competitor even decided to sell its business unit producing similar commodities because they had become so unpredictable.

In order for Alpha to decrease the uncertainty about the future evolution of the price of Delta, Alpha's management decided to hire external consultants to support them in better understanding the market, to support the forecasting process, and to increase forecasting accuracy.

Before they hired us, Alpha had already tried to solve the forecasting challenge with at least six other consultants. However, their use of statistics and other more mainstream methods to shed light into the forecasting market had not met Alpha's requirement of forecasting accuracy: according to the client, Alpha's internal forecasting team was still outperforming the other consultants' forecasts. This fact just proved the difficulty involved in forecasting the price for Delta.

In the following we will describe how we used System Dynamics tools, methods and insights to design a process that eventually proved successful.

4 Modeling the petrochemical market

4.1 The chosen modeling approach

As stated above, Alpha hired us to support it in understanding Delta's market better, and in particular in forecasting the short-term price movements of Delta. The project lasted for about twelve months, including six months of model building and six months of real-life model testing. The modeling team consisted of a five consultants, including a project director, a project manager, a lead model builder and data collectors. We applied the traditional System Dynamics modeling process described by STERMAN (2000) from problem articulation over the design of a dynamic hypothesis, formulation, testing to evaluation. During the modeling process we laid special focus on data collection and model validation. In the following we describe our general project plan and these two sub processes in more detail.

In the authors' experience, is it generally accepted on System Dynamics-based projects that the client's organization provides the subject matter-experts, and the consultants provide the process expertise (Lyneis, 1999). In this case, recognizing our lack of expertise in the petrochemical industry at the outset of the project, the client designated a team of company experts to work alongside us. We provided the frameworks, processes and analytical skills, and the market experts supplied their industry knowledge and insight about Delta. The potential damage to the project that might be caused by the consultants' initial lack of market knowledge credibility amongst Alpha's team of market experts was a necessary evil. Yet, it was more than compensated by our process experience and skills developed in many similar client projects in different situations and industries..

However, bearing in mind the potential risk that the lack of credibility of the modeling team represented for the success of the project, the project plan was designed so that each of its phases would produce two types of deliverables: a first one, which could be called 'long term insight', consisted of puzzle pieces that the modeling team required to build the forecasting model; a second one, which might be called 'immediate insight', consisted of new information about the price dynamics of Delta that the team hoped to identify. Latter was intended to support establishing our credibility in front of the client.

Advantages of this approach are obvious:: modelers are able to approach the managerial challenge without any preconceived industry or market assumptions that market experts have usually been burdened with. This allows modeling teams to start from a blank sheet of paper, gather facts and insights from experts, and ask challenging and sometimes 'stupid' questions that an insider might not think of (or dare) to ask, but which sometimes prove crucial in highlighting inconsistencies in the mental models of the experts. The combination of a careful application of a rigorous approach and being allowed to pursue the quest for information freely enable modeling teams to form a consistent, non-contradicting dynamic hypothesis (Forrester, 2007; Sterman, 2000) as to how a market works.

We further planned to use triangulation (Yin, 1994) in order to fully understand the managerial challenge. Generally, triangulation enables researchers to gain a broader scope of historical, attitudinal, and behavioral issues and hence balance the drawbacks of one method with the advantages of others. We also planned on conducting semi-structured, open-ended, and focused interviews (Andersen et al., 2012) with participants of Alpha's forecast team. We further intended to analyze the client's internal market analyses, archival records, and other documents of relevance throughout the project. The interviews to be conducted needed to be transcribed and analyzed. Regular meetings with the client were set-up to present and discuss the dynamic hypotheses and how we expected it to replicate the observed behavior of the market.

Historical data of the main market variables was crucial to understand the dynamics involved of the underlying managerial challenge. The authors' experience has shown that data collection is usually a particular challenge in projects using System Dynamics. Data was going to be required in order to identify and analyze past behavior of the main variable by establishing reference modes (Lyneis, 1999; Sterman, 2000) and to calibrate the model's behavior against past performance, as part of conducting a thorough model validation process (Sterman, 2000).

Calibrating the project's model against real data is important for such forecasting projects because of two aspects: Firstly, a fit to data can sometimes be the only way to test which part of the model structure is more likely to have generated an observed behavior, following MORECROFT's (1985) and HOMER's (2012) concept of conducting partial model tests. This is done in addition to assessing the appropriateness of underlying assumptions, the robustness, and the sensitivity of results to assumptions about the model boundary and the feedback structure in general (Sterman, 2000). Secondly, the client has also more confidence in the model and its structure if the model is able to generate past behavior (Sterman, 2000). Yet, we were also aware of the fact that we needed to be careful when handling data in the modeling process. FORRESTER (2007) emphasizes how misleading fit to data can be if done improperly. When a simulation model achieves great fit to data but does so based on simulating causal mechanisms that do not match those present in the real system, there is no guarantee that the model's simulations will continue to be accurate when extended into the future. In a project aimed at forecasting the future, 'curve-fitting' had to be strictly avoided.

Moreover, data was also going to be required to allow modelers to carry out 'backward-looking forecasts'. Backward-looking forecasting enable a modeling team to assess whether the model is able to provide accurate 'true' forecasts, i.e., to deliver simulations that would have accurately matched the future performance of the system, using only data that would have been known at the time of forecast.

Fehler! Verweisquelle konnte nicht gefunden werden. Figure 3 illustrates this principle. The x-axis shows the time. It is assumed that data for a certain variable, indicated by the dashed line, are available until 'today'. In this example we further assume three points in time for backward-looking forecasts: 1, 2, and 3. Backward-looking forecasts are carried out at each of these points. As mentioned above, if a backward-looking forecast is conducted at point 1, for example, it is assumed that data are only available until that point in time. In other words, data input to the model is stopped and the model continues to run without external data input for a limited amount of time. Consequently, the model produces the behavior for the period of backward-looking forecast – the forecast horizon – itself. This is repeated at times 2 and 3 in this example.

Benefits of this procedure are twofold: not only are modelers able to assess the model structure according to the model behavior reproduction test. It also enables modelers to ensure that the model structure is able to forecast the 'future behavior' of the variable.

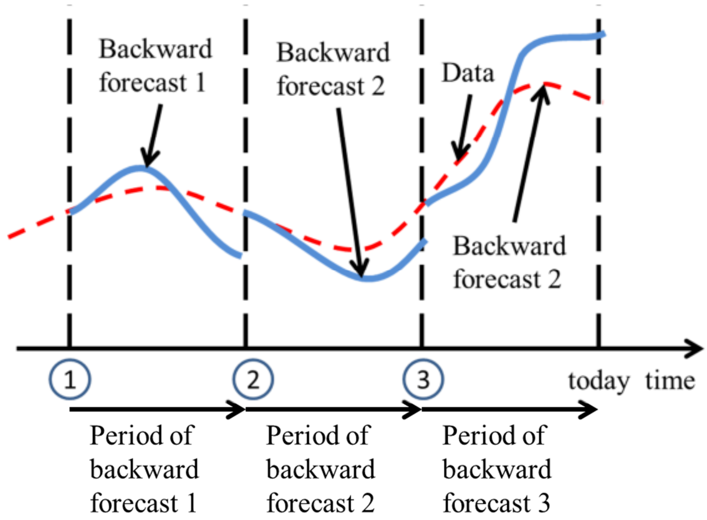

Figure 3: Process to conduct blind backward-looking forecasts

Doing so is not enough to prove that "a model is correct or reliable" (Sterman, 2000: 879). However, because of our prior experience we found that if carried out sensibly, a fit to data can sometimes be the only way to test which model structure is more likely to have generated the behavior.

4.2 On client site: Gathering information and analyzing data

As described above, the production of Delta is part of a highly complex system. It involves several chemical processes, and in many of them Delta is just a byproduct of more widely consumed commodities. Furthermore, Delta is not bought by end consumers, but instead it is used as an input to produce several further downstream petrochemical commodities. Our starting hypothesis about the structure of the market for Delta was that it had something to do with the balance of its supply and demand, as is generally the case in commodity markets (Sterman, 2000). It was thus was necessary for us to understand this complexity at least at a basic level.

For this reason, we structured the initial round of interviews as follows. During the first four weeks of the project we conducted fifteen focused one-on-one expert interviews with chemical experts, procurement managers, and commodity traders. We started by interviewing the experts who were able to provide us with a broader overview of the overall market. We then used the information thus obtained to plan the following interviews, identifying the questions to be asked to the interviewees with a more focused area of expertise.

Following the interview process we analyzed the transcribed interviews to identify data items that might support the interviewee's statements. The historical spot and contract prices for Delta had been the obvious initial targets of our analysis, followed by the price for the raw materials involved in Delta production, annual production volumes for these and other related commodities, inventory levels, etc.

This process could have derailed into an open-ended search for any data or information related to a large number of petrochemical commodities. Instead, we used our initial hypothesis about the workings of the market to structure the questions to be asked. We constantly updated our hypothesis with the information received in the interviews, and identified parts of the model that had already been documented by previous interviewees and other parts that required further evidentiary support or clarification.

The focus provided by this hypothesis-driven approach was validated by the client's manager at the end of the project. He commended the team for how quickly it had been able to come to grips with the complexities of Delta's market.

The structure in Figure 4 shows one of the earlier states of the initial 'market hypothesis'.

