

Die Karlsruher Kapitalmarktdatenbank – Bilanz und Ausblick –

Ralf Herrmann

Diskussionspapier Nr. 189

Stand: Mai 1996

Zusammenfassung

Schon seit fast 20 Jahren dient die KKMDB empirischer Kapitalmarktforschung in Deutschland als Datengrundlage. Nach einer kurzen Darstellung der Entstehungsgeschichte wird auf Inhalt und Umfang der Datenbank eingegangen. Es wird weiter beispielhaft aufgezeigt, wo die Daten Verwendung finden und welchen Beitrag die KKMDB zur empirischen Kapitalmarktforschung in Deutschland bereits geleistet hat. Die Arbeitsschritte und die Ausstattung die für die Aufbereitung der Rohdaten notwendig sind, werden im Anschluß daran dargestellt. Ein Ausblick auf die zukünftige Entwicklung und anstehende Veränderungen bildet den Abschluß der Arbeit.

Rückfragen und Kommentare an folgende Adresse:

Institut für Entscheidungstheorie und Unternehmensforschung
Universität Karlsruhe (TH)
Postfach 69 80
76 128 Karlsruhe
Tel. 0721/608-6033
Fax 0721/35 92 00

Inhaltsverzeichnis

1	Geschichte der KKMDB	2
2	Datenbankinhalt	3
2.1	Objekte	3
2.2	Attribute	3
2.3	Datenvolumen	9
3	Verwendung	10
3.1	DAFOX	10
3.2	Forschung	10
4	Arbeitsablauf	11
4.1	Tägliche Arbeiten	11
4.2	Monatliche Arbeiten	12
4.3	Umstellungsarbeiten	14
5	Notwendige Ausstattung	15
5.1	Personal	15
5.2	Hardware	15
5.3	Software	16
6	Ausblick	17

1 Geschichte der KKMDB

Die Wurzeln der KKMDB reichen zurück bis in die siebziger Jahre, in denen in Deutschland mit dem organisierten Aufbau sowie der regelmäßigen Pflege von Daten für empirische Kapitalmarktforschung begonnen wurde. Bis 1988 wurden sämtliche Daten manuell aus Tageszeitungen wie der Börsenzeitung, Handelsblatt, FAZ oder Hoppenstedt Monats-Kurstabellen erfaßt. Dieser Zustand änderte sich grundlegend als 1988 von der Deutschen Forschungsgemeinschaft (DFG) der Forschungsschwerpunkt "Empirische Kapitalmarktforschung" genehmigt wurde. Mit dem Aufbau von drei Datenbanken sollte eine adäquate Datenbasis für umfangreiche empirische Forschungsprojekte¹ geschaffen werden.

Basierend auf den bereits bestehenden Datenbeständen wurde durch die Förderung der DFG der Aufbau von drei dezentralen Datenbanken ermöglicht, die zusammen die sogenannte Deutsche Finanzdatenbank (DFDB) bilden. Die Standorte dieser drei Teildatenbanken sind die Universitäten Aachen (Bilanzdaten) bei Prof. Dr. Hans Peter Möller, Karlsruhe (Aktien-, Options- und Optionsscheindaten) bei Prof. Dr. Hermann Göppl und Mannheim (Rentendaten) bei Prof. Dr. Wolfgang Bühler. Über diesen von der DFG geförderten Teil² hinaus wurden aus Eigenmitteln des Instituts für Entscheidungstheorie und Unternehmensforschung der Universität Karlsruhe zusätzliche Datenbestände aufgebaut. Diese Institutsdatenbestände und die Karlsruher Teildatenbank der DFDB bilden zusammen die Karlsruher Kapitalmarktdatenbank³. Seit 1988 stammen alle Daten der KKMDB aus offiziellen Quellen. Datenlieferanten der bereits elektronisch erfaßten Daten sind die Deutsche Wertpapierdatenzentrale (DWZ), Wertpapiermitteilungen (WM) und die Deutsche Terminbörse (DTB). Die Datenbestände zeichnen sich durch Vollständigkeit und hohe Qualität aus.

Mit dem 31. Januar 1995 ist die Förderung der DFDB durch die DFG ausgelaufen. Die bisherige Weiterführung war durch eine großzügige Spende der Deutschen Börse AG möglich.

¹Vgl. beispielsweise Bühler/Hax/Schmidt[1993].

²Die DFDB-Daten umfassen nur Kurs- und Umsatzdaten inländischer Emittenten, die zum börsenmäßigen Handel an der Frankfurter Wertpapierbörse zugelassen sind.

³Der Begriff Datenbank ist insofern irreführend als nur ein relativ kleiner Teil der Daten (ca. 3 GB) tatsächlich in Form einer relationalen Datenbank verwaltet werden. Die übrigen Daten liegen bedingt durch das enorme Datenvolumen nur in SAS-Datasets bzw. sequentiellen Dateien vor.

2 Datenbankinhalt

Der Datenbankinhalt kann nach den betrachteten Märkten, den Wertpapieren und den verschiedenen gespeicherten Informationsarten eingeteilt werden.

2.1 Objekte

Die KKMDB enthält Informationen von allen deutschen *Wertpapierbörsen*. Neben den acht Präsenzbörsen Berlin, Bremen, Düsseldorf, Frankfurt, Hamburg, Hannover, München und Stuttgart, werden Daten aus den elektronischen Handelssystemen IBIS und IBIS-R sowie der Deutschen Terminbörse erfaßt.

Die Informationen betreffen alle an deutschen Börsen gehandelten *Wertpapiere*⁴. Dies umfaßt Informationen der Präsenzbörsen zu Aktien, Rentenzertifikaten, Optionsscheinen, Optionen, Genußscheinen, Bezugsrechten usw. sowie Daten von der elektronischen Terminbörse zu den verschiedenen derivativen Wertpapieren wie Futures und Optionen. Darüberhinaus werden Indizes, Währungsdaten und Rücknahmepreise von Investmentfonds verwaltet.

