J.P.Morgan/Reuters

RiskMetrics™—Technical Document

Fourth Edition, 1996

New York
December 17, 1996

- J.P. Morgan and Reuters have teamed up to enhance RiskMetrics™. Morgan will continue to be responsible for enhancing the methods outlined in this document, while Reuters will control the production and distribution of the RiskMetrics™ data sets.
- Expanded sections on methodology outline enhanced analytical solutions for dealing with nonlinear options risks and introduce methods on how to account for non-normal distributions.
- Enclosed diskette contains many examples used in this document. It allows readers to experiment with our risk measurement techniques.
- All publications and daily data sets are available free of charge on J.P. Morgan's Web page on the Internet at **http://www.jpmorgan.com/RiskManagement/RiskMetrics/RiskMetrics.html**. This page is accessible directly or through third party services such as CompuServe®, America Online™, or Prodigy®.

Morgan Guaranty Trust Company
Risk Management Advisory
Jacques Longerstaey
(1-212) 648-4936
riskmetrics@jpmorgan.com

Reuters Ltd
International Marketing
Martin Spencer
(44-171) 542-3260
martin.spencer@reuters.com

This *Technical Document* provides a detailed description of RiskMetrics™, a set of techniques and data to measure market risks in portfolios of fixed income instruments, equities, foreign exchange, commodities, and their derivatives issued in over 30 countries. This edition has been expanded significantly from the previous release issued in May 1995.

We make this methodology and the corresponding RiskMetrics™ data sets available for three reasons:

1. We are interested in promoting greater transparency of market risks. Transparency is the key to effective risk management.

2. Our aim has been to establish a benchmark for market risk measurement. The absence of a common point of reference for market risks makes it difficult to compare different approaches to and measures of market risks. Risks are comparable only when they are measured with the same yardstick.

3. We intend to provide our clients with sound advice, including advice on managing their market risks. We describe the RiskMetrics™ methodology as an aid to clients in understanding and evaluating that advice.

Both J.P. Morgan and Reuters are committed to further the development of RiskMetrics™ as a fully transparent set of risk measurement methods. We look forward to continued feedback on how to maintain the quality that has made RiskMetrics™ the benchmark for measuring market risk.

RiskMetrics™ is based on, but differs significantly from, the risk measurement methodology developed by J.P. Morgan for the measurement, management, and control of market risks in its trading, arbitrage, and own investment account activities. **We remind our readers that no amount of sophisticated analytics will replace experience and professional judgment in managing risks**. RiskMetrics™ is nothing more than a high-quality tool for the professional risk manager involved in the financial markets and is not a guarantee of specific results.

RiskMetrics™—Technical Document
Fourth Edition (December 1996)

This book

This is the reference document for RiskMetrics™. It covers all aspects of RiskMetrics and supersedes all previous editions of the *Technical Document*. It is meant to serve as a reference to the methodology of statistical estimation of market risk, as well as detailed documentation of the analytics that generate the data sets that are published daily on our Internet Web sites.

This document reviews

1. The conceptual framework underlying the methodologies for estimating market risks.
2. The statistics of financial market returns.
3. How to model financial instrument exposures to a variety of market risk factors.
4. The data sets of statistical measures that we estimate and distribute daily over the Internet and shortly, the Reuters Web.

Measurement and management of market risks continues to be as much a craft as it is a science. It has evolved rapidly over the last 15 years and has continued to evolve since we launched RiskMetrics in October 1994. Dozens of professionals at J.P. Morgan have contributed to the development of this market risk management technology and the latest document contains entries or contributions from a significant number of our market risk professionals.

We have received numerous constructive comments and criticisms from professionals at Central Banks and regulatory bodies in many countries, from our competitors at other financial institutions, from a large number specialists in academia and last, but not least, from our clients. Without their feedback, help, and encouragement to pursue our strategy of open disclosure of methodology and free access to data, we would not have been as successful in advancing this technology as much as we have over the last two years.

What is RiskMetrics?

RiskMetrics is a set of tools that enable participants in the financial markets to estimate their exposure to market risk under what has been called the “Value-at-Risk framework”. RiskMetrics has three basic components:

- A set of market risk measurement methodologies outlined in this document.
- Data sets of volatility and correlation data used in the computation of market risk.
- Software systems developed by J.P.Morgan, subsidiaries of Reuters, and third party vendors that implement the methodologies described herein.

With the help of this document and the associated line of products, users should be in a position to estimate market risks in portfolios of foreign exchange, fixed income, equity and commodity products.

J.P. Morgan and Reuters team up on RiskMetrics

In June 1996, J.P. Morgan signed an agreement with Reuters to cooperate on the building of a new and more powerful version of RiskMetrics. Since the launch of RiskMetrics in October 1994, we have received numerous requests to add new products, instruments, and markets to the daily volatility and correlation data sets. We have also perceived the need in the market for a more flexible VaR data tool than the standard matrices that are currently distributed over the Internet. The new partnership with Reuters, which will be based on the precept that both firms will focus on their respective strengths, will help us achieve these objectives.

Methodology

J.P. Morgan will continue to develop the RiskMetrics set of VaR methodologies and publish them in the quarterly *RiskMetrics Monitor* and in the annual *RiskMetrics—Technical Document*.

RiskMetrics data sets

Reuters will take over the responsibility for data sourcing as well as production and delivery of the risk data sets. The current RiskMetrics data sets will continue to be available on the Internet free of charge and will be further improved as a benchmark tool designed to broaden the understanding of the principles of market risk measurement.

When J.P. Morgan first launched RiskMetrics in October 1994, the objective was to go for broad market coverage initially, and follow up with more granularity in terms of the markets and instruments covered. This over time, would reduce the need for proxies and would provide additional data to measure more accurately the risk associated with non-linear instruments.

The partnership will address these new markets and products and will also introduce a new customizable service, which will be available over the Reuters Web service. The customizable RiskMetrics approach will give risk managers the ability to scale data to meet the needs of their individual trading profiles. Its capabilities will range from providing customized covariance matrices needed to run VaR calculations, to supplying data for historical simulation and stress-testing scenarios.

More details on these plans will be discussed in later editions of the *RiskMetrics Monitor.*

Systems

Both J.P. Morgan and Reuters, through its Sailfish subsidiary, have developed client-site RiskMetrics VaR applications. These products, together with the expanding suite of third party applications will continue to provide RiskMetrics implementations.

What is new in this fourth edition?

In terms of content, the Fourth Edition of the *Technical Document* incorporates the changes and refinements to the methodology that were initially outlined in the 1995–1996 editions of the *RiskMetrics Monitor*:

- **Expanded framework:** We have worked extensively on refining the analytical framework for analyzing options risk without having to perform relatively time consuming simulations and have outlined the basis for an improved methodology which incorporates better information on the tails of distributions related to financial asset price returns; we've also developed a data synchronization algorithm to refine our volatility and correlation estimates for products which do not trade in the same time zone;

- **New markets:** We expanded the daily data sets to include estimated volatilities and correlations of additional foreign exchange, fixed income and equity markets, particularly in South East Asia and Latin America.

- **Fine-tuned methodology:** We have modified the approach in a number of ways. First, we've changed our definition of price volatility which is now based on a total return concept; we've also revised some of the algorithms used in our mapping routines and are in the process of redefining the techniques used in estimating equity portfolio risk.

- **RiskMetrics products:** While we have continued to expand the list of third parties providing RiskMetrics products and support, this is no longer included with this document. Given the rapid pace of change in the availability of risk management software products, readers are advised to consult our Internet web site for the latest available list of products. This list, which now includes FourFifteen™, J.P. Morgan's own VaR calculator and report generating software, continues to grow, attesting to the broad acceptance RiskMetrics has achieved.

- **New tools to use the RiskMetrics data sets:** We have published an Excel add-in function which enables users to import volatilities and correlations directly into a spreadsheet. This tool is available from our Internet web site.

The structure of the document has changed only slightly. As before, its size warrants the following note: One need not read and understand the entire document in order to benefit from RiskMetrics. The document is organized in parts that address subjects of particular interest to many readers.

Part I: Risk Measurement Framework

This part is for the general practitioner. It provides a practical framework on how to think about market risks, how to apply that thinking in practice, and how to interpret the results. It reviews the different approaches to risk estimation, shows how the calculations work on simple examples and discusses how the results can be used in limit management, performance evaluation, and capital allocation.

Part II: Statistics of Financial Market Returns

This part requires an understanding and interest in statistical analysis. It reviews the assumptions behind the statistics used to describe financial market returns and how distributions of future returns can be estimated.

Part III: Risk Modeling of Financial Instruments

This part is required reading for implementation of a market risk measurement system. It reviews how positions in any asset class can be described in a standardized fashion (foreign exchange, interest rates, equities, and commodities). Special attention is given to derivatives positions. The purpose is to demystify derivatives in order to show that their market risks can be measured in the same fashion as their underlying.

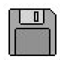
Part IV: RiskMetrics Data Sets

This part should be of interest to users of the RiskMetrics data sets. First it describes the sources of all daily price and rate data. It then discusses the attributes of each volatility and correlation series in the RiskMetrics data sets. And last, it provides detailed format descriptions required to decipher the data sets that can be downloaded from public or commercial sources.

Appendices

This part reviews some of the more technical issues surrounding methodology and regulatory requirements for market risk capital in banks and demonstrates the use of RiskMetrics with the example diskette provided with this document. Finally, Appendix H shows you how to access the RiskMetrics data sets from the Internet.

RiskMetrics examples diskette



This diskette is located inside the back cover. It contains an Excel workbook that includes some of the examples shown in this document. Such examples are identified by the icon shown here.

Future plans

We expect to update this *Technical Document* annually as we adapt our market risk standards to further improve the techniques and data to meet the changing needs of our clients.

RiskMetrics is a now an integral part of J.P. Morgan’s Risk Management Services group which provides advisory services to a wide variety of the firm’s clients. We continue to welcome any suggestions to enhance the methodology and adapt it further to the needs of the market. All suggestions, requests and inquiries should be directed to the authors of this publication or to your local RiskMetrics contacts listed on the back cover.

Acknowledgments

The authors would like to thank the numerous individuals who participated in the writing and editing of this document, particularly Chris Finger and Chris Athaide from J.P. Morgan’s risk management research group, and Elizabeth Frederick and John Matero from our risk advisory practice. Finally, this document could not have been produced without the contributions of our consulting editor, Tatiana Kolubayev. We apologize for any omissions to this list.

Table of contents

Part I Risk Measurement Framework

Chapter 1. Introduction **3**
1.1 An introduction to Value-at-Risk and RiskMetrics 6
1.2 A more advanced approach to Value-at-Risk using RiskMetrics 7
1.3 What RiskMetrics provides 16

Chapter 2. Historical perspective of VaR **19**
2.1 From ALM to VaR 22
2.2 VaR in the framework of modern financial management 24
2.3 Alternative approaches to risk estimation 26

Chapter 3. Applying the risk measures **31**
3.1 Market risk limits 33
3.2 Calibrating valuation and risk models 34
3.3 Performance evaluation 34
3.4 Regulatory reporting, capital requirement 36

Part II Statistics of Financial Market Returns

Chapter 4. Statistical and probability foundations **43**
4.1 Definition of financial price changes and returns 45
4.2 Modeling financial prices and returns 49
4.3 Investigating the random-walk model 54
4.4 Summary of our findings 64
4.5 A review of historical observations of return distributions 64
4.6 RiskMetrics model of financial returns: A modified random walk 73
4.7 Summary 74

Chapter 5. Estimation and forecast **75**
5.1 Forecasts from implied versus historical information 77
5.2 RiskMetrics forecasting methodology 78
5.3 Estimating the parameters of the RiskMetrics model 90
5.4 Summary and concluding remarks 100

Part III Risk Modeling of Financial Instruments

Chapter 6. Market risk methodology **105**
6.1 Step 1—Identifying exposures and cash flows 107
6.2 Step 2—Mapping cash flows onto RiskMetrics vertices 117
6.3 Step 3—Computing Value-at-Risk 121
6.4 Examples 134

Chapter 7. Monte Carlo **149**
7.1 Scenario generation 151
7.2 Portfolio valuation 155
7.3 Summary 157
7.4 Comments 159

Part IV RiskMetrics Data Sets

Chapter 8.	**Data and related statistical issues**	**163**
8.1	Constructing RiskMetrics rates and prices	165
8.2	Filling in missing data	170
8.3	The properties of correlation (covariance) matrices and VaR	176
8.4	Rebasing RiskMetrics volatilities and correlations	183
8.5	Nonsynchronous data collection	184
Chapter 9.	**Time series sources**	**197**
9.1	Foreign exchange	199
9.2	Money market rates	199
9.3	Government bond zero rates	200
9.4	Swap rates	202
9.5	Equity indices	203
9.6	Commodities	205
Chapter 10.	**RiskMetrics volatility and correlation files**	**207**
10.1	Availability	209
10.2	File names	209
10.3	Data series naming standards	209
10.4	Format of volatility files	211
10.5	Format of correlation files	212
10.6	Data series order	214
10.7	Underlying price/rate availability	214

Part V Backtesting

Chapter 11.	**Performance assessment**	**217**
11.1	Sample portfolio	219
11.2	Assessing the RiskMetrics model	220
11.3	Summary	223

Appendices

Appendix A.	Tests of conditional normality	227
Appendix B.	Relaxing the assumption of conditional normality	235
Appendix C.	Methods for determining the optimal decay factor	243
Appendix D.	Assessing the accuracy of the delta-gamma approach	247
Appendix E.	Routines to simulate correlated normal random variables	253
Appendix F.	BIS regulatory requirements	257
Appendix G.	Using the RiskMetrics examples diskette	263
Appendix H.	RiskMetrics on the Internet	267

Reference

Glossary of terms	271
Bibliography	275

List of charts

Chart 1.1	VaR statistics	6
Chart 1.2	Simulated portfolio changes	9
Chart 1.3	Actual cash flows	9
Chart 1.4	Mapping actual cash flows onto RiskMetrics vertices	10
Chart 1.5	Value of put option on USD/DEM	14
Chart 1.6	Histogram and scattergram of rate distributions	15
Chart 1.7	Valuation of instruments in sample portfolio	15
Chart 1.8	Representation of VaR	16
Chart 1.9	Components of RiskMetrics	17
Chart 2.1	Asset liability management	22
Chart 2.2	Value-at-Risk management in trading	23
Chart 2.3	Comparing ALM to VaR management	24
Chart 2.4	Two steps beyond accounting	25
Chart 3.1	Hierarchical VaR limit structure	33
Chart 3.2	Ex post validation of risk models: DEaR vs. actual daily P&L	34
Chart 3.3	Performance evaluation triangle	35
Chart 3.4	Example: comparison of cumulative trading revenues	35
Chart 3.5	Example: applying the evaluation triangle	36
Chart 4.1	Absolute price change and log price change in U.S. 30-year government bond	47
Chart 4.2	Simulated stationary/mean-reverting time series	52
Chart 4.3	Simulated nonstationary time series	53
Chart 4.4	Observed stationary time series	53
Chart 4.5	Observed nonstationary time series	54
Chart 4.6	USD/DEM returns	55
Chart 4.7	USD/FRF returns	55
Chart 4.8	Sample autocorrelation coefficients for USD/DEM foreign exchange returns	57
Chart 4.9	Sample autocorrelation coefficients for USD S&P 500 returns	58
Chart 4.10	USD/DEM returns squared	60
Chart 4.11	S&P 500 returns squared	60
Chart 4.12	Sample autocorrelation coefficients of USD/DEM squared returns	61
Chart 4.13	Sample autocorrelation coefficients of S&P 500 squared returns	61
Chart 4.14	Cross product of USD/DEM and USD/FRF returns	63
Chart 4.15	Correlogram of the cross product of USD/DEM and USD/FRF returns	63
Chart 4.16	Leptokurtotic vs. normal distribution	65
Chart 4.17	Normal distribution with different means and variances	67
Chart 4.18	Selected percentile of standard normal distribution	69
Chart 4.19	One-tailed confidence interval	70
Chart 4.20	Two-tailed confidence interval	71
Chart 4.21	Lognormal probability density function	73
Chart 5.1	DEM/GBP exchange rate	79
Chart 5.2	Log price changes in GBP/DEM and VaR estimates (1.65σ)	80
Chart 5.3	NLG/DEM exchange rate and volatility	87
Chart 5.4	S&P 500 returns and VaR estimates (1.65σ)	88
Chart 5.5	GARCH(1,1)-normal and EWMA estimators	90
Chart 5.6	USD/DEM foreign exchange	92
Chart 5.7	Tolerance level and decay factor	94
Chart 5.8	Relationship between historical observations and decay factor	95
Chart 5.9	Exponential weights for $T = 100$	95
Chart 5.10	One-day volatility forecasts on USD/DEM returns	96
Chart 5.11	One-day correlation forecasts for returns on USD/DEM FX rate and on S&P500	96
Chart 5.12	Simulated returns from RiskMetrics model	101
Chart 6.1	French franc 10-year benchmark maps	109