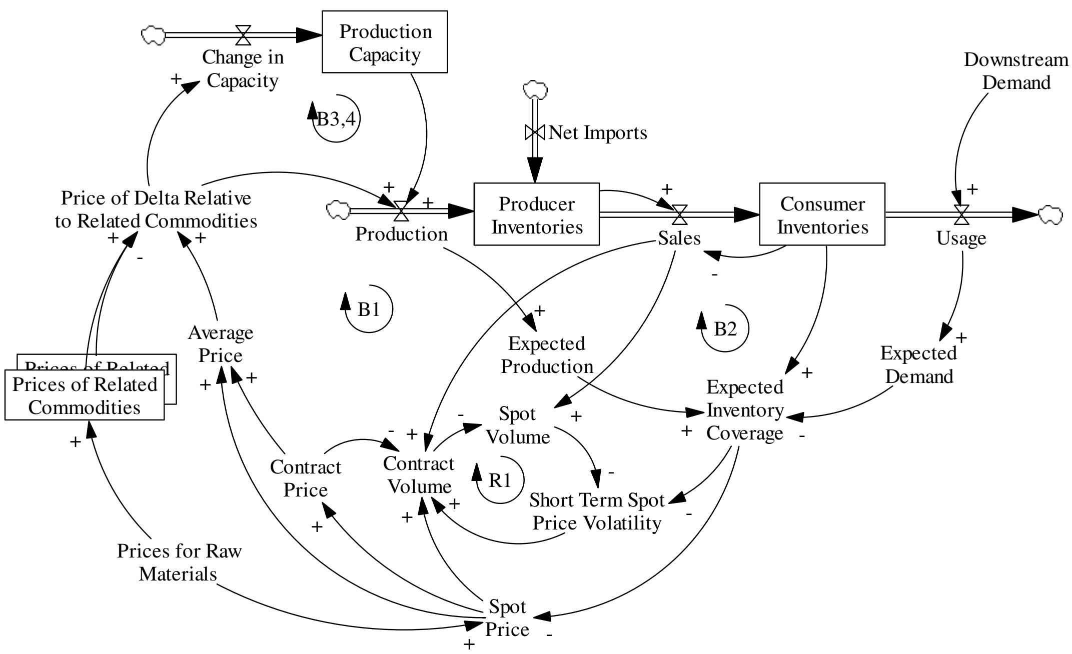

Figure 4: Initial dynamic hypothesis of Delta's market structure

This initial hypothesis was based on standard assumptions about how commodity markets typically work: commodities are produced and then stored first at production site, later at the consuming site – in Delta's case both sites are small and amounted to just a few days of production. Low inventory levels plus expected demand is assumed to increase spot prices, and consequently contract prices as well. When contract prices are lower than spot prices, more production tends to be supplied via contracts. The contracts allow for month-to-month variations in the amounts supplied, as long as the annual total remains fixed. This in turn leads the spot market to become tighter. The subsequent spot price volatility pushes further market participants in favor for contracts over spot arrangements, thus reducing the volume available on the spot market (R1 in Figure 4).

In case commodity prices stay high for a longer term, the industry becomes more profitable. With increasing profitability it becomes more attractive for companies to produce more of the product. Hence, they invest in expanding production capacity, and capacity slowly increases. The balancing loops B1 and B2 show how production rates react to market conditions and provide more supply when prices increase. Assuming small inventory levels, the actual increase in supply available for purchase (B1) and the expectations for this (B2) happen almost simultaneously, both countering the initial price increase.

The two balancing loops B3 and B4 describe how production capacities are involved in similar dynamics, although the delays involved are generally several orders of magnitude longer. Finally, all prices in the petrochemical industry are eventually influenced by the price movements of oil and its immediate derivatives, which provide the basic raw material.

As a next step in the data collection process we analyzed the interview transcripts to identify potential drivers for the price of Delta that the market experts had mentioned to us. We also cross-checked them against each other and against the historical data record, and discussed the apparent discrepancies that we found in group sessions with the market experts.

One of the most insightful findings in this phase was that prices of raw materials seemed to play a far more important role in determining the price of Delta than the experts had been actually expecting. During the interviews, we learned that raw materials had indeed played a fundamental role in establishing the price of Delta in the past (see Figure 5), but that in recent years the market had “gone crazy, and now nobody really understands what drives the price of Delta anymore [sic.]”, as one interviewee stated.

Figure 5 illustrates this change. The figure shows the price development of Delta and that of its raw material over a time period of nine years. As can be seen, the two lines follow very similar patterns. As expected (see also insights from the Beer Game described by STERMAN (1989, 2000)), the amplitude of the variations in the price of Delta are clearly larger than that of the raw material. Yet, the timing of the ups and downs in both curves seem a clear match. We learned that small inventories in Alpha’s supply chain as well as the speed with which production rates can be adjusted up and down, justified the absence of a distinct time lag between both curves. Also, the price of Delta is generally higher than that of the raw material, as would be expected due to the cost involved in the additional processing steps.

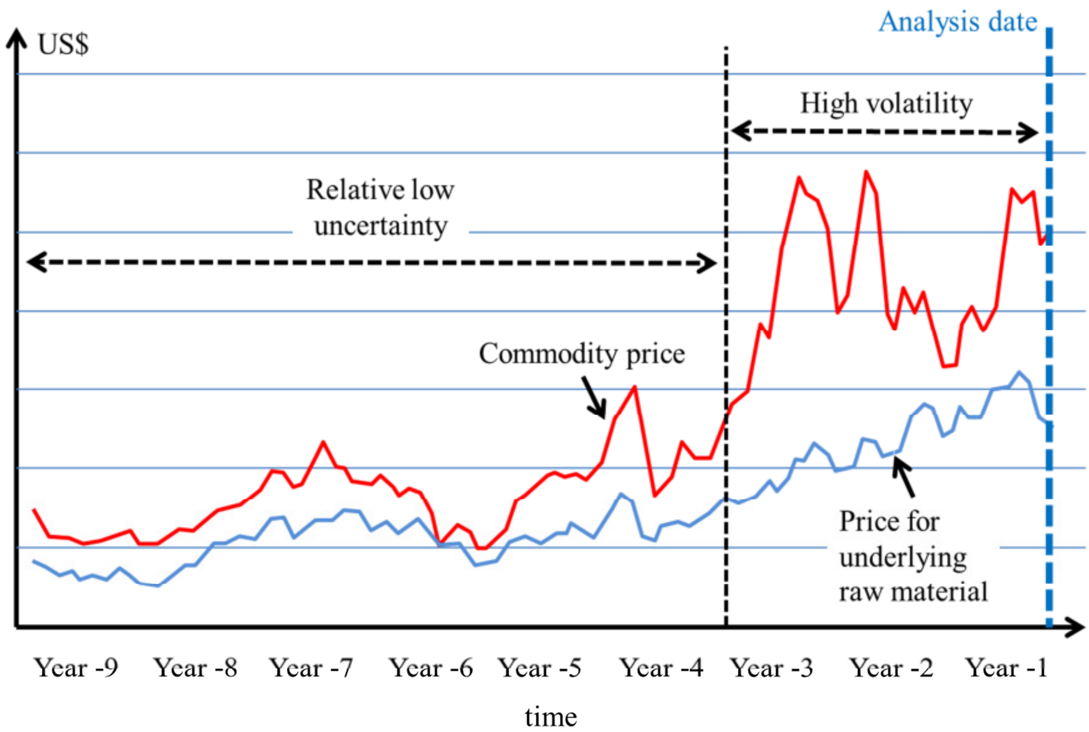

Figure 5: Past price development of petrochemical commodity Delta and its raw material (monthly prices, numbers slightly modified for anonymity reasons)

As can be seen Figure 5 the nine-year-period can be divided into two sections. The first period is a time of relative low uncertainty. The prices of the two commodities seemed to be connected and develop fairly parallel from Year -9 until Year -4. During this period forecasting seemed to be fairly straightforward: the short term price estimate for Delta could apparently be explained as a function of the variation in the price of the underlying raw material, with the spread between them depending on the 'market tightness', the difference between production capacity and actual production – and thus consumption[1] – rate of Delta.

Then, in the period from Year -4 to the date of the analysis, uncertainty increased significantly and the price of Delta experienced extreme volatility, while the price for its raw material followed a much steadier pattern. To the client, Figure 5 clearly illustrated the challenge that the forecasting of Alpha had become: the market had recently abandoned its traditional rules, and was now following new ones that had yet to be understood. However, despite the market experts' opinion it turned out that reality proved to be simpler, and a careful analysis of the fluctuations in the data showed that while the price volatility of Delta had indeed increased significantly after Year -4, its upward and downward movements still roughly matched those of its raw material.

Thus, the analysis apparently demonstrated that the correlation between variations in the prices for both commodities had held steady throughout the nine-year period for which data were available. The only thing that seemed to have changed during the last three years was Delta's *beta*, the strength with which its price reacted to the fluctuations in the price of the raw material.

Analyzing the interview records, we found a likely explanation for the step-change in the volatility of Delta's prices, which seemed to have been driven by a structural one-time change in the process that was followed to determine market prices. Thus, the team had obtained two main conclusions from the data gathering and analysis phase: first, discounting the one-time change in market structures just mentioned, the market seemed to be driven by essentially the same factors throughout the nine-year period reviewed – in other words, in spite of appearances the market was still essentially the same one that the industry experts had known from before. Second, the analysis allowed us to arrive at a consensus with Alpha's market experts regarding the likely two

[1] Due to the small inventory levels mentioned above.

key factors driving the dynamics of Delta's prices, which had not existed before the beginning of the project: the price of the raw material for Delta, and its market tightness.