2.2 Attribute

Inhaltlich läßt sich der Datenbestand der KKMDB in

- Kursinformationen⁵
- Umsatzdaten⁶,
- Termindaten⁷,
- Stammdaten⁸ und
- sonstige Informationen

unterteilen.

⁴Strenggenommen besitzen Optionen rechtlich keinen Wertpapiercharakter sondern sind lediglich Rechte. Aus Vereinfachungsgründen seien sie hier unter dem Begriff Wertpapier miterfaßt.

⁵Vgl. Göppl/Herrmann/Lüdecke[1995] und Göppl/Herrmann/Lüdecke[1993a].

⁶Vgl. Göppl/Herrmann/Lüdecke[1995].

⁷Vgl. Göppl/Herrmann/Lüdecke[1993c].

⁸Vgl. Göppl/Herrmann/Lüdecke[1993b].

Die *Kursdaten*⁹ umfassen Kassakurse, variable Notierungen und zeitgestempelte Intraday-Daten. Neben den reinen Kursdaten sind auch Zusatzinformationen, wie Börse und Marktsegment¹⁰ an denen der Kurs zustande kam, und der Kurszusatz verfügbar.

Umsätze sind in Form des täglichen stück- und wertmäßigen Umsatzes, teilweise auch in Form von Transaktionsdaten, vorhanden. Darüberhinaus ist die Anzahl der täglichen Geschäfte je Wertpapiernummer abfragbar.

Die *Stammdatensätze* umfassen die Ausstattungsmerkmale der verschiedenen Wertpapiere¹¹. Für Aktien sind dies beispielsweise die Wertpapierbezeichnung, die Wertpapierart, die Branche, der Emissionskurs, das Emissionsvolumen, die Höhe des Grundkapitals. Kontraktsspezifikationen von Optionscheinen wie die Laufzeit, das Basispapier, der Emissionsbetrag, der Emittent und ähnliches werden ebenso erfaßt wie die Höhe der Zinskupons, die Fälligkeitstermine, das Emissionsvolumen, die Nennwerte, die Kupon- und Zinsanpassungstermine von Rentenpapieren des Bundes.

Im Gegensatz zu Stammdatensätzen, die grundlegende Eigenschaften des Wertpapiers darstellen, sind *Terminaten* ereignisbezogene Informationen. Besondere Bedeutung haben sie für die Berechnung von sogenannten Bereinigungs-faktoren¹² mit denen man Kurszeitreihen um technisch bedingte Kursveränderungen bereinigt. Hauptsächlich handelt es sich um Informationen zu Dividendenzahlungen, Bezugsrechten und sonstigen kapitalverändernden Maßnahmen. Hierzu wird der Extag, die Art des Terminereignisses, das Bezugsverhältnis, der Bezugsrechtswert und die resultierenden Bereinigungs-faktoren in der Datenbank geführt.

Sonstige Informationen umfassen u.a. die Anzahl der vorhandenen Kurse, Handelszeiträume für spezifische Wertpapiere, die Zusammensetzung von Indizes, Anzahl der ausstehenden Kontrakte ("open interest"), Settlementpreise, Quotes, Verfalltermine der DTB-Produkte und einen Börsentagkalender für jede Börse.

Die vorhandenen Kurszeitreihen in der KKMDB reichen bis 1960 zurück. Im Zeitraum von 1960 bis 1974 liegen für 100 Unternehmen tägliche Kassakurse vor, für die restlichen an der Frankfurter Börse gehandelten Unternehmen Monatsultimokurse. Ab 1974 sind die täglichen Kassakurse aller an

⁹Es wird nicht zwischen Preisen und Kursen unterschieden, da der Unterschied akademischer Natur ist und hier keine Relevanz besitzt.

¹⁰Man unterscheidet seit dem 1. Mai 1985 den amtlichen Markt, den geregelten Markt und den Freiverkehr. Vor diesem Termin existierte zwischen dem 4. Mai 1987 und dem 30. April 1988 ein geregelter und ein ungeregelter Freiverkehr, somit also vier Marktsegmente. Vor dem 4. Mai 1987 gab es keinen geregelten Markt. Innerhalb des amtlichen Handels wird zudem zwischen variabel und nicht variabel notierten Wertpapieren unterschieden.

¹¹Jeder Wertpapiernummer ist somit genau ein Stammdatensatz zuordenbar.

¹²Das genaue Verfahren wird in Sauer[1991] beschrieben.

deutschen Wertpapierbörsen gehandelten Wertpapiere vorhanden. Ab dem gleichen Zeitpunkt wurden Umsatzdaten der Frankfurter Wertpapierbörse in die Datenbank integriert. Die Umsatzdaten der Regionalbörsen folgten sukzessiv. Seit 1994 stehen mit Aufnahme der Umsatzdaten der Bremer Börse die Umsatzdaten aller Präsenzbörsen zur Verfügung. Neben den Aktien- und Rentenkursen liegen Optionspreise des Frankfurter Optionsmarktes seit 1983 vor. Mit Einführung der DTB wurden auch Informationen von der elektronischen Terminbörse aufgenommen. Der DTB-Datenbestand umfaßt sekundliche Preis- und Umsatzinformationen sowie Informationen über die offenen Kontrakte. Ab 1989 liegen zeitgestempelte Intraday-Kurse der Dax-Werte vor. Fortlaufende Notierungen¹³ der restlichen variabel notierten Wertpapiere reichen zurück bis 1989. Transaktionsdaten aus IBIS wurden 1991, Informationen aus IBIS-R 1995 in die Datenbank aufgenommen. Für alle Wertpapiere liegen für den gesamten Zeitraum sämtliche Termindaten und korrespondierende Stammdaten vor.

Eine genaue Aufstellung der historischen Datenbestände, nach verschiedenen Marktsegmenten geordnet, findet sich in den nachfolgenden Tabellen.

¹³Allerdings ohne Zeitstempel.

