Virtuelle Modelle in der Logistik
und ihre praktische Bedeutung

Harald-Fritjof NELSON

Abstract

Ohne Modelle ist keine Optimierung möglich, es sei denn, man ist bereit, gigantische Investitionen vorzunehmen, um alles im "real life" auszuprobieren. Dabei muss ein gutes Modell keineswegs ein "Abbild der Wirklichkeit" sein, und auch nicht versuchen einen Vorgang 1:1 abzubilden, es muss lediglich im Rahmen der vorgesehenen Optimierung sich "richtig" verhalten. Man unterscheidet zwischen realen Modellen (die man auch "angreifen" kann) und virtuellen Modellen, die also nur in der Vorstellung existieren, oder in einem mathematischen Algorithmus ihren Niederschlag finden. Je mehr Erfahrung über einen Vorgang oder einen Zusammenhang existiert, umso eher kann man auf ein reales Modell verzichten und mit virtuellen Modellen arbeiten, was natürlich wesentlich effizienter und billiger ist. In der Technik ist fast nichts ohne Modelltheorie möglich, während in der Wirtschaft auch heute noch Modelle lediglich punktuell eingesetzt werden. Die Ursache dafür ist die geringere Akzeptanz von etwas komplexerer Mathematik im kaufmännischen Bereich, mehr aber vielleicht auch die Tatsache, dass sehr viele (auch publizierte) Modelle nur äußerst eingeschränkt "richtig" sind, weil schon im Ansatz gravierende Fehler gemacht werden, die dann den Wert des Modells in Frage stellen.

Der vorliegende Beitrag versucht die Theorie der virtuellen Modelle auf Basis der Regressionsanalyse anhand von praktischen Beispielen darzustellen und räumt mit den üblicherweise in der Literatur dargestellten starken Vereinfachungen, die zu den genannten Fehlern führen, auf. Mit den heutigen Möglichkeiten auf modernen Hard- und Software-Plattformen, ist es möglich, ohne selbst zu tief in die Mathematik eindringen zu müssen, lediglich aus dem Verständnis der eigenen Daten heraus, passende Modelle zu entwickeln, wobei insbesondere eine sehr stark verbesserte "Visualisierung" (also grafische Darstellung dieser Modelle) bei der Entwicklung hilft!

Virtuelle Regressionsmodelle

Virtuelle Modelle sind die Grundlage für jede rechnerische Optimierung und wären daher, würden sie mehr angewandt werden, von eminent wirtschaftlicher Bedeutung! Eine besondere Stellung nehmen dabei die Regressionsmodelle ein, also Modelle, mit denen Zusammenhänge zwischen interessierenden Parametern in Form von mathematischen Funktionen dargestellt werden. Hat man solche Modelle zur Verfügung, dann kann man im Planungs- oder Angebots-Stadium sehr schnell – praktisch auf „Fingerdruck“ - Kosten, Arbeitstunden, Materialgewichte etc. berechnen. In der Technik ist es schon lange selbstverständlich, dass man bei der Konstruktion eines technischen Werkes die gesamten bisherigen externen Erfahrungen nutzt, die in Form von Diagrammen, Nomogrammen etc., in technischen Handbüchern zu finden sind, wie auch die Erfahrungen aus den eigenen Projekten, bei denen auch aus Konstruktionsdaten Verhältniswerte gebildet werden, die man dann sozusagen verallgemeinern kann. Kein Konstrukteur wird heute ein technisches Werk planen, ohne nicht vorher „geschlossene Zusammenhänge“ aus schon erstellten „ähnlichen“ Werken zu bilden. Der Beginn seiner Planungsphase ist immer, aus existierenden Werken ähnlicher Art zu interpolieren und zu extrapolieren. Er bedient sich – manchmal ohne diesen modernen Ausdruck überhaupt zu kennen – zumindest ansatzweise, „virtueller Modelle“, die vielleicht nur in Form von einigen Punkten vorliegen, zwischen denen interpoliert wird, um so gewisse Kennwerte für das neue Werk abschätzen zu können.

In der Wirtschaft wird die gesamte Modelltheorie auch heute noch in einem weit geringeren Maße eingesetzt, als in der Technik, was auch ganz natürlich ist, denn die Wirtschaft hatte ja kaum etwas mit realen Modellen zu tun, also Modellen, die gebaut werden, um damit zu experimentieren. Damit war in der Technik auch der Übergang zu virtuellen Modellen, in die ja sehr viel Erfahrung mit realen Modellen einfloss, sehr viel einfacher. Ein weiterer Grund der relativ geringen Verwendung ist darin zu finden, dass virtuelle Modelle zumeist mit den rechnerischen Methoden der Statistik ermittelt werden, in der Literatur für Statistik aber auf die speziellen Bedürfnisse der Modelltheorie kaum eingegangen wird.

So ist in der Literatur, wenn man die Methoden der Regressionsrechnung studiert, sehr häufig von „linearen Regressionen“ die Rede, obwohl es lineare Zusammenhänge äußerst selten gibt. Bei der Darstellung von Funktionen mehrerer Veränderlicher (der multiplen Regression) ist fast immer nur von mehrfacher multipler linearer Regression die Rede, die man für virtuelle Modelle kaum brauchen kann und übliche Programmpakete, insbesondere Tabellenkalkulations-Programme geben sogar die Tendenzen der verwendbaren Funktionen fix vor, was es schlicht unmöglich macht ein brauchbares virtuelles Modell damit zu erstellen. Weiters wird in der statistischen Literatur im allgemeinen betont, dass für eine gute Regression der Korrelationskoeffizient hoch (nahe 1) sein sollte und die Summe der quadratischen Fehler möglichst klein. Wer sein Augenmerk allein darauf richtet, kommt bei der Erstellung von Modellen zwangsweise zu völlig unbrauchbaren Ergebnissen, da beides für Modelle nur scheinbar wichtig ist. Tatsächlich kommt es ausschließlich darauf an, eine – im allgemeinen bekannte (weil logische) - Funktionstendenz zu erreichen und unter genau dieser Randbedingung natürlich auch die Punkte unter Minimierung der Summe der quadratischen Fehler anzunähern, was ja die Methode der Regressionsrechnung ganz automatisch macht! Es wäre absolut unzulässig, um die Punkte noch „besser" anzunähern, die „natürliche Tendenz" der Funktion zu missachten, worauf in der Folge noch näher eingegangen wird! Um gute virtuelle Modelle zu erhalten muss man sozusagen auf die GEOMETRIE achten, das heißt, die entstehende Funktion muss ein ganz bestimmtes Aussehen haben.