Chart 6.2	Cash flow representation of a simple bond	109
Chart 6.3	Cash flow representation of a FRN	110
Chart 6.4	Estimated cash flows of a FRN	111
Chart 6.5	Cash flow representation of simple interest rate swap	111
Chart 6.6	Cash flow representation of forward starting swap	112
Chart 6.7	Cash flows of the floating payments in a forward starting swap	113
Chart 6.8	Cash flow representation of FRA	113
Chart 6.9	Replicating cash flows of 3-month vs. 6-month FRA	114
Chart 6.10	Cash flow representation of 3-month Eurodollar future	114
Chart 6.11	Replicating cash flows of a Eurodollar futures contract	114
Chart 6.12	FX forward to buy Deutsche marks with US dollars in 6 months	115
Chart 6.13	Replicating cash flows of an FX forward	115
Chart 6.14	Actual cash flows of currency swap	116
Chart 6.15	RiskMetrics cash flow mapping	118
Chart 6.16	Linear and nonlinear payoff functions	123
Chart 6.17	VaR horizon and maturity of money market deposit	128
Chart 6.18	Long and short option positions	131
Chart 6.19	DEM 3-year swaps in Q1-94	141
Chart 7.1	Frequency distributions for and	153
Chart 7.2	Frequency distribution for DEM bond price	154
Chart 7.3	Frequency distribution for USD/DEM exchange rate	154
Chart 7.4	Value of put option on USD/DEM	157
Chart 7.5	Distribution of portfolio returns	158
Chart 8.1	Constant maturity future: price calculation	170
Chart 8.2	Graphical representation	175
Chart 8.3	Number of variables used in EM and parameters required	176
Chart 8.4	Correlation forecasts vs. return interval	185
Chart 8.5	Time chart	188
Chart 8.6	10-year Australia/US government bond zero correlation	190
Chart 8.7	Adjusting 10-year USD/AUD bond zero correlation	194
Chart 8.8	10-year Japan/US government bond zero correlation	195
Chart 9.1	Volatility estimates: daily horizon	202
Chart 11.1	One-day Profit/Loss and VaR estimates	219
Chart 11.2	Histogram of standardized returns	221
Chart 11.3	Standardized lower-tail returns	222
Chart 11.4	Standardized upper-tail returns	222
Chart A.1	Standard normal distribution and histogram of returns on USD/DEM	227
Chart A.2	Quantile-quantile plot of USD/DEM	232
Chart A.3	Quantile-quantile plot of 3-month sterling	234
Chart B.1	Tails of normal mixture densities	238
Chart B.2	GED distribution	239
Chart B.3	Left tail of GED (ν) distribution	240
Chart D.1	Delta vs. time to expiration and underlying price	248
Chart D.2	Gamma vs. time to expiration and underlying price	249

List of tables

Table 2.1	Two discriminating factors to review VaR models	29
Table 3.1	Comparing the Basel Committee proposal with RiskMetrics	39
Table 4.1	Absolute, relative and log price changes	46
Table 4.2	Return aggregation	49
Table 4.3	Box-Ljung test statistic	58
Table 4.4	Box-Ljung statistics	59
Table 4.5	Box-Ljung statistics on squared log price changes (cv = 25)	62
Table 4.6	Model classes	66
Table 4.7	VaR statistics based on RiskMetrics and BIS/Basel requirements	71
Table 5.1	Volatility estimators	78
Table 5.2	Calculating equally and exponentially weighted volatility	81
Table 5.3	Applying the recursive exponential weighting scheme to compute volatility	82
Table 5.4	Covariance estimators	83
Table 5.5	Recursive covariance and correlation predictor	84
Table 5.6	Mean, standard deviation and correlation calculations	91
Table 5.7	The number of historical observations used by the EWMA model	94
Table 5.8	Optimal decay factors based on volatility forecasts	99
Table 5.9	Summary of RiskMetrics volatility and correlation forecasts	100
Table 6.1	Data provided in the daily RiskMetrics data set	121
Table 6.2	Data calculated from the daily RiskMetrics data set	121
Table 6.3	Relationship between instrument and underlying price/rate	123
Table 6.4	Statistical features of an option and its underlying return	130
Table 6.5	RiskMetrics data for 27, March 1995	134
Table 6.6	RiskMetrics map of single cash flow	134
Table 6.7	RiskMetrics map for multiple cash flows	135
Table 6.8	Mapping a 6x12 short FRF FRA at inception	137
Table 6.9	Mapping a 6x12 short FRF FRA held for one month	137
Table 6.10	Structured note specification	139
Table 6.11	Actual cash flows of a structured note	139
Table 6.12	VaR calculation of structured note	140
Table 6.13	VaR calculation on structured note	140
Table 6.14	Cash flow mapping and VaR of interest-rate swap	142
Table 6.15	VaR on foreign exchange forward	143
Table 6.16	Market data and RiskMetrics estimates as of trade date July 1, 1994	145
Table 6.17	Cash flow mapping and VaR of commodity futures contract	145
Table 6.18	Portfolio specification	147
Table 6.19	Portfolio statistics	148
Table 6.20	Value-at-Risk estimates (USD)	148
Table 7.1	Monte Carlo scenarios	155
Table 7.2	Monte Carlo scenarios—valuation of option	156
Table 7.3	Value-at-Risk for example portfolio	158
Table 8.1	Construction of rolling nearby futures prices for Light Sweet Crude (WTI)	168
Table 8.2	Price calculation for 1-month CMF NY Harbor #2 Heating Oil	169
Table 8.3	Belgian franc 10-year zero coupon rate	175
Table 8.4	Singular values for USD yield curve data matrix	182
Table 8.5	Singular values for equity indices returns	182
Table 8.6	Correlations of daily percentage changes with USD 10-year	184
Table 8.7	Schedule of data collection	186
Table 8.7	Schedule of data collection	187
Table 8.8	RiskMetrics closing prices	191
Table 8.9	Sample statistics on RiskMetrics daily covariance forecasts	191
Table 8.10	RiskMetrics daily covariance forecasts	192

Table 8.11	Relationship between lagged returns and applied weights	193
Table 8.12	Original and adjusted correlation forecasts	193
Table 8.13	Correlations between US and foreign instruments	196
Table 9.1	Foreign exchange	199
Table 9.2	Money market rates: sources and term structures	200
Table 9.3	Government bond zero rates: sources and term structures	201
Table 9.4	Swap zero rates: sources and term structures	203
Table 9.5	Equity indices: sources	204
Table 9.6	Commodities: sources and term structures	205
Table 9.7	Energy maturities	205
Table 9.8	Base metal maturities	206
Table 10.1	RiskMetrics file names	209
Table 10.2	Currency and commodity identifiers	210
Table 10.3	Maturity and asset class identifiers	210
Table 10.4	Sample volatility file	211
Table 10.5	Data columns and format in volatility files	212
Table 10.6	Sample correlation file	213
Table 10.7	Data columns and format in correlation files	213
Table 11.1	Realized percentages of VaR violations	220
Table 11.2	Realized “tail return” averages	221
Table A.1	Sample mean and standard deviation estimates for USD/DEM FX	228
Table A.2	Testing for univariate conditional normality	230
Table B.1	Parameter estimates for the South African rand	240
Table B.2	Sample statistics on standardized returns	241
Table B.3	VaR statistics (in %) for the 1st and 99th percentiles	242
Table D.1	Parameters used in option valuation	249
Table D.2	MAPE (%) for call, 1-day forecast horizon	251
Table D.3	ME (%) for call, 1-day forecast horizons	251

Part IV

RiskMetrics Data Sets

Chapter 8. Data and related statistical issues

8.1 Constructing RiskMetrics rates and prices 165
- 8.1.1 Foreign exchange 165
- 8.1.2 Interest rates 165
- 8.1.3 Equities 167
- 8.1.4 Commodities 167

8.2 Filling in missing data 170
- 8.2.1 Nature of missing data 171
- 8.2.2 Maximum likelihood estimation 171
- 8.2.3 Estimating the sample mean and covariance matrix for missing data 172
- 8.2.4 An illustrative example 174
- 8.2.5 Practical considerations 176

8.3 The properties of correlation (covariance) matrices and VaR 176
- 8.3.1 Covariance and correlation calculations 177
- 8.3.2 Useful linear algebra results as applied to the VaR calculation 180
- 8.3.3 How to determine if a covariance matrix is positive semi-definite 181

8.4 Rebasing RiskMetrics volatilities and correlations 183

8.5 Nonsynchronous data collection 184
- 8.5.1 Estimating correlations when the data are nonsynchronous 188
- 8.5.2 Using the algorithm in a multivariate framework 195

Chapter 8. Data and related statistical issues

Peter Zangari
Morgan Guaranty Trust Company
Risk Management Research
(1-212) 648-8641
zangari_peter@jpmorgan.com

This chapter covers the RiskMetrics underlying yields and prices that are used in the volatility and correlation calculations. It also discusses the relationship between the number of time series and the amount of historical data available on these series as it relates to the volatility and correlations.

This chapter is organized as follows:

- Section 8.1 explains the basis or construction of the underlying yields and prices for each instrument type.
- Section 8.2 describes the filling in of missing data points, i.e., expectation maximization.
- Section 8.3 investigates the properties of a generic correlation matrix since these determine whether a portfolio's standard deviation is meaningful.
- Section 8.4 provides an algorithm for recomputing the volatilities and correlations when a portfolio is based in a currency other than USD.
- Section 8.5 presents a methodology to calculate correlations when the yields or prices are sampled at different times, i.e., data recording is nonsynchronous.

8.1 Constructing RiskMetrics rates and prices

In this section we explain the construction of the underlying rates and prices that are used in the RiskMetrics calculations. Since the data represent only a subset of the most liquid instruments available in the markets, proxies should be used for the others. Recommendations on how to apply RiskMetrics to specific instruments are outlined in the paragraphs below.

8.1.1 Foreign exchange

RiskMetrics provides estimates of VaR statistics for returns on 31 currencies as measured against the US dollar (e.g., USD/DEM, USD/FRF) as well as correlations between returns. The datasets provided are therefore suited for estimating foreign exchange risk from a US dollar perspective.

The methodology for using the data to measure foreign exchange risk from a currency perspective other than the US dollar is identical to the one described (Section 6.1.2) above but requires the input of revised volatilities and correlations. These modified volatilities and correlations can easily be derived from the original RiskMetrics datasets as described in Section 8.4. Also refer to the examples diskette.

Finally, measuring market exposure to currencies currently not included in the RiskMetrics data set will involve accessing underlying foreign exchange data from other sources or using one of the 31 currencies as a proxy.

8.1.2 Interest rates

In RiskMetrics we describe the fixed income markets in terms of the price dynamics of zero coupon constant maturity instruments. In the interest rate swap market there are quotes for constant maturities (e.g., 10-year swap rate). In the bond markets, constant maturity rates do not exist therefore we must construct them with the aid of a term structure model.

The current data set provides volatilities and correlations for returns on money market deposits, swaps, and zero coupon government bonds in 33 markets. These parameters allow direct calculation of the volatility of cash flows. Correlations are provided between all RiskMetrics vertices and markets.

8.1.2.1 Money market deposits

The volatilities of price returns on money market deposits are to be used to estimate the market risk of all short-term cash flows (from one month to one year). Though they only cover one instrument type at the short end of the yield curve, money market price return volatilities can be applied to measure the market risk of instruments that are highly correlated with money market deposits, such as Treasury bills or instruments that reprice off of rates such as the prime rate in the US or commercial paper rates.[1]

8.1.2.2 Swaps

The volatilities of price returns on zero coupon swaps are to be used to estimate the market risk of interest rate swaps. We construct zero coupon swap prices and rates because they are required for the cashflow mapping methodology described in Section 6.2. We now explain how RiskMetrics constructs zero coupon swap prices (rates) from observed swap prices and rates by the method known as bootstrapping.

Suppose one knows the zero-coupon term structure, i.e., the prices of zero-coupon swaps $P_1, \ldots, P_n$, where each $P_i = 1 / (1 + z_i)^i$ $i = 1, \ldots, n$ and z_i is the zero-coupon rate for the swap with maturity *i*. Then it is straightforward to find the price of a coupon swap as

$$[8.1] \qquad P_{cn} = P_1 S_n + P_2 S_n + \ldots + P_n (1 + S_n)$$

where S_n denotes the current swap rate on the n period swap. Now, in practice we observe the coupon term structure, $P_{c1}, \ldots, P_{cn}$ maturing at each coupon payment date. Using the coupon swap prices we can apply Eq. [8.1] to solve for the implied zero coupon term structure, i.e., zero coupon swap prices and rates. Starting with a 1-period zero coupon swap, $P_{c1} = P_1 (1 + S_1)$ so that $P_1 = P_{c1} / (1 + S_1)$ or $z_1 = (1 + S_1) / P_{c1} - 1$. Proceeding in an iterative manner, given the discount prices $P_1, \ldots, P_{n-1}$, we can solve for P_n and z_n using the formula

$$[8.2] \qquad P_n = \frac{P_{cn} - P_{n-1} S_n - \ldots - P_1 S_n}{1 + S_n}$$

The current RiskMetrics datasets do not allow differentiation between interest rate risks of instruments of different credit quality; all market risk due to credit of equal maturity and currency is treated the same.

8.1.2.3 Zero coupon government bonds

The volatilities of price returns on zero coupon government bonds are to be used to estimate the market risk in government bond positions. Zero coupon prices (rates) are used because they are consistent with the cash flow mapping methodology described in Section 6.2. Zero coupon government bond prices can also be used as proxies for estimating the volatility of other securities when the appropriate volatility measure does not exist (corporate issues with maturities longer than 10 years, for example).

[1] See the fourth quarter, 1995 *RiskMetrics Monitor* for details.