SEGMENT	ZEITRAUM	VERFÜGBARE DATEN
Aktien	ab 1960	Tägliche Kassakurse für 100 Gesellschaften.
		Monatsultimokurse für die restlichen in Frankfurt notierten Gesellschaften
		Korrespondierende Termin- und Stammdaten.
	ab 1974	Tägliche Kassakurse aller an deutschen Wertpapierbörsen gehandelten Aktien deutscher Unternehmen
		Umsatzdaten für die Frankfurter Wertpapierbörse
		Zugehörige Termin- und Stammdaten.
	ab 01.11.88	Zeitgestempelte Intraday-Kurse der DAX-Werte
	ab 01.11.89	Fortlaufende Kurse aller deutschen Aktien, die zum variablen Handel zugelassen sind.
	ab 01.07.91	Transaktionsdaten aus IBIS
	ab 17.04.90	Umsatzdaten für die Regionalbörsen (außer Bremen u. Hamburg).
	ab 01.04.93	Umsatzdaten für die Hamburger Börse
ab 16.03.94	Umsatzdaten für die Bremer Börse	
Renten	ab 1974	Tägliche Kassakursnotierungen für Anleihen von Bund, Bahn, Post, FDE und Treuhand
		Umsatzdaten für die Frankfurter Wertpapierbörse
		Zugehörige Stammdaten.
	ab 01.07.91	Transaktionspreise aus IBIS
	ab 17.04.90	Umsatzdaten für die Regionalbörsen (außer Bremen u. Hamburg).
	ab 01.04.93	Umsatzdaten für die Hamburger Börse
	ab 16.03.94	Umsatzdaten für die Bremer Börse
	ab 05.12.95	Transaktionsdaten aus IBIS-R

Tabelle 1: Datenbestand - KASSAMÄRKTE I -

SEGMENT	ZEITRAUM	VERFÜGBARE DATEN
Optionsscheine	ab 1974	Tägliche Kassakurse aller an der Frankfurter Wertpapierbörse gehandelten Optionsscheine auf deutsche Aktien zzgl. Stammdaten
		Umsatzdaten für die Frankfurter Wertpapierbörse.
Investment	ab 1974	Tägliche Rücknahmepreise aller deutschen Investmentfonds.
Sonstiges	ab 1960	diverse Indizes für den deutschen Aktienmarkt.
	ab 1967	diverse Indizes für den deutschen Rentenmarkt.
	ab 1982	diverse internationale Aktienindizes.
	ab 1974	diverse Devisenkassakursnotierungen der wichtigsten Währungen.

Tabelle 2: Datenbestand - KASSAMÄRKTE II -

SEGMENT	ZEITRAUM	VERFÜGBARE DATEN
FWB Optionen	ab 01.04.83	Tägliche Kurse der an der Frankfurter Wertpapierbörse gehandelten Optionen auf deutsche Aktien.
	ab 01.11.89	Tägliche Kurse aller an deutschen Wertpapierbörsen gehandelten Optionen.
DTB Optionen	ab 26.01.90	Sekündliche Preise und Umsätze aller Aktien-Optionen.
		Settlement-Preise und Open Interest.
	ab 16.08.91	Sekündliche Preise und Umsätze der Optionen auf den DAX und den BUND-Future.
		Settlement-Preise und Open Interest.
	ab 24.01.92	Sekündliche Preise und Umsätze der Optionen auf den DAX-Future.
		Settlement-Preise und Open Interest.
ab 15.01.93	Sekündliche Preise und Umsätze der Optionen auf den BOBL-Future.	
	Settlement-Preise und Open Interest.	
DTB Futures	ab 23.11.90	Sekündliche Preise und Umsätze der DAX- und BUND-Futures.
		Settlement-Preise und Open Interest.
	ab 04.10.91	Sekündliche Preise und Umsätze aller BOBL-Futures.
		Settlement-Preise und Open Interest.
	ab 11.03.94	Sekündliche Preise und Umsätze aller BUXL-Futures.
		Settlement-Preise und Open Interest.
	ab 18.03.94	Sekündliche Preise und Umsätze aller FIBOR-Futures.
		Settlement-Preise und Open Interest.

Tabelle 3: Datenbestand - TERMINMÄRKTE -

2.3 Datenvolumen

Insgesamt sind in der ORACLE-Datenbank¹⁴ Informationen zu annähernd 13.000 verschiedenen Wertpapiernummern vorhanden. Insgesamt enthält sie ca. 22,5 Mio. Kassakurse. Auf die einzelnen Präsenzbörsenplätze verteilt ergibt sich folgende Tabelle¹⁵:

BÖRSE	ANZ. WP.	KASSAKURSE
Berlin	4.902	2.715.294
Bremen	414	638.742
Düsseldorf	7857	3.577.271
Frankfurt	11.905	7.558.433
Hamburg	4.027	2.639.068
Hannover	582	940.047
München	4.781	2.565.738
Stuttgart	4.741	1.815.546
	39.209	22.450.139

Daneben enthält die relationale Datenbank approximativ 6,5 Mio. variable Kursnotierungen, 2 Mio. Optionskurse, 13.000 Stammdatensätze, und 51.000 Termindatensätze. Neben dem in der ORACLE-Datenbank erfaßten Datenbestand existiert der größte Teil der Daten in Form von SAS-Datasets bzw. sequentiellen Dateien. Der gesamte *Datenbestand* umfaßt insgesamt ca. 300 GB Daten in annähernd 4.500 Dateien. Davon sind 30% DWZ-Daten mit Kurs- und Umsatzzahlen der acht deutschen Präsenzbörsen und der IBIS Handelssysteme, 8% Stamm- und Termindaten von WM und 60% DTB-Daten¹⁶. Die übrigen archivierten Daten beinhalten Sicherungen der Verarbeitungs- und Schnittstellenprogramme, verschiedene Datenabfragen und ähnliches. Insbesondere seien hier die manuell erfaßten Datenbestände erwähnt.

Momentan beträgt der monatliche *Zuwachs* des Datenbestands der KKMDB in etwa 160 MB DWZ-Daten, 40 MB WM-Daten und 2,5 GB DTB-Daten. Dies entspricht monatlich ungefähr 2 Mio. neuen Kurs- und Umsatzinformationen von Kassamärkten, 180.000 zusätzlichen Datensätzen von WM mit Updates auf ca. 1.200 verschiedene Felder und 25 Mio. neuen DTB-Datensätzen.