Der Autor hat schon vor fast 30 Jahren auf all diese Zusammenhänge in umfangreichen Arbeiten hingewiesen. Damals konnte man zwar schon ausgefeilte virtuelle Rechenmodelle erstellen, die Visualisierung (also die feine grafische Darstellung) war aber noch sehr aufwendig. Heute ist beides, also die rechnerische Beherrschung der Materie, wie auch die grafisch sehr ansprechende Darstellung – mit geeigneter Software – auf jedem aktuellen PC möglich!

Es soll deshalb nochmals in Form eines „Kochrezeptes" der Vorgang der Modellerstellung geschildert und an praktischen Beispielen erläutert werden.

Was ist neu?

Vorerst soll noch kurz darauf hingewiesen werden, welche Änderungen sich in der Methode der Modell-Erstellung ergeben haben und was daher wirklich neu ist!

Hinreichend bekannt ist jedenfalls:

- die Technik der prinzipiellen Modellerstellung mittels Regressionsanalyse, die man in jedem Lehrbuch für Statistik findet, allerdings meist in derart starken Vereinfachungen, dass damit eine Umsetzung in die Praxis nicht möglich ist.
- die Implementierung in Tabellenkalkulationsprogrammen, bei denen jedoch nur ein sehr beschränktes Set von Elementarfunktionen angeboten wird. Wählt man dort etwa Polynome, dann werden lediglich ganzzahlige und diese im allgemeinen als fortlaufende Reihe $\mathbf{x^0, x^1, x^2}$ etc. vorgeschlagen. Darin liegt schon der erste gravierende Fehler, denn diese willkürliche Auswahl ist absolut unzulässig, wie aus den späteren Ausführungen klar werden wird. Da bei Tabellen-Kalkulations-Programmen üblicherweise die konkrete Auswahl der Funktionen nicht möglich ist, und daher eine bestimmte Tendenz des resultierenden Modells nicht erreicht werden kann, sind Tabellen-Kalkulations-Programme für die Modellerstellung nur in sehr trivialen Fällen geeignet! Es gibt zwar von Software-Unternehmen eine Reihe

von „Add-Ins“ und „Plug-Ins“, beispielsweise für MS-Excel, die diesen Mangel zum Teil beheben, die volle Flexibilität erreicht man aber auch damit nicht!

Weiters bekannt sind:

- die prinzipielle Vorgehensweise bei der Modellerstellung, sowie die Technik der Transformationen, um schwierig darzustellende Modelle in beliebigen „anderen Räumen“ darzustellen. (siehe auch Literaturverzeichnis 3).

Obzwar also diese Methoden im Prinzip schon lange bekannt sind, wurden sie auf wirklich breiter Basis bislang nicht eingesetzt, weil folgende Voraussetzungen für den Einsatz gegeben waren:

- außer der genauen Kenntnis der darzustellenden Daten, war auch ein gewisses mathematisches Verständnis und insbesondere „geometrische Vorstellungskraft“ erforderlich
- die verlangte Flexibilität konnte mit starrer Programmierung nicht erreicht werden, womit auch heute noch alle am Markt erhältlichen Programme nur sehr bedingt dafür geeignet sind, bzw. irgendwelche gravierende Eingriffe für spezielle Zwecke erfordern

Das führte dazu, dass die Anwendung dieser Methoden bislang Spezialisten vorbehalten war, die meist ohne Eigenprogrammierung auch nicht auskamen.

Neu ist nunmehr:

- dass heute eine Verbindung zwischen Tabellenkalkulationsprogrammen (wie z.B. MS Excel) und einer sehr flexiblen Definitionsumgebung (Dyalog APL) möglich ist, die hinsichtlich der Flexibilität weit über das hinaus geht, was mit Visual Basic Makros und üblichen Plug-Ins bislang erreicht werden kann.
- dass die Visualisierung (also die grafische Darstellung der Modelle) auf den aktuellen Hard- und Software-Plattformen in sehr guter Auflösung sehr schnell durchgeführt werden kann, womit ein Mangel an mathematischen Kenntnissen und geometrischer Vorstellungskraft auf der Benutzerseite (wie er naturgemäß immer vorhanden ist), kompensiert werden kann.
- das kann soweit gehen, daß der Benutzer gar nicht bemerkt, dass er nicht nur mit einem Tabellenkalkulationsprogramm arbeitet, weil sogar die Rückrechnungsfunktionen automatisch direkt auf das Excel-Arbeitsblatt übertragen werden können.

