



Universität Augsburg
Prof. Dr. Hans Ulrich Buhl
Kernkompetenzzentrum
Finanz- & Informationsmanagement
Lehrstuhl für BWL, Wirtschaftsinformatik,
Informations- & Finanzmanagement

UNIA
Universität
Augsburg
University

Diskussionspapier WI-174

Ein Optimierungsansatz für ein fortlaufendes Datenqualitätsmanagement und seine praktische Anwendung bei Kundenkampagnen

von

Bernd Heinrich, Mathias Klier

Mai 2006

in: Zeitschrift für Betriebswirtschaft, 76, 6, 2006, S. 559-587

Ein Optimierungsansatz für ein fortlaufendes Datenqualitätsmanagement und seine praktische Anwendung bei Kundenkampagnen

Von Bernd Heinrich und Mathias Klier

Dr. Bernd Heinrich; Dipl.-Math. oec. Mathias Klier; wissenschaftliche Mitarbeiter am Lehrstuhl WI-IF sowie am Kernkompetenzzentrum Informationstechnologie & Finanzdienstleistungen, Universität Augsburg, Universitätsstraße 16, 86135 Augsburg.

E-Mail: {Bernd.Heinrich|Mathias.Klier}@wiwi.uni-augsburg.de

Ein Optimierungsansatz für ein fortlaufendes Datenqualitätsmanagement und seine praktische Anwendung bei Kundenkampagnen

Überblick:

- Die Sicherstellung und Verbesserung der Qualität von Daten bspw. für ein Kundenbeziehungsmanagement (KBM) werden in Wissenschaft und Praxis intensiv diskutiert. Ob die ergriffenen Datenqualitätsmaßnahmen jedoch ökonomisch überhaupt gerechtfertigt sind und wann und in welchem Umfang investiert werden soll, wird oftmals nicht analysiert.
- Im Beitrag wird zuerst die Frage nach den optimalen Umfängen und Zeitpunkten für Investitionen in ein fortlaufendes Datenqualitätsmanagement mittels modelltheoretischen Überlegungen untersucht, wobei die Domäne des KBM zugrunde gelegt wird. Zur Ableitung allgemeiner Ergebnisse wird auf eine mathematische Simulation mit numerischen Optimierungsverfahren zurückgegriffen. Neben der Entwicklung des Entscheidungsmodells, dessen praktische Anwendbarkeit eine Fallstudie verdeutlicht, werden so allgemeine Handlungsempfehlungen zur Planung von Datenqualitätsmaßnahmen gegeben, die neben informationstechnischen Aspekten vor allem die ökonomische Prüfung berücksichtigen.

A. Einleitung

Seit einigen Jahren gewinnt – insbesondere im Zuge von stark wachsenden Datenbeständen, dem verstärkten Einsatz von Data Warehouse (DW)-Systemen und der Relevanz, die dem Kundenbeziehungsmanagement (KBM) beigemessen wird – Datenqualität (DQ) sowohl in der Wissenschaft als auch in der Praxis immer mehr an Bedeutung. Dies überrascht nicht, denn der Nutzen einer Versorgung von Entscheidungsträgern und Mitarbeitern mit Daten steigt und fällt mit deren Vollständigkeit, Korrektheit, Konsistenz oder Aktualität, d. h. den Eigenschaften, die als Qualitätskriterien bekannt sind (vgl. z. B. Wang et al., 1995, S. 632) und deren Überprüfung einer spezifikationsorientierten Beurteilung der DQ zugrunde liegt. Im Gegensatz zur Ermittlung oder Verbesserung der Datenspezifikation im Rahmen oftmals diskutierter Datenbedarfsanalysen (vgl. Helfert, 2002), werden hierbei dem spezifizierten Datenangebot des Informationssystems (Datenschema) die tatsächlich enthaltenen Daten gegenüber gestellt. Für sehr viele Unternehmungen stellt dabei die Sicherstellung der DQ ein Problem dar (Helfert, 2000, S. 13; Ballou et al., 1998; Ballou, Tayi, 1999). Eine häufig zitierte Studie des DW Institute ergab, dass der US-Wirtschaft aufgrund von mangelhafter DQ ein jährlicher Schaden von über 600 Milliarden Dollar entsteht (Heijkers, 2003). Andere Zahlen, die leichter nachzuprüfen sind, besagen, dass im Durchschnitt zwischen 15% und 20% der Datenwerte in typischen Kundendatenbanken fehlerhaft sind (English, 1999, S. 9). Des Weiteren geben 67% der Marketingleiter in einer Umfrage der SAS Institute GmbH an, dass die Zufriedenheit ihrer Kunden durch schlechte DQ stark leide (SAS Institute, 2003). Deswegen ist es zum einen wenig überraschend, dass sich bei DW-Projekten der Anteil am Budget für (geplante und vor allem ungeplante) Datenqualitätsmaßnahmen (DQM) auf mehr als 50% beläuft (Matzer, 2004; Alt, 2003). Zum anderen ist es nicht verwunderlich, dass die Sicherstellung und Verbesserung der DQ in DW-Systemen in Wissenschaft und Praxis diskutiert werden. Dabei werden zwar der positive Beitrag, der durch DQ entsteht, und die Kosten, die mit einer schlechten DQ einhergehen, generell untersucht. Ob die deshalb ergriffenen DQM jedoch ökonomisch überhaupt gerechtfertigt sind, d. h. wann und in welchem optimalen Umfang in DQM investiert werden soll, wird oftmals nicht betrachtet.

Deshalb wird im vorliegenden Beitrag die Fragestellung nach einem effizienten und über mehrere Perioden fortlaufenden DQ-Management modellbasiert untersucht, wobei das Anwendungsgebiet des KBM zugrunde gelegt wird. Zur Bestimmung des optimalen Einsatzes von DQM und zur Ableitung allgemeiner Investitionsmuster und Handlungsempfehlungen wird dabei ein mehrperiodiges, quantitatives Entscheidungsmodell entwickelt, das den durch die Maßnahmen kreierte Einzahlungen den entsprechenden bewerteten Ressourceneinsatz

gegenüberstellt. Um die Anwendbarkeit und die Operationalisierung des Modells zu verdeutlichen, wird zudem eine praktische Umsetzung detailliert beschrieben.

Im Kapitel B werden der State-of-the-Art und ausgewählte Ansätze zur ökonomischen Analyse von DQM vorgestellt, bevor im darauf folgenden Kapitel ein quantitatives, mehrperiodiges Optimierungsmodell zur Planung des Einsatzes von DQM entwickelt wird, das die Grundlage eines effizienten, fortlaufenden DQ-Managements darstellen soll. Im Kapitel D werden anhand des Modells allgemeine Ergebnisse und Handlungsempfehlungen abgeleitet und erläutert, bevor im darauf folgenden Teil ausführlich die Operationalisierung und praktische Umsetzung des Modells am Beispiel des Kampagnenmanagements eines großen deutschen Mobilfunkanbieters dargestellt werden. Das letzte Kapitel fasst die Ergebnisse zusammen und würdigt diese kritisch.

B. Bisherige Beiträge zur ökonomischen Analyse von Datenqualitätsmaßnahmen

In der Literatur findet sich bereits eine ganze Reihe von Veröffentlichungen, in denen ökonomische Auswirkungen von DQ-Mängeln und Maßnahmen zu deren Behebung diskutiert werden (vgl. z. B. Cappelletto et al., 2003; Fisher et al., 2003; Feigenbaum, 1991; Helfert, 2002; Kreuzberg, 2000; Machowski, Dale, 1998; Shank, Govindarajan, 1994). Dabei wird i. d. R. das Ziel verfolgt, die Ursachen, insbesondere aber auch die Folgekosten, einer schlechten DQ zu analysieren, um somit die Höhe der Investitionen in DQM zu bestimmen und zu rechtfertigen.

In diesem Zusammenhang werden verschiedene Kategorien von Qualitätskosten unterschieden. Zum einen sind hier die sog. *internal* und *external failure costs* zu nennen, welche die Fehlerkosten für die Unternehmung infolge schlechter DQ, bspw. bei internen Abläufen oder bei der Zusammenarbeit mit Kunden, betreffen. Fehler in den Adress- und Kundendaten sind hier als einfache Beispiele zu nennen, die nicht nur Verärgerung beim Kunden und Korrekturbedarf nach sich ziehen, sondern ebenfalls den Erfolg einer Kampagne – bspw. durch Datendubletten oder falsch gespeicherte Produktdaten und einer damit verbundenen falschen Zielgruppenzuordnung – schmälern. DQM sollen hier die Interaktion mit dem Kunden verbessern und dadurch einen zusätzlichen Wertbeitrag generieren. Die *failure costs* sind in diesem Fall als Opportunitätskosten für einen nicht realisierten Wertbeitrag zu interpretieren. Den *failure costs* sind im Weiteren die Kosten gegenüber zu stellen, die für die Konzeption und Durchführung der DQM anfallen, um die Qualitätsmängel aufzudecken und zu bereinigen (*appraisal costs*), bzw. zu vermeiden (*prevention costs*). So fallen bei einer Intensivierung der DQM zwar einerseits die Fehlerkosten – andererseits steigen zugleich die Kosten, die den

ergriffenen Maßnahmen zuzurechnen sind (vgl. z. B. Campanella, 1999; Krishnamoorthi, 1989).

Auf Basis dieses grundlegenden Zusammenhangs findet sich zudem eine Reihe von Beiträgen, die den positiven Beitrag einer verbesserten DQ untersuchen. So diskutieren (Cappiello et al., 2003) die Notwendigkeit korrekter und interpretierbarer Daten, um Anwendungssysteme und Datenbanken überbetrieblich zu koppeln und damit eine IT-gestützte Zusammenarbeit zwischen Unternehmungen zu ermöglichen. Demgegenüber wird in (Fisher et al., 2003) das zweckmäßige Qualitätsniveau entscheidungsrelevanter Daten sowie von Metadaten im Kontext von Managementinformationssystemen analysiert.

Daneben existieren Veröffentlichungen, die sich mit der ökonomischen Wirkung von DQM im KBM-Bereich beschäftigen (vgl. z. B. Arndt, Langbein, 2002; Khalil, Harcar, 1999; Redman, 1996). Hier wird häufig unterstellt, dass eine erhöhte DQ per se – bspw. durch die Ermöglichung einer Individualisierung der Produkte und der Kundenansprache – die Zufriedenheit der Kunden steigert und somit die Geschäftsbeziehung mit diesem verbessert. Problematisch ist dabei, dass die Beeinflussung durch die Möglichkeiten zur Produktindividualisierung, zur gezielten Kundenansprache oder gar zur höheren Arbeitsmotivation des Verkaufspersonals an Beispielen argumentiert wird (wie z. B. in Redman, 1996) – auf eine konkrete Darstellung von Wirkungszusammenhängen oder eine modellgestützte Argumentation wird dagegen verzichtet.

In (Heinrich, Helfert, 2003a) und (Heinrich, Helfert, 2003b) werden die Wirkungen einer höheren DQ zur Verbesserung der Kundenbeziehung erstmals modellbasiert untersucht. Eine verbesserte Beziehung bedeutet dabei, ob und in welchen Fällen ein Kunde infolge von DQM die Geschäftsbeziehung mit der Unternehmung intensiviert, d. h. mehr Transaktionen in einem Zeitraum tätigt. Transaktionen werden deshalb verwendet, da diese im Gegensatz zu vielen anderen, sog. weichen Faktoren der Beziehung, wie bspw. Vertrauen oder Commitment, direkt und bewertungsunabhängig messbar sind und der Erfolg der DQM so ex post auch nachvollziehbar ist. Letztlich sollen die weichen Faktoren – unbenommen ihrer Bedeutung in der qualitativen Argumentation für die Ergreifung von DQM – auch zu einer intensiveren und längeren Geschäftsbeziehung führen, die sich in einer stärkeren Transaktionstätigkeit der Kunden widerspiegelt. Bei begrenztem Transaktionsvolumen des Kunden (z. B. Anlagevolumen im Finanzdienstleistungskontext) hat dies eine Verlagerung des Volumens zwischen den Anbietern zur Folge. Als Zielgröße wird dementsprechend der Transaktionsanteil λ bei der Unternehmung und dessen Veränderung infolge der Realisierung von DQM und einer damit verbundenen höheren DQ herangezogen. Ziel ist es dabei, DQM nicht als Selbstzweck zu

begreifen, sondern um dadurch – messbar und nachvollziehbar – zusätzliche Geschäftstransaktionen zu generieren. Die Frage, wann und in welchem Umfang unter ökonomischen Gesichtspunkten investiert werden soll, wird dagegen auch in diesen Beiträgen nicht untersucht, so dass eine Reihe wichtiger Fragestellungen unbeantwortet bleibt. Zudem ist zu berücksichtigen, dass die in einer Periode aufgebaute DQ natürlich in den Folgeperioden ebenso zur Verfügung steht. Dies ist schon vor dem Hintergrund wichtig, dass DQ-Probleme nicht in einem Schritt nachhaltig zu lösen sind, sondern vielmehr in übereinstimmender Literaturmeinung ein fortlaufendes DQ-Management notwendig ist. Deshalb sind insbesondere die folgenden Aspekte zu untersuchen:

- Wie wirkt sich die Durchführung von DQM auf das Qualitätsniveau aus und inwiefern kann die gesteigerte DQ in Folgeperioden zu zusätzlichen Geschäftstransaktionen beitragen?
- Zu welchen Zeitpunkten und wie viel soll eine Unternehmung unter ökonomischen Gesichtspunkten in DQM investieren?
- Welchen Einfluss auf den Zeitpunkt und die Höhe der Investitionen in DQM haben bspw. das bereits bestehende DQ-Niveau, die bereits vorhandene sowie die neu hinzukommende Kundendatenmenge und der Wirkungsgrad der Maßnahmen?
- Existieren Indikatoren und Richtlinien, die eine Unternehmung im Rahmen eines fortlaufenden DQ-Managements beachten sollte?