Zero coupon government bond yield curves cannot be directly observed, they can only be implied from prices of a collection of liquid bonds in the respective market. Consequently, a term structure model must be used to estimate a synthetic zero coupon yield curve which best fits the collection of observed prices. Such a model generates zero coupon yields for arbitrary points along the yield curve.

8.1.2.4 EMBI+

The J. P. Morgan Emerging Markets Bond Index Plus tracks total returns for traded external debt instruments in the emerging markets. It is constructed as a "composite" of its four markets: Brady bonds, Eurobonds, U.S. dollar local markets, and loans. The EMBI+ provides investors with a definition of the market for emerging markets external-currency debt, a list of the traded instruments, and a compilation of their terms. U.S dollar issues currently make up more than 95% of the index and sovereign issues make up 98%. A fuller description of the EMBI+ can be found in the J. P. Morgan publication *Introducing the Emerging Markets Bond Index Plus (EMBI+)* dated July 12, 1995.

8.1.3 Equities

According to the current RiskMetrics methodology, equities are mapped to their domestic market indices (for example, S&P500 for the US, DAX for Germany, and CAC40 for Canada). That is to say, individual stock betas, along with volatilities on price returns of local market indices are used to construct VaR estimates (see Section 6.3.2.2) of individual stocks. The reason for applying the beta coefficient is that it measures the covariation between the return on the individual stock and the return on the local market index whose volatility and correlation are provided by RiskMetrics.

8.1.4 Commodities

A commodity futures contract is a standardized agreement to buy or sell a commodity. The price to a buyer of a commodity futures contract depends on three factors:

1. the current spot price of the commodity,
2. the carrying costs of the commodity. Money tied up by purchasing and carrying a commodity could have been invested in some risk-free, interest bearing instrument. There may be costs associated with purchasing a product in the spot market (transaction costs) and holding it until or consuming it at some later date (storage costs), and
3. the expected supply and demand for the commodity.

The future price of a commodity differs from its current spot price in a way that is analogous to the difference between 1-year and overnight interest rates for a particular currency. From this perspective we establish a term structure of commodity prices similar to that of interest rates.

The most efficient and liquid markets for most commodities are the futures markets. These markets have the advantage of bringing together not only producers and consumers, but also investors who view commodities as they do any other asset class. Because of the superior liquidity and the transparency of the futures markets, we have decided to use futures prices as the foundation for modeling commodity risk. This applies to all commodities except bullion, as described below.

8.1.4.1 The need for commodity term structures

Futures contracts represent standard terms and conditions for delivery of a product at future dates. Recorded over time, their prices represent instruments with decreasing maturities. That is to say, if the price series of a contract is a sequence of expected values of a single price at a specific date in the future, then each consecutive price implies that the instrument is one day close to expiring.

RiskMetrics constructs constant maturity contracts in the same spirit that it constructs constant maturity instruments for the fixed income market. Compared to the fixed income markets, however, commodity markets are significantly less liquid. This is particularly true for very short and very long maturities. Frequently, volatility of the front month contract may decline when the contract is very close to expiration as it becomes uninteresting to trade for a small absolute gain, difficult to trade (a thin market may exist due to this limited potential gain) and, dangerous to trade because of physical delivery concerns. At the long end of the curve, trading liquidity is limited.

Whenever possible, we have selected the maturities of commodity contracts with the highest liquidity as the vertices for volatility and correlation estimates. These maturities are indicated in Table 9.6 in Section 9.6.

In order to construct constant maturity contracts, we have defined two algorithms to convert observed prices into prices from constant maturity contracts:

- Rolling nearby: we simply use the price of the futures contract that expires closest to a fixed maturity.

- Linear interpolation: we linearly interpolate between the prices of the two futures contracts that straddle the fixed maturity.

8.1.4.2 Rolling nearby futures contracts

Rolling nearby contracts are constructed by concatenating contracts that expire, approximately 1, 6, and 12 months (for instance) in the future. An example of this method is shown in Table 8.1.

Table 8.1

Construction of rolling nearby futures prices for Light Sweet Crude (WTI)

	Rolling nearby			**Actual contracts**					
	1st	6th	12th	Mar-94	Apr-94	Aug-94	Sep-94	Feb-95	Mar-95
17-Feb-94	13.93	15.08	16.17	13.93	14.13	15.08	15.28	16.17	16.3
18-Feb-94	14.23	15.11	16.17	14.23	14.3	15.11	15.3	16.17	16.3
19-Feb-94	14.21	15.06	16.13	14.21	14.24	15.06	15.25	16.13	16.27
23-Feb-94	**14.24**	15.23	16.33	**14.24**	14.39	15.23	15.43	16.33	16.47
24-Feb-94	***14.41***	15.44	16.46		***14.41***	15.24	15.44	16.32	16.46

Note that the price of the front month contract changes from the price of the March to the April contract when the March contract expires. (To conserve space certain active contracts were omitted).

The principal problem with the rolling nearby method is that it may create discontinuous price series when the underlying contract changes: for instance, from February 23 (the March contract) to February 24 (the April contract) in the example above. This discontinuity usually is the largest for very short term contracts and when the term structure of prices is steep.

8.1.4.3 Interpolated futures prices

To address the issue of discontinuous price series, we use the simple rule of linear interpolation to define constant maturity futures prices, P_{cmf}, from quoted futures prices:

[8.3] $$P_{cmf} = \omega_{NB1}P_{NB1} + \omega_{NB2}P_{NB2}$$

where

P_{cmf} = constant maturity futures prices

$\omega_{NB1} = \frac{\delta}{\Delta}$ = ratio of P_{cmf} made up by P_{NB1}

δ = days to expiration of $NB1$

Δ = days to expiration of constant maturity contract

P_{NB1} = price of $NB1$

$\omega_{NB2} = 1 - \omega_{NB1}$

= ratio of P_{cmf} made up by P_{NB2}

P_{NB2} = price of $NB2$

$NB1$ = nearby contract with a maturity < constant maturity contract

$NB2$ = first contract with a maturity < constant maturity contract

The following example illustrates this method using the data for the heating oil futures contract. On April 26, 1994 the 1-month constant maturity equivalent heating oil price is calculated as follows:

[8.4]
$$\begin{aligned} P_{1m,\ \text{April 26}} &= \left[\left(\frac{1\text{day}}{30\text{ days}}\right) \times Price_{\text{April}}\right] + \left[\left(\frac{29\text{ days}}{30\text{ days}}\right) \times Price_{\text{May}}\right] \\ &= \left[\left(\frac{1}{30}\right) \times 47.37\right] + \left[\left(\frac{29}{39}\right) \times 47.38\right] \\ &= 47.379 \end{aligned}$$

Table 8.2 illustrates the calculation over successive days. Note that the actual results may vary slightly from the data represented in the table because of numerical rounding.

Table 8.2
Price calculation for 1-month CMF NY Harbor #2 Heating Oil

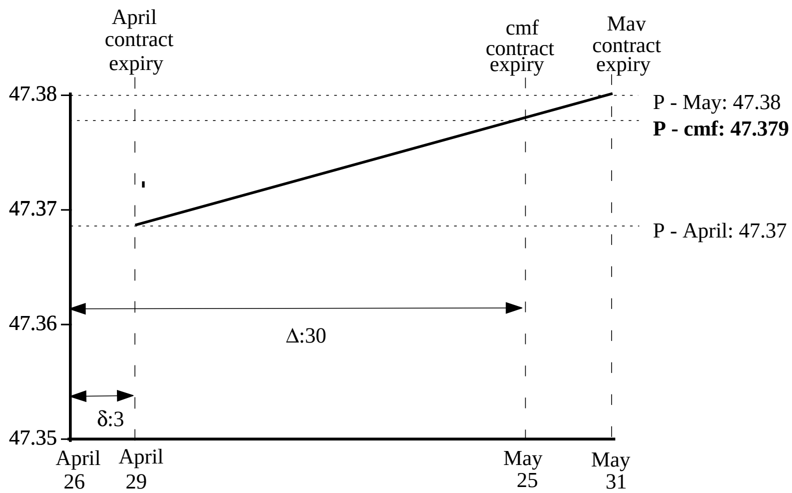
	Contract expiration			**Days to expiration**			**Weights (%)**		**Contract prices**			**cmf†**
Date	1 nb*	1m cmf†	2 nb*	1 nb*	1m cmf	2 nb*	1 nb*	2 nb*	Apr	May	Jun	
22-Apr-94	29-Apr	23-May	31-May	7	30	39	23.33	76.67	47.87	47.86	48.15	47.862
25-Apr-94	29-Apr	25-May	31-May	4	30	36	13.33	86.67	48.23	48.18	48.48	48.187
26-Apr-94	29-Apr	26-May	31-May	3	30	35	10.00	90.00	47.37	47.38	47.78	47.379
28-Apr-94	29-Apr	30-May	31-May	1	30	33	3.33	96.67	46.52	46.57	47.02	47.005
29-Apr-94	29-Apr	31-May	31-May	0	30	32	0.00	100.00	47.05	47.09	47.49	47.490
2-May-94	31-May	1-Jun	30-Jun	29	30	59	96.67	3.33	—	47.57	47.95	47.583
3-May-94	31-May	2-Jun	30-Jun	28	30	58	93.33	6.67	—	46.89	47.29	46.917
4-May-94	31-May	3-Jun	30-Jun	27	30	57	90.00	10.00	—	46.66	47.03	46.697

* 1 nb and 2 nb indicate first and second nearby contracts, respectively.

† cmf means constant maturity future.

Chart 8.1 illustrates the linear interpolation rule graphically.

Chart 8.1
Constant maturity future: price calculation



8.2 Filling in missing data

The preceding section described the types of rates and prices that RiskMetrics uses in its calculations. Throughout the presentation it was implicitly assumed that there were no missing prices. In practice, however, this is often not the case. Because of market closings in a specific location, daily prices are occasionally unavailable for individual instruments and countries. Reasons for the missing data include the occurrence of significant political or social events and technical problems (e.g., machine down time).

Very often, missing data are simply replaced by the preceding day's value. This is frequently the case in the data obtained from specialized vendors. Another common practice has simply been to exclude an entire date from which data were missing from the sample. This results in valuable data being discarded. Simply because one market is closed on a given day should not imply that data from the other countries are not useful. A large number of nonconcurrent missing data points across markets may reduce the validity of a risk measurement process.

Accurately replacing missing data is paramount in obtaining reasonable estimates of volatility and correlation. In this section we describe how missing data points are "filled-in"—by a process known as the EM algorithm—so that we can undertake the analysis set forth in this document. In brief, RiskMetrics applies the following steps to fill in missing rates and prices:

- Assume at any point in time that a data set consisting of a cross-section of returns (that may contain missing data) are multivariate normally distributed with mean μ and covariance matrix Σ.

- Estimate the mean and covariance matrix of this data set using the available, observed data.

- Replace the missing data points by their respective conditional expectations, i.e., use the missing data's expected values given current estimates of μ, Σ and the observed data.

8.2.1 Nature of missing data

We assume throughout the analysis that the presence of missing data occur randomly. Suppose that at a particular point in time, we have K return series and for each of the series we have T historical observations. Let **Z** denote the matrix of raw, observed returns. **Z** has T rows and K columns. Each row of **Z** is a Kx1 vector of returns, observed at any point in time, spanning all K securities. Denote the tth row of **Z** by $\mathbf{z}_t$ for t = 1,2,...T. The matrix **Z** may have missing data points.

Define a complete data matrix **R** that consists of all the data points **Z** plus the "filled-in" returns for the missing observations. The tth row of **R** is denoted $\mathbf{r}_t$. Note that if there are no missing observations then $\mathbf{z}_t=\mathbf{r}_t$ for all t=1,...,T. In the case where we have two assets (K=2) and three historical observations (T=3) on each asset, **R** takes the form:

[8.5]
$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \\ r_{31} & r_{32} \end{bmatrix} = \begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix}$$

where "T" denotes transpose.

8.2.2 Maximum likelihood estimation

For the purpose of filling in missing data it is assumed that at any period t, the return vector $\mathbf{r}_t$ (Kx1) follows a multivariate normal distribution with mean vector μ and covariance matrix Σ. The probability density function of $\mathbf{r}_t$ is

[8.6]
$$p(r_t) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{k}{2}} \exp\left[-\frac{1}{2}(r_t - \mu)^T \Sigma^{-1} (r_t - \mu)\right]$$

It is assumed that this density function holds for all time periods, t = 1,2,...,T. Next, under the assumption of statistical independence between time periods, we can write the joint probability density function of returns given the mean and covariance matrix as follows

[8.7]
$$\begin{aligned} p(r_1, \ldots r_T | \mu, \Sigma) &= \prod_{t=1}^{T} p(r_t) \\ &= (2\pi)^{-\frac{Tk}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(r_t - \mu)^T \Sigma^{-1} (r_t - \mu)\right] \end{aligned}$$

The joint probability density function $p(r_1, \ldots r_T | \mu, \Sigma)$ describes the probability density for the data **given** a set of parameter values (i.e., μ and Σ). Define the total parameter vector $\theta = (\mu,\Sigma)$. Our task is to estimate θ given the data matrix that contains missing data. To do so, we must derive the likelihood function of θ given the data. The likelihood function $|L(\mu, \Sigma | r_1, \ldots r_T)$ is similar in all respects to $p(r_1, \ldots r_T | \mu, \Sigma)$ except that it considers the parameters as random variables and takes the data as given. Mathematically, the likelihood function is equivalent to the probability density function. Intuitively, the likelihood function embodies the entire set of parameter values for an observed data set.

Now, for a realized sample of, say, exchange rates, we would want to know what set of parameter values most likely generated the observed data set. The solution to this question lies in maximum likelihood estimation. In essence, **the maximum likelihood estimates (*MLE*) θ_{MLE} are the parameter values that most likely generated the observed data matrix.**

θ_{MLE} is found by maximizing the likelihood function $L(\mu, \Sigma | r_1, \ldots r_T)$. In practice it is often easier to maximize natural logarithm of the likelihood function $l(\mu, \Sigma | r_1, \ldots r_T)$ which is given by

[8.8] $$l(\mu, \Sigma | R) = -\frac{1}{2} TK \ln(2\pi) - \frac{T}{2} \ln|\Sigma| - \frac{1}{2} \sum_{t=1}^{T} (r_t - \mu)^T \Sigma^{-1} (r_t - \mu)$$

with respect to θ. This translates into finding solutions to the following first order conditions:

[8.9] $$\frac{\partial}{\partial \mu} l(\mu, \Sigma | r_1, \ldots r_T) = 0, \quad \frac{\partial}{\partial \Sigma} l(\mu, \Sigma | r_1, \ldots r_T) = 0$$

The maximum likelihood estimators for the mean vector, $\hat{\mu}$ and covariance matrix $\hat{\Sigma}$ are

[8.10] $$\hat{\mu} = [\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_k]^T$$

[8.11] $$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (r_t - \hat{\mu})(r_t - \hat{\mu})^T$$

where $\bar{r}_i$ represents the sample mean taken over T time periods.

8.2.3 Estimating the sample mean and covariance matrix for missing data

When some observations of $\mathbf{r}_t$ are missing, the maximum likelihood estimates θ_{MLE} are not available. This is evident from the fact that the likelihood function is not defined (i.e., it has no value) when it is evaluated at the missing data points. To overcome this problem, we must implement what is known as the EM algorithm.