4.3 Simulating our understanding of the market

At this stage we had developed an initial dynamic hypothesis about the workings of the market for Delta, grounded in integrating the knowledge of the client's experts with information extracted from the historical record. Still, in order to be sure of its accuracy, the hypothesis had to be tested. Thus, in the following weeks of the project we built a first simulation model with the sole objective of replicating the historical performance of Delta's price. If our hypothesis about how the market worked was correct, the model should be able to reproduce the general historical pattern followed by Delta's price.

Since it was our intention to test the hypothesis about our understanding of the market basics, we concentrated our efforts on the two factors that had emerged during step one (the price of raw materials, and the general 'tightness' of the market for Delta), ignoring all other indications on price drivers. We were aware that these assumptions resulted in a fairly simplified simulation model that was not able to accurately explain all the ups and downs of Delta's price for the previous nine years. Yet, if the simulation model was able to reproduce the general patterns correctly, it would be a strong indication that we had understood the main market dynamics correctly, giving us a sound foundation on which to build a forecasting model.

Figure 6 shows the basic structure of this simple market explanation model: changes in the price for Delta are driven by changes in the price of its raw material, multiplied by *beta*, a constant factor modulated by two variable effects. The first of these effects is tightness of the market for Delta, measured by its annual capacity utilization.[2] This variable was chosen because data were available and it depicted the balance of supply and demand. The second effect includes the known one-time structural change happening in Year -3. This indicates the change of the way

[2] Capacity utilization = volume actually produced / maximum production capacity. Because of the low levels of inventories, the volume produced can in the first instance be used to approximate the volume purchased.

market prices were established., which we because of reasons for simplification, represented as causing a stepwise increase in volatility followed by a 3rd-order exponential decay.[3]

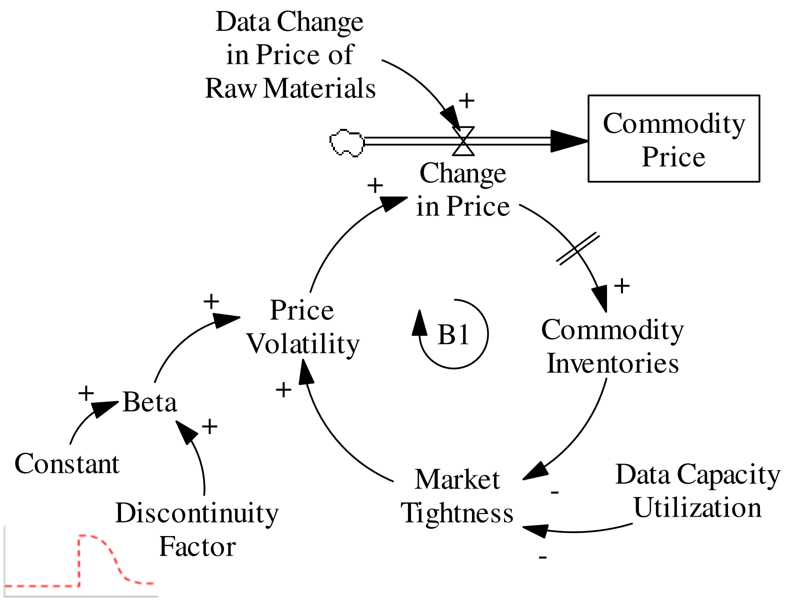

Figure 6: Basic structure of the simple market explanation model

The main feedback structure in this model is the loop labeled as B1 (see Figure 6). This balancing loop represents the impact that customer inventories have on prices: when inventories are full, new purchases are dragged out and/or reduced, thus reducing market tightness. This creates what is known as a ‘long’ market. Since Delta production volumes largely depend on factors other than its price, we decided that our first attempt at explaining the market ignores this factor altogether, initially implicitly assuming that short-term changes in production volumes have at most a secondary impact on price changes.[4] The reduction in market tightness then leads to a reduction in volatility, which reduces the amplitude with which Delta’s prices react to variations in the price for the raw material.

[3] Since the structural change was a one-time occurrence in the nine-year historical record and thus unlikely to happen again in the future, it was not deemed necessary to model their nature any more closely (this was decided jointly by the modeling team and by the client). The nature of the structural change cannot be disclosed in this article any further because of confidentiality reasons.
[4] All models developed for this project were produced following a spiral development process: the core dynamic hypothesis was modeled first, and then the team iteratively added as many additional structures as proved to be necessary in order for the model to be able to reproduce the expected market behavior with sufficient accuracy.

Note that while we label this as a balancing loop, assuming 'positive' links between market tightness, price volatility and the actual change in prices, this has been done for diagramming convenience. In reality, however, the relationships are more complex in nature. Essentially, market tightness amplifies the impact that other effects have on Delta's prices – and it does so in both directions, the effect is symmetrical. We initially expected market tightness to have a biased impact on volatility and thus on price changes. Yet, the calibration process produced unequivocal results as tight markets clearly led to more violent movements in prices. But this seemed to affect upward and downward price movements equally.

This structure was able to produce the results shown in Figure 7. As can be seen, the simulated output (green, solid line) matches fairly accurately the historical data (red, dashed line). The match, while not perfect, is surprisingly good for such a simple model [5]

Figure 7: Historical price data of Delta and simulation output

[5] Note that there is some increased short-term volatility of the simulation compared to the data; this can be explained at least in part by the fact that the simulation reproduced the daily variation in spot prices, whereas the data was updated much less frequently.

Yet, looking at Figure 7 it becomes evident that we were facing a serious problem. While the match between simulation results and historical data was fairly tight, it seemed to completely break down in Year -3. If this apparent breakdown were to be confirmed, this would obviously render the simulated hypothesis useless, since it would no longer apply in the present time and therefore be an unsound basis upon which to forecast the future.

However, when looking at the simulation and the data more carefully, a surprising conclusion emerged (see Figure 8): following a 1-2 month 'disruption' in which the simulation diverged from the data, the simulated prices again followed the same pattern as the data, just displaced a constant quantity downwards. In other words, the facts seemed to indicate that something had happened in year -2 which disrupted the market for a brief time and lead to lower prices, after which the market continued to operate according to its usual mechanisms and drivers, as if nothing had happened.

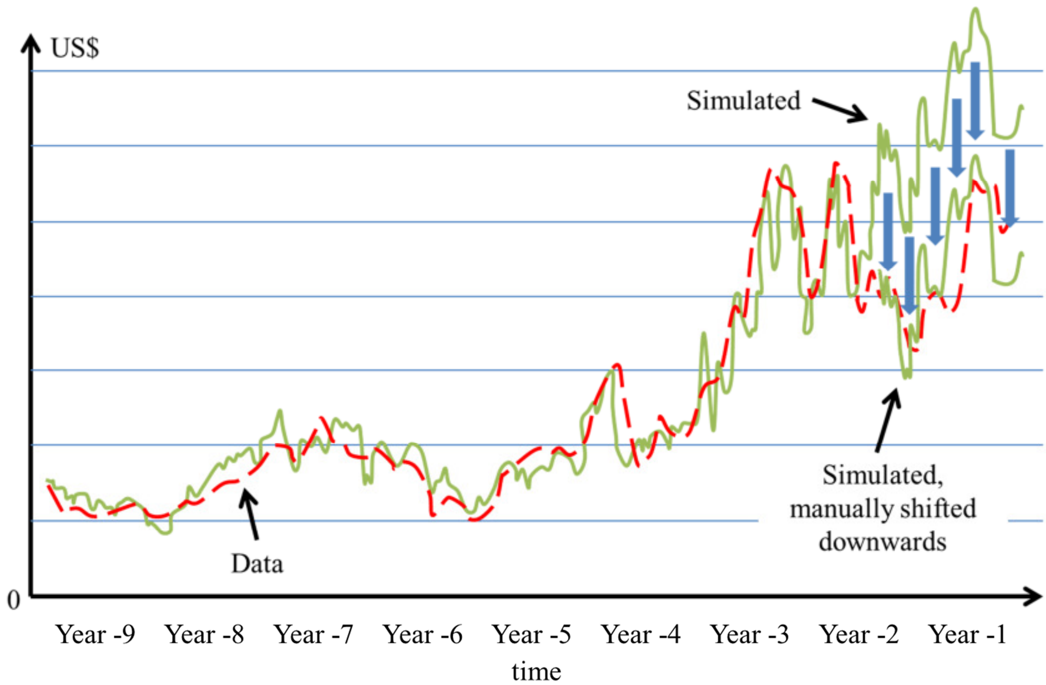

Figure 8: Model output started to diverge from reality in Year -2

This can be seen much more clearly in Figure 9 which highlights the final three years of the graph in more detail. Pricing data after the discontinuity can be matched by drawing a parallel curve to the simulated prices, just displaced downwards.