Diesbezüglich finden sich in der Literatur allerdings meist nur relativ vage formulierte Ausführungen. English führt bspw. aus, dass bei Einführung eines DQ-Managements meist einmalige Bereinigungsaktionen geeignet sind, wogegen für häufig stattfindende Prozesse Maßnahmen zur kontinuierlichen Verbesserung der DQ abzuleiten und umzusetzen sind (vgl. English, 1999). Eine ähnliche Argumentation wird auch von Redman dargestellt, der den Anteil der neu hinzukommenden Daten sowie den zeitlichen DQ-Verfall als Indikatoren dafür anführt, ob die Daten eher einmalig bereinigt oder deren Qualität fortlaufend mit Hilfe von Methoden der Prozesskontrolle und -verbesserung sichergestellt werden soll (vgl. Redman, 1996). Die wissenschaftliche Fundierung der Aussagen ist dabei jedoch problematisch, da oftmals Plausibilitäten verwendet werden, ohne Rahmenbedingungen, Ausgangssituationen oder ökonomische Zielsetzungen in notwendigem Umfang zu berücksichtigen.

Zur Beantwortung der obigen Investitionsfragestellung und der Ableitung von allgemeinen Investitionsmustern und Handlungsempfehlungen für ein fortlaufendes DQ-Management wird im nächsten Abschnitt ein mehrperiodiges Optimierungsmodell entwickelt.

C. Entwicklung eines quantitativen, mehrperiodigen Optimierungsmodells zur Planung des Einsatzes von Datenqualitätsmaßnahmen

Um den optimalen DQM-Einsatz bestimmen zu können, wird ein quantitatives Entscheidungsmodell entwickelt, das den durch die Maßnahmen kreierte Einzahlungen den entsprechenden bewerteten Ressourceneinsatz gegenüberstellt. Dabei werden im allgemeinen Modell bewusst keine einzelnen DQM und deren Kosten-/Nutzenwirkungen zugrunde gelegt. Vielmehr wird davon abstrahiert, da diese erst jeweils für die praktische Anwendung (wie auch das Fallbeispiel in Kapitel E verdeutlicht) zu operationalisieren und in der Höhe zu konkretisieren sind. Für das Modell werden dabei die folgenden Annahmen und Definitionen getroffen:¹

- A.1 Ein Kunde tätigt bei der Unternehmung in Periode t einen Anteil λ_t seines gesamten Transaktionsvolumens (mit $\lambda_t \in [0; 1]$), das auch das transaktionsorientierte, noch nicht realisierte Kundenpotenzial berücksichtigt². Da die Höhe des Kundenanteils λ_t durch DQM beeinflussbar ist, setzt sich dieser aus dem qualitätsinduzierten Anteil $\lambda_{DQM,t}$ und einem auch ohne DQM zu Stande kommenden, exogen vorgegebenen Anteil $\lambda_{-DQM,t}$ zusammen ($\lambda_{DQM,t}, \lambda_{-DQM,t} \in [0; 1]$ mit $\lambda_{-DQM,t} + \lambda_{DQM,t} = \lambda_t \leq 1$).
- A.2 Zahlungsorientierte Betrachtungsweise der Unternehmung: d. h. Verrechnung der diskontierten, DQ-induzierten Einzahlungen $E_t \in \mathfrak{R}^+$, die durch die Kundentransaktionen $\lambda_{DQM,t}$ generiert werden und der aus den DQM resultierenden Auszahlungen $K_t \in \mathfrak{R}^+$, bei gegebenem Zinsfuß $i_{kalk} \in \mathfrak{R}^+$.
- A.3 Die Einzahlungen E_t werden jeweils am Periodenende zahlungswirksam, wogegen die Auszahlungen K_t für die entsprechenden DQM bereits zu Beginn der Periode t anfallen.
- A.4 Abgesehen vom initialen Datenbestand ($D_{sup,t}$) resultieren alle Kundendaten der Unternehmung ($D_{sup,t}$) aus der Transaktions- bzw. Interaktionstätigkeit zwischen Kunde und Unternehmung. Externe Datenquellen werden folglich nicht berücksichtigt. Insofern handelt es sich neben Stamm- vor allem um Transaktionsdaten des Kunden. Da diese unterschiedliche Transaktionsvolumina besitzen, existiert ein exogen vorgegebenes $\tau_t \in [0; 1]$, welches festlegt, wie viele Daten ein Kunde bei der Durchführung des gesamten Transaktionsvolumens in Periode t maximal weitergibt. $D_{sup,t}$ bezeichnet im Folgenden die vorhandene Kundendatenmenge zu Beginn der Periode t , wohingegen $\Delta D_{sup,t}$ die Datenmenge beschreibt, die während Periode t hinzukommt. Sowohl $D_{sup,t}$ als auch $\Delta D_{sup,t}$ zielen dabei auf die Vollständigkeit der Daten in Bezug auf die Datenspezifikation ab, wobei $D_{sup,t}$ bzw. $\Delta D_{sup,t} \in [0; 1]$ gilt. $D_{sup,t} = 0$ bedeutet dabei, dass zu Beginn

der Periode t keine der in der Spezifikation geforderten Kundendaten vorhanden sind. Dagegen liegen bei $D_{sup,t} = 1$ alle laut Spezifikation benötigten Daten in der Datenbank vor (analog für $\Delta D_{sup,t}$).

- A.5 Aus Überschneidungen der Datenmengen $D_{sup,t}$ und $\Delta D_{sup,t}$ kann für die Unternehmung kein Mehrwert generiert werden, da beide Variablen jeweils nur auf die Vollständigkeit der Daten abzielen.
- A.6 Als weiteres Qualitätskriterium neben Vollständigkeit wird die Korrektheit der vorhandenen Daten zu Beginn der Periode t betrachtet. Diese drückt sich im Anteil der qualitätsgesicherten Daten $DQ_t \in [0; 1]$ aus.
- A.7 Die Unternehmung trifft zu Beginn jeder Periode t eine Entscheidung über den Umfang der eingesetzten reaktiven³ Maßnahmen DQM_t und setzt diese ohne Zeitverzögerung um. Die Variable DQM_t ist dabei auf das Intervall $[0; 1]$ normiert und stellt die Intensität der Maßnahmen dar.
- A.8 In jeder Periode steht ein unbegrenztes Budget zur Realisierung der DQM zur Verfügung.
- A.9 Für DQM_t , $D_{sup,t}$, $\Delta D_{sup,t}$, $\lambda_{DQM,t}$ und $\lambda_{-DQM,t}$ gilt jeweils die Eigenschaft der beliebigen Teilbarkeit.

Unter den obigen Annahmen gilt es durch ein geeignetes DQ-Management – d. h. durch die richtige Wahl der Maßnahmenintensitäten in den einzelnen Perioden – den Kapitalwert für die Unternehmung zu maximieren. Zweck der Ermittlung des optimalen Umfangs von reaktiven DQM ist es dabei, Software und Werkzeuge zur DQ-Sicherung mit diesen Werten zu konfigurieren, um damit bspw. den effizienten Umfang an Assoziationsregeln zu generieren. Werden Assoziationsregeln als DQM auf die Datenmenge der Unternehmung angewendet, so kann die Konsistenz der Daten überprüft und erhöht werden, indem identifizierte Fehlerkandidaten gezielt evaluiert und korrigiert werden. Unmittelbare Vorteile ergeben sich hier durch die Berichtigung von Fehlern bei Adress- und Kontaktdaten des Kunden, d. h. dieser kann nun fehlerlos angesprochen werden (z. B. für die Versendung von Produktinformationen). Daneben verbessert die Anwendung der Assoziationsregeln auch mittelbar die Beratungsqualität, da falsche Daten korrigiert werden können, die ansonsten zu Fehleinschätzungen geführt hätten. Für das Optimierungsmodell ergibt sich folgende Zielfunktion, wobei sich die Betrachtung auf $T \in \mathbb{N}$ Perioden bezieht:

$$\sum_{t=1}^T \left(\frac{E_t}{(1 + i_{\text{kalk}})^t} - \frac{K_t}{(1 + i_{\text{kalk}})^{t-1}} \right) = \max! \quad (1)$$

E_t bildet dabei die DQ-induzierten Einzahlungen ab, die gemäß der Annahmen A.2 und A.3 mit den entsprechenden Auszahlungen K_t , die für die Maßnahmen anfallen, verrechnet werden. Der Kapitalwert der getätigten DQ-Investitionen im betrachteten Zeitraum ergibt sich durch Summation der abgezinsten DQ-induzierten Einzahlungen abzüglich der abgezinsten Maßnahmenkosten je Periode (vgl. Annahme A.3).

Da die Einzahlungen E_t durch den qualitätsinduzierten Transaktionsanteil generiert werden, lassen sich diese in Abhängigkeit von $\lambda_{DQM,t}$ angeben. Dabei lässt sich die Steigerung des qualitätsinduzierten Transaktionsanteils für die Unternehmung ökonomisch bewerten. Diese Bewertung wird mit dem Faktor $u \in \mathfrak{R}^+$ vorgenommen⁴, der dem monetären Wertbeitrag entspricht, den die Unternehmung nach Abzug aller sonstigen Kosten außer den DQ-Kosten bei Erhöhung des Transaktionsanteils von null auf eins realisiert. Der Faktor u ist dabei, wegen der Betrachtung des einzelnen Kunden(segments) und den damit einhergehenden begrenzten Auswirkungen auf die Struktur der Leistungserstellung, konstant. Alternativ könnte der Faktor u bei steigendem Transaktionsanteil wegen der Fixkostendegression und den damit verbundenen fallenden Grenzkosten der Leistungserstellung stetig ansteigen. Dies würde jedoch zu keiner grundsätzlichen Änderung der Ergebnisse führen, sondern lediglich die Komplexität der Darstellung erhöhen. Folglich lassen sich die Einzahlungen, die aus dem (verbesserten) DQ-Niveau und den durchgeführten DQM in Periode t resultieren, wie folgt beschreiben:

$$E_t = u \cdot \lambda_{DQM,t} \quad (2)$$

Für das Zustandekommen des qualitätsinduzierten Transaktionsanteils der Periode t ($\lambda_{DQM,t}$) ist zu berücksichtigen, dass bei der Wirkung der DQM auf das Transaktionsverhalten des Kunden(segments) ceteris paribus (c. p.) ein abnehmender Grenzbeitrag vorliegt. Dies ist damit zu begründen, dass erste DQM, welche bspw. die größten Mängel beseitigen, vom Kunden stärker wahrgenommen und honoriert werden, als eine weitere Intensivierung bereits relativ starker Bemühungen. Um dies im Modell zu berücksichtigen wird der Exponent $(1 - \eta)$ mit dem Wirkungsgrad der Maßnahmen $\eta \in]0; 1[$ eingeführt. Zudem ist die Wirkung der DQM zur Steigerung des Transaktionsanteils umso stärker, je größer die zugrunde liegende, betroffene Datenmenge ist. Dies ist unmittelbar deshalb einsichtig, da zum einen reaktive DQM in Form von Assoziationsregeln „besser“ werden, je größer die Datenmenge ist, anhand derer sie generiert werden. Zum anderen ist bei einer größeren (noch nicht qualitätsgesicherten) Datenmenge und deren Sicherung die Kundenwahrnehmung stärker. So kann für den Kunden durch die Anwendung von DQM auf bisher nicht qualitätsgesicherte Daten ein Mehrwert entstehen, wogegen deren Anwendung auf bereits korrekte Daten vom Kunden nicht positiv wahrgenommen wird und somit keinen zusätzlichen Wertbeitrag erzeugt. Eine

Anwendung von Assoziationsregeln auf bereits qualitätsgesicherte Daten hätte bspw. zur Folge, dass keine potenziellen Fehlerkandidaten identifiziert werden oder sich zumindest kein solcher Fehlerverdacht bestätigt. Deshalb wird die Intensität der durchgeführten Maßnahmen, neben der Verknüpfung mit den kundensegmentspezifischen Parametern m und η , in Verbindung mit der bisher nicht qualitätsgesicherten Datenmenge gebracht: $m \cdot (1 - DQ_t) \cdot D_{sup,t} \cdot DQM_t^{(1-\eta)}$. Der lineare Faktor m ist dabei aus dem Intervall $]0; 1]$ und drückt aus, inwiefern ein Kunde(typ) die Durchführung von DQM honoriert.

Aber nicht nur aus der direkten Wirkung der Maßnahmen können zusätzliche Transaktionsanteile für die Unternehmung resultieren – vielmehr kann auch das bestehende DQ-Niveau genutzt werden. Die Auswirkungen einer guten DQ hinsichtlich des Transaktionsanteils sind dabei insbesondere vom Transaktionsanteil $\lambda_{-DQM,t}$ abhängig, der auch ohne DQ zustande kommt. Dies liegt daran, dass eine weitere Steigerung eines ohnehin schon hohen Transaktionsanteils, im Vergleich zu einem niedrigen oder mittleren Anteil, schwieriger ist. Mit dem kundensegmentspezifischen Parameter $\beta \in [0; 1]$ wird zudem berücksichtigt, inwiefern auf Basis des bereits zu Periodenbeginn bestehenden qualitätsgesicherten Datenbestands zusätzliche Transaktionsanteile generiert werden können – d. h. wie ein Kunde(ntyp) die (verbesserte) DQ beim Kontakt mit der Unternehmung (z. B. individuellere Beratung und Kampagnen) schätzt: $\beta \cdot (1 - \lambda_{-DQM,t}) \cdot D_{sup,t} \cdot DQ_t$.

Insgesamt ergibt sich folgender funktionaler Zusammenhang, der im Weiteren den qualitätsinduzierten Transaktionsanteil $\lambda_{DQM,t}$ repräsentiert:

$$\lambda_{DQM,t} = \min \left\{ \left(m \cdot (1 - DQ_t) \cdot D_{sup,t} \cdot DQM_t^{(1-\eta)} + \beta \cdot (1 - \lambda_{-DQM,t}) \cdot D_{sup,t} \cdot DQ_t \right) \cdot (1 - \lambda_{-DQM,t}) \right\} \quad (3)$$

Dieser setzt sich dabei zum einen aus einem Anteil zusammen, der auf der direkten Maßnahmenwirkung beruht (erster Summand). Zum anderen können auf Basis der bereits in den Vorperioden qualitätsgesicherten Daten zusätzliche Transaktionen generiert werden (zweiter Summand). Die Minimumfunktion in (3) stellt sicher, dass der gesamte Transaktionsanteil gemäß Annahme A.1 im Bereich $[0; 1]$ bleibt.