Since its formal exposition (Dempster, Laird and Rubin, 1977) the expectation maximization or EM algorithm (hereafter referred to as EM) has been on of the most successful methods of estimation when the data under study are incomplete (e.g., when some of the observations are missing). Among its extensive applications, the EM algorithm has been used to resolve missing data problems involving financial time series (Warga, 1992). For a detailed exposition of the EM algorithm and its application in finance see Kempthorne and Vyas (1994).

Intuitively, EM is an iterative algorithm that operates as follows.

- For a given set of (initial) parameter values, instead of evaluating the log likelihood function, (which is impossible, anyway) EM evaluates the conditional expectation of the latent (underlying) log likelihood function. The mathematical conditional expectation of the log-likelihood is taken over the observed data points.

- The expected log likelihood is maximized to yield parameter estimates θ^0_{EM}. (The superscript "0" stands for the initial parameter estimate). This value is then substituted into the log likelihood function and expectations are taken again, and new parameter estimates θ^1_{EM} are found. This iterative process is continued until the algorithm converges at which time final parameter estimates have been generated. For example, if the algorithm is iterated N+1 times then the sequence of parameter estimates $\left(\theta^0_{EM}, \theta^1_{EM}, \ldots, \theta^N_{EM}\right)$ is generated. The algorithm stops when adjacent parameter estimates are sufficiently close to one another, i.e., when θ_{EM}^{N-1} is sufficiently close to θ_{EM}^{N}.

The first step in EM is referred to as the expectation or E-Step. The second step is the maximization or M-step. EM iterates between these two steps, updating the E-Step from the parameter estimates generated in the M-Step. For example, at the *i*th iteration of the algorithm, the following equations are solved in the M-Step:

[8.12a] $$\hat{\mu}^{i+1} = \frac{1}{T}\sum_{t=1}^{T} E\left[r_t \middle| z_t, \theta^i\right] \quad \text{(the sample mean)}$$

[8.12b] $$\hat{\Sigma}^{i+1} = \frac{1}{T}\sum_{t=1}^{T} E\left[r_t r_t^T \middle| z_t, \theta^i\right] - \hat{\mu}^{i+1}\left(\hat{\mu}^{i+1}\right)^T \quad \text{(the sample covariance matrix)}$$

To evaluate the expectations in these expressions ($E\left[r_t \middle| z_t, \theta^i\right]$ and $E\left[r_t r_t^T \middle| z_t, \theta^i\right]$), we make use of standard properties for partitioning a multivariate normal random vector.

[8.13] $$\begin{bmatrix} R \\ \Re \end{bmatrix} \sim NID\left[\begin{bmatrix} \mu_R \\ \mu_\Re \end{bmatrix}, \begin{bmatrix} \Sigma_{RR} & \Sigma_{\Re R} \\ \Sigma_{R\Re} & \Sigma_{\Re\Re} \end{bmatrix}\right]$$

Here, one can think of $\Re$ as the sample data with missing values removed and R as the vector of the underlying complete set of observations. Assuming that returns are distributed multivariate normal, the distribution of R conditional on $\Re$ is multivariate normal with mean

[8.14] $$E[R|\Re] = \mu_R + \Sigma_{R\Re}\Sigma_{\Re\Re}^{-1}(\Re - \mu_\Re)$$

and covariance matrix

[8.15] $$\text{Covariance}(R|\Re) = \Sigma_{RR} - \Sigma_{R\Re}\Sigma_{\Re\Re}^{-1}\Sigma_{\Re R}$$

Using Eq. [8.14] and Eq. [8.15] we can evaluate the E- and M- steps. The E -Step is given by

[8.16] $$E-Step\begin{bmatrix} E[r_t|z_t,\theta] = \mu_r + \Sigma_{rz}\Sigma_{zz}^{-1}(z_t - \mu_z) \\ E\left[r_t r_t^T \middle| z_t, \theta\right] = \text{Covariance}\left(r_t^T \middle| z_t, \theta\right) + \left(E[r_t|z_t,\theta]\,E[r_t|z_t,\theta]^T\right) \end{bmatrix}$$

where

[8.17] $$\text{Covariance}\left(r_t^T \middle| z_t, \theta\right) = \Sigma_{rz} - \Sigma_{rz}\Sigma_{zz}^{-1}\Sigma_{zr}$$

Notice that the expressions in Eq. [8.17] are easily evaluated since they depend on parameters that describe the observed and missing data.

Given the values computed in the E-Step, the M-Step yields updates of the mean vector and covariance matrix.

[8.18] M-Step
$$\begin{cases} \hat{\mu}^{i+1} = \dfrac{1}{T}\left(\displaystyle\sum_{\text{Complete}\,t} r_t + \sum_{\text{Incomplete}\,t} E\left[r_t \middle| z_t, \theta^i\right]\right) \\ \hat{\Sigma}^{i+1} = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} E\left[r_t r_t^T \middle| z_t, \theta^i\right] - \hat{\mu}^{i+1}\left(\hat{\mu}^{i+1}\right)^T \\ \quad = \dfrac{1}{T}\left(\begin{array}{l}\left(\displaystyle\sum_{\text{Complete}\,t} r_t r_t^T\right) + \\ \displaystyle\sum_{\text{Incomplete}\,t} \text{Covariance}\left(r_t^T \middle| z_t, \theta\right) + E\,[r_t | z_t, \theta]\, E\,[r_t | z_t, \theta]^{\,T}\end{array}\right) \end{cases}$$

Notice that summing over t implies that we are adding "down" the columns of the data matrix **R**. For a practical, detailed example of the EM algorithm see Johnson and Wichern (1992, pp. 203–206).

A powerful result of EM is that when a global optimum exists, the parameter estimates from the EM algorithm converge to the ML estimates. That is, for a sufficiently large number of iterations, EM converges to θ_{MLE}. Thus, the EM algorithm provides a way to calculate the ML estimates of the unknown parameter even if all of the observations are not available.

The assumption that the time series are generated from a multivariate normal distribution is innocuous. Even if the true underlying distribution is not normal, it follows from the theory of pseudo-maximum likelihood estimation that the parameter estimates are asymptotically consistent (White, 1982) although not necessarily asymptotically efficient. That is, it has been shown that the pseudo-MLE obtained by maximizing the unspecified log likelihood as if it were correct produces a consistent estimator despite the misspecification.

8.2.4 An illustrative example

A typical application of the EM algorithm is filling in missing values resulting from a holiday in a given market. We applied the algorithm outlined in the section above to the August 15 Assumption holiday in the Belgian government bond market. While most European bond markets were open on that date, including Germany and the Netherlands which show significant correlation with Belgium, no data was available for Belgium.

A missing data point in an underlying time series generates two missing points in the log change series as shown below (from t–1 to t as well as from t to t + 1). Even though it would be more straightforward to calculate the underlying missing value through the EM algorithm and then generate the two missing log changes, this would be statistically inconsistent with our basic assumptions on the distribution of data.

In order to maintain consistency between the underlying rate data and the return series, the adjustment for missing data is performed in three steps.

1. First the EM algorithm generates the first missing percentage change, or –0.419% in the example below.
2. From that number, we can back out the missing underlying yield from the previous day's level, which gives us the 8.445% in the example below.
3. Finally, the second missing log change can be calculated from the revised underlying yield series.

Table 8.3 presents the underlying rates on the Belgian franc 10-year zero coupon bond, the corresponding EM forecast, and the adjusted “filled-in” rates and returns.

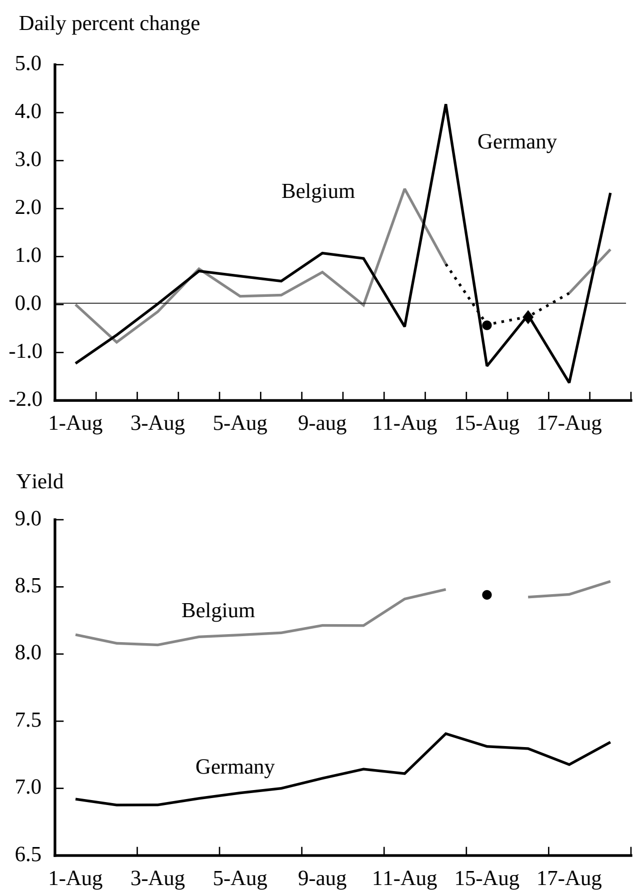
Table 8.3
Belgian franc 10-year zero coupon rate
application of the EM algorithm to the 1994 Assumption holiday in Belgium

	Observed			Adjusted	
Collection date	**10-year rate**	**Return (%)**	**EM forecast**	**10-year rate**	**Return (%)**
11-Aug-94	8.400	2.411		8.410	2.411
12-Aug-94	8.481	0.844		8.481	0.844
15-Aug-94	missing	missing	–0.419	8.445*	–0.419†
16-Aug-94	8.424	missing		8.424	–0.254‡
17-Aug-94	8.444	0.237		8.444	0.237
18-Aug-94	8.541	1.149		8.541	1.149

* Filled-in rate based on EM forecast.
† From EM.
‡ Return now available because prior rate (*) has been filled in.

Chart 8.2 presents a time series of the Belgian franc 10-year rate before and after the missing observation was filled in by the EM algorithm.

Chart 8.2
Graphical representation
10-year zero coupon rates; daily % change

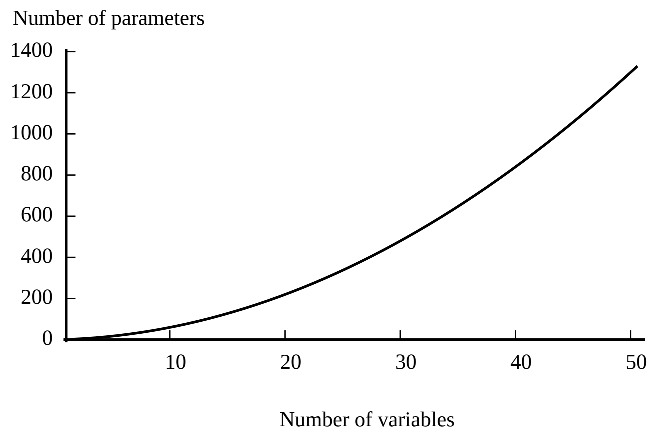


8.2.5 Practical considerations

A major part of implementing the EM algorithm is to devise the appropriate input data matrices for the EM. From both a statistical and practical perspective we do not run EM on our entire time series data set simultaneously. Instead we must partition the original data series into non-overlapping sub-matrices. Our reasons for doing so are highlighted in the following example.

Consider a TxK data matrix where T is the number of observations and K is the number of price vectors. Given this data matrix, the EM must estimate K+K(K+1)/2 parameters. Consequently, to keep the estimation practical K cannot be too large. To get a better understanding of this issue consider Chart 8.3, which plots the number of parameters estimated by EM (K +K(K+1)/2) against the number of variables. As shown, the number of estimated parameters grows rapidly with the number of variables.

Chart 8.3
Number of variables used in EM and parameters required
number of parameters (Y-axis) versus number of variables (X-axis)



The submatrices must be chosen so that vectors within a particular submatrix are highly correlated while those vectors between submatrices are not significantly correlated. If we are allowed to choose the submatrices in this way then EM will perform as if it had the entire original data matrix. This follows from the fact that the accuracy of parameter estimates are not improved by adding uncorrelated vectors.

In order to achieve a logical choice of submatrices, we classify returns into the following categories: (1) foreign exchange, (2) money market, (3) swap, (4) government bond, (5) equity, and (6) commodity.

We further decompose categories 2, 3, 4, and 6 as follows. Each input data matrix corresponds to a particular country or commodity market. The rows of this matrix correspond to time while the columns identify the maturity of the asset. Foreign exchange, equity indices, and bullion are the exceptions: all exchange rates, equity indices, and bullion are grouped into three separate matrices.

8.3 The properties of correlation (covariance) matrices and VaR

In Section 6.3.2 it was shown how RiskMetrics applies a correlation matrix to compute the VaR of an arbitrary portfolio. In particular, the correlation matrix was used to compute the portfolio's standard deviation. VaR was then computed as a multiple of that standard deviation. In this section we investigate the properties of a generic correlation matrix since it is these properties that will determine whether the portfolio's standard deviation forecast is meaningful.[2] Specifically, we will establish conditions[3] that guarantee the non-negativity of the portfolio's variance, i.e., $\sigma^2 \geq 0$.

At first glance it may not seem obvious why it is necessary to understand the conditions under which the variance is non-negative. However, the potential sign of the variance, and consequently the VaR number, is directly related to the relationship between (1) the number of individual price return series (i.e., cashflows) per portfolio and (2) the number of historical observations on each of these return series. In practice there is often a trade-off between the two since, on the one hand, large portfolios require the use of many time series, while on the other hand, large amounts of historical data are not available for many time series.

Below, we establish conditions that ensure the non-negativity of a variance that is constructed from correlation matrices based on equally and exponentially weighted schemes. We begin with some basic definitions of covariance and correlation matrices.

8.3.1 Covariance and correlation calculations

In this section we briefly review the covariance and correlation calculations based on equal and exponential moving averages. We do so in order to establish a relationship between the underlying return data matrix and the properties of the corresponding covariance (correlation) matrix.

8.3.1.1 Equal weighting scheme

Let X denote a T x K data matrix, i.e., matrix of returns. X has T rows and K columns.

[8.19]
$$X = \begin{bmatrix} r_{11} & \cdots & \cdots & \cdots & r_{1K} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & r_{JJ} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{T1} & \cdots & \cdots & \cdots & r_{TK} \end{bmatrix}$$

Each column of X is a return series corresponding to a particular price/rate (e.g., USD/DEM FX return) while each row corresponds to the time ($t = 1,...,T$) at which the return was recorded. If we compute standard deviations and covariances around a zero mean, and weigh each observation with probability $1/T$, we can define the covariance matrix simply by

[8.20]
$$\Sigma = \frac{X^T X}{T}$$

where X^T is the transpose of X.

Consider an example when $T = 4$ and $K = 2$.

[2] By properties, we mean specifically whether the correlation matrix is positive definite, positive semidefinite or otherwise (these terms will be defined explicitly below)

[3] All linear algebra propositions stated below can be found in Johnston, J. (1984).