¹⁴Die ORACLE-Datenbank umfaßt nur einen kleinen Teil des gesamten Datenbestands der KKMDB.

¹⁵Aus der Notierung des gleichen Wertpapiers an verschiedenen Börsen resultiert eine Mehrfachzählung von Wertpapieren in der Tabelle.

¹⁶Der volumenmäßige Vergleich der Präsenzbörsendaten mit den Daten der DTB wird bedingt durch die höhere Anzahl von Verarbeitungsstufen bei DWZ-Daten, die nach jedem Verarbeitungsschritt gesichert werden, zugunsten der DWZ-Daten verzerrt.

3 Verwendung

3.1 DAFOX

Im Jahr 1992 wurde die Reihe der deutschen Aktienindizes in der Datenbank um eine eigene, eigens für Forschungszwecke konzipierten Index-Familie erweitert. Der erste entwickelte Index der sogenannte Deutsche Aktien Forschungsindex (DAFOX) umfaßt den gesamten amtlichen Markt und ist von der Berechnungsweise mit dem CDAX vergleichbar¹⁷. Zur Zeit werden 21 verschiedene Indizes, die sich durch das betrachtete Marktsegment sowie die Berechnungsweise unterscheiden, regelmäßig aktualisiert. Es existieren Gesamtmarkt-, Teilmarkt- und Branchenindizes in Form von gleichgewichteten und marktkapitalisierungsgewichteten Performance-Indizes. Daneben existieren auch Kursindizes. Im Gegensatz zu anderen Aktienindizes¹⁸ die historisch künstlich verknüpft wurden und dadurch keine konsistent berechnete Zeitreihe aufweisen, konnten die DAFOX-Indizes Dank der vorhandenen Datengrundlage konsistent zurückgerechnet werden.

Der DAFOX¹⁹ findet deshalb in der Zwischenzeit in der empirischen Kapitalmarktforschung des akademischen wie auch des institutionellen Bereichs auf nationaler und internationaler Ebene breite Anwendung.

3.2 Forschung

Empirische Kapitalmarktforschung dient der Überprüfung von theoretisch abgeleiteten Zusammenhängen sowie dem Aufspüren empirischer Phänome²⁰. Dabei wird versucht deduktive Erkenntnisse anhand empirischer Daten zu validieren und bestehende Abweichungen zu analysieren. Diese Vorgehensweise führt in einem dynamischen Prozeß zu ständig modifizierten und neuen Modellansätzen. Im Mittelpunkt des Interesses stehen Bewertungszusammenhänge und damit eng verknüpft die Informationsverarbeitung auf den Kapitalmärkten.

Am Institut für Entscheidungstheorie und Unternehmensforschung der Universität Karlsruhe hat empirische Kapitalmarktforschung eine lange Tradition, die fast 20 Jahre zurückreicht.

¹⁷Er existierte aber bereits vor Einführung dieses Index durch die Deutsche Börse AG.

¹⁸Beispielsweise dem DAX oder CDAX.

¹⁹Vgl. auch <http://www-etu.wiwi.uni-karlsruhe.de/kapital/projekte/dafox.html> und Göppl/Schütz[1992].

²⁰Bespiele für solche Phänome sind Saisonalitäten (vgl. Frantzmänn[1989]), der Intervall-Effekt (vgl. Frantzmänn[1990], Schlag[1994]) oder der Overreaction-Effekt (vgl. Meyer[1994]).

Ein Großteil der am Institut erstellten *Dissertationen* hat empirische Fragestellungen zum Inhalt. So wurden umfangreiche Tests von Bewertungsmodellen durchgeführt sowie Bewertungsanomalien, Saisonalitäten und die Qualität von Kapitalmärkten untersucht²¹. Aktuelle empirische Forschungsprojekte umfassen die Bewertung von Zinsderivaten, Interdependenzen zwischen den verschiedenen Märkten und Handelssystemen, die Analyse der Kapitalstruktur von Unternehmen sowie die Effizienz und die Performance von Handelsstrategien. Neben den Dissertationen wurden eine große Zahl von *Diplomarbeiten* mit empirischem Inhalt erstellt²². Hinzu kommen eine Vielzahl von Veröffentlichungen in diversen Fachzeitschriften und Fachbüchern²³.

Die Datenbank ermöglicht aber nicht nur institutseigene Forschungsvorhaben sondern die Daten werden auch anderen externen akademischen Forschungseinrichtungen zu Verfügung gestellt. Auf diese Weise wurden in den vergangenen Jahren hunderte externer Forschungsvorhaben an mehr als dreißig Lehrstühlen ermöglicht, die ohne die Daten der KKMDB überhaupt nicht durchführbar gewesen wären²⁴.

4 Arbeitsablauf

Von den Rohdaten bis zu den für Forschungszwecke aufbereiteten Kapitalmarktdaten sind einige Arbeitsschritte notwendig, die sich nach der Regelmäßigkeit der Durchführung einteilen lassen.

4.1 Tägliche Arbeiten

Die täglich anfallenden Arbeiten gliedern sich in

- Einlesen der Daten,
- Aufbereitung der Rohdaten,
- Bearbeitung von Datenanfragen und
- sonstige Tätigkeiten.

²¹Vgl. exemplarisch Winkelmann[1984], Frantzmann[1989], Müller[1992], Sauer[1994] für den Aktienmarkt, Margana[1986], Bußmann[1988], Frantzmann[1989], Lassak[1993] für den Rentenmarkt und Trautmann[1986] für den Optionsmarkt.

²²Vgl. beispielsweise Sauer[1989], Hanel[1991] oder Meyer[1994].

²³Beispielsweise Göppl et al.[1996].

²⁴Exemplarisch Bauer[1992], Entrup[1995], Hecker[1993], Iversen[1994], Röder[1994], Uhrig[1996] und Walter[1996].

Nach dem Entpacken der angelieferten Rohdaten, werden die Daten eingelesen und aufbereitet²⁵. Dabei wird global die Vollständigkeit der Informationen für jeden Börsentag geprüft.