Mit Ausdruck (3) sind die DQ-induzierten Einzahlungen E_t konkretisiert. Deshalb werden im Folgenden die Kosten K_t , welche für die in Periode t ergriffenen Maßnahmen resultieren, genauer spezifiziert. Dabei wird zwischen fixen und variablen Kostenanteilen unterschieden, da mit DQM einerseits Kosten K_{fix} verbunden sind, die vom Umfang der ergriffenen Maßnahmen unabhängig sind (z. B. für die Institutionalisierung eines DQ-Projektteams). Diese fallen dann an, wenn das Projekt überhaupt durchgeführt wird (d. h. $DQM_t > 0$). Andererseits werden variable Kosten berücksichtigt, die abhängig vom Umfang der Maßnahmen und der zugrunde

liegenden Datenmenge $D_{sup,t}$ sind (z. B. für die Überprüfung und Anwendung von Assoziationsregeln). Für die Bestimmung der variablen Kosten stellt dabei k_{var} den linearen Kostenfaktor dar. Dass die variablen Maßnahmenkosten in Abhängigkeit von der Maßnahmenintensität überproportional ansteigen beruht darauf, dass erste DQ-Verbesserungen im Verhältnis schneller und kostengünstiger realisiert werden können, wohingegen die Beseitigung der letzten, latenten Qualitätsmängel kostenintensiver ist. Den steigenden Grenzkosten bei zunehmender Maßnahmenintensität wird deshalb mittels des Exponenten $\mu \in]1; \infty[$ im Modell Rechnung getragen. Insgesamt ergibt sich somit folgender funktionaler Zusammenhang, wobei $\text{sgn}(DQM_t)$ die Signumfunktion darstellt, die den Fixkostenanteil genau dann berücksichtigt, wenn in der betrachteten Periode DQM ergriffen werden:

$$K_t = \text{sgn}(DQM_t) \cdot K_{\text{fix}} + k_{\text{var}} \cdot DQM_t^\mu \cdot D_{\text{sup},t} \quad (4)$$

Für eine mehrperiodige, dynamische Betrachtung werden des Weiteren Übergangsfunktionen für die Zustandsvariablen $D_{sup,t}$ und DQ_t benötigt, die beschreiben, wie sich die vorhandene Datenmenge der Unternehmung sowie der Anteil der qualitätsgesicherten Daten von einer Periode zur nächsten entwickeln.

Deshalb wird eine Funktion hergeleitet, die den Übergang von der Datenmenge $D_{sup,t}$ nach $D_{sup,t+1}$ in Abhängigkeit von den während der Periode t (neu) hinzukommenden Daten $\Delta D_{sup,t}$ schafft. Eine graphische Illustration liefert Abb. 1:

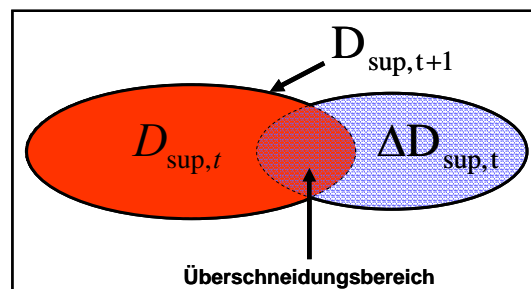


Abb. 1: Zusammenhang zwischen den Datenmengen $D_{sup,t}$, $\Delta D_{sup,t}$ und $D_{sup,t+1}$

Die Datenmengen $D_{sup,t}$ und $\Delta D_{sup,t}$ sind dabei als Mengen im mathematischen Sinne zu verstehen, d. h. jede ist für sich gesehen duplikatfrei – allerdings kann es sehr wohl Überschneidungen zwischen den beiden Datenmengen geben. Im Überschneidungsbereich befinden sich die Daten, die der Unternehmung bereits zu Beginn der Periode t zur Verfügung standen und während den Transaktionen der laufenden Periode erneut erfasst worden sind. Da beide Variablen auf die Vollständigkeit des Datenbestands hinsichtlich der Spezifikation abzielen, lässt sich aus diesen Duplikaten kein Mehrwert generieren (vgl. Annahme A.5). Im Gegensatz dazu erweitern die restlichen Daten aus $\Delta D_{sup,t}$ die Kundendatenmenge $D_{sup,t}$, wobei dieser Anteil

abhängig von der Größe der bereits vorhandenen Kundendatenmenge $D_{sup,t}$ ist. Dies ist damit zu begründen, dass bei einer kleinen Datenbasis noch verhältnismäßig viele der hinzukommenden Daten diese Ausgangsdatenbasis erweitern, wohingegen bei einer großen Datenmenge tendenziell viele bereits bekannt und somit in obigem Überschneidungsbereich anzusiedeln sind. Je mehr Daten $D_{sup,t}$ demnach bereits vorhanden sind, desto weniger sind von den neu hinzukommenden Daten $\Delta D_{sup,t}$ im Sinne der Datenspezifikation als neu anzusehen

($\frac{\partial(D_{sup,t+1} - D_{sup,t})}{\partial D_{sup,t}} < 0$) und umgekehrt. Dies soll an den beiden Extrembeispielen $D_{sup,t} = 0$

und $D_{sup,t} = 1$ verdeutlicht werden: Sind zu Beginn der Periode t keinerlei Kundendaten im System (d. h. $D_{sup,t} = 0$), so gilt $D_{sup,t+1} = \Delta D_{sup,t}$, da alle hinzukommenden Daten für die Unternehmung neu sind. Dagegen tragen bei $D_{sup,t} = 1$ die hinzukommenden Daten nicht zur Erweiterung der Datenbasis bei ($D_{sup,t+1} = D_{sup,t}$), da schon vorher alle benötigten Daten im System vorhanden sind. Der mit (5) gewählte Ausdruck für die vorhandene Kundendatenmenge zu Beginn der Periode $t+1$ erfüllt dabei nicht nur die obigen Anforderungen. Vielmehr gewährleistet er implizit ebenfalls $D_{sup,t+1} = f(D_{sup,t}, \Delta D_{sup,t}) \rightarrow [0;1]$ für $D_{sup,t}, \Delta D_{sup,t} \in [0;1]$:

$$D_{sup,t+1} = D_{sup,t} + (1 - D_{sup,t}) \cdot \Delta D_{sup,t} \quad (5)$$

Diese ist demnach mindestens so groß wie die zu Beginn der Vorperiode vorhandene sowie die während der Periode durch die Kundeninteraktionen hinzukommende Datenmenge ($D_{sup,t+1} \geq \max\{D_{sup,t}, \Delta D_{sup,t}\}$), da ein Datenverlust, bspw. bedingt durch einen technischen Defekt, ausgeschlossen wird. Zudem gilt $D_{sup,t+1} \leq D_{sup,t} + \Delta D_{sup,t}$, da externe Datenquellen, wie z. B. der Zukauf von Daten, im Modell gemäß Annahme A.4 nicht berücksichtigt werden.

Die während der Periode t hinzukommende Datenmenge $\Delta D_{sup,t}$ kann dabei aufgrund von Annahme A.4 als funktionaler Zusammenhang in Abhängigkeit von τ_t , $\lambda_{DQM,t}$ und $\lambda_{-DQM,t}$ dargestellt werden. Dies bedeutet, dass Daten aus der Interaktionstätigkeit mit dem Kunden gewonnen werden. Da hier vor allem Transaktionsdaten fokussiert werden, steigt die hinzukommende Datenmenge je mehr Transaktionen der Kunde bei der Unternehmung in Periode t tätigt. Da die Kunden(segmente) unterschiedliche absolute Transaktionsvolumina und verschiedene Präferenzen für die Datenweitergabe besitzen, wird durch τ_t festgelegt, wie viele Daten bei der Durchführung des gesamten Transaktionsvolumens eines Kunden in Periode t maximal hinzukommen können. Insofern soll der Zusammenhang folgendermaßen beschrieben werden⁵:

$$\Delta D_{sup,t} = \tau_t \cdot (\lambda_{-DQM,t} + \lambda_{DQM,t}) \quad (6)$$

Mit den Termen (5) und (6) ist die formale Repräsentation, wie sich die Datenmenge, die der Unternehmung zur Verfügung steht, von einer Periode zur nächsten entwickelt, konkretisiert⁶.

Im Folgenden wird der entsprechende Periodenübergang für den Anteil der qualitätsgesicherten Daten DQ_t betrachtet. Dabei wird berücksichtigt, dass die Maßnahmen lediglich auf nicht qualitätsgesicherte Daten wirken, da eine Anwendung auf bereits korrekte Daten realistischer Weise keinerlei Auswirkungen hinsichtlich DQ_{t+1} hat. Der Tatsache, dass qualitätsgesicherte Daten im Laufe der Zeit inkorrekt werden, wird dabei mittels des Faktors $\theta \in [0; 1]$ Rechnung getragen. Dieser gibt an, welcher Anteil der qualitätsgesicherten Datenmenge pro Periode inaktuell wird:⁷

$$DQ_{t+1} = \frac{(1-\theta) \cdot D_{sup,t} \cdot (DQ_t + (1-DQ_t) \cdot DQM_t)}{D_{sup,t+1}} \quad (7)$$

Der Anteil der qualitätsgesicherten Daten zu Beginn der Periode $t+1$ ist somit – abgesehen vom zeitlichen Verfall – neben der Maßnahmenintensität in Periode t vom DQ-Niveau zu Beginn der Periode t sowie den vorhandenen Kundendatenmengen zu Beginn der Perioden t und $t+1$ abhängig. Da DQ_{t+1} den Anteil der qualitätsgesicherten Daten repräsentiert, wird die im Zähler beschriebene qualitätsgesicherte Datenmenge⁸ ins Verhältnis zur insgesamt vorhandenen Datenmenge $D_{sup,t+1}$ gesetzt.

Das dynamische Optimierungsproblem zur Bestimmung des effizienten Einsatzes von DQM unter Berücksichtigung von mehreren Planungsperioden stellt sich basierend auf den Zusammenhängen (1) bis (7) wie folgt dar:

$$\begin{aligned} & \text{maximiere } \sum_{t=1}^T \left(\frac{E_t}{(1+i_{kalk})^t} - \frac{K_t}{(1+i_{kalk})^{t-1}} \right) \\ & \text{unter} \\ & E_t = u \cdot \lambda_{DQM,t} \\ & K_t = \text{sgn}(DQM_t) \cdot K_{fix} + k_{var} \cdot DQM_t^m \cdot D_{sup,t} \\ & \lambda_{DQM,t} = \min \left\{ m \cdot (1-DQ_t) \cdot D_{sup,t} \cdot DQM_t^{(1-n)} + \beta \cdot (1-\lambda_{-DQM,t}) \cdot D_{sup,t} \cdot DQ_t, (1-\lambda_{-DQM,t}) \right\} \\ & D_{sup,t+1} = D_{sup,t} + (1-D_{sup,t}) \cdot \Delta D_{sup,t} \\ & \Delta D_{sup,t} = \tau_t \cdot (\lambda_{-DQM,t} + \lambda_{DQM,t}) \\ & DQ_{t+1} = \frac{(1-\theta) \cdot D_{sup,t} \cdot (DQ_t + (1-DQ_t) \cdot DQM_t)}{D_{sup,t+1}} \\ & DQM_t \in [0;1] \end{aligned} \quad (8)$$

Alternativ kann im Optimierungsproblem (8) auch ein Endwert für die aufgebaute DQ eingehen, wie in Anhang 1 gezeigt wird. Dieser findet im Basismodell jedoch keine Berücksichtigung, da der Endwert zum einen keinen Vorteil bei der allgemeinen Simulation des Modells

bietet (siehe nächstes Kapitel) und zum anderen auch in der praktischen Anwendung problematisch⁹ sein kann.

Die obige detaillierte Formulierung wurde deshalb gewählt, um das Modell auch für reale Problemstellung einsetzen zu können, wie das Fallbeispiel in Kapitel E verdeutlicht. Zuvor wird jedoch noch auf eine Reihe allgemeiner Ergebnisse eingegangen.

D. Darstellung und Interpretation der Ergebnisse des Modells

Zur Ermittlung allgemeiner Ergebnisse wird aufgrund der Modelldetaillierung auf eine stetige Simulation mit numerischen Optimierungsverfahren zurückgegriffen, da sich das Modell für mehrere Perioden nicht algebraisch lösen lässt. Die Simulation erlaubt es dagegen allgemeine Investitionsszenarien für mehrere Perioden zu ermitteln, die bei gegebenen Rahmenbedingungen darstellen, zu welchem Zeitpunkt und in welcher Höhe in DQM zu investieren ist.