[8.21] $$X = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \\ r_{31} & r_{32} \\ r_{41} & r_{42} \end{bmatrix} \quad X^T = \begin{bmatrix} r_{11} & r_{21} & r_{31} & r_{41} \\ r_{12} & r_{22} & r_{32} & r_{42} \end{bmatrix}$$

An estimate of the covariance matrix is given by

[8.22] $$\Sigma = \frac{X^T X}{T} = \begin{bmatrix} \frac{1}{4}\sum_{i=1}^{4} r_{i1}^2 & \frac{1}{4}\sum_{i=1}^{4} r_{i1} r_{i2} \\ \frac{1}{4}\sum_{i=1}^{4} r_{i1} r_{i2} & \frac{1}{4}\sum_{i=1}^{4} r_{i2}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{bmatrix}$$

Next, we show how to compute the correlation matrix R. Suppose we divide each element of the matrix *X* by the standard deviation of the series to which it belongs; i.e., we normalize each series of *X* to have a standard deviation of 1. Call this new matrix with the standardized values *Y.*

The correlation matrix is

[8.23] $$Y = \begin{bmatrix} \frac{r_{11}}{\sigma_1} & \dots & \dots & \dots & \frac{r_{1K}}{\sigma_K} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \frac{r_{JJ}}{\sigma_J} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \frac{r_{T1}}{\sigma_1} & \dots & \dots & \dots & \frac{r_{TK}}{\sigma_K} \end{bmatrix}$$

where

$$\sigma_j = \frac{1}{T}\sqrt{\sum_{i=1}^{T} r^2{}_{ij}} \qquad j = 1, 2, \dots k$$

[8.24] $$R = \frac{Y^T Y}{T}$$

As in the previous example, if we set $T = 4$ and $K = 2$, the correlation matrix is

[8.25] $$R = \frac{Y^T Y}{T} = \begin{bmatrix} \frac{1}{4}\sum_{i=1}^{4} \frac{r_{i1}^2}{\sigma_1} & \frac{1}{4}\sum_{i=1}^{4} \frac{r_{i1} r_{i2}}{\sigma_1 \sigma_2} \\ \frac{1}{4}\sum_{i=1}^{4} \frac{r_{i1} r_{i2}}{\sigma_1 \sigma_2} & \frac{1}{4}\sum_{i=1}^{4} \frac{r_{i2}^2}{\sigma_2} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}$$

8.3.1.2 Exponential weighting scheme

We now show how similar results are obtained by using exponential weighting rather than equal weighting. When computing the covariance and correlation matrices, use, instead of the data matrix X, the augmented data matrix $\tilde{X}$ shown in Eq. [8.26].

$$[8.26] \qquad \tilde{X} = \begin{bmatrix} r_{11} & \dots & \dots & \dots & r_{1K} \\ \sqrt{\lambda} r_{21} & \dots & \dots & \dots & \sqrt{\lambda} r_{21} \\ \dots & \dots & \sqrt{\lambda^{J-1}} r_{JJ} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \sqrt{\lambda^{T-1}} r_{T1} & \dots & \dots & \dots & \sqrt{\lambda^{T-1}} r_{TK} \end{bmatrix}$$

Now, we can define the covariance matrix simply as

$$[8.27] \qquad \tilde{\Sigma} = \left(\sum_{i=1}^{T} \lambda^{i-1} \right)^{-1} \cdot \tilde{X}^T \tilde{X}$$

To see this, consider the example when $T = 4$ and $K = 2$.

$$X = \begin{bmatrix} r_{11} & r_{12} \\ \sqrt{\lambda} r_{21} & \sqrt{\lambda} r_{22} \\ \sqrt{\lambda^2} r_{31} & \sqrt{\lambda^2} r_{32} \\ \sqrt{\lambda^3} r_{41} & \sqrt{\lambda^3} r_{42} \end{bmatrix} \quad X^T = \begin{bmatrix} x_{11} & \sqrt{\lambda} r_{21} & \sqrt{\lambda^2} r_{31} & \sqrt{\lambda^3} r_{41} \\ x_{12} & \sqrt{\lambda} r_{22} & \sqrt{\lambda^2} r_{32} & \sqrt{\lambda^3} r_{42} \end{bmatrix}$$

$$[8.28] \qquad \tilde{\Sigma} = \left(\sum_{i=1}^{T} \lambda^{i-1} \right)^{-1} \tilde{X}^T \tilde{X}$$

$$= \left(\sum_{i=1}^{T} \lambda^{i-1} \right)^{-1} \begin{bmatrix} \sum_{i=1}^{4} \lambda^{i-1} r_{i1}^2 & \sum_{i=1}^{4} \lambda^{i-1} r_{i1} r_{i2} \\ \sum_{i=1}^{4} \lambda^{i-1} r_{i1} r_{i2} & \sum_{i=1}^{4} \lambda^{i-1} r_{i2}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{bmatrix}$$

The exponentially weighted correlation matrix is computed just like the simple correlation matrix. The standardized data matrix and the correlation matrix are given by the following expressions.

$$[8.29] \qquad \tilde{Y} = \begin{bmatrix} \frac{\tilde{r}_{11}}{\sigma_1} & \dots & \dots & \dots & \frac{\tilde{r}_{1K}}{\sigma_K} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \frac{\tilde{r}_{JJ}}{\sigma_J} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\tilde{r}_{T1}}{\sigma_1} & \dots & \dots & \dots & \frac{\tilde{r}_{TK}}{\sigma_K} \end{bmatrix}$$

where

$$\sigma_j = \left(\sum_{i=1}^{T} \lambda^{i-1}\right)^{-1} \sqrt{\sum_{i=1}^{T} \lambda^{i-1} \tilde{r}_{ij}^2} \qquad j = 1, 2, \ldots K$$

and the correlation matrix is

[8.30] $$\tilde{R} = \left(\sum_{i=1}^{T} \lambda^{i-1}\right)^{-1} \cdot \tilde{Y}^T \tilde{Y}$$

which is the exact analogue to Eq. [8.25]. Therefore, all results for the simple correlation matrix carry over to the exponential weighted matrix.

Having shown how to compute the covariance and correlation matrices, the next step is to show how the properties of these matrices relate to the VaR calculations.

We begin with the definition of positive definite and positive semidefinite matrices.

[8.31] If $\vec{z}^T C \vec{z} > (<) \ 0$ for all nonzero vectors $\vec{z}$, then C is said to be positive (negative) definite.

[8.32] If $\vec{z}^T C \vec{z} \ \geq \ (\leq) \ 0$ for all nonzero vectors $\vec{z}$, then C is said to be positive semidefinite (nonpositive definite).

Now, referring back to the VaR calculation presented in Section 6.3.2, if we replace the vector $\vec{z}$ by the weight vector $\vec{\sigma}_{t|t-1}$ and C by the correlation matrix, $R_{t|t-1}$, then it should be obvious why we seek to determine whether the correlation matrix is positive definite or not. Specifically,

- If the correlation matrix R is positive definite, then VaR will always be positive.
- If R is positive semidefinite, then VaR could be zero or positive.
- If R is negative definite,[4] then VaR will be negative.

8.3.2 Useful linear algebra results as applied to the VaR calculation

In order to define a relationship between the dimensions of the data matrix X (or $\tilde{X}$) (i.e., the number of rows and columns of the data matrix) and the potential values of the VaR estimates, we must define the rank of X.

The rank of a matrix X, denoted r(X), is the maximum number of linearly independent rows (and columns) of that matrix. The rank of a matrix can be no greater than the minimum number of rows or columns. Therefore, if X is T x K with T > K (i.e., more rows than columns) then r(X) $\leq$ K. In general, for an T x K matrix X, r(X) $\leq$ min(T,K).

[4] We will show below that this is not possible.

A useful result which equates the ranks of different matrices is:

[8.33] $$r(X) = r\left(X^T X\right) = r\left(X X^T\right)$$

As applied to the VaR calculation, the rank of the covariance matrix $\Sigma = X^T X$ is the same as the rank of X.

We now refer to two linear algebra results which establish a relationship between the rank of the data matrix and the range of VaR values.

[8.34] If X is T x K with rank K < T, then $X^T X$ is positive definite and XX^T is positive semidefinite.

[8.35] If X is T x K with rank J < min(T,K) then $X^T X$ and XX^T is positive semidefinite.

Therefore, whether Σ is positive definite or not will depend on the rank of the data matrix X.

Based on the previous discussion, we can provide the following results for RiskMetrics VaR calculations.

- Following from Eq. [8.33], we can deduce the rank of R simply by knowing the rank of Y, the standardized data matrix.
- The rank of the correlation matrix R can be no greater than the number of historical data points used to compute the correlation matrix, and
- Following from Eq. [8.34], if the data matrix of returns has more rows than columns and the columns are independent, then R is positive definite and VaR > 0. If not, then Eq. [8.35] applies, and R is positive semidefinite and $\text{VaR} \geq 0$.

In summary, a covariance matrix, by definition, is **at least** positive semidefinite. Simply put, positive semidefinite is the multi-dimensional analogue to the definition, $\sigma^2 \geq 0$.

8.3.3 How to determine if a covariance matrix is positive semi-definite[5]

Finally, we explain a technique to determine whether a correlation matrix is positive (semi) definite. We would like to note at the beginning that due to a variety of technical issues that are beyond the scope of this document, the suggested approach described below known as the singular value decomposition (SVD) is to serve as a general guideline rather than a strict set of rules for determining the "definiteness" of a correlation matrix.

The singular value decomposition (SVD)

The T x K standardized data matrix Y ($T \geq K$) may be decomposed as[6] $Y = UDV'$ where $U'U = V'V = I_K$ and D is diagonal with non-negative diagonal elements $(\iota_1, \iota_2, \ldots, \iota_K)$, called the singular values of Y. All of the singular values are $\geq (0)$.

[5] This section is based on Belsley (1981), Chapter 3.

[6] In this section we work with the mean centered and standardized matrix Y instead of X since Y is the data matrix on which an SVD should be applied.

A useful result is that the number of non-zero singular values is a function by the rank of Y. Specifically, if Y is full rank, then all K singular values will be non zero. If the rank of Y is J=K-2, then there will be J positive singular values and two zero singular values.

In practice, it is difficult to determine the number of zero singular values. This is due to that fact that computers deal with finite, not exact arithmetic. In other words, it is difficult for a computer to know when a singular value is **really** zero. To avoid having to determine the number of zero singular values, it is recommended that practitioners should focus on the condition number of Y which is the ratio of the largest to smallest singular values, i.e.,

[8.36] $$\upsilon = \frac{\iota_{max}}{\iota_{min}} \text{ (condition number)}$$

Large condition numbers point toward 'ill-condition' matrices, i.e., matrices that are nearly not full rank. In other words, a large υ implies that there is a strong degree of collinearity between the columns of Y. More elaborate tests of collinearity can be found in Belsley (1981).

We now apply the SVD to two data matrices. The first data matrix consists of time series of price returns on 10 USD government bonds for the period January 4, 1993–October 14, 1996 (986 observations). The columns of the data matrix correspond to the price returns on the 2yr, 3yr, 4yr, 5yr, 7yr, 9yr, 10yr, 15yr, 20yr, and 30yr USD government bonds. The singular values for this data matrix are given in Table 8.4.

Table 8.4
Singular values for USD yield curve data matrix

3.045	0.051
0.785	0.043
0.271	0.020
0.131	0.017
0.117	0.006

The condition number, υ, is 497.4. We conduct a similar experiment on a data matrix that consists of 14 equity indices.[7] The singular values are shown in Table 8.5. The data set consists of a total number of 790 observations for the period October 5, 1996 through October 14, 1996.

Table 8.5
Singular values for equity indices returns

2.329	0.873	0.696
1.149	0.855	0.639
0.948	0.789	0.553
0.936	0.743	0.554
0.894	0.712	

For this data matrix, the condition number, υ, is 4.28. Notice how much lower the condition number is for equities than it is for the US yield curve. This result should not be surprising since we expect the returns on different bonds along the yield curve to move in a similar fashion to one another relative to equity returns. Alternatively expressed, the relatively large condition number for the USD yield curve is indicative of the near collinearity that exists among returns on US government bonds.

[7] For the countries Austria, Australia, Belgium, Canada, Switzerland, Spain, France, Finland, Great Britain, Hong Kong, Ireland, Italy, Japan and the Netherlands.

The purpose of the preceding exercise was to demonstrate how the interrelatedness of individual time series affects the condition of the resulting correlation matrix. As we have shown with a simple example, highly correlated data (USD yield curve data) leads to high condition numbers relative to less correlated data (equity indices).

In concluding, due to numerical rounding errors it is not unlikely for the theoretical properties of a matrix to differ from its estimated counterpart. For example, covariance matrices are real, symmetric and non-positive definite. However, when estimating a covariance matrix we may find that the positive definite property is violated. More specifically, the matrix may not invert. Singularity may arise because certain prices included in a covariance matrix form linear combinations of other prices. Therefore, if covariance matrices fail to invert they should be checked to determine whether certain prices are linear functions of others. Also, the scale of the matrix elements may be such that it will not invert. While poor scaling may be a source of problems, it should rarely be the case.

8.4 Rebasing RiskMetrics volatilities and correlations

A user's base currency will dictate how RiskMetrics standard deviations and correlations will be used. For example, a DEM-based investor with US dollar exposure is interested in fluctuations in the currency USD/DEM whereas the same investor with an exposure in Belgium francs is interested in fluctuations in BEF/DEM. Currently, RiskMetrics volatility forecasts are expressed in US dollars per foreign currency such as USD/DEM for all currencies. To compute volatilities on cross rates such as BEF/DEM, users must make use of the RiskMetrics provided USD/DEM and USD/BEF volatilities as well as correlations between the two. We now show how to derive the variance (standard deviation) of the BEF/DEM position. Let $r_{1,t}$ and $r_{2,t}$ represent the time t returns on USD/DEM and USD/BEF, respectively, i.e.,

[8.37] $$r_{1t} = \ln\left[\frac{(USD/DEM)_t}{(USD/DEM)_{t-1}}\right] \text{ and } r_{2t} = \ln\left[\frac{(USD/BEF)_t}{(USD/BEF)_{t-1}}\right]$$

The cross rate BEF/DEM is defined as

[8.38] $$r_{3t} = \ln\left[\frac{(BEF/DEM)_t}{(BEF/DEM)_{t-1}}\right] = r_{1t} - r_{2t}$$

The variance of the cross rate r_{3t} is given by

[8.39] $$\sigma^2_{3,t} = \sigma^2_{1,t} + \sigma^2_{2,t} - 2\sigma^2_{12,t}$$

Equation [8.39] holds for any cross rate that can be defined as the arithmetic difference in two other rates.

We can find the correlation between two cross rates as follows. Suppose we want to find the correlation between the currencies BEF/DEM and FRF/DEM. It follows from Eq. [8.38] that we first need to define these cross rates in terms of the returns used in RiskMetrics.