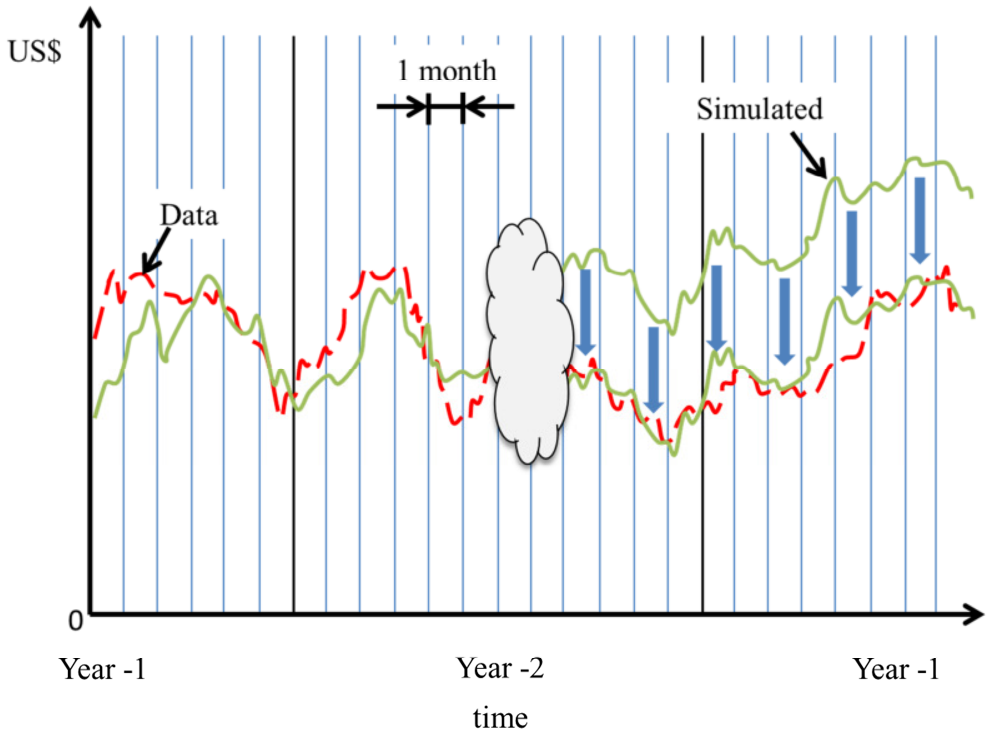

Figure 9: Detailed view on how the 'disruption' led to depressing prices without changing actual market dynamics

Discovering the displacement did not yet mean that we identified an explanation – or dynamic hypothesis – for this behavior. Yet when investigating the historical record, we did indeed find an unprecedented major global disturbance that did not have anything in particular to do with the petrochemical industry. The incident happened just at the time when simulated prices began to diverge from historical data. It was important enough for the general press to report on it globally, and Alpha's market experts were also able to explain its influence on Delta's price. With this final piece of information, what at first had seemed a problem, at this point became valuable market insight.

First, the hypothesis about the factors that drove the price of Delta seemed to be validated, since outside of two months in which clearly non-standard conditions applied, the model behaved quite accurately. Second, the dynamics originated by the disturbance – or, rather the lack thereof – provided us with valuable information about how Delta's market functioned. It did not show any memory, as it continued to react to current events after the disturbance as if it had never happened.[6] And third, this issue validated our structural choice for simulating price changes: portraying a market with no memory (no price anchor) had actually been quite a controversial choice at the time when we were building the model. This was because it assumed a market behavior that, while observed previously[7], was still somewhat extreme. However, during the development we had tried several different pricing structures (Sterman, 2000), and the 'no memory' one was the only that allowed us to reproduce the historical record with any degree of acceptable accuracy. It outperformed other alternative dynamic hypotheses for price formation mechanisms.

Recapitulating, roughly six weeks into the project, the team believed to have identified and quantified the price forming mechanisms of Delta's market. It further had produced quite a bit of new information on how this market worked. However, while this may have been valuable in itself, the client's main interest remained in forecasting the price of Delta a month into the future.

4.4 Forecasting the market

The sound understanding of the market structure allowed us to develop our 'market explanatory' model further into a forecasting model in the subsequent project phase. The first step in the development process required us to build structures into the model that switch off the use of data inputs, and turn on forecasts for such inputs when simulating future periods; the second one was to acknowledge that the accuracy of the explanatory model was still lacking, especially regarding short-term forecasts, and to explore additional dynamic hypotheses that were to be added to the model in order to improve its forecasting accuracy to meet client requirements.

Regarding the second point, while the model captured the medium-term to long-term pricing trends relatively well, it was still often off on a month-to-month basis. Thus, while continuing to

[6] In other words, the market was (and presumably still is), path dependent.
[7] This same lack of market memory was also detected in a previous project by the authors, where they were asked to determine the driving factors of a totally different market. For confidentiality reasons we are not allowed to disclose the commodity in question, nor the identity of the client.

believe in the validity of the major price-driving factors identified in the previous phase, we reanalyzed the interview transcripts and group meeting minutes of phase one to produce a list of potential other factors, which might help explain and thus eliminate the short-term deviations.

Regarding the first point, we added structures that would switch off the use of input data at the forecasting date, and then continue to simulate the price of Delta up to the forecasting horizon by using as model inputs only data known at the forecasting date. We further decided to work with 'backward-looking forecasts' in order to calibrate the relative strength of these potential additional factors. To support the backward-looking test-forecasting process, the team had a dataset comprising 30 months.[8]

While conducting backward-looking forecasts in a true 'blind' fashion significantly lowers the risk of falling into 'curve fitting', the tests conducted here were not truly 'blind'. According to theory we should have used half of the dataset, in other words, data for the first 15 months of the simulation time, to calibrate the model, and the remaining half to test the model's forecasting accuracy on an *ex post* basis. However, the team realized that the number of data points available was too small. So, 15 months' worth of data were not able to generate enough information to support an adequate calibration, since some of the price-driving factors that we needed to consider were annual in nature, i.e., they would have only appeared once or twice in the calibrating half of the dataset. Thus, we instead decided to follow the only other viable path open to us, and use the entire dataset for calibration.

By violating the 'blindness' of the backward-looking forecasts, there was the danger to force-fit the model to historical data (Forrester, 2007). Since there was no other option than to take this risk, we tried to at least reduce it as much as possible. We kept the number of price-driving factors as small as possible, accepting only factors that seemed to have a clear and explainable influence throughout the historical dataset, and which were not correlated to other factors used. We also used a few factors which showed a clear intermittent yet equally explainable pattern.[9]

---

[8] By automating this process we were able to develop what was semi-jokingly known as the 'hairy snake': this referred to the picture produced by the base simulation of historical prices, on which were superimposed the 30 into-the-future simulation extensions that produced the different test one-month forecasts.
[9] This would include the annual effects mentioned earlier in the text.

At this point we faced two major modeling challenges. Firstly, there existed no reliable forecasts on the price of Delta's raw material – consequently we considered producing our own. Secondly, the dataset provided by the client was not complete. Some key data items, like capacity utilization, for example, were only provided on an annual basis even though monthly or weekly data was required to conduct the forecasts. In the following we will focus on the former issue, since it became insurmountable and thus the second challenge became (at least temporarily) irrelevant.[10]

As mentioned, we could not use commercially available forecasts for the price of the raw material as they were basically flat, adopting the logical position that if reliable information about the future was not available, the only possible unbiased forecast price was the current price. However, actual raw material prices did vary from month to month, and thus caused the price of Delta to move as well – so we clearly needed a better forecast for the price of raw materials.

We decided to accept this unplanned challenge. In order to adequately forecast the price of the raw material using System Dynamics, though, it became obvious that we then needed to simulate the worldwide oil market – which laid clearly beyond the scope of the project. MAKRIDAKIS ET AL. (2009, 2010) analyze the difficulty of forecasting the oil price. Oil price development heavily depends on the actions of many different interdependent market players and factors, including human actions in finding the equilibrium price between supply and demand: OPEC that is interested in high prices, oil-importing countries that are interested in low prices and speculators that follow their own objectives. Because of these market participants acting simultaneously and news being overreacted to, it is difficult to forecast oil price development in the long, medium, and short term. For being able to conduct forecasts, the price's distribution of errors needs to be normally distributed, constant, and the errors need to be independent of each other. According to their findings, "uncertainty in forecasting future oil prices is not constant, therefore, and depends on whether we are in a usual or high-volatility period" (Makridakis et al., 2009:808). As a consequence, referring to whether uncertainties can be quantified, it is not possible to quantify the uncertainty of the oil price.

As an attempt to substitute causal, system dynamics forecasts, we also tried to produce stochastic and curve-fitting forecasts, but these approaches were not able to produce results accurate enough

[10] This point did eventually become relevant - see the subsequent section for details.

to support meeting the forecast accuracy required for Delta's price. Consequently, this modeling attempt failed, which put project success at risk: against the expectations of the team and of the client, the dynamics of supply and demand had proven to be much less dominant than the influence of the price of the raw material, and forecasting the latter with any degree of accuracy had (less surprisingly) proven to be impossible within the scope of the project.

4.5 Simulating expectations – could the solution be a model without feedback?

According to the client's viewpoint, our attempt of forecasting Delta's short-term price development had not been accurate enough. Therefore, at this point the client demanded a decision: it would no longer financially support the project if the team was not able to quickly propose a solution that might lead to expecting satisfying results.

At the same time, we wondered why we had not been able to meet the client's forecasting requirements. So far, the client's expert team had been able to soundly beat our model's forecasts accuracy. And yet, the client must clearly have believed that its current accuracy levels could be improved upon, as it was hiring us to achieve just that.