Die Durchführung der Simulation erfolgte mit Hilfe einer Sensitivitätsanalyse mit mehreren Variablen und sollte, unter Abdeckung der Definitionsbereiche der 13 Modellparameter, Auskunft darüber geben, welche Investitionsszenarien eintreten können und welchen Einfluss die einzelnen Modellparameter für das Auftreten eines Szenarios besitzen. Hierzu wurden für die Simulation die Definitionsbereiche eines jeden Modellparameters in Intervalle unterteilt. Hierbei ergaben sich insgesamt 69 Intervalle für die weiteren Untersuchungen (so stellten sich für die Analyse des Einflusses der initialen Datenbasis bspw. acht gleich große Intervalle über deren Definitionsbereich als ausreichend heraus). Danach wurde jeweils ein Parameter einzeln herausgegriffen (im Weiteren als Analysevariable bezeichnet) und über den gesamten Definitionsbereich systematisch variiert, um mittels einer Sensitivitätsanalyse zu ermitteln, welche Änderungen sich für den optimalen Umfang von DQM in den verschiedenen Perioden ergeben. Dabei erfolgten für jede Variation der Analysevariable 50 separate Simulationsläufe. Um den Einfluss der restlichen Modellparameter zu analysieren, wurden für jeden Simulationslauf mittels Zufallsoperator einzelne Werte für die Parameter aus den festgelegten Intervallen bestimmt und aufgezeichnet. Die sich dabei ergebende Vielzahl von unterschiedlichen Kombinationen an Intervallen wurde jeweils getrennt untersucht, d. h. es wurden automatisiert eigene Simulationsläufe durchgeführt. In jedem Simulationslauf wurden die resultierenden dynamischen Optimierungsprobleme mittels numerischer Verfahren (Nelder Mead, Differential Evolution, Simulated Annealing und Random Search) für insgesamt fünf Perioden ($T = 5$) gelöst, um die optimalen DQ-Maßnahmen DQM_t zu bestimmen und zu indizieren. In der Folge werden dabei für die Interpretation jedoch lediglich die Ergebnisse für die Perioden $t \in \{1, 2, 3, 4\}$ zugrunde gelegt. Dies liegt daran, dass bei dynamischen Entscheidungsmodellen die Optimierung der letzten Periode (hier die Höhe von DQM_5) stark davon abhängig ist, ob ein Endwert berücksichtigt wird oder nicht. Für den Fall, dass kein Endwert verrechnet

wird, fällt der Umfang der DQM in der letzten Periode c. p. geringer aus, da im Kalkül der monetäre Beitrag der aufgebauten DQ für spätere Perioden nicht eingeht. Wird im Gegensatz dazu ein Endwert berücksichtigt, ergibt sich in der letzten Periode c. p. ein höherer Umfang an DQM, da die Optimierungslogik versucht, die mangelnden Investitionsmöglichkeiten in den Perioden $T+1, T+2, \dots$ (d. h. $DQM_{T+1} = DQM_{T+2} = \dots = 0$) zu kompensieren.

Eine graphische Aufbereitung des Ergebnisses einer systematischen Variation der Analysevariablen $D_{sup,1}$ (entspricht der initialen Datenmenge) bei gegebenen Intervallen der anderen Modellparameter¹⁰ zeigt Abb. 2. In den Graphiken ist auf der Ordinate jeweils der optimale Maßnahmenumfang in den einzelnen Perioden DQM_t^* und auf der Abszisse die Analysevariable $D_{sup,1}$ angetragen.

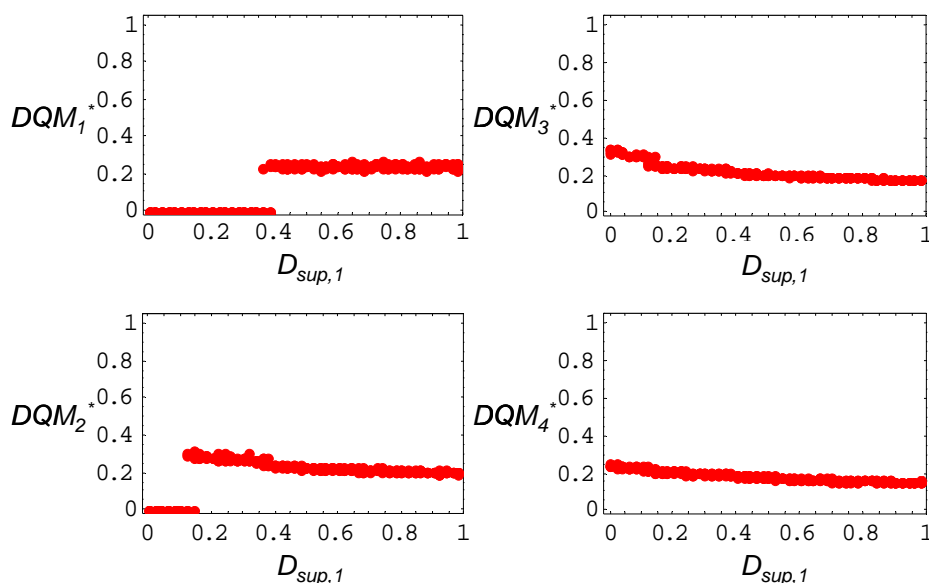


Abb. 2: Optimaler Maßnahmeneinsatz bei Variation der initialen Datenmenge

Die einzelnen Bilder verdeutlichen dabei die Wahl des optimalen Umfangs an DQM_t^* in den einzelnen Perioden. Sind nur wenige initiale Daten vorhanden (d. h. $D_{sup,1} \in [0; 0,15]$), so ist in den ersten beiden Perioden nicht zu investieren. Erst ab Periode 3 sind DQM zu ergreifen. Demgegenüber ist bspw. bei einer mittleren und höheren initialen Datenmenge (d. h. $D_{sup,1} \in [0,5; 1]$) nachhaltig über alle vier Perioden in DQM zu investieren.

Ergänzend ist zu erwähnen, dass aufgrund der unstetigen Kostenfunktion (keine Berücksichtigung der Fixkosten K_{fix} falls $DQM_t = 0$) eine Aufteilung in insgesamt 2^T stetige, differenzierbare Teilprobleme vorzunehmen war, um die Zuverlässigkeit der numerischen Verfahren zu gewährleisten. Aus diesem Grund wurden u. a. die zuvor genannten numerischen Verfahren nebeneinander verwendet.

Bevor detailliert auf die Investitionsszenarien eingegangen wird, lassen sich für eine Reihe von Parametern auf Basis der mittels Simulation vorgenommenen Sensitivitätsanalyse folgende allgemeine Ergebnisse festhalten. Dabei bezeichnet $\sum_t DQM_t^*$ die Summe über die in

den einzelnen Perioden optimalen Maßnahmenintensitäten, d. h. den optimalen Gesamteinsatz an DQM im Betrachtungszeitraum. Des Weiteren ist im Folgenden der optimale Zielfunktionswert des dynamischen Optimierungsmodells und somit der Kapitalwert, der mit der optimalen Wahl der Maßnahmenintensitäten in den einzelnen Perioden (DQM_t^*) erzielt werden kann, mit ZF^* beschrieben:

- Für den Wertbeitrag u gilt: $\partial\left(\sum_t DQM_t^*\right)/\partial u, \partial(ZF^*)/\partial u \geq 0$, d. h.: Können für den durch DQM induzierten Transaktionsanteil höhere/niedrige Einzahlungen u erzielt werden (ansonsten c. p.), so ergreift die Unternehmung DQM in höherem/niedrigerem Umfang (betrachtet über alle Perioden). Ebenso können auch höhere/niedrigere Kapitalwerte erzielt werden.
- Für den Kostenparameter k_{var} gilt: $\partial\left(\sum_t DQM_t^*\right)/\partial k_{var}, \partial(ZF^*)/\partial k_{var} \leq 0$, d. h. mit steigenden/fallenden variablen Kosten verringern/erhöhen sich in der Summe die ergriffenen DQM sowie der Kapitalwert.
- Für den Kostenparameter μ gilt: $\partial(ZF^*)/\partial \mu \geq 0$, d. h. mit steigendem/fallendem μ erhöht/verringert sich der Kapitalwert. Die Ursache für diesen Zusammenhang ist, dass die Maßnahmenintensität auf das Intervall $[0; 1]$ normiert ist und sich somit der Einsatz von DQM (wegen $(DQM_t)^{\mu_1} \leq (DQM_t)^{\mu_2}$ für $\mu_1 \geq \mu_2$ und $DQM_t \in [0; 1]$) verbilligt/verteuert. Ein weiterer Zusammenhang, der sich zwar bei Investitionsszenarien, in denen nur eine einmalige Investition in DQM erfolgte, nicht bestätigte, ansonsten jedoch regelmäßig beobachtet werden konnte, ist $\partial\left(\sum_t DQM_t^*\right)/\partial \mu \geq 0$. Folglich wird mit wachsendem μ , aufgrund der Verbilligung der Maßnahmen, deren Ergreifung vorteilhaft. Bei den Einmalinvestitionen ist dagegen zu beobachten, dass die stärker anwachsenden Grenzkosten für steigendes μ c. p. nur bis zu einem gewissen Grad toleriert werden. Danach wird dazu übergegangen, weniger in DQM zu investieren als noch bei kleinerem μ und so geringere DQ-induzierte Einzahlungen in Kauf zu nehmen. Trotzdem wird dabei insgesamt ein höherer Kapitalwert ZF^* erzielt, da die geringeren Kosten für die Durchführung der Maßnahmen den Rückgang an DQ-induzierten Einzahlungen überkompensieren.
- Für die kundensegmentspezifischen Parameter m und β gilt:
 $\partial\left(\sum_t DQM_t^*\right)/\partial m, \partial(ZF^*)/\partial m, \partial\left(\sum_t DQM_t^*\right)/\partial \beta, \partial(ZF^*)/\partial \beta \geq 0$. Mit den Parametern wird beschrieben, in welcher Höhe die Durchführung von DQM und das (resultierende)

DQ-Niveau zu einer Steigerung des Transaktionsanteils λ_{DQM} beiträgt. Mit steigenden/fallenden Parametern werden der Umfang an ergriffenen DQM im Betrachtungszeitraum sowie der Kapitalwert ZF^* erhöht/verringert.

- Für den Wirkungsgrad η gilt: $\partial(ZF^*)/\partial\eta \geq 0$, $\partial\left(\sum_t DQM_t^*\right)/\partial\eta \leq 0$, d. h. mit steigendem/fallendem η erhöht/verringert sich der Kapitalwert. Dies ist damit zu begründen, dass die Erhöhung/Verringerung des Parameters η einen besseren/schlechteren Wirkungsgrad der DQM bedeutet. Zudem ist, auf den ersten Blick überraschend, der Zusammenhang $\partial\left(\sum_t DQM_t^*\right)/\partial\eta \leq 0$ zu beobachten. Als Begründung hierfür kann die Kompensation eines schlechteren Wirkungsgrads ($\eta \downarrow$) durch einen höheren DQM-Einsatz angeführt werden.
- Für den Verfallsparameter der DQ θ gilt: $\partial\left(\sum_t DQM_t^*\right)/\partial\theta \geq 0$, $\partial(ZF^*)/\partial\theta \leq 0$, d. h. bei steigendem/fallendem Verfallsparameter werden höhere/niedrigere DQM induziert, da der DQ-Verfall (für den Fall eines höheren Verfalls) zu kompensieren ist. Eine Ausnahme davon bildet ein sehr geringer Verfall $\theta \in [0; 0,05]$. Hier ist zu beobachten, dass ausschließlich in der ersten Periode (relativ intensiv) investiert wird, um in den folgenden Perioden die (sich kaum oder nicht reduzierende) DQ auszunutzen.

Demgegenüber ist die Interpretation der bisher noch nicht betrachteten Größen D_{sup} (Datenmenge), DQ (DQ-Niveau) und $\lambda_{-DQM,t}$ (nicht DQ-induzierter Transaktionsanteil) hinsichtlich der zu ergreifenden DQM schwieriger. Zugleich ist diese interessant, da die Parameter in wechselseitiger Abhängigkeit die Modellergebnisse beeinflussen. Es lassen sich hier insgesamt vier allgemeine Investitionsmuster identifizieren, die nachfolgend – wie auch die Parameterkonstellationen, in welchen sie eintreten – beschrieben sind¹¹:

- [I] *[Nachhaltige Investition in konstanter Höhe]* Über die untersuchten Perioden hinweg sind Investitionen in DQM in gleicher Höhe vorzunehmen ($0 < DQM_t \cong DQM_{t+1}$ für $t \in \{1, 2, \dots, T-2\}$). Dieses Investitionsszenario tritt dann ein, wenn die Unternehmung zum einen bereits Qualitätssicherungen (konkret $DQ \in [0,3; 0,5]$) durchgeführt hat. Zudem müssen permanent infolge der durchzuführenden Basistransaktionen ($\lambda_{-DQM,t} \in [0,3; 0,6]$) Daten hinzukommen, deren Sicherung die Investitionen rechtfertigt. Die bereits vorhandenen Daten des Kunden D_{sup} beeinflussen hier nicht den Umfang und die Nachhaltigkeit der Investitionen. Interessant ist dieses Szenario bspw. für

Unternehmungen, die bereits eine etablierte Beziehung zum Kunden besitzen und bis zum Zeitpunkt der Investitionsentscheidung in DQ bereits mäßig investiert hatten.