[8.40a] $$r_{1,t} = \ln\left[\frac{(USD/DEM)_t}{(USD/DEM)_{t-1}}\right], \qquad r_{2,t} = \ln\left[\frac{(USD/BEF)_t}{(USD/BEF)_{t-1}}\right],$$

[8.40b] $$r_{3,t} = \ln\left[\frac{(BEF/DEM)_t}{(BEF/DEM)_{t-1}}\right] = r_{1,t} - r_{2,t}, \qquad r_{4,t} = \ln\left[\frac{(USD/FRF)_t}{(USD/FRF)_{t-1}}\right]$$

and

[8.40c]
$$r_{5,t} = \ln\left[\frac{(FRF/DEM)_t}{(FRF/DEM)_{t-1}}\right] = r_{1,t} - r_{4,t}$$

The correlation between BEF/DEM and USD/FRF ($r_{3,t}$ and $r_{4,t}$) is the covariance of $r_{3,t}$ and $r_{4,t}$ divided by their respective standard deviations, mathematically,

[8.41]
$$\begin{aligned}\rho_{34,t} &= \frac{\sigma^2_{34,t}}{\sigma_{4,t}\sigma_{3,t}}\\ &= \frac{\sigma^2_{1,t} - \sigma^2_{12,t} - \sigma^2_{14,t} + \sigma^2_{24,t}}{\sqrt{\sigma^2_{1,t} + \sigma^2_{4,t} - 2\sigma^2_{14,t}}\ \sqrt{\sigma^2_{1,t} + \sigma^2_{2,t} - 2\sigma^2_{12,t}}}\end{aligned}$$

Analogously, the correlation between USD/DEM and FRF/DEM is

[8.42]
$$\begin{aligned}\rho_{35,t} &= \frac{\sigma^2_{15,t}}{\sigma_{5,t}\sigma_{1,t}}\\ &= \frac{\sigma^2_{1,t} - \sigma^2_{14,t}}{\sqrt{\sigma^2_{1,t} + \sigma^2_{4,t} - 2\sigma^2_{14,t}}\ \sqrt{\sigma^2_{1,t}}}\end{aligned}$$

8.5 Nonsynchronous data collection

Estimating how financial instruments move in relation to each other requires data that are collated, as much as possible, consistently across markets. The point in time when data are recorded is a material issue, particularly when estimating correlations. When data are observed (recorded) at different times they are known to be nonsynchronous.

Table 8.7 (pages 186–187) outlines how the data underlying the time series used by RiskMetrics are recorded during the day. It shows that most of the data are taken around 16:00 GMT. From the asset class perspective, we see that potential problems will most likely lie in statistics relating to the government bond and equity markets.

To demonstrate the effect of nonsynchronous data on correlation forecasts, we estimated the 1-year correlation of daily movements between USD 10-year zero yields collected every day at the close of business in N.Y. with two series of 3-month money market rates, one collected by the British Bankers Association at 11:00 a.m. in London and the other collected by J.P. Morgan at the close of business in London (4:00 p.m.). This data is presented in Table 8.6.

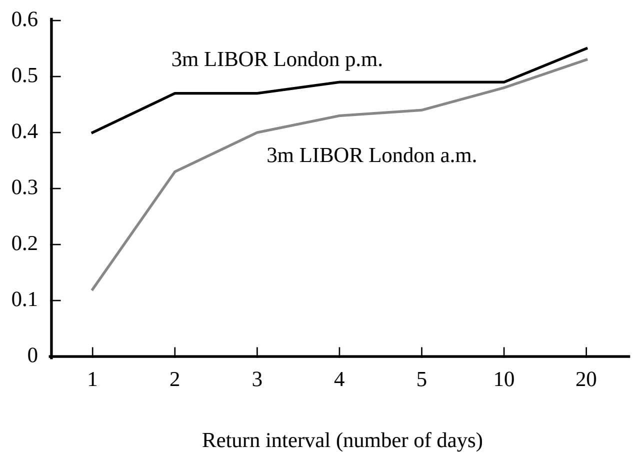
Table 8.6
Correlations of daily percentage changes with USD 10-year
August 1993 to June 1994 – 10-year USD rates collated at N.Y. close

	Correlation at London time:	
LIBOR	**11 a.m.**	**4 p.m.**
1-month	–0.012	0.153
3-month	0.123	0.396
6-month	0.119	0.386
12-month	0.118	0.622

None of the data series are synchronous, but the results show that the money market rates collected at the London close have higher correlation to the USD 10-year rates than those collected in the morning.

Getting a consistent view of how a particular yield curve behaves depends on addressing the timing issue correctly. While this is an important factor in measuring correlations, the effect of timing diminishes as the time horizon becomes longer. Correlating monthly percentage changes may not be dependent on the condition that rates be collected at the same time of day. Chart 8.4 shows how the correlation estimates against USD 10-year zeros evolve for the two money market series mentioned above when the horizon moves from daily changes to monthly changes. Once past the 10-day time interval, the effect of timing differences between the two series becomes negligible.

Chart 8.4
Correlation forecasts vs. return interval
3-month USD LIBOR vs. 10-year USD government bond zero rates



In a perfect world, all rates would be collected simultaneously as all markets would trade at the same time. One may be able to adapt to nonsynchronously recorded data by adjusting either the underlying return series or the forecasts that were computed from the nonsynchronous returns. In this context, data adjustment involves extensive research. The remaining sections of this document present an algorithm to adjust correlations when the data are nonsynchronous.

Table 8.7
Schedule of data collection

		London time, a.m.											
Country	**Instrument summary**	**1:00**	**2:00**	**3:00**	**4:00**	**5:00**	**6:00**	**7:00**	**8:00**	**9:00**	**10:00**	**11:00**	**12:00**
Australia	FX/Eq/LI/Sw/Gv						Eq	Gv					
Hong Kong	FX/Eq/LI/Sw			LI		Eq		Sw					
Indonesia	FX/Eq/LI/Sw								Eq	LI/Sw			
Japan	FX/Eq/LI/Sw/Gv						Gv	Eq					
Korea	FX/Eq								Eq				
Malaysia	FX/Eq/LI/Sw								Eq	LI/Sw			
New Zealand	FX/Eq/LI/Sw/Gv			Eq		LI/Gv	Sw						
Philippines	FX/Eq								Eq				
Singapore	FX/Eq/LI/Sw/Gv									LI/Eq			
Taiwan	FX/Eq/												
Thailand	FX/Eq/LI/Sw								Eq	LI/Sw			
Austria	FX/Eq/LI												Eq
Belgium	FX/Eq/LI/Sw/Gv												
Denmark	FX/Eq/LI/Sw/Gv												
Finland	FX/Eq/LI/Sw/Gv												
France	FX/Eq/LI/Sw/Gv												
Germany	FX/Eq/LI/Sw/Gv												
Ireland	FX/Eq/LI/Sw/Gv												
Italy	FX/Eq/LI/Sw/Gv												
Netherlands	FX/Eq/LI/Sw/Gv												
Norway	FX/Eq/LI/Sw/Gv												
Portugal	FX/Eq/LI/Sw/Gv												
South Africa	FX/Eq/LI//Gv												
Spain	FX/Eq/LI/Sw/Gv												
Sweden	FX/Eq/LI/Sw/Gv												
Switzerland	FX/Eq/LI/Sw/Gv												
U.K.	FX/Eq/LI/Sw/Gv												
ECU	FX/ /LI/Sw/Gv												
Argentina	FX/Eq												
Canada	FX/Eq/LI/Sw/Gv												
Mexico	FX/Eq/LI												
U.S.	FX/Eq/LI/Sw/Gv												

FX = Foreign Exchange, Eq = Equity Index, LI = LIBOR, Sw = Swap, Gv = Government

Table 8.7 (continued)
Schedule of data collection

London time, p.m.												Instrument summary	Country
1:00	**2:00**	**3:00**	**4:00**	**5:00**	**6:00**	**7:00**	**8:00**	**9:00**	**10:00**	**11:00**	**12:00**		
			FX/LI/Sw									FX/Eq/LI/Sw/Gv	Australia
			FX									FX/Eq/LI/Sw	Hong Kong
			FX									FX/Eq/LI/Sw	Indonesia
			FX/LI/Sw									FX/Eq/LI/Sw/Gv	Japan
			FX									FX/Eq	Korea
			FX									FX/Eq/LI/Sw	Malaysia
			FX									FX/Eq/LI/Sw/Gv	New Zealand
			FX									FX/Eq	Philippines
			FX									FX/Eq/LI/Sw/Gv	Singapore
			FX									FX/Eq	Taiwan
			FX									FX/Eq/LI/Sw	Thailand
			FX/LI									FX/Eq/LI	Austria
		Eq	FX/LI/Sw/Gv									FX/Eq/LI/Sw/Gv	Belgium
	Eq	Gv	FX/LI/Sw									FX/Eq/LI/Sw/Gv	Denmark
		Eq	FX/LI									FX/Eq/LI/Sw/Gv	Finland
		Gv	FX/LI/Sw/Eq									FX/Eq/LI/Sw/Gv	France
			FX/LI/Sw/Gv/Eq									FX/Eq/LI/Sw/Gv	Germany
			FX/LI/Sw/Gv	Eq								FX/Eq/LI/Sw/Gv	Ireland
			FX/LI/Sw/Gv/Eq									FX/Eq/LI/Sw/Gv	Italy
			FX/LI/Sw/Gv/Eq									FX/Eq/LI/Sw/Gv	Netherlands
	Eq		FX/LI									FX/Eq/LI/Sw/Gv	Norway
			FX/LI/Eq									FX/Eq/LI/Sw/Gv	Portugal
	Eq	Gv	FX/LI									FX/Eq/LI//Gv	South Africa
			FX/LI/Sw	Gv/Eq								FX/Eq/LI/Sw/Gv	Spain
		Gv	FX/LI/Sw/Eq									FX/Eq/LI/Sw/Gv	Sweden
			FX/LI/Sw/Eq									FX/Eq/LI/Sw/Gv	Switzerland
			FX/LI/Sw/Eq	Gv								FX/Eq/LI/Sw/Gv	U.K.
			FX/LI/Sw	Gv								FX/ /LI/Sw/Gv	ECU
			FX					Eq				FX/Eq	Argentina
			FX/LI/Sw				Gv	Eq				FX/Eq/LI/Sw/Gv	Canada
			FX/LI					Eq				FX/Eq/LI	Mexico
			FX/LI/Sw				Gv	Eq				FX/Eq/LI/Sw/Gv	U.S.
FX = Foreign Exchange, Eq = Equity Index, LI = LIBOR, Sw = Swap, Gv = Government													

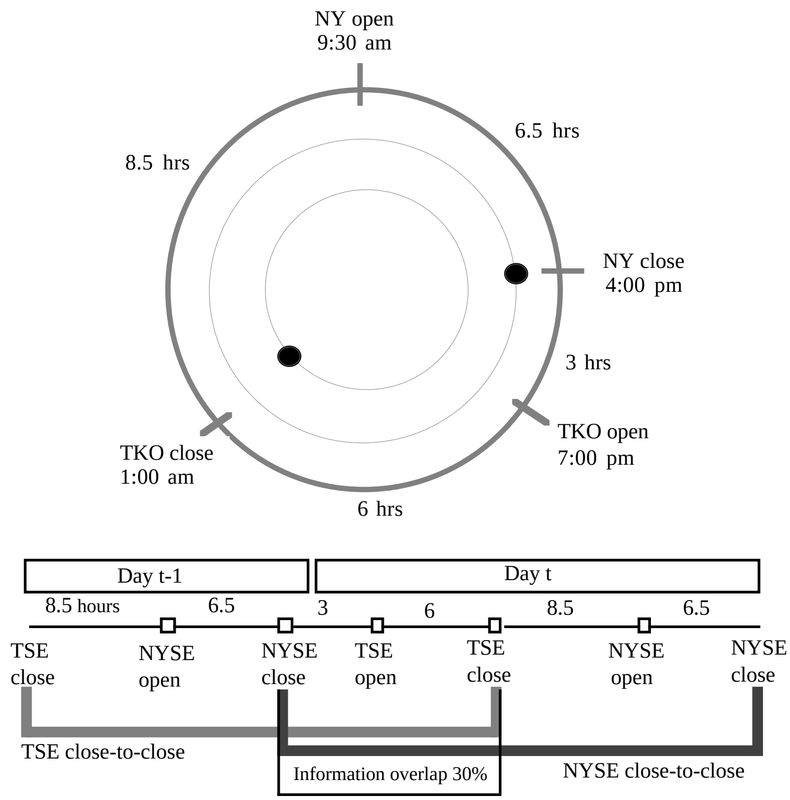
8.5.1 Estimating correlations when the data are nonsynchronous

The expansion of the RiskMetrics data set has increased the amount of underlying prices and rates collected in different time zones. The fundamental problem with nonsynchronous data collection is that correlation estimates based on these prices will be underestimated. And estimating correlations accurately is an important part of the RiskMetrics VaR calculation because standard deviation forecasts used in the VaR calculation depends on correlation estimates.

Internationally diversified portfolios are often composed of assets that trade in different calendar times in different markets. Consider a simple example of a two stock portfolio. Stock 1 trades only on the New York Stock Exchange (NYSE 9:30 am to 4:00 pm EST) while stock 2 trades exclusively on the Tokyo stock exchange (TSE 7:00 pm to 1:00 am EST). Because these two markets are never open at the same time, stocks 1 and 2 cannot trade concurrently. Consequently, their respective daily closing prices are recorded at different times and the return series for assets 1 and 2, which are calculated from daily close-to-close prices, are also nonsynchronous.[8]

Chart 8.5 illustrates the nonsynchronous trading hours of the NYSE and TSE.

Chart 8.5
Time chart
NY and Tokyo stock markets



[8] This terminology began in the nonsynchronous trading literature. See, Fisher, L. (1966) and Sholes, M. and Williams (1977). Nonsynchronous trading is often associated with the situation when some assets trade more frequently than others [see, Perry, P. (1985)]. Lo and MacKinlay (1990) note that "the nonsynchronicity problem results from the assumption that multiple time series are sampled simultaneously when in fact the sampling is nonsynchronous." For a recent discussion of the nonsynchronous trading issue see Boudoukh, et. al (1994).

We see that the Tokyo exchange opens three hours after the New York close and the New York exchange reopens 81/2 hours after the Tokyo close. Because a new calendar day arrives in Tokyo before New York, the Tokyo time is said to precede New York time by 14 hours (EST).

RiskMetrics computes returns from New York and Tokyo stock markets using daily close-to-close prices. The black orbs in Chart 8.5 mark times when these prices are recorded. Note that the orbs would line up with each other if returns in both markets were recorded at the same time.

The following sections will:

1. Identify the problem and verify whether RiskMetrics really does underestimate certain correlations.
2. Present an algorithm to adjust the correlation estimates.
3. Test the results against actual data.

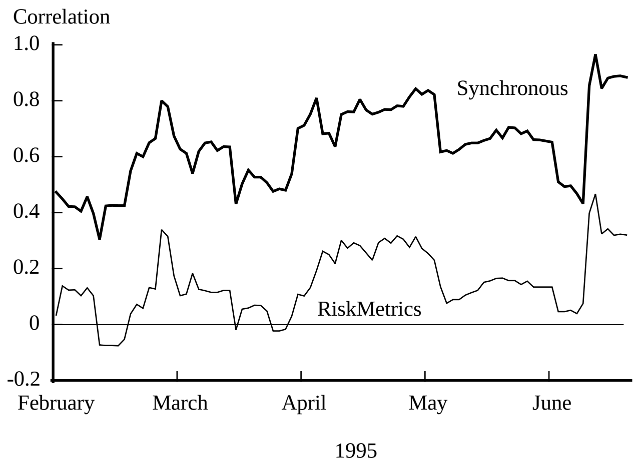
8.5.1.1 Identifying the problem: correlation and nonsynchronous returns

Whether different return series are recorded at the same time or not becomes an issue when these data are used to estimate correlations because the absolute magnitude of correlation (covariance) estimates may be underestimated when calculated from nonsynchronous rather than synchronous data. Therefore, when computing correlations using nonsynchronous data, we would expect the value of observed correlation to be below the true correlation estimate. In the following analysis we first establish the effect that nonsynchronous returns have on correlation estimates and then offer a method for adjusting correlation estimates to account for the nonsynchronicity problem.