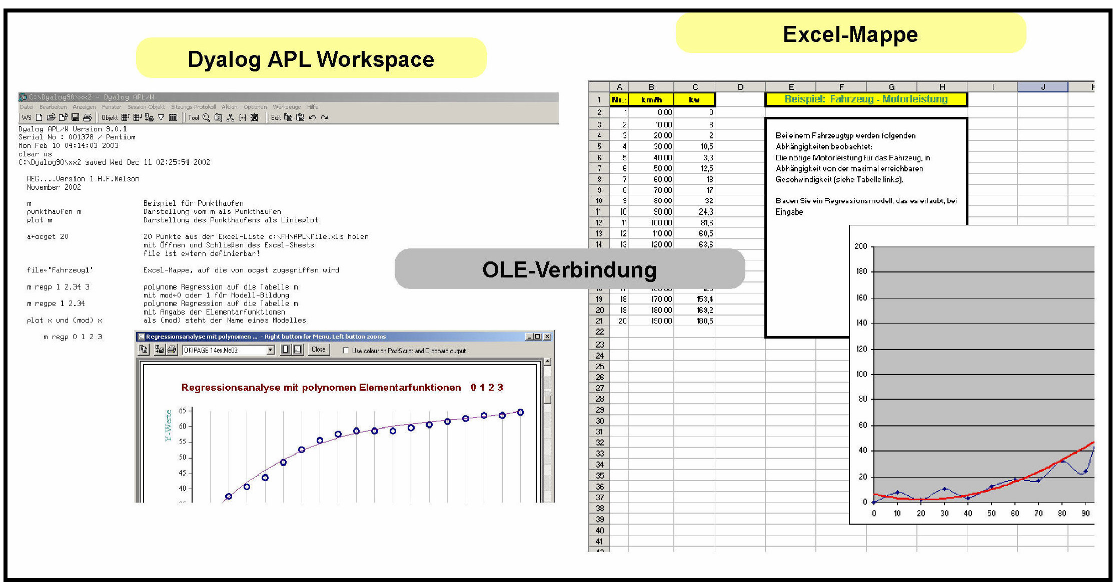
Die „APL – Excel – Brücke“

Von den Bequemlichkeiten einer guten Tabellenkalkulation, die dem Benutzer häufig tatsächlich das Denken abnimmt und ihm in jeder Situation versucht vernünftige Ergebnisse zu liefern, auch wenn er sie gar nicht konkret verlangt hat, „hoffnungslos verwöhnt“, ist es schwierig die Erstellung von virtuellen Modellen in ähnlicher Weise „bequem“ zu gestalten!

Dies kann jedoch weitgehend realisiert werden, indem man dem Benutzer seine gewohnte Tabellenkalkulations-Umgebung belässt, für jene Berechnungen aber, die flexible mathematische Ansätze erfordern – im Hintergrund - auf eine „Definitions-Umgebung“ ausweicht, wie es **Dyalog-APL** ist. (Siehe Literaturverzeichnis Punkt 1) Dies ist nicht zu verwechseln mit einer Makro-Programmierung wie beispielsweise Visual Basic, mit der sich komplexere Mathematik kaum realisieren lässt.

Dazu ist es weiters erforderlich, für die Modelle vollautomatisch Rückrechnungsfunktionen zu erstellen, die so, wie sie sind, in Programmpakete eingebaut werden können.

Schließlich müssen diese Modelle sehr schnell und bequem optisch dargestellt werden, wofür die exzellenten Funktionen von **Causeway RainPro** eingestzt wurden! (siehe Literaturverzeichnis Punkt 2)



Wiederholung der Theorie

Um einen besseren Einstieg in das Thema zu gewährleisten, sei kurz die Theorie der Modell-Erstellung mittels Regressionsanalyse wiederholt:

Unter Regressionsanalyse versteht man ganz allgemein die Darstellung der Erfahrungen, die in einem „Punkthaufen" liegen, als eine geschlossene mathematische Funktion, die – im Idealfall - über alle unabhängigen Variablen vom Wert 0 bis ∞ seine Gültigkeit hat. Diese geschlossene Funktion bildet immer nur eine abhängige Variable über eine oder mehrere unabhängige Variable ab und wird als Summe von „Elementarfunktionen" dargestellt. Eine Elementarfunktion kann jede beliebige Funktion sein, solange sie nur Elemente der unabhängigen Variablen enthält.

Damit kann man also beispielsweise die folgenden Modelle erstellen:

- die Herstellungskosten eines Fahrzeuges in Abhängigkeit von der Dienstgeschwindigkeit und der Ladekapazität
- das Stahlgewicht einer Brücke in Abhängigkeit von der Länge, der Breite und der Tragfähigkeit
- den Durchmesser des Turbinenrades einer Förderpumpe in Abhängigkeit von der Fördermenge und der Förderhöhe
- die Steigerung des Umsatzes in Abhängigkeit von den Ausgaben für Werbung

Diese Beispiele sind ganz beliebig herausgegriffen. Es sei aber betont, dass es praktisch KEIN Unternehmen gibt, in dem nicht in irgendeinem Bereich solche Modelle sinnvoll und im Bereich der Vorkalkulation, Planung und Optimierung wirtschaftlich interessant sind!

Will man virtuelle Modelle als Rechenmodelle implementieren, die nicht nur in einem gewissen Bereich, sondern weitgehend im Bereich von 0 bis ∞ (der unabhängigen Variablen) Gültigkeit haben, dann sind folgende Phasen zu durchlaufen:

- Festlegung der (richtigen) unabhängigen Variablen
- Ermittlung der „natürlichen Zusammenhänge" zwischen den Variablen
- Sammlung von Erfahrungsdaten
- Abbildung der Erfahrungsdaten mit einer Funktion, die exakt dem unter 2). ermittelten „natürlichen Zusammenhang" entspricht.

Dabei ist weiters zu beachten:

- die Modelle sollen so klein wie möglich sein, also Abhängigkeiten mit extrem geringem Einfluß sollten weggelassen werden
- insbesondere die Zahl der unabhängigen Variablen sollte ein Minimum sein
- die Zahl der „Elementarfunktionen", mit denen die Zusammenhänge beschrieben werden soll möglichst klein sein (in der Praxis ist jedes Modell mit mehr als 3 Elementarfunktionen falsch, weil das bedeutet, daß nicht die richtigen Elementarfunktionen gewählt wurden!)

Um diese Vorgänge und Regeln etwas zu erläutern seien in der Folge einige Beispiele gebracht.

Festlegung der richtigen unabhängigen Variablen

Die unabhängige Variable ist immer jene, die ich erreichen will, die abhängige die, was ich dafür aufwenden muss.