Zusätzlich zur Datenaufbereitung fallen täglich Datenanfragen von externen Forschungseinrichtungen an. Im letzten Jahr wurden von mehreren hundert Datenanfragen ca. 80 beantwortet²⁶. Die angefragten Datenmengen übersteigen dabei immer häufiger die 100 MB Grenze, bei DTB-Daten bewegen sich die Datenanfrage im GB-Bereich. Diese Datenmengen müssen aus den vorhandenen Datenbeständen extrahiert²⁷, gepackt und auf Datenträger kopiert oder via ftp²⁸ übertragen werden.

Sonstige Tätigkeiten umfassen administrative Aufgaben wie die Vergabe von Benutzernummern, Zugriffsrechten, Erstellung und Pflege von World-Wide-Web (WWW) Seiten²⁹ sowie Datenbeschreibungen oder ähnliches.

4.2 Monatliche Arbeiten

Basierend auf den täglich aufbereiteten Dateien des Vormonats werden zu Beginn eines jeden Monats von der DWZ mitausgelieferte Korrektursätze in die Rohdaten eingearbeitet und sequentielle Dateien erstellt. Aus diesen werden die Daten selektiert, die in die ORACLE Datenbank aufgenommen werden und Eingabedateien erstellt.

Während dieser Verarbeitung erfolgt mehrfach eine Überprüfung auf Korrektheit und Vollständigkeit der Daten. Im Gegensatz zur täglichen Verarbeitung, bei der das Augenmerk auf der globalen Korrektheit liegt, wird bei der monatlichen Überprüfung auf Wertpapiererebene die Vollständigkeit und Korrektheit der Zeitreihen überprüft. Fehlende Informationen und fehlerhafte Daten werden manuell nacherfaßt bzw. korrigiert. Der Ablauf der Tätigkeiten läßt sich vereinfacht wie folgt strukturieren.

- Erstellung von sequentiellen Monatsdateien,

²⁵Hierzu müssen Dateiformate, Dateistrukturen und Variableninhalte konvertiert werden.

²⁶Datenanfragen können aus den verschiedensten Gründen, z.B. aus Ermangelung der angefragten Daten oder aufgrund mangelnder Form (Datenanfragen werden beispielsweise nur von Hochschullehrern bearbeitet) nicht bearbeitet werden.

²⁷Abhängig von den angefragten Informationen aus der ORACLE-Datenbank abgefragt, oder aus SAS-Datasets bzw. sequentiellen Dateien selektiert werden.

²⁸File Transfer Protocol – ein Internetdienst, der zur Übertragung von Dateien dient.

²⁹WWW – ebenfalls ein Internetdienst, der auf multimedialer Ebene neben vielen anderen Möglichkeiten auch zur Darstellung und Verbreitung von Informationen verwendet werden kann. Die WWW-Seite der KKMDB hat die URL <http://www-etu.wiwi.uni-karlsruhe.de/kapital/projekte/kkmdb.html>.

- Selektion der Daten und Erstellung von Eingabedateien für die Datenbank,
- Laden der Kurs- und Umsatzdaten in ORACLE,
- Verarbeitung der WM-Daten,
- Berechnung von Bereinigungsfaktoren.
- Verarbeitung der DTB-Daten und
- Archivierung der Daten.

Nachdem die neuen Kurs- und Umsatzzeitreihen für den vergangenen Monat in die Datenbank aufgenommen sind, werden Stamm- und Termindaten von WM verarbeitet und relevante Daten in ORACLE-Tabellen geladen. Basierend auf den Termindaten von WM und den Kurszeitreihen der DWZ werden dann Bereinigungsfaktoren berechnet und ebenfalls in die Datenbank geladen. Diese bilden u.a. die Grundlage für die Berechnung der DAFOX-Indizes zu Beginn eines jeden Quartalsmonats.

Unabhängig davon erfolgt die DTB-Daten Verarbeitung. Dabei werden die von der DTB gelieferten sequentiellen Protokolldateien zerlegt und die für jeden Kontrakt selektierten relevanten Informationen in SAS-Datasets abgespeichert. Diese Datasets, die nach den verschiedenen Kontrakten getrennt alle Informationen enthalten, sind die Ausgangsbasis für die Erstellung von weiteren SAS-Datasets, die beispielsweise nur Transaktionsdaten enthalten. Diese Vorgehensweise ist notwendig, um die enormen Datenmengen für Datenanfragen bzw. eigene Forschungsprojekte überhaupt noch effizient handhaben zu können.

Nach erfolgreich abgeschlossener Verarbeitung werden alle Daten gesichert. Dies erfolgt aus Sicherheitsgründen auf den verschiedenen Verarbeitungsstufen auf verschiedenen Speichermedien redundant. Alle Dateien werden im ASM2-Archiv³⁰ des Rechenzentrums der Universität Karlsruhe archiviert. Zunächst wurden zusätzlich Sicherungskopien auf Magnetbänder kopiert. Dieses Vorgehen hat sich aber im Zeitablauf aufgrund des stark angestiegenen Datenvolumens als nicht mehr praktikabel erwiesen. Stattdessen ist geplant nach erfolgter Umstellung der KKMDB die Rohdaten und nachfolgende Verarbeitungsschritte auf CD-Rom abzuspeichern. Die Historie wird sukzessiv auf diese Weise ebenfalls gesichert, dabei wird und wurde insbesondere immer darauf geachtet, daß die gleichen Daten auf den verschiedenen Verarbeitungsstufen nach Möglichkeit³¹ auf verschiedenen Datenträgern gesichert

³⁰Zukünftig im Nachfolgearchivsystem ADSM.

³¹Die Vergabe der Kassetten, auf die eine Datei gespeichert wird, erfolgt im Archivsystem des Rechenzentrums automatisch. Man hat also nur indirekte Möglichkeiten der Einflußnahme, beispielsweise durch die zeitliche Streuung des Archivierungsvorgangs.

wurden. Dieser Vorgang erfolgt auf jeder Verarbeitungsstufe mehrfach. Neben den Rohdaten werden so die aufbereiteten Monatsdateien und die ORACLE Datenbank gesichert.