The first step in checking for downward bias is estimating what the “true” correlation should be. This is not trivial since these assets do not trade in the same time zone and it is often not possible to obtain synchronous data. For certain instruments, however, it is possible to find limited datasets which can provide a glimpse of the true level of correlation; this data would then become the benchmark against which the methodology for adjusting nonsynchronous returns would be tested.

One of these instruments is the US Treasury which has the advantage of being traded 24 hours a day. While we generally use nonsynchronous close-to-close prices to estimate RiskMetrics correlations, we obtained price data for both the US and Australian markets quoted in the Asian time zone (August 1994 to June 1995). We compared the correlation based on synchronous data with correlation estimates that are produced under the standard RiskMetrics data (using the nonsynchronous US and Australian market close). Plots of the two correlation series are shown in Chart 8.6.

Chart 8.6
10-year Australia/US government bond zero correlation
based on daily RiskMetrics close/close data and 0:00 GMT data



While the changes in correlation estimates follow similar patterns over time (already an interesting result in itself), the correlation estimates obtained from price data taken at the opening of the markets in Asia are substantially higher. One thing worth noting however, is that while the synchronous estimate appears to be a better representation of the “true” level of correlation, it is not necessarily equal to the true correlation. While we have adjusted for the timing issue, we may have introduced other problems in the process, such as the fact that while US Treasuries trade in the Asian time zone, the market is not as liquid as during North American trading hours and the prices may therefore be less representative of “normal trading” volumes. Market segmentation may also affect the results. Most investors, even those based in Asia put on positions in the US market during North American trading hours. U.S. Treasury trading in Asia is often the result of hedging.

Nevertheless, from a risk management perspective, this is an important result. Market participants holding positions in various markets including Australia (and possibly other Asian markets) would be distorting their risk estimates by using correlation estimates generated from close of business prices.

8.5.1.2 An algorithm for adjusting correlations

Correlation is simply the covariance divided by the product of two standard errors. Since the standard deviations are unaffected by nonsynchronous data, correlation is adversely affected by nonsynchronous data through its covariance. This fact simplifies the analysis because under the current RiskMetrics assumptions, long horizon covariance forecasts are simply the 1-day covariance forecasts multiplied by the forecast horizon.

Let us now investigate the effect that nonsynchronous trading has on correlation estimates for historical rate series from the United States (USD), Australian (AUD) and Canadian (CAD) government bond markets. In particular, we focus on 10-year government bond zero rates. Table 8.8 presents the time that RiskMetrics records these rates (closing prices).

Table 8.8
RiskMetrics closing prices
10-year zero bonds

Country	EST	London
USD	3:30 p.m.	8:00 p.m.
CAD	3:30 p.m.	8:00 p.m.
AUD	2:00 a.m.	7:00 a.m.

Note that the USD and CAD rates are synchronous while the USD and AUD, and CAD and AUD rates are nonsynchronous. We chose to analyze rates in these three markets to gain insight as to how covariances (correlations) computed from synchronous and nonsynchronous return series compare with each other. For example, at any time t, the observed return series, $r_{USD,t}^{obs}$ and $r_{AUD,t}^{obs}$ are nonsynchronous, whereas $r_{USD,t}^{obs}$ and $r_{CAD,t}^{obs}$ are synchronous. We are interested in measuring the covariance and autocovariance of these return series.

Table 8.9 provides summary statistics on 1-day covariance and autocovariance forecasts for the period May 1993 to May 1995. The numbers in the table are interpreted as follows: over the sample period, the average covariance between USD and AUD 10-year zero returns, $\mathrm{cov}\left(r_{USD,t}^{obs}, r_{AUD,t}^{obs}\right)$ is 0.16335 while the average covariance between current USD 10-year zero returns and lagged CAD 10-year zero returns (autocovariance) is –0.0039.

Table 8.9
Sample statistics on RiskMetrics daily covariance forecasts
10-year zero rates; May 1993 – May 1995

Daily forecasts	Mean	Median	Std. dev.	Max	Min
$\mathrm{cov}\left(r_{USD,t}^{obs}, r_{AUD,t}^{obs}\right)$	0.1633*	0.0995	0.1973	0.8194	–0.3396
$\mathrm{cov}\left(r_{USD,t-1}^{obs} r_{AUD,t}^{obs}\right)$	0.5685	0.4635	0.3559	1.7053	0.1065
$\mathrm{cov}\left(r_{USD,t}^{obs}, r_{AUD,t-1}^{obs}\right)$	0.0085	–0.0014	0.1806	0.5667	–0.6056
$\mathrm{cov}\left(r_{USD,t}^{obs}, r_{CAD,t}^{obs}\right)$	0.6082	0.4912	0.3764	1.9534	0.1356
$\mathrm{cov}\left(r_{USD,t-1}^{obs}, r_{CAD,t}^{obs}\right)$	0.0424	0.0259	0.1474	0.9768	–0.2374
$\mathrm{cov}\left(r_{USD,t}^{obs}, r_{CAD,t-1}^{obs}\right)$	–0.0039	–0.0003	0.1814	0.3333	–0.7290

* All numbers are multiplied by 10,000.

The results show that when returns are recorded nonsynchronously, the covariation between lagged 1-day USD returns and current AUD returns (0.5685) is larger, on average, than the covariance (0.1633) that would typically be reported. Conversely, for the USD and CAD returns, the autocovariance estimates are negligible relative to the covariance estimates. This evidence points to a typical finding: first order autocovariances of returns for assets that trade at different times are larger than autocovariances for returns on assets that trade synchronously.[9]

[9] One possible explanation for the large autocovariances has to do with information flows between markets. The literature on information flows between markets include studies analyzing Japanese and US equity markets (Jaffe and Westerfield (1985), Becker, et.al, (1992), Lau and Diltz, (1994)). Papers that focus on many markets include Eun and Shim, (1989).

As a check of the results above and to understand how RiskMetrics correlation forecasts are affected by nonsynchronous returns, we now focus on covariance forecasts for a specific day. We continue to use USD, CAD and AUD 10-year zero rates. Consider the 1-day forecast period May 12 to May 13, 1995. In RiskMetrics, these 1-day forecasts are available at 10 a.m. EST on May 12. The most recent USD (CAD) return is calculated over the period 3:30 pm EST on 5/10 to 3:30 pm EST on 5/11 whereas the most recent AUD return is calculated over the period 1:00 am EST on 5/10 to 1:00 am EST on 5/11. Table 8.10 presents covariance forecasts for May 12 along with their standard errors.

Table 8.10
RiskMetrics daily covariance forecasts
10-year zero rates; May 12, 1995

Return series	Covariance	T-statistic†
$r^{\text{obs}}_{USD,5/12}r^{\text{obs}}_{AUD,5/12}$	0.305	-
$r^{\text{obs}}_{USD,5/11}r^{\text{obs}}_{AUD,5/12}$	0.629 (0.074)*	8.5
$r^{\text{obs}}_{USD,5/12}r^{\text{obs}}_{AUD,5/11}$	0.440 (0.074)	5.9
$r^{\text{obs}}_{USD,5/11}r^{\text{obs}}_{CAD,5/12}$	0.530	-
$r^{\text{obs}}_{USD,5/12}r^{\text{obs}}_{CAD,5/12}$	0.106 (0.058)	1.8
$r^{\text{obs}}_{USD,5/12}r^{\text{obs}}_{CAD,5/11}$	0.126 (0.059)	2.13

* Asymptotic standard errors are reported in parentheses.
† For a discussion on the use of the t-statistic for the autocovariances see Shanken (1987).

In agreement with previous results, we find that while there is strong covariation between lagged USD returns $r^{obs}_{USD,5/11}$ and current AUD returns $r^{obs}_{USD,5/12}$ (as shown by large t-statistics), the covariation between lagged USD and CAD returns is not nearly as strong. The results also show evidence of covariation between lagged AUD returns and current USD returns.

The preceding analysis describes a situation where the standard covariances calculated from nonsynchronous data do not capture all the covariation between returns. By estimating autocovariances, it is possible to measure the 1-day lead and lag effects across return series. With nonsynchronous data, these lead and lag effects appear quite large. In other words, current and past information in one return series is correlated with current and past information in another series. If we represent information by returns, then following Cohen, Hawawini, Maier, Schwartz and Whitcomb, (CHMSW 1983) we can write observed returns as a function of weighted unobserved current and lag true returns. The weights simply represent how much information in a specific true return appears in the return that is observed. Given this, we can write observed (nonsynchronous) returns for the USD and AUD 10-year zero returns as follows:

[8.43]
$$\begin{aligned} r^{\text{obs}}_{USD,t} &= \theta_{USD,t}R_{USD,t} + \theta_{USD,t-1}r_{USD,t-1} \\ r^{\text{obs}}_{AUD,t} &= \theta_{AUD,t}R_{USD,t} + \theta_{AUD,t-1}r_{AUD,t-1} \end{aligned}$$

The $\theta_{j,t-i}$'s are random variables that represent the proportion of the true return of asset *j* generated in period *t*-i that is actually incorporated in observed returns in period *t*. In other words, the $\theta_{j,t}$'s are weights that capture how the true return generated in one period impacts on the observed returns in the same period and the next. It is also assumed that:

[8.44]
$$\begin{aligned}&\theta_{AUD,t} \text{ and } \theta_{USD,\tau} \text{ are independent for all t and } \tau\\&\theta_{AUD,t} \text{ and } \theta_{USD,\tau} \text{ are independent of } R_{AUD,t} \text{ and } R_{USD,\tau}\\&E(\theta_{AUD,t}) = E(\theta_{USD,t}) \text{ for all t and } \tau\\&E(\theta_{j,t}+\theta_{j,t-1}) = 1 \text{ for j = AUD, USD and for all t and } \tau\end{aligned}$$

Table 8.11 shows, for the example given in the preceding section, the relationship between the date when the true return is calculated and the weight assigned to the true return.

Table 8.11
Relationship between lagged returns and applied weights
observed USD and AUD returns for May 12, 1995

Date	5/9–5/10	5/9–5/10	5/10–5/11	5/10–5/11
Weight	$\theta_{AUD,t-1}$	$\theta_{USD,t-1}$	$\theta_{AUD,t}$	$\theta_{USD,t}$

Earlier we computed the covariance based on observed returns, $\mathrm{cov}\left(r^{obs}_{USD,t}, r^{obs}_{AUD,t}\right)$ However, we can use Eq. [8.43] to compute the covariance of the true returns $\mathrm{cov}(r_{USD,t}, r_{AUD,t})$, i.e.,

[8.45]
$$\begin{aligned}\mathrm{cov}(r_{USD,t}, r_{AUD,t}) &= \mathrm{cov}\left(r^{obs}_{USD,t}, r^{obs}_{AUD,t-1}\right)\\&+\mathrm{cov}\left(r^{obs}_{USD,t}, r^{obs}_{AUD,t}\right)+\mathrm{cov}\left(r^{obs}_{USD,t-1}, r^{obs}_{AUD,t}\right)\end{aligned}$$

We refer to this estimator as the **“adjusted” covariance. Having established the form of the adjusted covariance estimator, the adjusted correlation estimator for any two return series *j* and *k* is:**

[8.46]
$$\rho_{jk,t} = \frac{\mathrm{cov}\left(r^{obs}_{j,t}, r^{obs}_{k,t-1}\right)+\mathrm{cov}\left(r^{obs}_{j,t}, r^{obs}_{k,t}\right)+\mathrm{cov}\left(r^{obs}_{j,t-1}, r^{obs}_{k,t}\right)}{\mathrm{std}\left(r^{obs}_{j,t}\right)\mathrm{std}\left(r^{obs}_{k,t}\right)}$$

Table 8.12 shows the original and adjusted correlation estimates for USD-AUD and USD-CAD 10-year zero rate returns.

Table 8.12
Original and adjusted correlation forecasts
USD-AUD 10-year zero rates; May 12, 1995

Daily forecasts	**Original**	**Adjusted**	**% change**
$\mathrm{cov}(r_{USD,5/12}, r_{AUD,5/12})$	0.305	0.560	84%
$\mathrm{cov}(r_{USD,5/12}, r_{CAD,5/12})$	0.530	0.573	8%

Note that the USD-AUD adjusted covariance increases the original covariance estimate by 84%. Earlier (see Table 8.10) we found the lead-lag covariation for the USD-AUD series to be statistically significant. Applying the adjusted covariance estimator to the synchronous series USD-CAD, we find only an 8% increase over the original covariance estimate. However, the evidence from Table 8.10 would suggest that this increase is negligible.

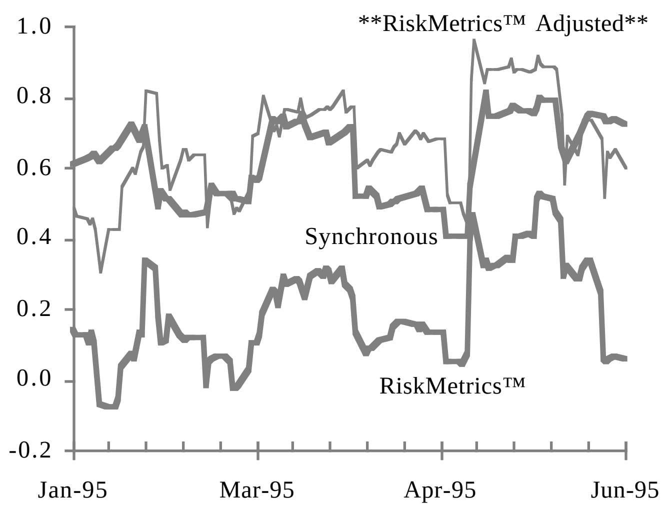
8.5.1.3 Checking the results

How does the adjustment algorithm perform in practice? Chart 8.7 compares three daily correlation estimates for 10-year zero coupon rates in Australia and the United States: (1) Standard RiskMetrics using nonsynchronous data, (2) estimate correlation using synchronous data collected in Asian trading hours and, (3) RiskMetrics Adjusted using the estimator in Eq. [8.46].

Chart 8.7

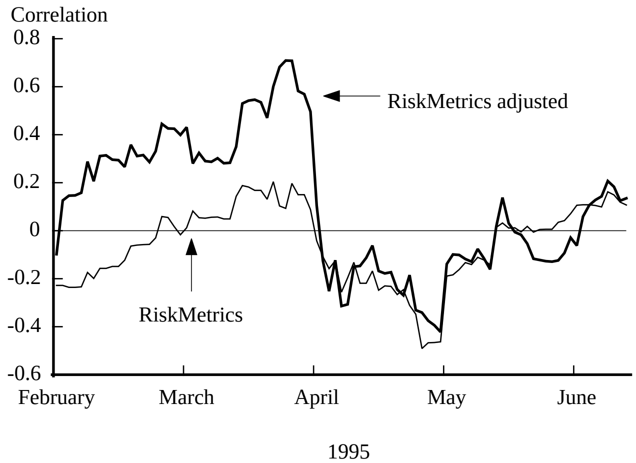
Adjusting 10-year USD/AUD bond zero correlation

using daily RiskMetrics close/close data and 0:00 GMT data



The results show that the adjustment factor captures the effects of the timing differences that affect the standard RiskMetrics estimates which use nonsynchronous data. A potential drawback of using this estimator, however, is that the adjusted series displays more volatility than either the unadjusted or the synchronous series. This means that in practice, choices may have to be made as to when to apply the methodology. In the Australian/US case, it is clear that the benefits of the adjustment in terms of increasing the correlation to a level consistent with the one obtained when using synchronous data outweighs the increased volatility. The choice, however, may not always be that clear cut as shown by Chart 8.8 which compares adjusted and unadjusted correlations for the US and Japanese 10-year zero rates. In periods when the underlying correlation between the two markets is significant (Jan-Feb 1995, the algorithm correctly adjusts the estimate). In periods of lower correlation, the algorithm only increases the volatility of the estimate.