- Es ist daher sinnvoller, die nötige Motorleistung in Abhängigkeit von der Geschwindigkeit aufzutragen, die ich mit einem Fahrzeug erreichen will, als umgekehrt.
- Will ich beispielsweise ein Modell erstellen, für die Herstellungskosten von Pumpen einer bestimmten Bauart, dann wäre es nicht sinnvoll diese in Abhängigkeit von den Hauptabmessungen der Pumpen aufzutragen, weil das keine primären Parameter sind. Den Auftraggeber interessieren die Fördermenge und die Förderhöhe, alles andere sind abgeleitete Parameter, daher nicht unabhängig.
- Genauso wäre es sinnlos, ein Modell für die Herstellungskosten eines Schiffes zu erstellen, bei dem die Kosten über der Länge, der Breite und dem Tiefgang aufgetragen sind. Auch das sind abgeleitete Parameter. Die eigentlichen unabhängigen Größen sind die Dienstgeschwindigkeit und die Ladekapazität. Alles andere sind – für einen bestimmten Schiffstyp – abgeleitete Größen.

Bei der Regressionsanalyse wird jeweils eine abhängige Variable aus einer oder mehreren unabhängigen Variablen erklärt. Dabei können Modellansätze – ohne an Richtigkeit einzubüßen – oft sehr stark vereinfacht werden, wenn man nur die Zusammenhänge logisch durchleuchtet. Dies sei an einem extrem trivialen Beispiel erklärt:

Jemand will ein Modell erstellen, mit dem er „auf Knopfdruck" für eine Kartonschachtel beliebiger Größe die Herstellungskosten (und natürlich den Materialverbrauch) berechnen will.

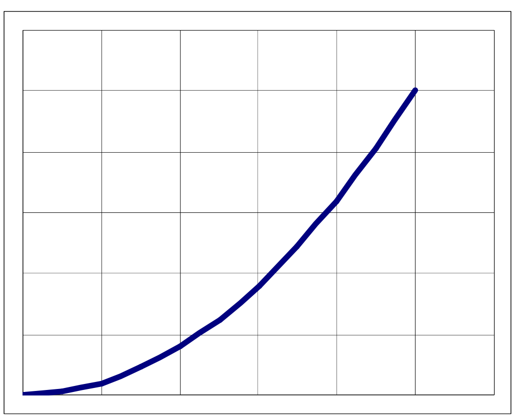
Die Kosten sind offensichtlich abhängig von der Länge, der Breite, der Höhe und der Wandstärke des Kartons. Es handelt sich daher bei der ersten Betrachtung um ein 5-dimensionales Modell. Überlegt man aber, dass man natürlich alle Schachteln mit derselben relativen Festigkeit erzeugen möchte, dann fällt die Wandstärke als unabhängige Variable weg, sie wird zur abhängigen Variablen. Bedenkt man weiter, dass der Materialverbrauch (abgesehen von einer eventuellen Börtelung) exakt die Oberfläche ist und die notwendige Wandstärke – um eine bestimmte Festigkeit zu erreichen – vom Verhältnis zwischen Volumen und Oberfläche abhängig ist, dann genügen diese beiden unabhängigen Variablen, nämlich Volumen und Oberfläche, um die notwendige Wandstärke einerseits und die Kosten andererseits zu berechnen. Aus dem 5-dimensionalen Problem ist somit ein 3-dimensionales geworden!

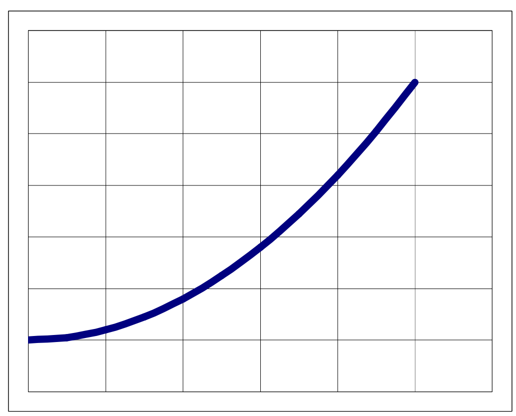
Ermittlung der „natürlichen Tendenzen“

Bei der Erstellung von virtuellen Modellen geht es im allgemeinen darum, Zusammenhänge, die im Prinzip bekannt sind, darzustellen, das heißt, es werden zwar „Erfahrungswerte“ oder eine „Punktwolke“ als eine Eingabe genommen, es gibt aber in diesem Fall eine zweite Eingabe, nämlich die ganz konkrete Form der Funktion, die sich ergeben soll. Diese „natürliche Tendenz“ zu umgehen, nur um die Erfahrungswerte besser anzupassen, wäre völlig falsch.

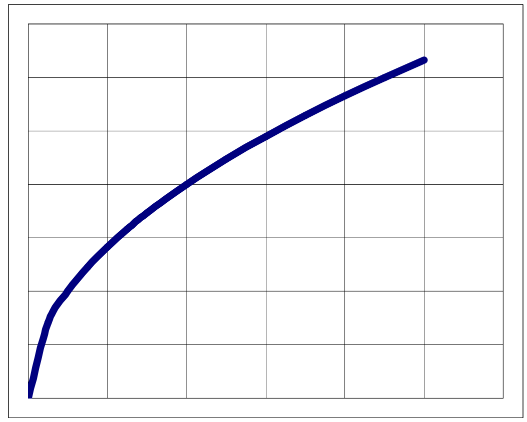
In der Folge werden einige dieser „natürlichen Tendenzen“ angeführt!