- [II] [*Nachhaltige Investition in zunehmender Höhe*] Bei diesem Szenario steigen die Investitionen monoton an, d. h. es gilt $0 < DQM_t \leq DQM_{t+1}$ für $t \in \{1, 2, \dots, T-2\}$. Im Gegensatz zu Szenario [I] ist hier keine oder nur eine geringe DQ ($DQ \in [0; 0,3]$) vorhanden, jedoch der Basistransaktionsanteil hoch bis sehr hoch ($\lambda_{-DQM,t} \in [0,6; 0,9]$). Die spezifizierten Kundendaten D_{sup} beeinflussen hier wiederum nicht den Umfang und die Nachhaltigkeit der Investitionen. Bei intensiven Kundenbeziehungen, bei welchen die zugehörigen Daten bisher noch nicht gesichert wurden, ist dieses Investitionsszenario von großem Interesse.
- [III] [*Nachhaltige Investition in abnehmender Höhe*] Bei diesem Szenario nehmen die Investitionen monoton ab, d. h. es gilt $DQM_t \geq DQM_{t+1} > 0$ für $t \in \{1, 2, \dots, T-2\}$. Das Muster tritt auf, wenn eine Unternehmung bereits über eine umfangreiche Kundendatenbasis verfügt ($D_{sup} \in [0,5; 1]$), die jedoch wenig qualitätsgesichert ist ($DQ \in [0; 0,4]$). Daneben zeichnet sich die Konstellation durch einen niedrigen Anteil an Basistransaktionen ($\lambda_{-DQM,t} \in [0,1; 0,3]$) aus. Zusammen mit der bereits größeren Menge an vorhandenen Daten kommen hierdurch in den späteren Perioden wenig neue Daten hinzu. Dies ist auch so, wenn der qualitätsinduzierte Transaktionsanteil und die daraus resultierenden Kundendaten größeren Umfangs wären. Da jedoch bereits umfangreiche Kundendaten vorhanden sind, wird kontinuierlich (und bei gleich bleibendem Basistransaktionsanteil auch rückläufig) in die Qualitätssicherung investiert. Bei ehemals intensiven Kundenbeziehungen, die sich inzwischen verschlechtert haben (niedriger Basistransaktionsanteil im Zeitverlauf), ist dieses Szenario interessant, da es zwar zukünftige Investitionen vorsieht, jedoch bei einem sich insgesamt nicht verändernden, geringen Kunden- transaktionsanteil, die Investitionsrate kontinuierlich zurückgefahren wird.
- [IV] [*Einmalinvestition*] Im Unterschied zu den bisherigen Investitionsmustern findet hier nur eine einmalige Investition statt. Einmalinvestitionen charakterisieren sich dadurch, dass in mindestens einer Vor- und Folgeperiode keine DQM ergriffen werden (DQM_t ist eine Einmalinvestition falls gilt: $DQM_t > 0 \wedge DQM_{t-1} = 0 \wedge DQM_{t+1} = 0$). Im Gegensatz zu den Szenarien [II] und [III] stellt sich der Fall der Einmalinvestition in einer Konstellation mit wenig Kundendaten ($D_{sup} \in [0,15; 0,4]$), einer geringen, vorhandenen DQ ($DQ \in [0; 0,3]$) sowie einem sehr geringen Basistransaktionsanteil mit $\lambda_{-DQM,t} \in [0; 0,15]$ ein. Kennzeichnend für die Einmalinvestition ist dabei insbesondere der Basistransaktionsanteil, der äußerst wenig neu hinzukommende Daten zur Folge hat.

Zwar könnte demgegenüber qualitätsinduziert der Umfang an neu hinzukommenden Daten erhöht werden. Allerdings wären hierzu höhere Investitionen in DQM notwendig, um den Transaktionsanteil $\lambda_{DQM,t}$ und damit $\Delta D_{sup,t}$ zu steigern. Diese Investitionen werden jedoch deshalb nicht mehrperiodig getätigt, da in diesem Fall das Kosten-/ Nutzenverhältnis der DQM in den einzelnen Perioden zu schlecht wäre und insbesondere die resultierende mehrmalige Fixkostenbelastung nicht erwirtschaftet werden könnte. Somit kommt eine Einmalinvestition und keine nachhaltige DQ-Investition zu Stande. Letzgenannter Fall kann in einem dynamischen Modell auch (implizit) beim Vergleich von Investitionsmöglichkeiten zwischen Perioden auftreten. Dies bedeutet, dass in einer Periode zwar ein Nettoüberschuss erzielt werden könnte, es jedoch durchaus sinnvoll sein kann, in dieser Periode nicht zu investieren, um in der Folgeperiode eine (dann höhere) Einmalinvestition durchzuführen. Derzeit und auch zukünftig wenig intensive Kundenbeziehungen, über die zudem nur Daten schlechter Qualität verfügbar sind, fallen in dieses Investitionsszenario. Hier ist die Logik, mit einer einmaligen Investition zu versuchen, die Beziehung (ggf. wieder) zu aktivieren, ohne jedoch nachhaltig bei einem sich nicht verändernden, geringen Kundentransaktionsanteil zu investieren.

Analysiert man die genannten Investitionsmuster, so lässt sich Folgendes festhalten:

- Der Umfang der Datenmenge D_{sup} steuert maßgeblich, ob in einer Periode investiert werden soll (und damit den Startzeitpunkt der Investitionen). Nicht eine vorhandene schlechte DQ sollte der Auslöser sein, damit eine Unternehmung DQM ergreift. Vielmehr beeinflusst D_{sup} zusammen mit dem Umfang der Basistransaktionen λ_{-DQM} den Beginn der Investitionen im Zeitverlauf. Gilt $D_{sup,t} \gg \lambda_{-DQM,t}, \lambda_{-DQM,t+1}, \lambda_{-DQM,t+2},$ usw. (d. h. die vorhandene Datenmenge ist im Vergleich zur zukünftig zu erwartenden Datenmenge viel größer), so sollte c. p. in der Gegenwart (Periode t) investiert werden. Analog gilt für $D_{sup,t} \ll \lambda_{-DQM,t}, \lambda_{-DQM,t+1}, \lambda_{-DQM,t+2},$ usw., dass die Investitionen erst zukünftig zu tätigen sind.
- Der Umfang der Basistransaktionen λ_{-DQM} gibt Auskunft darüber, ob einmalig oder nachhaltig investiert werden soll. Bei einem geringen Basistransaktionsanteil $\lambda_{-DQM,t}$ wird eine Einmalinvestition getätigt. Bei einem dann kontinuierlich steigenden Anteil wird zuerst nachhaltig abnehmend (Szenario [III]), dann nachhaltig konstant (Szenario [I]) und schließlich nachhaltig zunehmend (Szenario [II]) investiert.
- Das DQ-Niveau ist, wie oben beschrieben, nicht ausschlaggebend dafür, ob Investitionen zu tätigen sind. Jedoch steuert DQ (sozusagen als Abbruchkriterium), wann die Investitionen in DQ eingestellt werden sollen, wobei dies durchaus (im Zusammenwirken mit dem

Basistransaktionsanteil λ_{-DQM}) auch schon bei einem mittleren DQ-Niveau erfolgen kann (vgl. Szenario [IV]).

Wenn in der Praxis die Modellanwendung und die damit verbundene Messung und Schätzung aller notwendigen Parameter (bspw. aufgrund der Erhebungskosten) nicht durchgeführt wird, so können die obigen Sensitivitäten und Investitionsmuster wichtige Anhaltspunkte und Handlungsempfehlungen für ein fortlaufendes DQ-Management geben und zur Entscheidungsunterstützung dienen. Strebt die Unternehmung dagegen im Rahmen eines ökonomisch orientierten DQ-Managements eine Anwendung des obigen Modells zur effizienten Planung des Einsatzes von DQM an, so sind die einzelnen Modellparameter zu quantifizieren.

Auf diese Messung und Schätzung wird neben der prinzipiellen Operationalisierung des Modells im nächsten Kapitel eingegangen.

E. Operationalisierung und Anwendung des Modells am Beispiel eines deutschen Mobilfunkanbieters

Zur Anwendung des obigen Modells wurden im Kampagnenmanagement-Prozess eines großen deutschen Mobilfunkanbieters spezifische DQ-Probleme in Zusammenarbeit mit der Fachseite und der DW-Abteilung identifiziert.

Dieser Prozess wird unten kurz dargestellt, um die beiden, hier im Weiteren diskutierten DQ-Probleme und die resultierenden ökonomischen Auswirkungen zu verdeutlichen. Danach wird detailliert auf die Operationalisierung und Anwendung des Modells eingegangen. Abschließend wird eine Auswahl an Handlungsempfehlungen diskutiert, die auf Basis des Modells abgeleitet werden konnten. Die im Folgenden verwendeten Zahlen stellen dabei eine Vereinfachung dar.

Der Kampagnenmanagement-Prozess setzt sich aus den einzelnen Teilprozessen *Planung & Design*, *Beauftragung & Durchführung* und *Monitoring & Steuerung* zusammen. Grundsätzlich wurden im Prozess in Zusammenarbeit mit den Fachseiten zwei Hauptansatzpunkte identifiziert, bei welchen insbesondere DQ-Probleme existierten. Zum einen handelt es sich um die *Kundenselektion* und zum anderen um die *Kundenansprache*.

Die *Kundenselektion* ist im Teilprozess *Planung & Design* angesiedelt. Ziel ist es, auf Basis der bisher im Planungssystem vorliegenden Informationen eine oder mehrere Zielgruppen für die Kampagne zu ermitteln und die entsprechenden Kunden zu selektieren. Dazu werden Auswertungen und Analysen auf den vorhandenen Kundendaten und Kampagneninformationen durchgeführt, um die Selektionskriterien für die Segmente zu bestimmen und zu verfeinern. Sind bei der Selektion relevante Attribute wie bspw. das Alter des Kunden, die Laufzeit des Vertrages oder gar der Vertragstyp selbst nicht oder nicht korrekt befüllt, hat dies zur Fol-

ge, dass Kunden den falschen Segmenten zugeordnet werden – daraus wiederum resultieren geringere Kampagnenerfolgsquoten, da „falsche“ Kunden dem Kreis der teilnehmenden Kunden zugewiesen werden.

Daneben stellt die vorhandene Kundendatenmenge auch die Grundlage für die *Kundenansprache* dar, d. h. die Kontaktierung des Kunden im Rahmen der Kampagne. Eine mangelhafte DQ führt hier bspw. dazu, dass keine oder keine korrekte und individuelle Ansprache des Kunden möglich ist. Die Folgen einer mangelhaften Qualität der Kontaktdaten können dabei von der Verärgerung des Kunden bis zu dessen Nichterreichbarkeit reichen. Insgesamt sind somit weniger (erfolgreiche) Kundenkontakte sowie geringere Erfolgsquoten auf Kampagnenebene das Resultat. Die Einspielung der Daten ist im Teilprozess *Beauftragung & Durchführung*, direkt vor dem Durchführungsstart der Kampagne, angesiedelt.

Zwei konkrete DQ-Probleme, die in den identifizierten Bereichen auftraten, sind zum einen fehlerhafte oder nicht vorhandene Kundenkontaktdaten (z. B. Adressdaten) und zum anderen eine mangelhafte Rollenzuordnung von Vertragspartner und Nutzer, was Probleme beim Kundenkontakt sowie der Kundenselektion nach sich zieht. Bei ersterem handelt es sich um ein allgemein bekanntes DQ-Problem, mit dem viele Unternehmungen branchenübergreifend zu kämpfen haben (vgl. z. B. English, 1999). Demgegenüber stellt das zweite ein Problem dar, das konkret vor Ort identifiziert wurde. So existiert sowohl im Geschäftskunden- (z. B. Rahmenvertragsgeschäft) als auch im Privatkundenbereich (z. B. Prepaidgeschäft) eine große Anzahl von Vertragsverhältnissen, in denen der Vertragspartner und somit Entscheider eines Mobilfunkvertrages nicht mit dem Nutzer der SIM-Karte übereinstimmt. Für den Mobilfunkanbieter sind diese Rolleninformationen deshalb von so großer Bedeutung, da lediglich der Entscheider eines Mobilfunkvertrages das unterbreitete Angebot annehmen kann und deshalb in den Kampagnen zu kontaktieren ist.

Ein Kampagnentyp, der von den vorgestellten DQ-Problemen besonders betroffen war, umfasst die *Prepaid2Postpaid*-Kampagnen, da hier neben den (allgemeinen) Kontaktdaten insbesondere die Rolleninformationen von Bedeutung sind, weil Prepaid-Karten häufig verschenkt, weitergegeben oder verkauft werden. Ziel der Kampagnen ist es, gezielt Prepaid-Kunden zu kontaktieren und dazu zu bewegen, auf Postpaid-Tarife und somit einen Laufzeitvertrag mit festem Mindestumsatz umzusteigen. Bei mangelhaft gepflegten Rolleninformationen ist dabei jedoch keine zielgerichtete Ansprache des Entscheiders des Vertrages möglich, was zwangsläufig zu Angebotsablehnungen führt. Dies spiegelte sich auch relativ deutlich in den Auswertungen des Kampagnenreportings wider: So wurde in einem dreimonatigen Test festgestellt, dass bei bis zu 20% der kontaktierten Personen das Kampagnenangebot nicht angenommen

werden konnte, da der entsprechende Vertragspartner nicht bekannt war. Insgesamt konnten die angestrebten Erfolgsquoten bei den Kampagnen nicht erreicht werden.

Da jedoch von Seiten des Mobilfunkanbieters Prepaid-Nutzer als Zielgruppe mit hohem Potenzial angesehen werden, weil diese Verträge einen typischen Einsteigertarif darstellen, sollten DQM durchgeführt werden, um die Erfolgsquote der Kampagnen zu steigern. Wegen der nicht zu unterschätzenden Kosten, die mit DQ-Projekten einhergehen, sollte zuvor untersucht werden, in welchem Ausmaß DQM unter ökonomischen Gesichtspunkten gerechtfertigt werden können.

Im hier dargestellten Teil des Anwendungsfalls wird dabei der Einsatz von Assoziationsregeln betrachtet, um mögliche Fehlerkandidaten im Datenbestand zu identifizieren, diese zu überprüfen und Fehler zu bereinigen (insgesamt wurden jedoch noch weitere DQM mit dem Modell auf ihren Einsatz hin untersucht und in der Wirkung miteinander verglichen). Assoziationsregeln beschreiben häufig auftretende Zusammenhänge zwischen Attributwerten durch Regeln, wie z. B. $a \text{ AND } b \Rightarrow c$, die nach einer Überprüfung als Integritätsbedingungen für den Bestand dienen können: $PLZ \Rightarrow Ort$ und $(Aktuelles\ Datum - Geburtsdatum < 16 \text{ Jahre}) \Rightarrow (Familienstand = \text{“ledig“})$ können hier als einfache Beispiele angeführt werden. Die Intensität der Maßnahmen (hier Umfang der Assoziationsregeln) ist dabei variierbar, d. h. es können bspw. mehr oder weniger Regeln entwickelt, überprüft und angewendet werden. Somit stellte sich die Frage, wann und in welchem Umfang in DQM bzw. in DQ-Projekte investiert werden soll.