Zur reinen Datenverarbeitung kommen permanente Anpassungen der Verarbeitungsprogramme und Schnittstellen an neue Gegebenheiten hinzu. Solche notwendigen Veränderungen resultieren beispielsweise aus der Einführung neuer Variablen oder neuer Satzarten bei WM, Einführung neuer Handelssysteme oder neuer Kontrakte an den Deutschen Börsen. Desweiteren verändern sich im Zeitablauf Dateiattribute, wie Dateinamen oder Dateiformate³² sowie die Datenträger auf denen die Rohdaten geliefert werden.

4.3 Umstellungsarbeiten

Zusätzlich zu den normalen Arbeiten bedingt die Abschaffung der IBM3090 des Rechenzentrums der Universität Karlsruhe umfangreiche Umstellungsarbeiten. Vereinfacht läßt sich dies in folgende Tätigkeitsbereiche aufspalten.

- Konvertierung der ORACLE-Datenbank,
- Konvertierung des ASM2-Archivs,
- Sicherung des Archivdatenbestands auf CD-Rom,
- Umstellung der Verarbeitungsprogramme und
- Umstellung der Datenbankschnittstellen.

Neben der relativ problemlosen Konvertierung der aktuellen ORACLE Datenbanktabellen³³ von MVS auf UNIX, müssen sämtliche Archivdatenbestände aus dem alten MVS-Archivsystem ASM2 in ein neues ADASM-Archiv unter UNIX übertragen werden. Es handelt sich dabei um ca. 300 GB Daten in ungefähr 4.500 Dateien. Dieser Vorgang kann zudem nicht automatisiert erfolgen, da abhängig von der spezifischen Dateiart verschiedene Übertragungsschritte notwendig sind. Zusätzlich erfordert die Übertragung umfangreiche Kontrollen der Korrektheit und Vollständigkeit des technischen Übertragungsvorgangs. Vor allem die Konvertierung der umfangreichen DTB-Daten in Form von SAS-Dateien erweist sich als außerordentlich zeit- und arbeitsintensiv, da die gesamten Datenbestände via ftp von der einen in die andere Rechnerwelt übertragen werden müssen. Hierzu müssen die gesamten SAS-Datasets in ein Transportformat konvertiert, übertragen und wieder

³²Es verändern sich beispielsweise ab 1. Juni 1996 die Datenformate der DWZ-Daten grundlegend.

³³Diese Tabellen umfassen ca. 3 GB Daten.

zurückkonvertiert werden. Parallel wird die Umstellung des Archivs dazu genutzt die gesamte Historie auf CD-Rom's zu sichern³⁴.

Der größte Arbeitsanfall erfolgt durch die notwendige Neuprogrammierung sämtlicher Schnittstellen- und Verarbeitungsprogramme, die innerhalb der letzten sieben Jahre erstellt wurden³⁵. Diese Gelegenheit wird genutzt, um sämtliche Programme³⁶ auf eine einheitliche, auch im Hinblick auf zukünftige notwendige Umstellungen optimale, Softwareplattform³⁷ zu stellen.

Neben diesen technischen Umstellungen bedarf es auch umfangreicher organisatorischer Veränderungen, erwähnt seien an dieser Stelle beispielhaft die unterschiedliche Vergabe von Zugriffsrechten und die verschiedenen Dateikonvention auf beiden Rechnersystemen.

Voraussetzung für die Durchführung dieser Tätigkeiten ist eine ausreichende Ausstattung sowohl im personellen wie auch technischen Bereich.

5 Notwendige Ausstattung

5.1 Personal

Die KKMDB wird zur Zeit von drei wissenschaftlichen Mitarbeitern³⁸ betreut. Ihnen zur Seite stehen für die tägliche Datenverarbeitung, Datenanfragen, Programmieraufgaben usw. zusätzlich neun studentische Hilfskräfte.

5.2 Hardware

Neben der Hardwareausstattung die das Rechenzentrum der Universität Karlsruhe in Form eines IBM3090 Mainframe, diversen Workstations und eines Archivierungssystem zur Verfügung stellt, wurden aus Institutsmitteln drei Arbeitsplatzrechner für die verantwortlichen wissenschaftlichen Mitarbeiter angeschafft.

Daneben wurden drei weitere Pentium-Rechner mit zusammen 13 GB Festplatten für die Datenkonvertierung und die Datensicherung auf CD-Rom beschafft. Neben einem CD-Rom Recorder wird auch ein institutseigener Tape-Streamer zur Verarbeitung der Daten eingesetzt. Diese Rechner und weitere

³⁴Das Rechenzentrum bietet nur das ADSM als Sicherungsmedium für große Datenvolumen an.

³⁵Mehr als 15 Mannjahre Programmieraufwand.

³⁶Es existieren umfangreiche Cobol, Fortran, Pascal, SAS und SQL-Programme und zusätzlich komfortable Benutzeroberflächen auf ISPF-Panel Basis.

³⁷Voraussichtlich SAS und C.

³⁸Für die Verwaltung der KKMDB sind Dipl. Wirt.-Ing. Ralf Herrmann (DWZ, WM), Dr. Torsten Lüdecke (ORACLE Datenbank) und Dipl. Wirt.- Ing. Marco Neumann (DTB) zuständig.

IBM-kompatible Rechner dienen den studentischen Hilfskräften als Arbeitsplatzrechner. Zusätzlich zu den institutsinternen Geräten wurden insgesamt 10 GB Festplatten aus Institutsmitteln angeschafft, die im Rechenzentrum der Universität Karlsruhe installiert wurden und den dringend notwendigen Plattenplatz für die Datenkonvertierung und die Umstellungsarbeiten schafften.

5.3 Software

Die Datenverarbeitung findet unter MVS, UNIX und DOS statt. Standardprogramme die regen Einsatz finden sind vor allem SAS und ORACLE, desweiteren diverse Fortran, Cobol, Pascal und C-Compiler. Die Programme die speziell für die KKMDB entwickelt wurden lassen sich funktional in

- Verarbeitungsprogramme,
- Schnittstellenprogramme und
- Benutzeroberflächen

einteilen. Die *Datenverarbeitung* läuft zum größten Teil³⁹ über Benutzeroberflächen⁴⁰ ab. Organisatorisch lassen sich vier Bereiche unterscheiden.