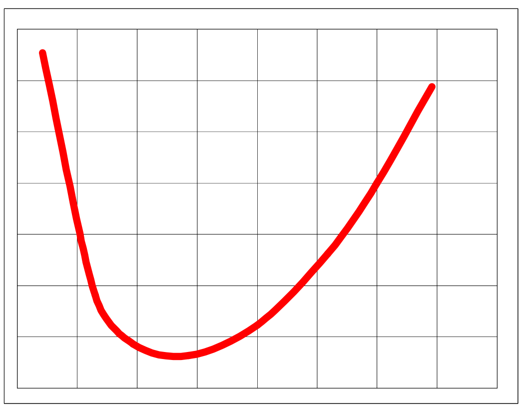
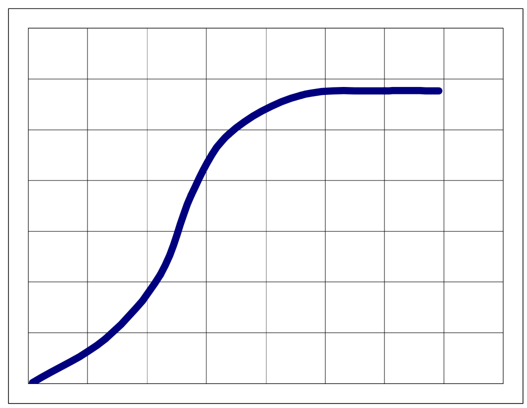
(der Ursprung der unabhängigen Variablen bei den folgenden Modellen ist immer NULL!)





abhängige Variable:	**Motorleistung**	abhängige Variable:	**Fahrzeugpreis**
unabhängige Variable:	**Geschwindigkeit**	unabhängige Variable:	**Geschwindigkeit**

			
abhängige Variable:	**Einnahmen** (Transportvolumen)	abhängige Variable:	**Einnahmen** (Transportvolumen)
unabhängige Variable:	**Ladekapazität**	unabhängige Variable:	**Geschwindigkeit**

			
abhängige Variable:	**Kosten pro transportierter Ladungseinheit**	abhängige Variable:	**Umsatz**
unabhängige Variable:	**Geschwindigkeit**	unabhängige Variable:	**Werbung**

Es ist wichtig, dass das Regressionsmodell die vorgegebene Funktionstendenz exakt trifft und der Einlauf bei $\mathbf{x = 0}$ und Auslauf bei $\mathbf{x = \infty}$ gesichert ist! Die Funktion soll also über den gesamten „denkbaren“ Bereich gültig sein! Dies soll auch dann der Fall sein, wenn nur über einen sehr limitierten Bereich Punktbelegungen vorhanden sind, weil ansonsten das Modell der Tendenz nach nicht stimmt.

Es wäre daher völlig sinnlos ein Modell zu erstellen, für die eingegebenen Punkte die Modellwerte zurückzurechnen und – weil die konkreten Abweichungen gering sind – damit zufrieden zu sein. Das Modell ist schließlich nicht dafür gedacht gegebene Werte rückzurechnen, sondern für die Interpolation und die

Extrapolation. Darüber sagen aber die statistischen Kenndaten, wie auch die Summe der quadratischen Abweichungen überhaupt nichts aus. Die Rückrechnung zu den gegebenen x-Werten kann durchaus befriedigend sein, und das Modell kann dennoch für Interpolationen, aber insbesondere für Extrapolationen unbrauchbar sein!

Ein konkretes Beispiel soll das zeigen:

Beispiel: Fahrzeug - Kosten

Nr.:	km/h	Kosten
1	0,00	4,8
2	10,00	4,7
3	20,00	4,4
4	30,00	4
5	40,00	5,5
6	50,00	6,7
7	60,00	19,3
8	70,00	6,3
9	80,00	9
10	90,00	9,6
11	100,00	12,8
12	110,00	2,8
13	120,00	13,6
14	130,00	24,1
15	140,00	20,8
16	150,00	23,5
17	160,00	23,7
18	170,00	32,6
19	180,00	33,5
20	190,00	39

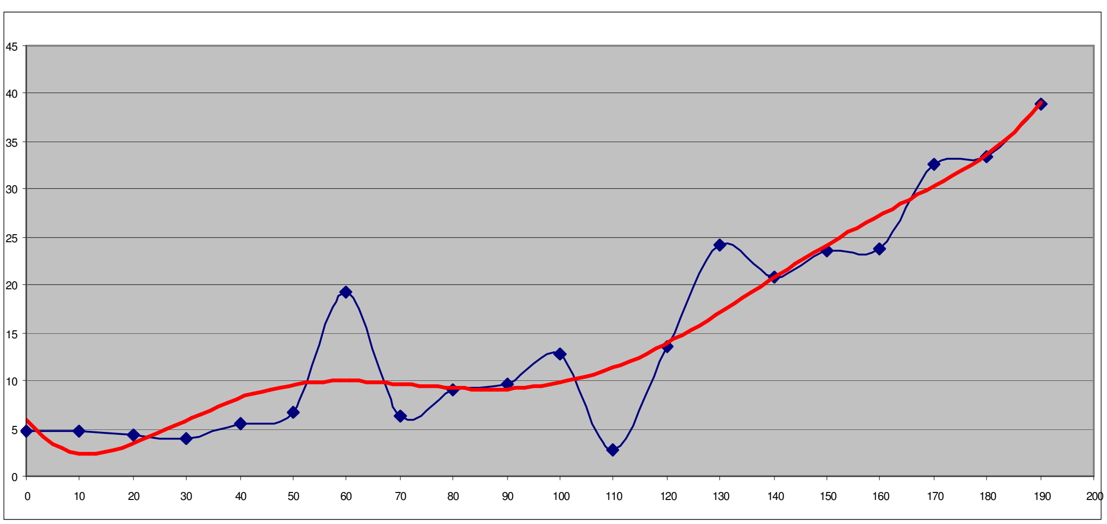
Bei einem Fahrzeugtyp werden folgenden Abhängigkeiten beobachtet:

Die Herstellungskosten für das Fahrzeug, in Abhängigkeit von der maximal erreichbaren Geschwindigkeit (siehe Tabelle links).