Die Quantifizierung des Wertbeitrags, der mit den DQM und dem verbesserten DQ-Niveau in den einzelnen Perioden einhergeht, wurde mittels der Erfolgsquote λ_t vorgenommen. Der Grund für die Verwendung der Erfolgsquote lag in der direkten Zurechenbarkeit ihrer Verbesserung durch die DQM begründet. So lässt sich aus der erhöhten Erfolgsquote zwar wiederum der Umfang an zusätzlichen Transaktionen und damit die Steigerung des Transaktionsanteils ermitteln (weswegen auch der Beitrag u ohne größere Probleme zu berechnen war). Jedoch sollte der Transaktionsanteil nicht direkt Eingang finden, da im vorliegenden Fall neben den DQM auch andere Maßnahmen zu dessen Veränderung durchgeführt wurden, die nicht mit den Kampagnen in Bezug standen. Somit war es wesentlich geeigneter zu untersuchen, welche zurechenbare Steigerung der Erfolgsquote der Kampagnen aufgrund einer verbesserten Kundenselektion und -ansprache durch DQM erzielt werden kann. Der Beitrag u , der dabei mit einer Erhöhung der Erfolgsquote entsteht, konnte in Zusammenarbeit mit der Vertriebsabteilung ermittelt werden. Diesbezüglich wurde analysiert, welche Überschüsse die Unternehmung (nach Abzug aller sonstigen Kosten außer den DQ-Kosten) generiert, falls 10% der

Kunden des fokussierten Segments von Prepaid- auf Postpaid-Verträge umsteigen. Hier wurde ein Wert von 1,9 bis 1,95 Mio. Euro prognostiziert: Kann demzufolge die Erfolgsquote der betrachteten Kampagnen infolge der DQM bspw. um 1% gesteigert werden, so ergibt sich ein Wertbeitrag von 190 bis 195 Tsd. Euro. Da in der Praxis bei der Bestimmung der entscheidungsrelevanten Größen i. d. R. auf Schätzungen und Prognosen zurückgegriffen werden muss, wurde bei der Operationalisierung des Modells auf Intervalle abgestellt, in welchen sich die zu quantifizierenden Größen befinden sollen. Mittels Simulation können dann auf Basis dieser Entscheidungsgrundlage praktische Handlungsempfehlungen abgeleitet und der effiziente Einsatz von DQM bestimmt werden.

Für die Erfolgsquote $\lambda_{\sim DQM,t}$, die auch ohne die Durchführung von DQM und verbesserte DQ im entsprechenden Kundensegment zustande kommt, wurde das Kampagnenreporting der zurückliegenden Monate zugrunde gelegt. So ergab sich im Anwendungsbeispiel ein Schätzwert von 15%.

Um eine verbesserte Erfolgsquote zu erreichen, wurde die Anwendung von Assoziationsregeln als DQM fokussiert. Die Maßnahmenintensität und somit der Umfang an Regeln, die im Rahmen der Qualitätssicherung zum Einsatz kommen, können dabei über die Variation von Mindestsupport und Mindestkonfidenz der Regeln variiert werden. Beim Mindestsupport handelt es sich um die Häufigkeit, mit welcher der entsprechende Zusammenhang in der betrachteten Datenbasis enthalten sein muss, um als Assoziationsregel Verwendung finden zu können. Bei der Mindestkonfidenz wird dagegen die Untergrenze für die relative Häufigkeit festgelegt, mit der die entsprechende Implikation erfüllt sein muss – d. h., in wie vielen Fällen aus *a* AND *b* tatsächlich *c* folgen muss. Die Ermittlung von möglichen Assoziationsregeln kann dabei mittels geeigneter Software (z. B. WizRule® der Firma Wizsoft) toolbasiert erfolgen. Über die Vorgabe von Mindestsupport und -konfidenz können der Umfang der Assoziationsregeln und somit die Intensität der Maßnahmen DQM_t stufenlos eingestellt werden.

Um den Einsatz von DQM zu planen, sind insbesondere die bereits vorhandene Datenmenge und deren DQ-Niveau zu berücksichtigen. Dabei wurde zum einen die Vollständigkeit des Datenbestandes hinsichtlich der Spezifikation des Datenschemas zweckorientiert untersucht und bewertet. Dies bedeutet, dass mittels einer Metrik für das DQ-Merkmal Vollständigkeit (vgl. z. B. Parssian et al., 2004; Heinrich, Klier, 2005) die vorhandenen den für die Durchführung der Kampagnen benötigten Attributwerten gegenübergestellt wurden, um eine Kennzahl zu berechnen. Der relativen Bedeutung der einzelnen Attribute im Hinblick auf den Kampagnenzweck wurde dabei mittels Gewichtungsfaktoren Rechnung getragen – hier ist insbesondere das Attribut *Rolle* von Bedeutung. In der Anwendung ergab sich erwartungsgemäß ein eher

geringer Wert $D_{sup,1}$ von 0,40, da es sich beim Prepaid-Tarif um einen Einsteigertarif handelt und sich demzufolge i. d. R. noch nicht viele Informationen angesammelt haben. Zum anderen müssen bei Abschluss eines Prepaid-Vertrages gewisse Informationen und Stammdaten nicht angegeben werden. Da Kunden mit einer höheren Affinität für die unterbreiteten Angebote i. d. R. mehr telefonieren, höhere Umsätze tätigen und häufiger mit dem Mobilfunkanbieter in Kontakt treten, steigt die bekannte Kundendatenmenge, die in den zukünftigen Beratungen und Kampagnen eingesetzt werden kann.¹²

Daneben musste das Qualitätsniveau der bereits vorhandenen Kundendaten quantifiziert werden. Auch hier kann eine DQ-Metrik Verwendung finden (vgl. z. B. Parssian et al., 2004; Heinrich, Klier, 2005), die eine Kennzahl für die Korrektheit des relevanten Datenausschnittes lieferte. Für das in der Kampagne fokussierte Kundensegment ergab sich dabei ein verhältnismäßig geringer Wert von $DQ_I = 0,35$, da es sich bei der zugrunde liegenden Zielgruppe um ein Jugend- und Einsteigersegment handelt. So besteht mit dieser Kundengruppe i. d. R. nur selten Kundenkontakt, sodass Änderungen, die sich bei den Kundendaten ergeben, nur in Ausnahmefällen nachgepflegt werden können (z. B. veränderte Rolleninformationen).

Dass die vorhandenen Daten im Laufe der Zeit nicht mehr aktuell sind und die DQ demzufolge sinkt, muss in einer mehrperiodigen Planung ebenfalls berücksichtigt werden. Im Anwendungsfall wurde hier ein Schätzwert für den Verfallsparameter der relevanten Datenbasis gebildet. Dazu wurden jeweils Verfallsparameter auf Attributebene geschätzt (auf Basis von Erfahrungswerten und Stichproben) und dann zu einem Schätzer für die betroffene Datengrundlage aggregiert. Hierbei ergab sich ein Wert von 14%, da sich im betrachteten Segment die relevanten Kundendaten (z. B. Vertragsdaten und Rolleninformationen) verhältnismäßig oft ändern. Der Wert von 0,14 drückt dabei aus, dass innerhalb einer Planungsperiode – im Beispiel wurde ein Jahr zugrunde gelegt – ungefähr ein Siebtel der Daten inaktuell und somit inkorrekt werden.

Des Weiteren blieb zu untersuchen, inwiefern sich die DQM bei der Durchführung der Kampagnen auszahlen – d. h. inwiefern sich die Erfolgsquoten der Kampagnen verbessern. Um diesen Zusammenhang zu quantifizieren, wurden Vergangenheits- und Erfahrungswerte bei unterschiedlichen Kundengruppen – mit mehr oder weniger vorhandenen und qualitätsgesicherten Daten – analysiert und die jeweiligen Erfolgsquoten verglichen. So konnte eine Abschätzung vorgenommen werden, inwiefern bei Kunden, bei denen die Unternehmung eine umfangreichere und bessere Datenbasis zur Verfügung hat und stärker um die Qualität der Daten bemüht ist, eine höhere Erfolgsquote erzielt werden kann. Für die praktische Anwendung sollte dabei eine gewisse Fehlertoleranz möglich sein, d. h., für die Operationalisierung

des Modells konnten Intervalle für die Parameter m , η und β angegeben werden. Folgende Intervalle wurden hier ermittelt: $m \in [0,1; 0,2]$, $\eta \in [0,7; 0,8]$, $\beta \in [0,1; 0,2]$.

Daneben sind Kostenschätzungen für die Umsetzung der DQM vorzunehmen. Dabei sind zum einen die vom Umfang der Maßnahmen unabhängigen Fixkosten K_{fix} zu ermitteln, die bei Projektdurchführung anfallen. Im vorliegenden Beispiel wurden diese mit 450 bis 500 Tsd. Euro veranschlagt, da Kosten in entsprechender Höhe für die Einrichtung und Organisation eines Projektteams zur Umsetzung der DQM kalkuliert wurden. Zum anderen müssen die variablen Kosten, die in Abhängigkeit vom Umfang der Maßnahmen entstehen, quantifiziert werden. In der praktischen Anwendung wurden diesbezüglich die Kostenwirkung einer erhöhten Anzahl von Assoziationsregeln (durch Verminderung von Mindestsupport und -konfidenz) sowie die resultierenden Kosten für deren Evaluierung und Anwendung untersucht. Eine hohe Maßnahmenintensität DQM_t führt dabei dazu, dass relativ komplexe Regeln mit einbezogen werden, die einen relativ geringen Support- und Konfidenzwert besitzen und deren Prüfung und Anwendung demzufolge hohe Zeit- und Mitarbeiterressourcen beanspruchen. Gerade deswegen wird bei der Entwicklung des Modells auch ein konvexer Kostenverlauf gefordert. Im Praxisfall wurde der Parameter μ im Intervall $[1,7; 1,9]$ geschätzt. Der variable Kostensatz k_{var} ergibt sich dabei aus den insgesamt anfallenden Kosten, falls alle Assoziationsregeln umgesetzt werden würden. Im Beispiel resultierte ein Wert von $k_{var} \in [1,8; 2,0]$ (in Mio. Euro).

Mittels der bisherigen Analysen und Ausführungen konnte das Modell für die praktische Anwendung operationalisiert werden¹³. Welche Handlungsempfehlungen sich daraus ergaben, wird im Weiteren diskutiert.

Auf Basis des Modells soll bei oben beschriebener Konstellation in der ersten Periode noch nicht in DQ investiert werden ($DQM_1 = 0$). Vielmehr sollen weitere Kundendaten gesammelt werden, um die vorhandene Datenbasis von $D_{sup,1} = 0,4$ zu erweitern und den Beginn von Investitionen in DQM zu rechtfertigen. Dies ist nach Ablauf der ersten Planungsperiode (d. h. nach einem Jahr) der Fall. So soll zu Beginn der zweiten Periode eine stärkere Startinvestition erfolgen ($DQM_2 \approx 0,3$). Danach wird – gemäß Investitionsszenario [III] – kontinuierlich weiter in DQ-Projekte investiert, um dem Datenverfall von $\theta = 0,14$ entgegenzuwirken und den erfolgreichen Einsatz der Daten im Kampagnenmanagement auch in zukünftigen Perioden zu gewährleisten. Die Investitionshöhe ist hierbei aufgrund der relativ geringen Erfolgsquote von $\lambda_{-DQM,t} = 0,15$ leicht rückläufig ($DQM_3 \approx 0,27$, $DQM_4 \approx 0,23$).

Nach der beschriebenen Analyse und der Präsentation der Ergebnisse wurde seitens des Fachbereichs die Frage aufgeworfen, inwiefern sich das Ergebnis der Untersuchung verändern

würde, falls man beim betrachteten Segment – wie es bei anderen Zielgruppen der Fall ist – mehr Daten zur Verfügung hätte. Zur Illustration der entsprechenden Konsequenzen wurde die nachfolgende Abb. 3 erstellt. Hierbei wurde die initiale Datenmenge $D_{sup,1}$, die der Unternehmung zu Beginn der Betrachtung zur Verfügung steht, über den gesamten Definitionsbereich $[0; 1]$ variiert, um den Einfluss dieser Größe auf die zu ergreifenden DQM zu demonstrieren. Alle anderen Modellparameter wurden in den Simulationsläufen per Zufallsstichprobe aus den oben ermittelten Intervallen gezogen. In den Graphiken ist dabei die optimale Maßnahmenintensität in den einzelnen Perioden jeweils auf der Ordinate, die initiale Datenmenge $D_{sup,1}$ auf der Abszisse abgetragen. Der Untersuchung liegen wiederum vier Planungsperioden zugrunde (optimiert wurde wiederum für $T = 5$ Perioden):

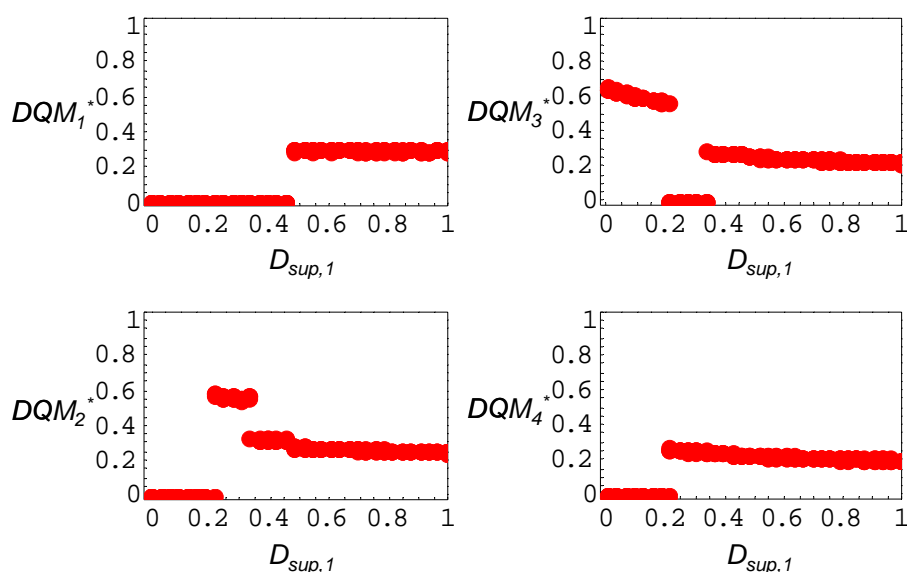


Abb. 3: Investitionsszenarien in Abhängigkeit von der initialen Datenmenge

Die Graphik zeigt, dass bei einer größeren vorhandenen Ausgangsdatenbasis ($D_{sup,1} > 0,5$) bei sonst gleichen Voraussetzungen sofort mit der Investition in DQM begonnen wird – der Aufbau eines Datenstocks kann somit im Gegensatz zu oben ($D_{sup,1} = 0,4$) entfallen. Stattdessen wird beginnend ab Periode 1 nachhaltig in abnehmender Höhe investiert. Legt die Unternehmung jedoch im Gegensatz dazu statt der Anwendung des Modells die Gestaltungsempfehlungen aus der Literatur zugrunde (vgl. z. B. English, 1999; Redman, 1996), so würde diese aufgrund der großen vorhandenen Datenmenge (z. B. bei $D_{sup,1} = 0,8$), dem relativ geringen DQ-Niveau und dem begrenzten neu hinzukommenden Datenvolumen auf eine einmalige Großinvestition in DQM setzen, um den Datenbestand (einmalig) zu bereinigen. Die tatsächliche Realisierung dieses Vorgehens hätte jedoch einen deutlich geringeren Kapitalwert im Vergleich zur nachhaltigen Investition in DQM zur Folge.