Developing a different – and the most insightful – hypothesis

At this crucial point in the project, following the *lex parsimoniae*[11] proved to be the road to follow. We hypothesized that the simplest explanation for the surprising accuracy of the experts' forecasts might also be the likeliest: the experts' forecasts could in fact be self-fulfilling prophecies. In spite of the large volumes being traded, the market for Delta included a relatively small number of buyers and sellers, all making similar decisions and looking at similar and sometimes even identical data. Thus, even though market participants are mostly competitors and thus hardly share any information amongst them, it was still reasonable to expect that they just might react to new data in roughly similar ways.

Figure 10 illustrates our hypothesis. Expectations of market participants about future price movements actually lead to the participants to adjust their buying and selling volumes 'now', to take advantage of said movement: for example, buyers buy higher volumes before prices are ex-

[11] This is more widely known as 'Occam's [or Ockham's] razor'.

pected to increase, thus creating additional market tightness that pushes actual prices. Similarly, if prices are expected to decrease, buyers wait to buy at a later point in time, creating a situation of oversupply that actually depresses the market.

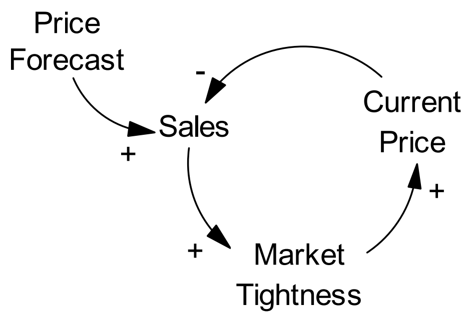

Figure 10: The team's hypothesis of short-term self-fulfilling prophecies

Then, in the longer run we assumed market fundamentals to assert their dominance by forcing market participants to adjust their expectations to match the evolving actual conditions in the market. Our confidence was that price expectations would not change sufficiently *within* a month 'under normal market conditions'. In other words, we hoped that the dynamics generated by the forecast (done at the beginning of each month) would continue to drive prices throughout the month (and thus the price at the end of the month, the actual objective of the forecast.)

To summarize, in order to be able to forecast the market we believed that we did not need to forecast the evolution of the fundamentals affecting the market, but rather how this evolution affected the perceptions of the market participants. By including humans' bounded rationality (Camerer, 2003; Granovetter, 1985), we were following an oftentimes postulated tenet of System Dynamics models (Größler, 2004; Morecroft, 1982; Sterman, 1987, 2000). This assumption also follows FORRESTER's (2007) argument explained above (see **Fehler! Verweisquelle konnte nicht gefunden werden.**), stating that market fundamentals have inertia.

Following this insight, we had only two weeks in which to validate its hypothesis, and produce acceptable forecasting results before the next scheduled meeting with the client. This required an update of the forecasting model to drive it with market participants' expectations. And, we also had to find a source for additional data about a range of potential short-term impacts on Delta's supply and demand, which the client had thus far not supplied.[12]

To find more short term information we decided to review the last two years' worth of a pair of well-known industry weekly bulletins. We analyzed them systematically applying content analysis. Via capturing and categorizing relevant information we gained information on expected plant shutdowns, prices for a whole number of upstream and downstream commodities, and qualitative information regarding the sentiment of the market, abnormal conditions, and backwards-looking information explaining some of the price movements in previous weeks.

In order to test whether the new hypothesis worked as expected, we needed to update the model with the new hypothesis, include the data, and then run 'backward-looking' test forecasts. This 'standard' process meant that we had had the first inkling about the capabilities of the new model only after modeling and testing – i.e., right before the end of the two-week deadline. In other words, we no longer had time to update the model if initial results were not as planned.

Developing a qualitative model to forecast Delta's short-term price development

To reduce the risk described above, we decided to move testing ahead, and perform a critical first round of testing *before* the simulation model was updated. Since we did not yet have a computer model available, we had to use mental models to perform the test. Under normal circumstances a mental model would be a poor substitute for a simulation model (Sterman, 1989, 2000), yet in this instance the incidence of feedback was expected to be so low as to make the tests accurate enough to provide a good indication of the validity of the new modeling approach.

The mental model used was the one developed during the earlier stages of the project: we expected the price of Delta to react strongly to expected changes in the price of the raw material (approximately 1:1). The *beta*, the relative amplitude of the variation in Delta; depends on the tightness of the market; this variable was estimated based on the annual capacity utilization data, adjusted by any known (planned) plant shutdowns and other expected supply-reducing factors. Knowing the overall size of the market and the production capacity of all of its production plants,

[12] This problem had already been noticed in earlier stages of the project, as described earlier in this article. While the team had dedicated some resources to try to solve it, this was never a top priority since there had hitherto always existed more fundamental problems to contend with.

we were able to estimate the relative size of any ‘supply event’, and thus transform it into percentage points of capacity that could be consistent with the expected structure of the simulation model.

To prepare for the mental testing, we used the dataset comprising the 24 months at-hand. For each of these 24 months we plotted, on a separate sheets of paper, the historical price curves for Delta, the raw material and two other commodities that had been identified in the project’s initial interview rounds as especially relevant. On each of these sheets of paper the curves stopped at the date that each forecast would have been conducted. We then looked at the trend in the price for the raw material and extrapolated it one month into the future, thus establishing a ‘base forecast’ for the month. We then also read the two last newsletters published prior to the forecast date, and learned about any other potential issues that might need to be contemplated in the forecast. Based on this information we just adjusted the ‘base forecast’ up ‘or down a bit’ – it was deemed that the qualitative nature of the process would not warrant any attempt at higher accuracy at this point.

The results of this ‘manual forecast’ test phase looked promising. As the mental model applied seemed to work very well in most months, we decided to build the underlying simulation model. Within one week we had developed the quantified simulation model, calibrated it, and prepared the results to be shared with the client. As expected, based on the mental model described above, the final simulation model was hardly dynamic (see Figure 11).

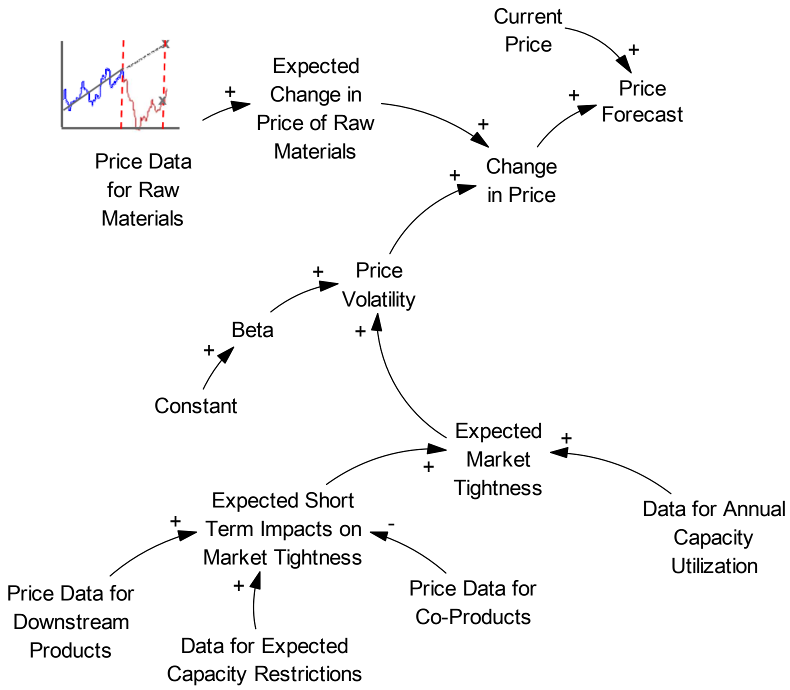

Figure 11: The final, simple model to forecast Delta's short-term price (simplified view)

Developing the quantified model to forecast Delta's short-term price development

The quantified model read current prices for Delta, and forecast its values one month into the future by simulating a series of 'pressures' on Delta's price. These pressures resulted from the expectations for changes in the price of the raw material and a series of related upstream and downstream commodities, expected plant shortages for these commodities due to ongoing shut-downs to take into account expected restrictions in supply and demand, and a few other factors.[13] The interesting point to note here, as stated above, is that the only dynamic elements left in the model were the structures that reproduced the formation of the expectations, which used 'dynamic' third-order delays to estimate the different variables' trends (Sterman, 2000). Apart from that, there was no feedback in this simulation model.

The diagram in Figure 11 is a simplified representation of the final model delivered to the client. However, the model depicted here was not the first 'expectations-based' model that we built: based on the expert interviews conducted during the earliest stage of the project, there were at

[13] We are not at liberty to mention these since they are very particular to this market.

roughly 15 potentially significant price-driving factors that we needed to consider in the model. However, since our model calibration dataset only contained 24 data points, it became evident that we could not use all these factors, since that would almost inevitably lead to curve-fitting – we thus focused on the most relevant.

This set off an iterative test stage, in which we tested many different combinations of influencing factors. There was a general consensus amongst us, the client's experts and our review of the historical record and of the results from our previous models, that three of these potential factors were certain to play a significant enough role as to require their inclusion in the model. To these we ended up adding another two, which we had found to significantly and consistently contribute to improving the accuracy in the 'backward-looking' test forecasts. Finally, we added another two factors that were known to happen at most once every year: it was not clear whether the model would be actually able to forecast them accurately in the future, but their strong (yet short-lived) impact on Delta's prices required that we at least attempt to do so, and include these factors in the model.