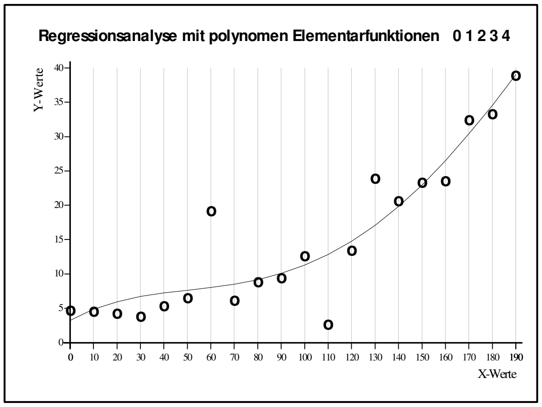
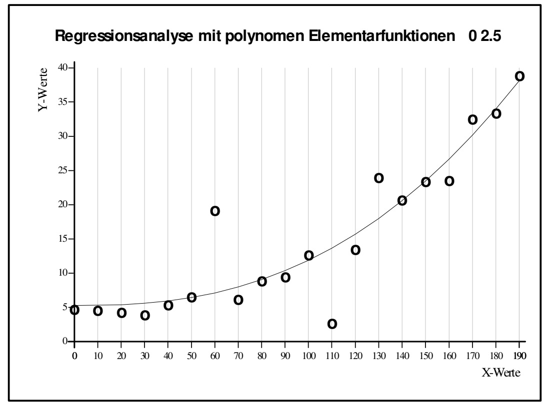
Bauen Sie ein Regressionsmodell, das es erlaubt, bei Eingabe der gewünschten Dienstgeschwindigkeit die Herstellungskosten für das Fahrzeug zu berechnen!

(Bei der „Punktwolke“ handelt es sich bewusst um „schlechte“ Daten, die eben nicht gut der tatsächlichen Tendenz entsprechen, um die Fehlerquellen aufzuzeigen! Auch ist die unabhängige Variable bewusst „gleichverteilt“ und die um eine deutlichere Grafik zu erreichen!

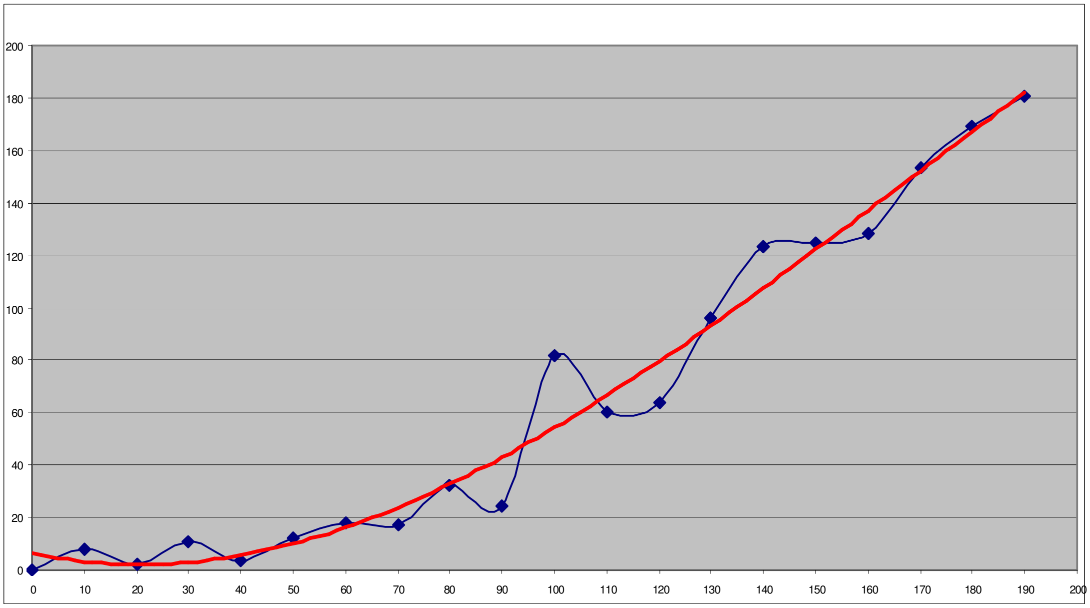
Im Excel, mit den dortigen Funktionen „modelliert“, wäre man geneigt, eine Funktion zu erstellen, die tatsächlich so aussieht:



Tatsächlich ist die Funktion natürlich falsch, wie man auch schon aus der Excel-Grafik deutlich sieht!

	
die Punkte werden zwar „besser“ angeglichen, das Modell ist aber gesamtheitlich falsch!	die Punkte werden weniger genau angeglichen, das Modell ist aber gesamtheitlich richtig! Durch die Möglichkeit beliebige Potenzen zu verwenden (auch reale Zahlen, negative Zahlen (hyperbolische Glieder), Potenzen unter 1 (Wurzelglieder) etc. ist die „natürliche Tendenz“ leichter zu erreichen!

Das oben gezeigte Beispiel ist extrem einfach und könnte mit den Möglichkeiten eines Tabellenkalkulationsprogrammes durchaus richtig gelöst werden, wenn man sich nur dessen bewusst ist, dass bei der Regressionsrechnung stets 2 Eingaben vorgenommen werden müssen, nämlich die Beobachtungen (die „Punktwolke“) und die natürliche Tendenz (also die Elementarfunktionen) und die Auswahl dieser besondere Sorgfalt erfordert. Das „Hineinwerfen“ eines Punkthaufens in ein Regressionsprogramm ohne die Elementarfunktionen selbst auszuwählen, kann nie richtige Ergebnisse zeitigen! Beim oben gezeigten Beispiel haben wir das Glück, dass die Funktion nicht durch den Punkt 0/0 geht, da natürlich die Herstellungskosten für ein Fahrzeug – auch ohne Motor – ungleich 0 sind. Wollten wir eine Funktion darstellen, die unbedingt durch 0 gehen muss, wie beispielsweise die – um eine bestimmte Geschwindigkeit zu erreichen - nötige Motorleistung, dann kommen wir mit den Möglichkeiten der Tabellenkalkulation nicht mehr zu Rande, weil in jedem Fall das Polynomglied $\mathbf{x^0}$ gewählt wird, das einen absoluten Abstand auf der y-Achse erlaubt, und man kann NICHT erwarten, dass der dazugehörige Koeffizient Null gesetzt wird, sondern es kann durchaus sein, dass die resultierende Funktion nahe x=0 zuerst negativ wird und dann hochkrümmt, wie die untenstehende Abbildung zeigt.