Anhand von Abb. 3 wird zudem der Einfluss der initialen Datenmenge auf den Startzeitpunkt der Investition deutlich: So erfolgt für $D_{sup,1} \geq 0,5$ die erste Investition schon in Periode 1, wohingegen für eine initiale Datenmenge von $0,2 < D_{sup,1} < 0,5$ erst in der zweiten Periode DQM ergriffen werden. Für $D_{sup,1} \leq 0,2$ ist lediglich eine Einmalinvestition in der dritten Periode zu tätigen, da aufgrund der geringen Wirkungsgrundlage der Maßnahmen die Fixkosten, die mit einer nachhaltigen Investition in DQM einhergehen, nicht erwirtschaftet werden können.

Die resultierenden Kapitalwerte ZF^* , die dabei mit der Realisierung der in Abb. 3 illustrierten Maßnahmen in den Perioden 1 bis 4 einhergehen, sind in Abhängigkeit von der vorhandenen Datenmenge $D_{sup,1}$ in Abb. 4 (in Mio. Euro) dargestellt:

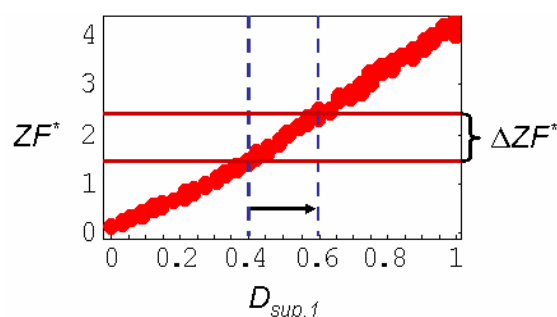


Abb. 4: Kapitalwert in Abhängigkeit von der initialen Datenmenge

Würde der Unternehmung in diesem Fall c. p. tatsächlich eine breitere Ausgangsdatenbasis zur Verfügung stehen, so könnte mittels des oben dargestellten optimalen Einsatzes von DQM ein deutlich höherer (in Höhe von ΔZF^*) barwertiger Einzahlungsüberschuss erwirtschaftet werden. Gemäß der Graphik hätte bspw. eine Erhöhung von $D_{sup,1} = 0,4$ auf $D_{sup,1} = 0,6$ ein ΔZF^* von ca. einer Mio. Euro zur Folge. Wird in diesem Zusammenhang ein Zukauf externer Daten (bspw. Adressdaten) in Erwägung gezogen, um die Ausgangsdatenmenge $D_{sup,1}$ zu erhöhen (die Annahme A.4 bleibt davon unbeeinflusst, da es sich um die Ausgangsdatenmenge handelt), so ist zu beachten, dass i. d. R. neue Simulationsläufe durchzuführen sind: So ändert sich in diesem Falle nicht nur die initiale Datenmenge ($D_{sup,1}$), sondern insbesondere auch deren Qualitätsniveau (DQ_I). Aus diesem Grund müssen für derartige Erwägungen überarbeitete Parameterkonstellationen Verwendung finden (vgl. hierzu Heinrich, Klier, 2005).

F. Zusammenfassung

Im Beitrag wurde ein quantitatives Optimierungsmodell zur mehrperiodigen Planung des Einsatzes von DQM entwickelt. Mit Hilfe des Modells wird die Entscheidung über die Investitionshöhen und -zeitpunkte in DQM und damit die Nachhaltigkeit des DQ-Managements unterstützt. Dies ist insofern wesentlich, da sich die in einer Periode aufgebaute DQ in den Folgeperioden natürlich ebenso positiv auswirkt. Um den hohen Bezug zu praktischen Problemstel-

lungen und die Operationalisierung des Modells zu verdeutlichen, wurde seine Anwendung im Kampagnenmanagement bei einem großen deutschen Mobilfunkanbieter erläutert. Wichtige Ergebnisse des Beitrags sind:

- Da das entwickelte Modell wichtige Einflussgrößen für die theoretische und praktische Planung von DQM berücksichtigt, kann eine Operationalisierung und Anwendung auch über das geschilderte Beispiel hinaus erfolgen. Dies bedeutet, dass mit dem Modell ein Instrumentarium zur Verfügung gestellt wird, das neben informationstechnischen auch ökonomische Einflussgrößen berücksichtigt. Wissenschaftliche Untersuchungen können auf Basis der obigen allgemeinen Ergebnisse nunmehr detailliert auf die Analyse einzelner DQM und die Beeinflussung verschiedener Parameter und Zusammenhänge eingehen. In der praktischen Nutzung ist es Unternehmungen im Gegensatz zu bisherigen Empfehlungen der Literatur möglich, ihre DQ-Anstrengungen zielorientiert und im Detail durchzuführen, indem die Umfänge einzelner DQM ermittelt werden.
- Bei den allgemeinen Ergebnissen sind vor allem die vier diskutierten Investitionsszenarien zu nennen. Die Einordnung in eines der Szenarien kann dabei anhand der Ausgangssituation der Kundenbeziehungen erfolgen (z. B.: Wie viele Kundendaten sind vorhanden? Wie intensiv sind die Beziehungen, d. h. wie viele neue Daten sind zu erwarten?). Auf dieser Basis können szenariospezifisch Handlungsempfehlungen für ein DQ-Management unter Berücksichtigung von mehreren Planungsperioden gegeben werden. So ist bspw. bei ehemals intensiven Kundenbeziehungen, die sich inzwischen verschlechtert haben (nur geringer Basistransaktionsanteil), zwar über mehrere Perioden zu investieren, jedoch bei einem sich nicht verändernden, geringen Transaktionsanteil, in abnehmender Höhe. Demgegenüber sollte bspw. bei intensiven Kundenbeziehungen, bei welchen die zugehörigen, umfangreichen Daten bisher noch nicht gesichert wurden über mehrere Perioden und in zunehmender Höhe in DQM investiert werden.
- Szenarioübergreifend ließen sich mit dem *Umfang der Datenmenge*, dem *Basistransaktionsanteil* sowie dem *DQ-Niveau* drei wesentliche Faktoren für die Investitionsentscheidung identifizieren. So ist der *Umfang der Datenmenge* (und nicht das existierende DQ-Niveau) maßgeblich für die Entscheidung, ob überhaupt in DQ investiert werden soll. Übersteigt die vorhandene Datenmenge die durch den Basistransaktionsanteil zukünftig hinzukommende Datenmenge, so ist c. p. sofort mit der Investition zu beginnen. Im umgekehrten Fall sind die Investitionen in die Zukunft zu verschieben. Die Fragestellung, inwiefern eine einmal begonnene Investition fortzuführen ist, wird primär durch den *Basistransaktionsanteil* beeinflusst. So ist bei einem sehr geringen Basistransaktionsanteil eine Einmalinvestition zu tätigen, wohingegen mit einem dann kontinuierlich steigenden

Anteil zuerst nachhaltig abnehmend, dann nachhaltig konstant und schließlich in der Höhe zunehmend investiert werden sollte. Als geeignetes Abbruchkriterium für die Investitionen ist das *DQ-Niveau* heranzuziehen, wobei die Investitionen in DQ dabei (im Zusammenwirken mit dem Basistransaktionsanteil) durchaus auch schon bei einem mittleren DQ-Niveau eingestellt werden können. Interessanterweise ist das DQ-Niveau jedoch nicht ausschlaggebend dafür, ob Investitionen in DQM überhaupt zu tätigen sind.

- Es gibt eine Reihe von Sensitivitäten, die hinsichtlich des Investitionsumfangs sowie der generierbaren Kapitalwerte allgemein abgeleitet werden können. Insbesondere ist bei höherem zeitlichem DQ-Verfall nicht – wie ggf. erwartet – weniger in DQM zu investieren, da sich die Wirkung der Maßnahmen schneller reduziert. Vielmehr sind i. d. R. umfangreichere DQM zu ergreifen, um dem DQ-Verfall entgegenzuwirken. Des Weiteren ist ein schlechter Wirkungsgrad der Maßnahmen kein Entscheidungskriterium dafür, weniger in DQM zu investieren. Stattdessen soll auch hier eine Kompensation durch einen höheren Maßnahmeneinsatz erfolgen.
- Unternehmungen, die für ihre Problemstellung detaillierte Empfehlungen zum Einsatz einzelner DQM ableiten wollen, können das Modell operationalisieren und anwenden. Damit können spezifische DQM, die im allgemeinen Modell aufgrund der notwendigen Abstraktion nicht berücksichtigt wurden, im Anwendungskontext analysiert werden.

Einen kritischen Punkt der Operationalisierung des Modells stellt die Schätzung der enthaltenen Parameter (insbesondere m , β , η und μ) dar. Bei der praktischen Anwendung wurde dem so begegnet, dass keine Einzelwerte, sondern vielmehr Intervalle für die Parameter zugrunde gelegt wurden, die wegen der Fehlertoleranz besser abgeschätzt werden konnten. Hintergrund war es, zu ermitteln, ob der optimale Umfang an DQM in den betrachteten Perioden (DQM_t^*) über die Parameterintervalle hinweg stark streut (Sensitivitätsanalyse bei Veränderung mehrerer Parameter). Wie die obigen Graphiken belegen, ist das Simulationsergebnis für die zugrunde gelegten Intervalle stabil – d. h., selbst wenn die praktische Schätzung der Parameter mit leichten Fehlern behaftet ist, bleibt das empfohlene Investitionsszenario i. d. R. gleich und der optimale Umfang der DQM schwankt nur gering.

Daneben ist die ex ante Messung der Wirkung der DQ problematisch (für die ex post DQ-Messung vgl. Heinrich, Klier, 2005). Während Größen wie Kosten, positiver Wertbeitrag, Datenmengen oder Transaktionsanteile im Ist abzuschätzen sind, liegt das Hauptproblem in der schwierigen Isolierbarkeit und Zurechenbarkeit der Wirkung der DQM, da in Unternehmungen eine Vielzahl von Eingriffen gleichzeitig stattfindet. So verändert sich bspw. der Transaktionsanteil des Kunden(segments), der zudem auch durch nicht zu eliminierende unternehmensexterne Eingriffe (z. B. von Wettbewerbern) beeinflusst wird.

Als kritischer Punkt der Modellierung ist die Annahme A.5 anzuführen, die besagt, dass aus Überschneidungen der bereits vorhandenen ($D_{sup,t}$) und der neu hinzukommenden ($\Delta D_{sup,t}$) Datenmenge kein Mehrwert für die Unternehmung generiert wird. Diese Annahme ist hinsichtlich der Vollständigkeit bzgl. der Spezifikation richtig. Dagegen ist es durchaus vorstellbar, dass aus einer erneuten Erfassung eines bereits im Informationssystem enthaltenen Attributwertes für die Unternehmung insofern Vorteile resultieren können, dass ein Abgleich, bzw. eine Aktualisierung der Daten vorgenommen werden kann – der Mehrwert würde dabei bspw. in einer Prüfung der Aktualität bestehen.

Neben der Bearbeitung der kritischen Punkte wird derzeit noch an einer Erweiterung des Modells um proaktive DQM gearbeitet. Dieser Maßnahmentyp unterscheidet sich in der Wirkungsgrundlage von den reaktiven Maßnahmen. Reaktive DQM wirken auf die bereits vorhandene Datenmenge (z. B. Anwendung von Assoziationsregeln), wohingegen sich proaktive DQM auf die Qualität zukünftig zu erfassender Daten auswirken (z. B. Überarbeitung der Erfassungsprozesse der Kundendaten). Daneben werden die Erweiterung des obigen Modells hinsichtlich einer gleichzeitigen Betrachtung mehrerer Kundensegmente untersucht und der praktische Anwendungsfall weiter verfolgt, um die ex post Analyse über mehrere Perioden hinweg durchzuführen.

Anmerkungen

¹ Für den Fall, dass bei der Optimierung ein Kundensegment betrachtet wird, sind die einzelnen Parameter für das Segment zu bestimmen.

² Um das gesamte Transaktionsvolumen für verschiedene Kundengruppen zu ermitteln, kann bspw. auf allgemeine Studien und Umfragen, wie die Allensbacher Computer- und Technik-Analyse (ACTA), die Allensbacher Markt- und Werbeträger-Analyse (AWA) oder auf Online-Offline 2 des Spiegel-Verlags zurückgegriffen werden. Darüber hinaus führen viele Unternehmen im Rahmen einer Potenzialanalyse für Märkte und Kundengruppen Erhebungen hierzu durch bzw. lassen diese erstellen.

³ Reaktive Maßnahmen besitzen analytischen Charakter (Qualitätssicherung „im nachhinein“) und beeinflussen die DQ bereits vorhandener Kundendaten.

⁴ Der Parameter kann hierbei durchaus auch zeit- bzw. periodenabhängig als u_t ins Modell eingehen, falls wirtschaftliche Entwicklungen des Kunden dies erfordern. Dadurch ändern sich die allgemeinen Ergebnisse nicht wesentlich.