As stated, a dataset of only 24 months was available to support the calibration of the forecasting model. Considering the minimum number of pricing factors that we needed to consider, we did not have enough data points to split the dataset into two sets and produce real 'blind' backward-looking forecasts. Instead we decided to produce 24 'test' forecasts, and reduce the danger of 'curve fitting' by following the same procedures used for testing our earlier simulation model.

Figure 12 shows the results of the backwards-looking forecast tests after the conclusion of the iterative model-development stage. The x-axis depicts the time of 24 months, indicated by the 24 vertical grey lines. The y-axis indicates Delta's price with the solid green line indicating Delta's actual contract prices development, and the corridor around them shows the forecast deviation allowed by the client. Finally, the red diamonds show the 'forecasted' contract price. As can be seen, the simulation model was able to meet the client's strict accuracy targets in all but three of the 24 cases. This is more clearly shown by the red line in the lower part of the graphic, which indicates the monthly forecast deviation in percent. Furthermore, the model was also able to forecast the direction of price movement, indicating whether prices would go up or down, in all but two instances – this was also of high importance, since the costliest mistake a forecaster could make in this market was to get the direction of the price movement wrong.

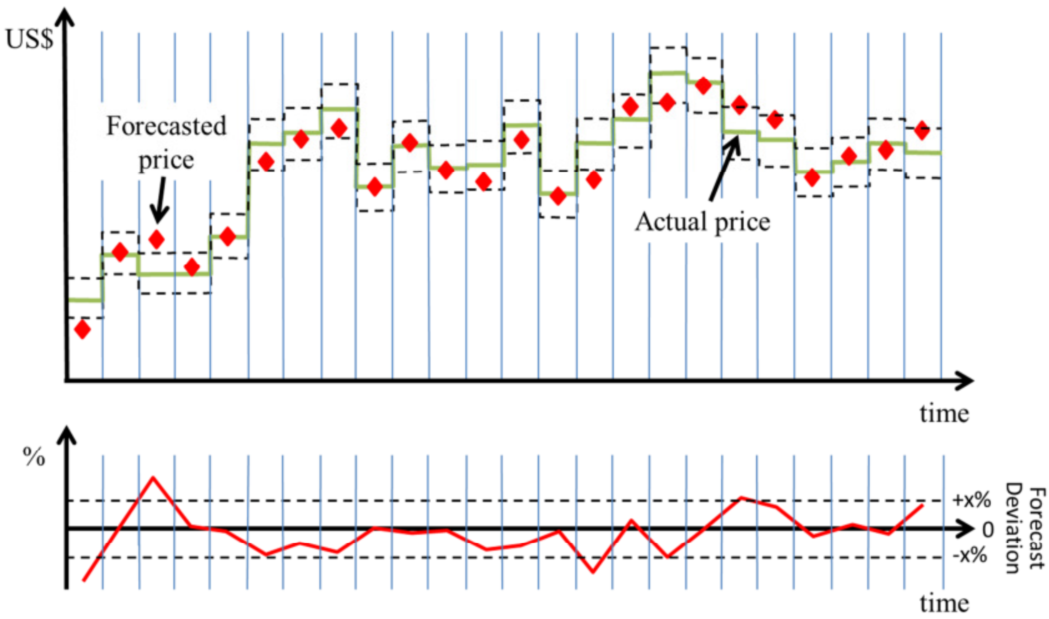

Figure 12: The forecast's deviation from Delta's actual contract price

4.6 Forecasting accuracy: the proof of the pudding is still in the eating

While the simulation model produced 'backward-looking' forecasts that fell within the client's forecasting accuracy requirements, the model still needed to be tested for real. Given the promising results of 'backward-looking' forecasts, the client agreed to continue to fund the project into a real-life testing stage, where, for the next six months the model was used to produce actual forecasts, and the results checked against actual price performance.

First indications about the model's forecasting ability

We did not expect actual forecasting performance to be as good as that evidenced by the 'test' forecasts for two reasons: firstly, and as stated earlier in this paper, any model can only be based upon information from conditions in the past, and this data will be incomplete, and conditions in the past will not match perfectly those in the future. Secondly, our backwards-looking forecast tests had not been truly 'blind'. Therefore, in spite of all our precautions to make them as 'blind' as possible, it was still possible that we might inadvertently have incurred in some degree of curve-fitting.

The problems encountered proved some of the difficulties that plague the use of forecasting in the real world: in one month the data updating procedures had not yet been refined and the wrong data was updated into the model, thus producing a wrong forecast. Later in the year, the world-wide economic climate changed, and the model failed to capture this since its expectations on demand were still partially based on annual data. However, in the end the forecast tests proved the simulation model's ability to support the client's forecasting process successfully: while the model was not able to achieve the desired degree of accuracy 100% of the time, in the words of the client's manager "its performance surpassed all our initial expectations."

At the end of the testing phase and after a thorough review of the model and its results, the client considered the project a clear success, decided to take the forecasting model in-house, and to fully integrate it into its monthly forecasting process for Delta.

Ongoing use and improvement

It was not the objective of the project to replace market experts in the forecasting process, but rather to provide them with a consistent and unbiased baseline against which to check their expectations. Thus, the objective for the simulation model was rather to support the monthly 'delphi' (Linstone, 1975) forecasting process run by Alpha's experts, which so far had consisted in a series of virtual monthly meetings in which these experts exchanged views on the market until they reached a consensus forecast.

The simulation-based forecasting system added to this 'delphi' process consistency and quantification. Every month the model proposed a forecast, clearly showing the influence on Delta's price that was expected to be exerted by the different factors considered in the model. Against these, the process also reminded the participants in the process about the strengths that they had given to these same factors in the past, so that as time went by they had an ever longer record of computer and 'human' forecasting decisions and performance. Armed with all this information, the market expert team ran through the 'delphi' process and produced a consensus forecast – the inputs to which (the reasons argued by the experts to justify the forecast figures) were also duly recorded in the system.

Thus, the resulting forecasting process was designed so that the computer and the people learned from each other, setting off a virtuous forecasting improvement loop: better simulator forecast performance reinforced the forecasters' belief in the process and in the model, which led to stricter adherence to the process, more reliance on the model's output, and better recordkeeping. All of these factors combined allowed the in-house modelers to further improve their forecasts and to further improve the simulation model.

To conclude the story about the Delta project, the authors of this paper checked with their former client when starting to write this paper, years after the original project had been completed. The client confirmed that in spite of the difficulties inherent in proceeding with model use and improvement without external support, the organization was able to develop enough internal expertise (Canovi, Els, Graham, & Voigt, 2002) and was still using the system just as described above.

5 Discussion and Conclusion

In this paper we described how we conducted short-term forecasts for a petrochemical commodity product. We developed an internally consistent simulation model in multiple iterative steps that were capable of forecasting the price of an oil derivative for a four-week time frame. In this section we discuss and highlight insights from the project in the following four areas: the wider usefulness of System Dynamics when part of a broader analytical process, the crucial use of data, the influence of combining different forecasting methods on forecasting accuracy, and the importance of modeling people's perceptions.

5.1 The broader applicability of System Dynamics tools, methods and insights

It has often been claimed that System Dynamics should be used to understand the basic drivers and behaviors surrounding the critical issues that shape our time (Forrester, 2007).

Yet, the definition of the applicability of the System Dynamics methodology has shown to be sometimes narrow, especially when seeking answers to specific questions posed by industry. The project described in this paper is one example of a general trend experienced by the authors throughout their careers: when applied carefully, System Dynamics can sometimes deliver answers unobtainable through other means, even in areas that were not initially envisaged by the pioneers of the field.

Thus, when faced with Alpha's request for decision-making support we were confronted with the question whether it is acceptable to use the tools, methods and insights developed in System Dynamics to complement other methodologies, in ways not originally envisaged to fall within System Dynamic's field of application. In the case described in this paper, the client had attempted to solve its managerial challenge via six other methodologies with the help of six other organizations. They then turned to System Dynamics only after all prior attempts had failed to deliver acceptable results. While the 'true' initial purpose of System Dynamics was indeed not to support the development of forecasting tools, should the authors have rejected the clients' plea for help because of this? Or, suspecting how their tools, methods and insights might help this company, were we not right to use them?

Thus, we propose that the crucial question to be asked regarding the applicability of System Dynamics in addressing a given issue should not just concern the nature of the issue, but go beyond it and focus on the way in which the tools, methods and insights could be applied to specifically address it. In other words, the critical question should shift from where to apply System Dynamics to how to do so, since only the latter question can help us to determine whether an approach can be effective or not.

We applied the standard System Dynamics method to gather and structure initial market information, and used widely accepted model-building processes (Sterman, 2000) to develop a market-explanatory simulation model. We based the final forecasting model on insight of a clearly dynamic nature: the commodity is produced based on a single raw material, and most of its annual production volume is achieved as a byproduct in production processes for other commodities. Despite the strong external influences we needed to consider that the price for any commodity is still determined by the balance between demand and supply. Shedding light on the underlying structure actually was modeling Delta's sellers' and buyers' perception of future market development including their bounded rationality (Camerer, 2003; Granovetter, 1985), which has a long tradition in System Dynamics modeling (Größler, 2004; Morecroft, 1982; Sterman, 1987, 2000). Latter is expressed in this instance by the decision-makers' expectations about future changes in key market drivers. Modeling such market perceptions entails modeling subway uncertainties, or known unknowns.[14] Summarizing, actions and decisions are based on perceptions, and these shape reality.