Was dem flüchtigen Betrachter hier vielleicht als „vernachlässigbarer" kleiner Fehler erscheint, ist in einem Gesamtmodell, das aus sehr vielen solcher Einzelmodelle bestehen kann, und das zur Optimierung, (beispielsweise eines Fahrzeuges auf einer bestimmten Route), dienen soll, absolut unbrauchbar, weil dann alle Tendenzen nicht stimmen, und es eben auf das Zusammenspiel dieser Tendenzen ankommt. Man erreicht eher ein richtiges Optimierungsergebnis, wenn alle Modelle die eingegebene Punktwolke zwar numerisch nicht konkret darstellen, aber alle Tendenzen richtig sind!

Transformationen

Wie schon oben deutlich gemacht, ist es absolut notwendig, die resultierenden Funktionen bei **x = 0** und **x = ∞** richtig „anzubinden". Dies kann schwierig sein, wenn die Funktion sich im Unendlichen einer Asymptote annähern soll (wie es beim früher erwähnten Beispiel mit der Werbung der Fall ist!)

Dafür gibt es die Möglichkeit den Punkthaufen „in einem anderen Raum" abzubilden und nach der Regression zurück zu transformieren.

Dabei kommt uns die Funktion des Tangens-Hyperbolicus entgegen, der bei y = 1 eine horizontale Asymptote hat. Damit kann man ein Polynom unter eine „Hüllkurve" drücken und sogar einen „S-Schlag" im Kurvenverlauf erreichen, der mit normalen Polynomen kaum möglich ist!

Im wesentlichen funktioniert die Transformation folgendermaßen:

- wenn der übliche Regressionsansatz lautet:
  $y = a.x^{0} + b.x^{1} + c.x^{2,34} + d.x^{3,78}+..$
- wir dieser umgewandelt zu:
  $\text{arth } y = a \cdot x^{0} + b \cdot x^{1} + c.\ x^{2,34} + d.\ x^{3,78}+..$
- dann wird die Regression ausgeführt (die Regressionskoeffizienten a, b, c, d .. werden bestimmt)
- die resultierende Funktion heißt dann:
  $y = \tanh (a \cdot x^{0} + b \cdot x^{1} + c.\ x^{2,34} + d.\ x^{3,78}+..)$
- die Funktion liegt somit zur zur Gänze innerhalb einer umhüllenden Funktion, die den Wert 1 nie überschreiten kann (es ist selbstverständlich möglich – durch Multiplikation mit einem bestimmten Wert – die absolute Höhe der Asymptote zu verändern und ihr auch eine beliebige Neigung zu geben.

Weitere Transformationen sind beispielsweise:

- die logarithmische Transformation, mit der man Funktionen beliebig „dehnen“ kann, um zu starke und schwer abbildbare Krümmungen zu vermeiden
- die Regression mit Splines (also die abschnittsweise Definition von Modellen, mit identischen Koordinaten und identischer Tangente und Krümmung beim Übergang)

Alle diese (und weitere Transformationen) sind beim eingangs erwähnten Interface zwischen Dyalog APL und Excel implementiert.

Mehrdimensionale Modelle

Die vorerwähnten Ausführungen, die sich auf 2-dimensionale Modelle beziehen, sollen dem prinzipiellen Verständnis dienen. Weit häufiger haben wir es aber mit 3- und 4-dimensionalen Modellen zu tun (während mehr als 4 Dimensionen kaum benötigt werden!)

Für solche Modelle gibt es die Methode der multiplen Regressionsanalyse, bei der jede einzelne Elementarfunktion von allen unabhängigen Variablen (derer gibt es ja jetzt mehrere) abhängen kann.

Der Nachteil dieser Methode ist jedoch, dass die geometrische Vorstellungsmöglichkeit meist verloren geht, man es mit sehr vielen Elementarfunktionen zu tun hat, was zu Rechenungenauigkeiten führt und die resultierenden Rechenmodelle unübersichtlich und schwer zu handhaben sind.

Eleganter ist eine mehr „geordneten Methode“, die vom Autor bereits 1974 vorgeschlagen wurde, und die ebenfalls in der erwähnten „APL-Excel-Brücke“ implementiert ist! Die Methode besteht im wesentlichen darin, dass man mehr als

2-dimensionale Modelle auf 2-dimensionale zurückführt, was die Übersichtlichkeit, Rechnengenauigkeit und praktische Anwendbarkeit ganz wesentlich verbessert.

Dies geschieht dermaßen, dass über die virtuell im Raum aufgespannte Fläche ein Netz mit konstanten x bzw. y-Werten gelegt wird. (Dazu müssen durch eine örtlich sehr begrenzte Interpolation die Punkte der Punktwolke etwas verschoben werden.) Sodann werden nach einer Richtung alle Netzlinien mit demselben Regressionsansatz regrediert, wobei man für jede dieser Netzlinien natürlich andere Regressionskoeffizienten a, b, c … erhält. Diese werden sodann in die andere Richtung regrediert.