⁵ Der Faktor τ_t geht hier vereinfachend linear ein, könnte jedoch auch als Exponent Berücksichtigung finden. Dies würde jedoch an den Modellergebnissen nichts Grundlegendes ändern.

⁶ Neben den Transaktionsdaten können in der neu hinzukommenden Datenmenge $\Delta D_{sup,t}$ ohne Weiteres ebenfalls Interaktionsdaten berücksichtigt werden. Sie sind demnach Teil des spezifizierten Datenschemas (spezifikationsorientierte Datenqualität) und damit auch Teil der Menge $D_{sup,t}$. Hier sind zwei Fälle zu unterscheiden:

1) Die Mengen an Transaktionsdaten und Interaktionsdaten korrelieren nicht. Zwar erhöhen dann im ersten Schritt die Interaktionsdaten die Datenmenge $D_{sup,t}$, die mit den DQM gesichert werden, was sich positiv auf den qualitätsinduzierten Anteil $\lambda_{DQM,t}$ auswirkt. Jedoch

wird – wegen der fehlenden Korrelation – die Menge an Interaktionsdaten nicht wieder durch $\lambda_{DQM,t}$ erhöht. Es ergibt sich nur ein einmaliger Effekt. Deshalb können die Interaktionsdaten im Ausdruck (6) bspw. als Variable γ_t (mit $\gamma_t \in [0; 1]$ in Bezug auf die in einem Schema spezifizierte Interaktionsdatenbasis als Teil der Menge $D_{sup,t}$) eingehen, wobei gilt: $\gamma_t \neq f(\lambda_{DQM,t}, \lambda_{-DQM,t})$ und somit: $\Delta D_{sup,t} = \tau_t(\lambda_{-DQM,t} + \lambda_{DQM,t}) + \gamma_t$ mit $\tau_t + \gamma_t \leq 1$. Da γ_t im Modell exogen ist, ändern sich dadurch die allgemeinen Ergebnisse nicht.

2) Die Mengen an Transaktionsdaten und Interaktionsdaten sind (positiv) korreliert (wie z. B. bei Filialbanken, weswegen diese u. a. auch den Kundenkontakt und eine Beratung aktiv anstreben). Falls dies der Fall ist, können Interaktionsdaten analog zu Transaktionsdaten behandelt werden, d. h. ein höherer Transaktionsanteil führt nicht nur zu mehr neu hinzukommenden Transaktionsdaten, sondern ebenfalls zu mehr neu hinzukommenden Interaktionsdaten. Beide erhöhen die Datenmenge $\Delta D_{sup,t}$ und sind damit Teil der Datenbasis $D_{sup,t}$. Die Menge $D_{sup,t}$ an Transaktions- und Interaktionsdaten steigt – wegen der vorhandenen Korrelation – mit dem Transaktionsanteil (ausgedrückt durch die Variable δ_t), d. h. der bisherige Ausdruck (6) kann folgendermaßen ergänzt werden:

$\Delta D_{sup,t} = (1 + \delta_t) \tau_t(\lambda_{-DQM,t} + \lambda_{DQM,t}) + \gamma_t$ mit $(1 + \delta_t) \tau_t + \gamma_t \leq 1$. Die Berücksichtigung der Interaktionsdaten, die nicht in Zusammenhang mit den durchgeführten Transaktionen stehen, können dabei natürlich wieder wie in 1) mit γ_t eingehen.

⁷ Derartige Verfallparameter lassen sich in der Praxis (siehe Fallbeispiel) relativ gut mittels Stichproben ermitteln.

⁸ Die Wirkung des Eingriffs geht hier vereinfachend linear ein. Alternativ könnte ein konkaver Zusammenhang argumentiert werden, was an den Modellergebnissen jedoch nichts Grundlegendes verändern würde.

⁹ Stellt sich bei der Modellanwendung in der Praxis unter Berücksichtigung des Endwerts heraus, dass zwar die Investitionen/Auszahlungen in den ersten Perioden erfolgen müssen, jedoch die Nutzenwirkung der DQ (infolge des Endwerts) primär in sehr viel späteren Perioden auftritt, so kann dies trotz Diskontierung schwerlich gegenüber dem Investitionsentscheider argumentiert werden. Hier würde man sich wohl dem Vorwurf des „Schönrechnens“ von Investitionen ausgesetzt fühlen.

¹⁰ Die Parameterkonstellation ist im Beispiel wie folgt: $u \in [1,90 \text{ Mio.}; 1,95 \text{ Mio.}]$; $\lambda_{-DQM,t} = 0,20$; $DQ_I = 0,30$; $\theta = 0,15$; $m \in [0,12; 0,15]$; $\eta \in [0,70; 0,75]$; $\beta \in [0,08; 0,12]$; $K_{fix} \in [450 \text{ Tsd.}; 500 \text{ Tsd.}]$; $k_{var} \in [1,9 \text{ Mio.}; 2,0 \text{ Mio.}]$; $\mu \in [1,6; 1,7]$; $\tau_t = 1$ ($t = 1, 2, \dots, 5$); $i_{kalk} = 0,10$

¹¹ Bei den Parametern mit nach oben offenem Definitionsbereich wurden bei der Simulation Obergrenzen festgelegt, um unnötige Simulationsläufe auf Basis unrealistischer Ausprägungen zu vermeiden.

¹² In diesem Zusammenhang musste auch festgelegt werden, wie viele der relevanten Daten ein Kunde maximal weitergibt. Aus Erfahrungswerten ergab sich hier, dass für das Kampagnenangebot affine Kunden, bereitwillig diese Daten bereitstellten. Deswegen wurde der Parameter τ_t für alle Perioden t jeweils mit eins festgelegt.

¹³ Der kalkulatorische Zinssatz i_{kalk} wurde dabei vereinfachend mit 0,1 angesetzt.

Literatur

- Alt, G. (2003): Sehr geehrter Frau Müller – Falsche Daten sind nicht nur peinlich, sondern verursachen auch hohe Kosten. In: FAZ Nr. 244 vom 21.10.2003, S. B2.
- Arndt, D.; Langbein, N. (2002): Data Quality in the Context of Customer Segmentation. In: Proceedings of the 7th International Conference on Information Quality. Cambridge 2002, S. 47-60.

- Ballou, D. P.; Wang, R.; Pazer, H.; Tayi, G. K. (1998): Modeling Information Manufacturing Systems to Determine Information Product Quality. In: *Management Science* 44 (1998) 4, S. 462-484.
- Ballou, D. P.; Tayi, G. K. (1999): Enhancing data quality in data warehouse environments. In: *Communications of the ACM* 42 (1999) 1, S. 73-78.
- Campanella, J. (1999): *Principles of quality cost*. Milwaukee 1999.
- Cappiello, C.; Francalanci, C.; Pernici, B.; Plebani, P.; Scannapieco, M. (2003): Data Quality Assurance in Cooperative Information Systems: A multi-dimensional Quality Certificate. In: Catarci, T. (Hrsg.): *International Workshop on Data Quality in Cooperative Information Systems*. Siena 2003, S. 64-70.
- English, L. (1999): *Improving Data Warehouse and Business Information Quality*. New York 1999.
- Feigenbaum, A. V. (1991): *Total Quality Control*. New York 1991.
- Fisher, C. W.; Chengalur-Smith, I.; Ballou, D. P. (2003): The Impact of Experience and Time on the Use of Data Quality Information in Decision Making. In: *Information Systems Research* 14 (2003) 2, S. 170-188.
- Heijkers, M. (2003): Ohne Datenhygiene kein CRM. In: *Computerwoche* Nr. 41 vom 10.10.2003, S. 40.
- Heinrich, B.; Helfert, H. (2003a): Nützt Datenqualität wirklich im CRM? – Wirkungszusammenhänge und Implikationen. In: Uhr, W; Esswein, W; Schoop, E. (Hrsg.): *Wirtschaftsinformatik 2003, Band II*. Heidelberg 2003, S. 231-253.
- Heinrich, B.; Helfert, H. (2003b): Analyzing Data Quality Investments in CRM – a model based approach. In: *Proceedings of the 8th International Conference on Information Quality*. Cambridge 2003, S. 80-95.
- Heinrich, B.; Klier, M. (2005): *Metriken und Messverfahren zur Bewertung der Datenqualität als Basis für ein ökonomisch orientiertes Datenqualitätsmanagement*. Diskussionspapier des Lehrstuhls WI-IF der Universität Augsburg. Augsburg 2005.
- Helfert, M. (2000): *Eine empirische Untersuchung von Forschungsfragen beim Data Warehousing aus Sicht der Unternehmenspraxis*. Arbeitsbericht BE HSG/CC DWS/05 des Instituts für Wirtschaftsinformatik der Universität St. Gallen. St. Gallen 2000.
- Helfert, M. (2002): *Planung und Messung der Datenqualität in Data-Warehouse-Systemen*. Dissertation. Bamberg 2002.
- Khalil, O. E. M.; Harcar, T. D. (1999): Relationship Marketing and Data Quality Management. In: *SAM Advanced Management Journal* 64 (1999) 2, S. 26-33.
- Krishnamoorthi, K.S. (1989): Predict quality cost changes using regression. In: *Quality Progress* 22 (1989) 12, S. 52-55.
- Kreutzberg, J. H. (2000): Has Quality Management Any Effect on Quality? – Analysis of Quality Management by a Non-linear Model. In: *Proceedings of the 5th International Conference on Information Quality*. Cambridge 2000, S. 242-257.
- Machowski, F.; Dale, B. G. (1998): Quality costing: An examination of knowledge, attitudes, and perceptions. In: *Quality Management Journal* 5 (1998) 3, S. 84-95.
- Matzer, M. (2004): Datenqualität frisst die Hälfte des Data-Warehouse-Etats. In: *Computerzeitung* Nr. 3 vom 12.01.2004, S. 12.
- Parssian, A.; Sarkar, S.; Jacob, V. S. (2004): Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product. In: *Management Science* 50 (2004) 7, S. 967-982.
- Redman, T. C. (1996): *Data Quality for the Information Age*. Norwood 1996.
- SAS Institute (2003): *Europäische Unternehmen leiden unter Profitabilitätseinbußen und niedriger Kundenzufriedenheit durch schlechte Datenqualität*. Studie der SAS Institute GmbH. Heidelberg 2003.
- Shank, J.M.; Govindarajan, V. (1994): Measuring the cost of quality: A strategic cost management perspective. In: *Journal of Cost Management* 8 (1994) 2, S. 5-17.

Wang R. Y.; Storey, V. C.; Firth, C. P. (1995): A Framework for analysis of data quality research. In: IEEE Transaction on Knowledge and Data Engineering 7 (1995) 4, S. 623-640.

Anhang 1: Berücksichtigung des Endwertes der Datenqualität im Modell

Will man im entwickelten Modell berücksichtigen, dass die aufgebaute DQ noch in weiteren Folgeperioden genutzt wird (Endwert), so kann dem wie folgt Rechnung getragen werden:

Die Erfolgsbeiträge aus der Umsetzung der DQM resultieren aus der (DQ-induzierten) Erhöhung des Transaktionsanteils $\lambda_{DQM,t}$. Da der Endwert im Fokus steht, sind für die Perioden $T+1$, $T+2$, usw. keine (weiteren) DQM zu berücksichtigen. Somit ergibt sich die Erhöhung des Transaktionsanteils lediglich aus dem zweiten Summanden der Gleichung (3) $\{\beta (1 - \lambda_{-DQM,t}) \cdot D_{sup,t} \cdot DQ_t\}$, d. h. aus den zum betrachteten Zeitpunkt gesicherten Daten $D_{sup,t} \cdot DQ_t$ ($t = T+1, T+2, \dots$). Der Endwert der aufgebauten DQ lässt sich demnach wie folgt angeben:

$$EW = \sum_{t=T+1}^{\infty} \frac{u \cdot \min\{\beta \cdot (1 - \lambda_{-DQM,t}) D_{sup,t} \cdot DQ_t; (1 - \lambda_{-DQM,t})\}}{(1+i)^t}$$

Da hierbei in den Perioden $T+1$, $T+2$, usw. keine DQM durchgeführt werden, ergibt sich der zu berechnende Term $D_{sup,t} \cdot DQ_t$ mit Gleichung (7) für $t = T+2, T+3$, usw. zu

$$D_{sup,t} \cdot DQ_t = D_{sup,t} \cdot \frac{(1-\theta) \cdot D_{sup,t-1} \cdot (DQ_{t-1} + (1-DQ_{t-1}) \cdot DQM_{t-1})}{D_{sup,t}} = (1-\theta) \cdot D_{sup,t-1} \cdot DQ_{t-1} = \dots = (1-\theta)^{t-(T+1)} \cdot D_{sup,T+1} \cdot DQ_{T+1}$$

Insofern lässt sich der Endwert wie folgt darstellen:

$$EW = \sum_{t=T+1}^{\infty} \frac{u \cdot \min\{\beta \cdot (1 - \lambda_{-DQM,t}) (1-\theta)^{t-(T+1)} D_{sup,T+1} \cdot DQ_{T+1}; (1 - \lambda_{-DQM,t})\}}{(1+i)^t}$$

Dass dieser Ausdruck konvergiert, lässt sich dabei leicht mit Hilfe des Majorantenkriteriums nachweisen. Unterstellt man zusätzlich, dass der nicht DQ-induzierte Transaktionsanteil $\lambda_{-DQM,t}$ ab Periode $T+1$ konstant ist und verzichtet wegen $\beta \leq 1$ auf die Minimumfunktion, so lässt sich der Ausdruck noch weiter vereinfachen mit:

$$EW = u \cdot \beta \cdot D_{sup,T+1} \cdot DQ_{T+1} \cdot (1 - \lambda_{-DQM,T+1}) \sum_{t=T+1}^{\infty} \frac{(1-\theta)^{t-(T+1)}}{(1+i)^t} = \frac{u \cdot \beta \cdot D_{sup,T+1} \cdot DQ_{T+1} \cdot (1 - \lambda_{-DQM,T+1})}{(i + \theta) \cdot (1+i)^T}$$