- Karlsruher Datenbank-Updatesystem (KADBUS),
- Karlsruher Interaktives Datenzugriffssystem (KIDZ),
- Karlsruher Wertpapierinformationssystem (KWISS) und
- Sonstiges.

KADBUS ermöglicht die tägliche und monatliche Verarbeitung der anfallenden DWZ- und WM-Daten. Daneben umfaßt es eine Archivverwaltung und Update-Menüs zur Korrektur von Bereinigungsdaten und Kursen. Technisch basiert die Verarbeitung der Rohdaten auf diversen SAS, Cobol und Fortran-Programmen. KIDZ dient der Datenabfrage aus der ORACLE-Datenbank. Sie ermöglicht sowohl interaktiv als auch im Batchbetrieb die komfortable Abfrage von Kurs- und Umsatzdaten. Unter anderem ist es möglich diskrete und kontinuierliche Renditen und Umsätze auf täglicher, wöchentlicher oder monatlicher Basis abzurufen. Dies kann sowohl bereinigt als auch unbereinigt erfolgen. Zusätzlich sind Stamm- und Termindaten nach verschiedenen

³⁹Die Datenverarbeitung der DTB-Daten erfolgt mittels von den Benutzeroberflächen unabhängigen SAS-Programmen.

⁴⁰ISPF-Panels.

Kriterien abfragbar. Im Gegensatz zu KIDZ, das dateorientiert arbeitet⁴¹, ermöglicht KWISS die komfortable bildschirmorientierte Abfrage und Pflege von Daten. Die bei Datenanfragen notwendigen Berechnungen und die benötigten Datenbankschnittstellen sind in Fortran realisiert. Bei KWISS werden die Daten in SQL-Forms-Masken dargestellt. Unter Sonstiges fallen vor allem die Programme zum Laden der ORACLE Datenbank und die SAS-Programme zur Verarbeitung der DTB-Daten.

In den umfangreichen selbstentwickelten Programmen und Benutzeroberflächen stecken mehr als 15 Mannjahre Programmieraufwand.

6 Ausblick

Die Zukunft der KKMDB ist gekennzeichnet durch

- aufwendige Umstellungsarbeiten,
- übergangsweise parallele Datenverarbeitung auf zwei Rechnersystemen,
- weiteres starkes Anwachsen des täglichen Datenvolumens,
- steigende Anzahl von Datenanfragen,
- steigende Anforderung an Plattenplatz und CPU-Leistung sowie
- unsichere Finanzierung.

Die zur Zeit vordringlichste Aufgabe ist die Umstellung der KKMDB auf die neue UNIX-Welt. Dazu ist es notwendig in einer Übergangsphase parallel auf beiden Systemen die Daten zu verarbeiten und die Schnittstellen und Verarbeitungsprogramme vollständig neu zu erstellen.

Daneben wird der allgemeine Trend hin zu elektronischen Handelssystemen auch in *Zukunft* weiterhin zu stark wachsendem täglich anfallendem Datenvolumen führen, da solche Handelssysteme es erlauben das Handelsgeschehen viel genauer zu erfassen. Die Tatsache, daß heute schon mehr als 90% des Datenzuwachses von nur einer elektronischen Börse, der DTB, stammt und deren 6 jähriger Datenbestand die Daten der 8 Präsenzbörsen, die 36 Jahre zurückreichen, bereits um mehr als das Vierfache übersteigt, läßt erahnen, welche Konsequenzen die zunehmende Einführung von elektronischen Handelssystemen in bezug auf das anfallende Datenvolumen haben wird.

Diese Datenflut, vor allem bedingt durch den damit einhergehenden Detaillierungsgrad der enthaltenen Informationen, erschließt völlig neue empirische

⁴¹Das heißt die Eingabedaten werden aus einer Eingabedatei gelesen und das Ergebnis der Datenanfrage in eine Ausgabedatei geschrieben.

Forschungsbereiche⁴². Der Trend hin zu immer detaillierteren Daten spiegelt sich somit auch in der empirischen Kapitalmarktforschung wider.

Um diesen Anforderungen zu genügen, bietet die DWZ ab April 1996 für alle an Deutschen Wertpapierbörsen gehandelten Wertpapiere zeitgestempelte Intraday-Informationen an, Vorbote einer Entwicklung, die bereits zur Vervierfachung des täglichen DWZ-Datenvolumens⁴³ führt. Um mit internationaler Forschung Schritt halten zu können und um den Wert der bereits geleisteten Arbeit, der vor allem in der Vollständigkeit und der Qualität der Daten liegt, zu erhalten, ist es dringend notwendig die neu zu Verfügung stehenden Information in die KKMDB aufzunehmen.

Für die KKMDB resultieren aus dieser Entwicklung zwei Konsequenzen. Zum einen wird das zu verarbeitende Datenvolumen weiter stark anwachsen, zum anderen werden zukünftig auch immer mehr Datenanfragen mit immer größerem Datenvolumen⁴⁴ von externen Forschungseinrichtungen eingehen. Als Lösung für die KKMDB bietet sich zur Befriedigung von Datenanfragen verstärkt die Verwendung von standardisierten Datenpaketen an.

Für den reibungslosen weiteren Betrieb der Datenbank ist die effiziente monatliche Verarbeitung, sowohl im Hinblick auf die CPU-Leistung wie auch den Plattenplatz eine existentielle Notwendigkeit. Parallel zu dieser Datenverarbeitung muß es auch weiterhin möglich sein externe Datenanfragen zu beantworten, um die Grundlage für die meisten deutschen empirischen Forschungsprojekte im Bereich der Kapitalmärkte zu erhalten.