In this case, the process based on SD tools and processes delivered useful results that exceeded client expectations.

Therefore we update the first research proposition:

> *Research proposition 1a: The System Dynamics methodology can be successfully applied for conducting short-term forecasting of commodity price movements.*

A further conclusion from the project was that the use of the System Dynamics approach can be effective in answering complex managerial questions even when the final analysis is not performed with a true System Dynamics model.

Based on the outcome of project Delta, as well as on other relevant (unpublished) prior professional experience, we would like to propose that a rigorous implementation of the System Dynamics modeling process will generally help to better understand the internal workings of most complex systems. The requirement to produce a detailed causal system description forces an analyst to follow a systemic approach, casting a wider net than will be allowed by other methodologies. This also focuses the attention of the analyst on the variables that play a key role in the system. It drives and focuses data and information gathering efforts, and it ensures that no potentially relevant aspects of the system are left uninvestigated. Furthermore, the experience gathered by system dynamicists during years of model building and analysis allows them to understand the potential dynamic relevance of the structures being discovered, thus helping to set the direction and priorities for further analysis.

[14] The importance of people's expectations was highlighted when meeting the client at the review session following the six-month forecasting trial period. In order to investigate potential avenues to improve the model's accuracy, we decided to test the influence of the mismatch between the expectations of market participants and actual variations in market factors. When we substituted the simulated expectations about the variations in key factors with future data about the real variations in those same factors, the forecasting accuracy of the model was actually reduced, and consistently so throughout the whole forecasting period. This was the simulation model was recalibrated to obtain the highest possible backward-looking forecasting accuracy. The forecasting period comprised the initial period of 24 months' worth of data that enabled initial backward-looking test forecasts, plus the recently concluded six-month period of actual forecasting.

Thus, System Dynamics tools, methods and insights do not only add value by themselves. They also contribute significantly to establishing an unbiased and consistent foundation upon which to build an analytical structure of the system – without the need to culminate in the development of a dynamic simulation model.

Thus, at this point we would like to expand our first research proposition with the following two corollaries:

> *Research proposition 1b: System Dynamics tools, methods and insights can be applied to a much broader range of managerial questions than that originally envisioned for the field of System Dynamics – if they are applied in a rigorous way.*
>
> *Research proposition 1c: System Dynamics tools, methods and insights can be used for answering managerial questions in systems whose behavior is not dominated by feedback.*

5.2 The crucial use of data

The use of historical data about the performance of the system was crucial to the success of forecasting the short-term price development of Delta. The use of data allowed us to check, validate and refine the information provided by client experts. It also enabled us to validate the different hypotheses offered to explain the behavior of the market, and it eventually played a critical role in quantifying the influence of each key market driver on the monthly forecast of Delta's price.

While all three uses of data were critical to the success of the project, we especially highlight the importance of model calibration (Sterman, 2000). Calibrating a model to match historical performance of Delta's price allowed us to gain a crucial degree of confidence in our understanding of how the market works. Before building the simple model that explained the past behavior of the price for Delta, we had gathered contradictory opinions from the client's experts as to which factors played a more important role in driving prices. Furthermore, the market experts believed that the market had undergone a radical change in recent years, possibly even making it unpredictable. After showing to the client the output of our small initial simulation model and explaining how it worked, the situation changed dramatically: the market experts agreed that the market had probably only undergone one very specific (and one-time) change, and they also agreed on the most likely factors driving the price of Delta.

As shown and described with Figure 7 and Figure 8, the outcome of the model did not perfectly match the historical performance of the market. The simulation showed some short-term variability not present in the data, and it clearly deviated from historical data in the three final years of the simulation. However, as proposed by (Barlas, 1996; Sterman, 2000), the objective of comparing the simulation output to the historical data should not be a perfect match, but rather to see whether the model's structures can clearly reproduce the main patterns of behavior shown in the data that cannot be explained away *a priori* by events external to the system.

The match of model output to data hinged on the market drivers selected to drive the simulation and on how the model represented the mechanisms of market price formation. As mentioned above, when building this particular model we considered several different price forming mechanisms – for example, wet considered the two ones:

$$\boldsymbol{p_{\Delta}} = \int \boldsymbol{f}(\overline{\boldsymbol{p_{\Delta}}}, \boldsymbol{p_{rm}}, \boldsymbol{mt}) \qquad \text{[Equation 1]}$$

$$\Delta \boldsymbol{p_{\Delta}} = \boldsymbol{f}(\Delta \boldsymbol{p_{rm}}, \boldsymbol{mt}) \qquad \text{[Equation 2]}$$

With: p_{Δ} = price of Delta

$\overline{p_{\Delta}}$ = reference price of Delta

p_{rm} = price of the raw material

mt = market tightness

Both of the formulations explained in Equations 1 and 2 create similar patterns of behavior, and the client's experts were unable to suggest whether one of these was more likely than the other. Yet, the behaviors driven by these two formulations are not identical. Choosing one or the other can thus have different implications for policy decisions, or in this case, for forecasting accuracy.

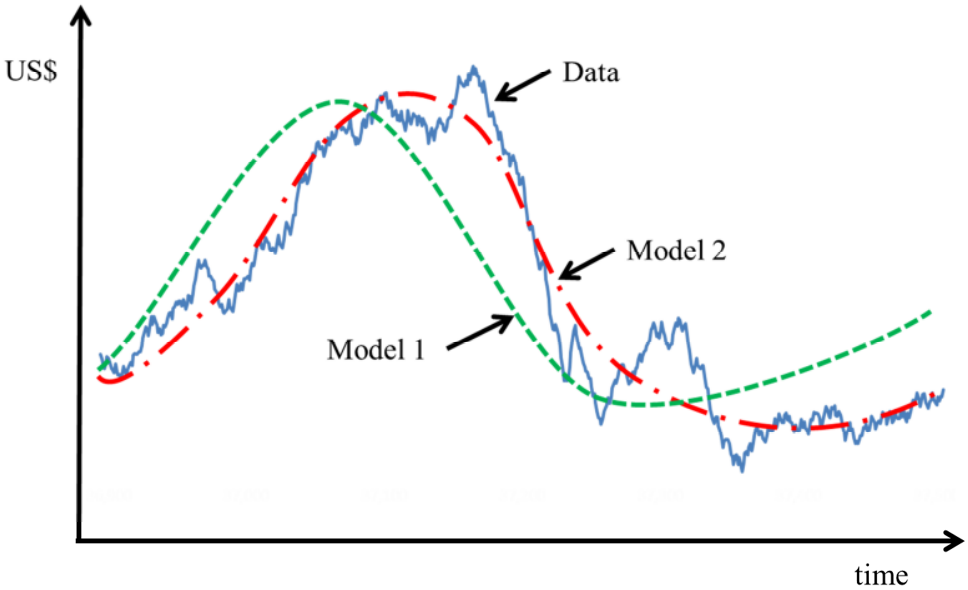

Figure 13

Figure 13depicts a situation like the one described above. The blue solid line shows representative historical price data. The green dashed and red dot-dashed lines represent outputs of two fictitious competing model structures to explain historical behavior. Outputs of both models show similar patterns of behavior. On a purely qualitative basis it is not possible to distinguish between them. Yet, both models cannot be correct, understanding the definition of 'model correctness' to be 'fitness for purpose', as proposed by STERMAN (Sterman, 2000).

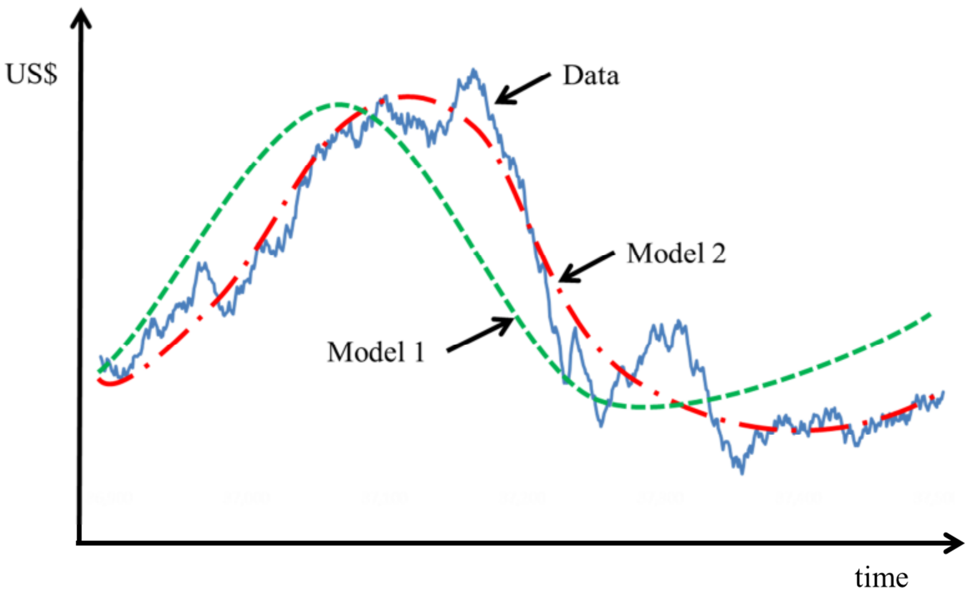

Figure 13: Example for different models creating different behaviors for curve fitting

Market participants are influenced by some factors or by others, they either have a reference price in mind when deciding to buy or they will not, etc. Since qualitative information cannot sufficiently inform this critical choice, a quantitative approach (like model calibration to historical data) is required (Sterman, 2000). So, it is crucial to understand what patterns need to be matched and why.