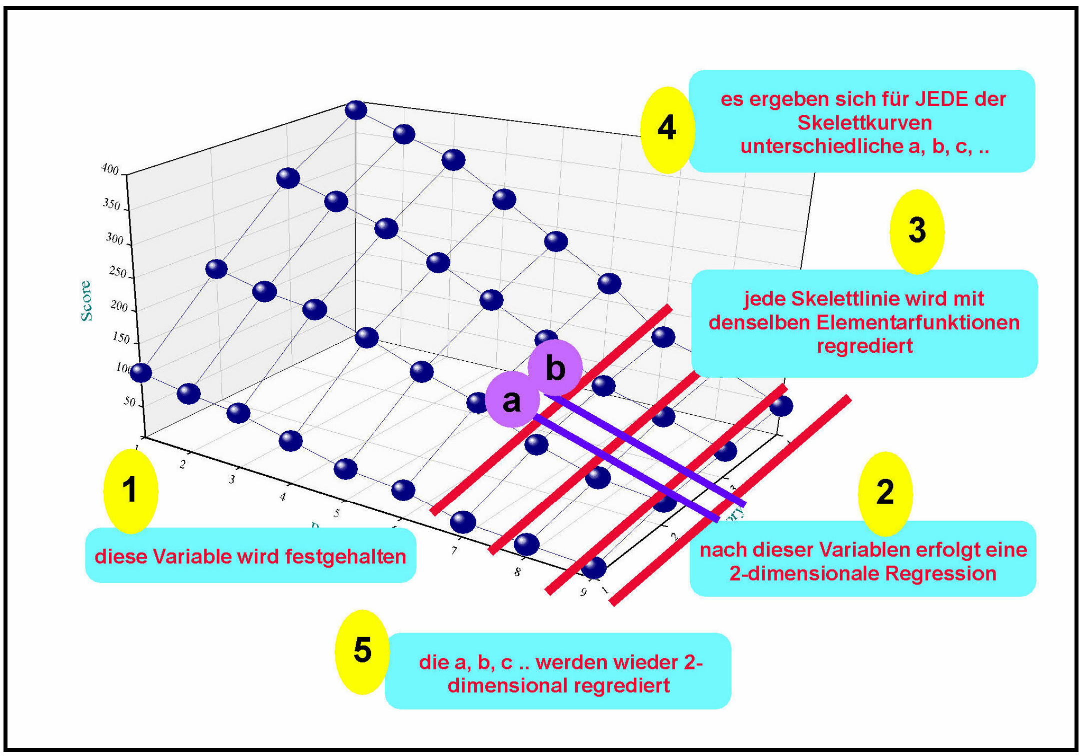
Während der Ansatz für eine 2.dimensionale Regression so lautet:

$y= a.f1(x) + b.f2(x) + ...$ wobei a, b, c etc. Konstante sind

lautet der für eine 3-dimensionale:

$z= a.f1(x) + b.f2(x) + ...$ wobei a, b, c etc. jeweils Funktionen von y sind

Das folgende Diagramm verdeutlicht nochmals die Vorgehensweise:



Zusammenfassung

Beim Einsatz der Methoden der Regressionsrechung für die Erstellung von virtuellen Modellen, steht die Erreichung einer „natürlichen Tendenz“, den der abzubildende Zusammenhang tatsächlich hat, vor dem Wunsch die eingegebene Punktwolke auch tatsächlich möglichst genau zu erreichen. Daher benötigen wir ein Werkzeug, mit dem es möglich ist, in sehr flexibler Art die Geometrie von Funktionen zu kontrollieren und an ihr zu experimentieren.

Tabellenkalkulationsprogramme bieten dafür zu wenig Flexibilität und liefern daher bei komplexeren Modellen falsche Ergebnisse. Spezielle Statistikprogramme, die sich dabei naturgemäß besser verhalten, sind oftmals zu komplex zu bedienen, liefen zuviel „unnötiges Beiwerk“ und sind ebenfalls im allgemeinen zu wenig flexibel. Es bietet sich daher eine Formulierung des Lösungsweges in einer „Definitionsumgebung“ (= Dyalog APL) an, die mit MS-Excel in der Weise kommuniziert, dass der Benutzer die Vorteile der bekannten Tabellenkalkulation nutzen können und dennoch die für die richtige Modellerstellung nötige Flexibilität nicht verliert.

Mit den vorgeschlagenen Methoden kann man in einfacher Weise virtuelle Modelle auf Basis der Regressionsrechnung erstellen und zwar – unter Beibehaltung des Überblickes über die Geometrie – auch für mehr für als 2 Dimensionen. Die meisten Modelle in der Logistik sind 3 oder 4-dimensional. Bis zu 4 Dimensionen kann man das resultierende Modell auch auf einem Bildschirm optisch deutlich darstellen! (auf dem Papier leider nur 3, denn die vierte Dimension ist die Bewegung in Abhängigkeit von der Zeit).

Für Leser, die sich mit „Modelling“ noch nicht beschäftigt haben, ist der vorliegende Beitrag sicher zu komplex, für solche, die mit diesen Methoden bereits gearbeitet haben, wieder zu rudimentär. In beiden Fällen steht jedoch der Autor für Rückfragen gerne zur Verfügung! (hfn@chello.at)

Literatur:

1	Dyalog	Introduction to Dyalog APL	http://www.dyadic.com/
2	Causeway	Causeway Graphical Systems RainPro	http://www.causeway.co.uk/products/index.htm
3	H.-F. NELSON	Procedure for developing economically optimum ships for given traffic relations	Sondernummer, Oktober 1973, der "Maierform News", Kundenzeitschrift der Maierform S.A. Genf
4	H.-F. NELSON	Wirtschaftlichkeitsvergleiche zwischen Bargecarriern und Vollcontainerschiffen (eine Systemanalyse) Dissertation 1974	Dissertation 1974, Technische Hochschule Wien
5	H.-F. NELSON	Hat der Bargecarrier Zukunft?	SGKV, Berichte der Studiengesellschaft für den kombinierten Verkehr 1974
6	H.-F. NELSON	Schiffahrt hat Ölpreisprobleme - optimale Geschwindigkeit verändert!	Internationale Wirtschaft (Mitteilungen der Bundeswirtschaftskammer) Juni 1975
7	H.-F. NELSON	A revolution in ship design; The Maierform Fairing Program "Totalfair"	Sondernummer, Februar 1972, der "Maierform News", Kundenzeitschrift der Maierform S.A. Genf
8	H.-F. NELSON	EDV-gesteuerte Verkehrsmitteleinsatzplanung	Vorlesung an der Wirtschaftsuniversität Wien (1974-2001)
9	H.-F. NELSON	Simulations - und Optimierungstechniken im Bereich Transport - Verkehr - Logistik.	Übungen an der Wirtschaftsuniversität Wien (ab 2002)
10	H.-F. NELSON	Statistik, Markt- und Trendanalysen	Vorlesung an der Fachhochschule des bfi in Wien (ab 2002)
11	H.-F. NELSON	Prozessmodellierung, Simulation und Optimierung	Vorlesung an der Donau-Universität Krems (2002)