Desweiteren sollte genügend CPU-Leistung und Speicherplatz für die verschiedenen institutseigenen Forschungsprojekte zur Verfügung stehen. Bei denen auch mit drastisch zunehmenden Anforderungen in bezug auf Plattenplatz und CPU-Leistung zu rechnen ist. Dies ist dringend notwendig um im internationalen Kontext mit der Entwicklung im Bereich der empirischen Kapitalmarktforschung Schritt halten zu können.

Über allen diesen technischen Problemen steht die Frage der weiteren Finanzierung dieser einzigartigen Quelle deutscher Kapitalmarktdaten.

⁴²Beispielsweise im Marktstrukturbereich.

⁴³Ab April 1996 wächst voraussichtlich das monatlich zu verarbeitende DWZ-Datenvolumen auf mehr als 500 MB an.

⁴⁴Nicht zuletzt bedingt durch einen neuen DFG-Schwerpunkt, bei dem unter anderem Fragestellungen aus dem Marktstrukturbereich untersucht werden sollen.

Literatur

- [1] **Bauer, C. (1992)** *Das Risiko von Aktienanlagen*. Müller Botermann Verlag, Reihe: Finanzierung, Steuer, Wirtschaftsprüfung Band 15, Köln.
- [2] **Bühler, W., Hax, H. Schmidt R. (Hrsg.)(1993)** *Empirische Kapitalmarktforschung*. Zeitschrift für betriebswirtschaftliche Forschung, Sonderheft 31, Verlagsgruppe Handelsblatt.
- [3] **Bußmann, Johannes (1988)** *Das Management von Zinsänderungsrisiken*. Peter Lang Verlag, Frankfurt am Main.
- [4] **Entrup, U. (1995)** *Kapitalmarktreaktionen auf Optionsanleihen*. Deutscher Universitätsverlag, Wiesbaden.
- [5] **Frantzmann, H.-J. (1989)** *Saisonalitäten und Bewertung am deutschen Aktien- und Rentenmarkt*. Fitz Knapp Verlag, Frankfurt am Main.
- [6] **Frantzmann, H.-J. (1990)** *Zur Messung des Marktrisikos deutscher Aktien*. Zeitschrift für betriebswirtschaftliche Forschung, Januar 1990, S. 67-83.
- [7] **Göppl, H., Herrmann, R., Lüdecke, T. (1995)** *Datenbank-Handbuch Teil I : Beschreibung der Kursdaten für Aktien und Optionscheine*. Universität Karlsruhe.
- [8] **Göppl, H., Herrmann, R., Lüdecke, T. (1993a)** *Datenbank-Handbuch Teil II : Beschreibung der Kursdaten für Optionen*. Universität Karlsruhe.
- [9] **Göppl, H., Herrmann, R., Lüdecke, T. (1993b)** *Datenbank-Handbuch Teil III : Beschreibung der Stammdaten von Aktien*. Universität Karlsruhe.
- [10] **Göppl, H., Herrmann, R., Lüdecke, T. (1993c)** *Datenbank-Handbuch Teil V : Beschreibung der Termindaten*. Universität Karlsruhe.
- [11] **Göppl, H., et al. (1996)** *The German Equity Market: Risk, Return, and Liquidity*. in Bühler, W., Hax, H., Schmidt, R. (Hrsg.) *Empirical Research on the German Capital Market*. Springer Verlag, Berlin, in Veröffentlichung.
- [12] **Göppl, H., Schütz, H. (1992)** *Die Konzeption eines Deutschen Aktienindex für Forschungszwecke (DAFOX)*. Diskussionspapier, Universität Karlsruhe.

- [13] **Hanel, M. (1991)** *Aktienauswahl auf der Grundlage des Konzeptes der Relativen Stärke*. Fitz Knapp Verlag, Frankfurt am Main.
- [14] **Hecker, R. (1993)** *Informationsgehalt von Optionspreisen*. Physica-Verlag, Heidelberg.
- [15] **Iversen, P. (1994)** *Geld-Brief-Spannen deutscher Standardwerte*. Deutscher Universitäts-Verlag, Wiesbaden.
- [16] **Lüdecke, T. (1996)** *Struktur und Qualität von Finanzmärkten*. Gabler Verlag, in Veröffentlichung.
- [17] **Margana, I. (1986)** *Zinsstruktur und Zinsentwicklung auf effizienten Märkten*. Dissertation, Universität Karlsruhe.
- [18] **Meyer, B. (1994)** *Der Overreaction Effekt am deutschen Aktienmarkt*. Fitz Knapp Verlag, Frankfurt am Main.
- [19] **Müller, W. (1992)** *Bilanzinformation und Aktienbewertung*. Fitz Knapp Verlag, Frankfurt am Main.
- [20] **Röder, K. (1994)** *Der DAX-Future*. Verlag Josef Eul, Bergisch Gladbach, Köln.
- [21] **Sauer, A. (1989)** *Arbitragemöglichkeiten am deutschen Rentenmarkt*. Fitz Knapp Verlag, Frankfurt am Main.
- [22] **Sauer, A. (1991)** *Die Bereinigung von Aktienkursen – Ein kurzer Überblick über Konzept und praktische Umsetzung*. Diskussionspapier, Universität Karlsruhe.
- [23] **Sauer, A. (1994)** *Faktormodelle und Bewertung am deutschen Aktienmarkt*. Fitz Knapp Verlag, Frankfurt am Main.
- [24] **Schlag, C. (1994)** *Neues zum Intervalling-Effekt*. Kredit und Kapital 27, S. 437-460.
- [25] **Trautmann, S. (1986)** *Finanztitelbewertung bei arbitragefreien Finanzmärkten*. Unveröffentlichte Habilitationsschrift, Universität Karlsruhe.
- [26] **Uhrig, M. (1996)** *Bewertung von Zinsoptionen bei stochastischer Zinsvolatilität: ein Inversionsansatz*. Gabler Verlag, in Veröffentlichung.
- [27] **Walter, U. (1996)** *Die Bewertung von Zinsoptionen unter Berücksichtigung der Zins- und Volatilitätsstruktur*. Gabler Verlag, in Veröffentlichung.
- [28] **Winkelmann, M. (1984)** *Aktienbewertung in Deutschland*. Verlag Anton Hain, Königstein/Ts.