Based on the above, we find support for

Research Proposition 2: Carefully analyzing data about the past performance of the system, and carefully comparing model output to it, are crucial steps in ensuring forecasting accuracy.

5.3 Combining different methods increases forecasting accuracy

As mentioned before, the project never expected to replace the qualitative, expectations-based forecasts run by company experts by the output of a computer simulation model. It was always acknowledged that a simulation model would never be able to reflect the full complexity of the market, and that thus some degree of human input into the forecasting process would always remain necessary. The objective of project Delta was rather to provide a simulation model that could be used to provide an unbiased consistency check on people's forecasts, with the double aim of increasing short-term accuracy, and also the long-term one through accelerated learning.

The ongoing forecasting activity using this combination of inputs showed that forecasting accuracy of the combined approach was clearly better than that of any of the two individual inputs used (expert and simulation model). This finding supports MAKRIDAKIS ET AL. (2009) and

Research Proposition 3: Combining the predictions of different methods and individuals increases forecasting accuracy.

5.4 Perceptions shape reality

Project Delta also confirmed an unanticipated conclusion, which actually constitutes one of the central tenets of the field of System Dynamics: actions and decisions are based on perceptions, and thus these shape reality.

6 Summary

Having a profound understanding of possible future developments has been shown to be crucial for many businesses. Conducting forecasts is thus part of many businesses' planning processes. Especially since the ascent of computer modeling power in the 1970s social scientists seek to base estimations on insights from computer models. While scholars realize that many forecasts actually fail due to many unknown unknowns in the underlying managerial challenge to be studied, insights from forecast simulations may still be necessary and valuable for policy-designers and decision-makers since it reduces the uncertainty from known unknowns.

In this paper we have explained the reasons for forecasting failures. We further evaluated whether the System Dynamics methodology is suitable for conducting forecasts, and we challenged this proposition by discussing a client project for a global petrochemical company. We then described how we required three modeling steps in order to produce reliable forecasts. Each of thesewas delivered insights into how the market worked, and the final one eventually delivered the results expected by the client.

This lessons learned on this project led us to insightful research propositions. Most importantly, we postulate the application of System Dynamics tools, methods and insights to answer research questions which the field has so far been reluctant to address. When applied carefully, System Dynamics can sometimes deliver answers unobtainable through other ways, even in areas that were not initially envisaged by the pioneers of the field – and which may not even culminate in the use of a feedback-based simulation model.

We also postulate the importance of making good use of historical data about the system. Data can be (and thus in our view should be) used to check the consistency of the information provided by system experts. It can further be used to calibrate a simulation model in order to better validate dynamic hypotheses about the workings of the system. And sometimes the use of data can be the only way to accurately quantify the influence of competing key factors on system behavior.

Bibliography

Andersen, D. L., Luna-Reyes, L. F., Diker, V. G., Black, L., Rich, E., & Andersen, D. F. 2012. The disconfirmatory interview as a strategy for the assessment of system dynamics models. ***System Dynamics Review (Wiley)***, 28(3): 255-275.

Barlas, Y. 1996. Formal Aspects of Model Validity and Validation in System Dynamics. ***System Dynamics Review***, 12(3): 183 - 210.

Camerer, C. F. 2003. ***Behavioral Game Theory: Experiments in Strategic Interaction***. Princeton, New York: Princeton University Press.

Canovi, A., Els, S., Graham, A., & Voigt, A. 2002. ***Building A Sustainable, In-house System Dynamics Capability: How an Internal System Dynamics Group Can Develop a Distinctive Service and Become an Enduring Component of the Company's Operations***. Paper presented at the 20th International System Dynamics Conference, Palermo, Italy.

Croson, R., & Donohue, K. 2005. Upstream versus downstream information and its impact on the bullwhip effect. ***System Dynamics Review (Wiley)***, 21(3): 249-260.

Croson, R., & Donohue, K. 2006. Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information. ***Management Science***, 52(3): 323-336.

Fildes, R., & Goodwin, P. 2007. Against Your Better Judgment? How Organizations Can Improve Their Use of Management Judgment in Forecasting. ***Interfaces***, 37(6): 570-576.

Forrester, J. W. 1961. ***Industrial Dynamics***. Waltham, MA: Pegasus Communications.

Forrester, J. W. 2007. System dynamics—the next fifty years. ***System Dynamics Review (Wiley)***, 23(2/3): 359-370.

Granovetter, M. 1985. Economic Action and Social Structure: The Problem of Embeddedness. ***The American Journal of Sociology***, 91(3): 481 - 510.

Größler, A. 2004. A Content and Process View on Bounded Rationality in System Dynamics. ***Systems Research and Behavioral Science***, 21(4): 319 - 330.

Homer, J. B. 2012. Partial-model testing as a validation tool for system dynamics (1983). ***System Dynamics Review (Wiley)***, 28(3): 281-294.

Kapmeier, F., Tilebein, M., Voigt, A., & Dillerup, R. 2011. ***Applying System Dynamics to Overcome Unsuccessful Success Factor Research***. Paper presented at the 25th International System Dynamics Conference, Boston, MA.

Khin, A. A., Zainalabidin, M., & Mad. Nasir, S. 2011. Comparative Forecasting Models Accuracy of Short-Term Natural Rubber Prices. ***Trends in Agriculture Economics***, 4(1): 1-17.

Lee, H. L., Padmanabhan, V., & Whang, S. 1997. The Bullwhip Effect in Supply Chains. ***Sloan Management Review***, 38(3): 112-112.

Libby, R., & Rennekamp, K. 2012. Self-Serving Attribution Bias, Overconfidence, and the Issuance of Management Forecasts. ***Journal of Accounting Research***, 50(1): 197-231.

Linstone, H. A. (Ed.). 1975. ***The Delphi Method: Techniques and Applications***. Reading, Mass. and other: Addison-Wesley.

Lyneis, J. M. 1999. System dynamics for business strategy: a phased approach. ***System Dynamics Review (Wiley)***, 15(1): 37-70.

Lyneis, J. M. 2000. System dynamics for market forecasting and structural analysis. ***System Dynamics Review (Wiley)***, 16(1): 3-25.

Makridakis, S. 1996. Forecasting: its role and value for planning and strategy. ***International Journal of Forecasting***, 12(4): 513-537.

Makridakis, S., Hogarth, R. M., & Gaba, A. 2009. Forecasting and uncertainty in the economic and business world. ***International Journal of Forecasting***, 25(4): 794-812.

Makridakis, S., Hogarth, R. M., & Gaba, A. 2010. Why Forecasts Fail. What to Do Instead? . ***MIT Sloan Management Review***, 51 (2): 83-90.

Makridakis, S., & Taleb, N. 2009. Decision making and planning under low levels of predictability, ***International Journal of Forecasting***, Vol. 25: 716-733.

Makridakis, S., & Wheelwright, S. C. 1977. Forecasting: Issues & Challenges for Marketing Management. ***Journal of Marketing***, 41(4): 24-38.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. 1998. ***Forecasting : methods and applications*** (3rd ed.). New York, NY; Chichester; Weinheim and other: Wiley.

Micheli, P., & Manzoni, J.-F. 2010. Strategic Performance Measurement: Benefits, Limitations and Paradoxes. ***Long Range Planning***, 43(4): 465-476.

Morecroft, J. D. 1982. System Dynamics: Portraying Bounded Rationality: 39: Working paper of the Sloan School of Management 1332-82A.

Morecroft, J. D. W. 1985. Rationality in the Analysis of Behavioral Simulation Models. ***Management Science***, 31(7): 900 - 916.

Peters, T. J., & Waterman, R. H. 1982. ***In Search of Excellence: Lessons from America's Best-Run Companies***. New York: Harper & Row.

Randers, J., & Göluke, U. 2007. Forecasting turning points in shipping freight rates: lessons from 30 years of practical effort. ***System Dynamics Review (Wiley)***, 23(2/3): 253-284.

Sherden, W. A. 1998. ***The fortune sellers: the big business of buying and selling predictions***: John Wiley.

Sterman, J. D. 1987. Testing Behavioral Simulation Models By Direct Experiment. ***Management Science***, 33(12): 1572 - 1592.

Sterman, J. D. 1989. MODELING MANAGERIAL BEHAVIOR: MISPERCEPTIONS OF FEEDBACK IN A DYNAMIC DECISION MAKING EXPERIMENT. ***Management Science***, 35(3): 321-339.

Sterman, J. D. 2000. ***Business Dynamics: Systems Thinking and Modeling for a Complex World***. Boston, MA and other: Irwin/McGraw-Hill.

Taleb, N. N. 2010. ***The black swan*** (2., rev. ed. ed.). London [u.a.]: Penguin Books.

Yin, R. K. 1994. ***Case Study Research: Design and Models***. Thousand Oaks, London und New Delhi.

Zahn, E., Hülsmann, O., Kapmeier, F., & Tilebein, M. 2007. Management von Wechselwirkungen in Supply Chain-Systemen, ***Management am Puls der Zeit – Strategien, Konzepte und Methoden. Festschrift für Prof. Dr. Dr. hc. mult. Horst Wildemann***, Vol. 2 (Produktion und Logistik): 1183-1203. München: TCW.