Chart 8.8
10-year Japan/US government bond zero correlation
using daily RiskMetrics close/close data and 0:00 GMT data



Also, in practice, estimation of the adjusted correlation is not necessarily straightforward because we must take into account the chance of getting adjusted correlation estimates above 1. This potential problem arises because the numerator in Eq. [8.46] is being adjusted without due consideration of the denominator. An algorithm that allows us to estimate the adjusted correlation without obtaining correlations greater than 1 in absolute value is given in Section 8.5.2.

Table 8.13 on page 196 reports sample statistics for 1-day correlation forecasts estimated over various sample periods for both the original RiskMetrics and adjusted correlation estimators. Correlations between United States and Asia-Pacific are based on non-synchronous data.

8.5.2 Using the algorithm in a multivariate framework

Finally, we explain how to compute the adjusted correlation matrix.

1. Calculate the unadjusted (standard) RiskMetrics covariance matrix, Σ. (Σ is an N x N, positive semi-definite matrix).
2. Compute the nonsynchronous data adjustment matrix K where the elements of K are

[8.47] $$k_{k,j} = \begin{cases} \operatorname{cov}(r_{k,t}, r_{j,t-1}) + \operatorname{cov}(r_{k,t-1}, r_{j,t}) & \text{for } k \neq j \\ 0 & \text{for } k = j \end{cases}$$

3. The adjusted covariance matrix M, is given by $M = \Sigma + fK$ where $0 \leq f \leq 1$. The parameter *f* that is used in practice is the largest possible *f* such that *M* is positive semi-definite.

Table 8.13
Correlations between US and foreign instruments

Correlations between USD 10-year zero rates and JPY, AUD, and NZD 10-year zero rates.*
Sample period: May 1991–May 1995.

	Original			Adjusted		
	JPY	AUD	NZD	JPY	AUD	NZD
mean	0.026	0.166	0.047	0.193	0.458	0.319
median	0.040	0.155	0.036	0.221	0.469	0.367
std dev	0.151	0.151	0.171	0.308	0.221	0.241
max	0.517	0.526	0.613	0.987	0.937	0.921
min	–0.491	–0.172	–0.389	–0.762	–0.164	–0.405

Correlations between USD 2-year swap rates and JPY, AUD, NZD, HKD 2-year swap rates.*
Sample period: May 1993–May 1995.

	Original				Adjusted			
	JPY	AUD	NZD	HKD	JPY	AUD	NZD	HKD
mean	0.018	0.233	0.042	0.139	0.054	0.493	0.249	0.572
median	0.025	0.200	0.020	0.103	0.065	0.502	0.247	0.598
std dev	0.147	0.183	0.179	0.217	0.196	0.181	0.203	0.233
max	0.319	0.647	0.559	0.696	0.558	0.920	0.745	0.945
min	–0.358	–0.148	–0.350	–0.504	–0.456	–0.096	–0.356	–0.411

Correlations between USD equity index and JPY, AUD, NZD, HKD, SGD equity indices.*
Sample period: May 1993–May 1995.

	Original					Adjusted				
	JPY	AUD	NZD	HKD	SGD	JPY	AUD	NZD	HKD	SGD
mean	0.051	0.099	-0.023	0.006	0.038	0.124	0.330	–0.055	–0.013	0.014
median	0.067	0.119	-0.021	-0.001	0.028	0.140	0.348	–0.053	0.056	–0.024
std dev	0.166	0.176	0.128	0.119	0.145	0.199	0.206	0.187	0.226	0.237
max	0.444	0.504	0.283	0.271	0.484	0.653	0.810	0.349	0.645	0.641
min	–0.335	–0.345	–0.455	–0.298	–0.384	–0.395	–0.213	–0.524	–0.527	–0.589

* JPY = Japanese yen, AUD = Australian dollar, NZD = New Zealand dollar, HKD = Hong Kong dollar, SGD = Singapore dollar

Chapter 9. Time series sources

9.1 Foreign exchange 199
9.2 Money market rates 199
9.3 Government bond zero rates 200
9.4 Swap rates 202
9.5 Equity indices 203
9.6 Commodities 205

Chapter 9. Time series sources

Scott Howard
Morgan Guaranty Trust Company
Risk Management Advisory
(1-212) 648-4317
howard_james_s@jpmorgan.com

Data is one of the cornerstones of any risk management methodology. We examined a number of data providers and decided that the sources detailed in this chapter were the most appropriate for our purposes.

9.1 Foreign exchange

Foreign exchange prices are sourced from WM Company and Reuters. They are mid-spot exchange prices recorded at 4:00 p.m. London time (11:00 a.m. EST). All foreign exchange data used for RiskMetrics is identical to the data used by the J.P. Morgan family of government bond indices. (See Table 9.1.)

Table 9.1
Foreign exchange

Currency Codes					
Americas		**Asia Pacific**		**Europe and Africa**	
ARS	Argentine peso	AUD	Australian dollar	ATS	Austrian shilling
CAD	Canadian dollar	HKD	Hong Kong dollar	BEF	Belgian franc
MXN	Mexican peso	IDR	Indonesian rupiah	CHF	Swiss franc
USD	U.S. dollar	JPY	Japanese yen	DEM	Deutsche mark
EMB	EMBI+*	KRW	Korean won	DKK	Danish kroner
		MYR	Malaysian ringgit	ESP	Spanish peseta
		NZD	New Zealand dollar	FIM	Finnish mark
		PHP	Philippine peso	FRF	French franc
		SGD	Singapore dollar	GBP	Sterling
		THB	Thailand baht	IEP	Irish pound
		TWD	Taiwan dollar	ITL	Italian lira
				NLG	Dutch guilder
				NOK	Norwegian kroner
				PTE	Portuguese escudo
				SEK	Swedish krona
				XEU	ECU
				ZAR	South African rand

* EMBI+ stands for the J.P. Morgan Emerging Markets Bond Index Plus.

9.2 Money market rates

Most 1-, 2-, 3-, 6-, and 12-month money market rates (offered side) are recorded on a daily basis by J.P. Morgan in London at 4:00 p.m. (11:00 a.m. EST). Those obtained from external sources are also shown in Table 9.2.

Table 9.2
Money market rates: sources and term structures

	Source		Time	Term Structure			
Market	J.P. Morgan	Third Party*	U.S. EST	1m	3m	6m	12m
Australia	•		11:00 a.m.	•	•	•	•
Hong Kong		•	10:00 p.m.	•	•	•	•
Indonesia†	•		5:00 a.m.	•	•	•	•
Japan	•		11:00 a.m.	•	•	•	•
Malaysia†	•		5:00 a.m.	•	•	•	•
New Zealand		•	12:00 a.m.	•	•	•	
Singapore		•	4:30 a.m.	•	•	•	•
Thailand†	•		5:00 a.m.	•	•	•	•
Austria		•	11:00 a.m.	•	•	•	•
Belgium	•		11:00 a.m.	•	•	•	•
Denmark	•		11:00 a.m.	•	•	•	•
Finland		•	11:00 a.m.	•	•	•	•
France	•		11:00 a.m.	•	•	•	•
Ireland		•	11:00 a.m.	•	•	•	•
Italy	•		11:00 a.m.	•	•	•	•
Netherlands	•		11:00 a.m.	•	•	•	•
Norway		•	11:00 a.m.	•	•	•	•
Portugal		•	11:00 a.m.	•	•	•	•
South Africa			11:00 a.m.	•	•	•	•
Spain	•		11:00 a.m.	•	•	•	•
Sweden	•		11:00 a.m.	•	•	•	•
Switzerland	•		11:00 a.m.	•	•	•	•
U.K.	•		11:00 a.m.	•	•	•	•
ECU	•		11:00 a.m.	•	•	•	•
Canada	•		11:00 a.m.	•	•	•	•
Mexico‡	•		12:00 p.m.	•	•	•	•
U.S.	•		11:00 a.m.	•	•	•	•

* Third party source data from Reuters Generic except for Hong Kong (Reuters HIBO), Singapore (Reuters MASX), and New Zealand (National Bank of New Zealand).

† Money market rates for Indonesia, Malaysia, and Thailand are calculated using foreign exchange forward-points.

‡ Mexican rates represent secondary trading in Cetes.

9.3 Government bond zero rates

Zero coupon rates ranging in maturity from 2 to 30 years for the government bond markets included in the J.P. Morgan Government Bond Index as well as the Irish, ECU, and New Zealand markets. (See Table 9.3.)

Table 9.3
Government bond zero rates: sources and term structures

	Source		Time	Term structure									
Market	J.P. Morgan	Third Party	U.S. EST	2y	3y	4y	5y	7y	9y	10y	15y	20y	30y
Australia	•		1:30 a.m.	•	•	•	•	•	•	•	•		
Japan	•		1:00 a.m.	•	•	•	•	•	•	•			
New Zealand		•	12:00 a.m.	•	•	•	•	•	•	•	•		
Belgium	•		11:00 a.m.	•	•	•	•	•	•	•	•	•	
Denmark		•	10:30 a.m.	•	•	•	•	•	•	•	•	•	•
France	•		10:30 a.m.	•	•	•	•	•	•	•	•	•	•
Germany	•		11:30 a.m.	•	•	•	•	•	•	•	•	•	•
Ireland		•	10:30 a.m.	•	•	•	•	•	•	•	•	•	
Italy	•		10:45 a.m.	•	•	•	•	•	•	•	•	•	•
Netherlands	•		11:00 a.m.	•	•	•	•	•	•	•	•	•	•
South Africa	•		11:00 a.m.	•	•	•	•	•	•	•	•	•	
Spain	•		11:00 a.m.	•	•	•	•	•	•	•	•		
Sweden		•	10:00 a.m.	•	•	•	•	•	•	•	•		
U.K.	•		11:45 a.m.	•	•	•	•	•	•	•	•	•	•
ECU	•		11:45 a.m.	•	•	•	•	•	•	•			
Canada	•		3:30 p.m.	•	•	•	•	•	•	•	•	•	•
U.S.	•		3:30 a.m.	•	•	•	•	•	•	•	•	•	•
Emerging Mkt.†	•		3:00 p.m.										

* Third party data sourced from Den Danske Bank (Denmark), NCB Stockbrokers (Ireland), National Bank of New Zealand (New Zealand), and SE Banken (Sweden).

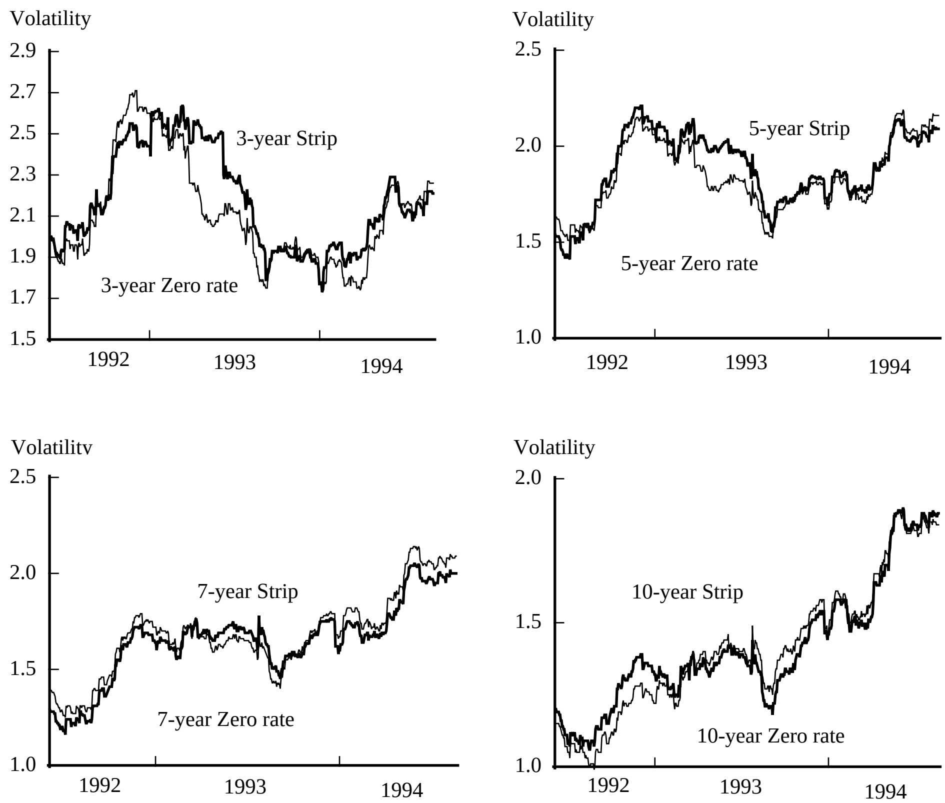
† J. P. Morgan Emerging Markets Bond Index Plus (EMBI+).

If the objective is to measure the volatility of individual cash flows, then one could ask whether it is appropriate to use a term structure model instead of the underlying zero rates which can be directly observed from instruments such as Strips. The selection of a modeled term structure as the basis for calculating market volatilities was motivated by the fact that there are few markets which have observable zero rates in the form of government bond Strips from which to estimate volatilities. In fact, only the U.S. and French markets have reasonably liquid Strips which could form the basis for a statistically solid volatility analysis. Most other markets in the OECD have either no Strip market or a relatively illiquid one.

The one possible problem of the term structure approach is that it would not be unreasonable to assume the volatility of points along the term structure may be lower than the market's real volatility because of the smoothing impact of passing a curve through a universe of real data points.

To see whether there was support for this assumption, we compared the volatility estimates obtained from term structure derived zero rates and actual Strip yields for the U.S. market across four maturities (3, 5, 7, and 10 years). The results of the comparison are shown in Chart 9.1.

Chart 9.1
Volatility estimates: daily horizon
1.65 standard deviation—6-month moving average



The results show that there is no clear bias from using the term structure versus underlying Strips data. The differences between the two measures decline as maturity increases and are partially the result of the lack of liquidity of the short end of the U.S. Strip market. Market movements specific to Strips can also be caused by investor behavior in certain hedging strategies that cause prices to sometimes behave erratically in comparison to the coupon curve from which the term structure is derived.

9.4 Swap rates

Swap par rates from 2 to 10 years are recorded on a daily basis by J.P. Morgan, except for Ireland (provided by NCB Stockbrokers), Hong Kong (Reuters TFHK) and Indonesia, Malaysia and Thailand (Reuters EXOT). (See Table 9.4.) The par rates are then converted to zero coupon equivalents rates for the purpose of inclusion within the RiskMetrics data set. (Refer to Section 8.1 for details).

Table 9.4
Swap zero rates: sources and term structures

	Source		Time	Term structure					
Market	J.P. Morgan	Third Party*	US EST	2y	3y	4y	5y	7y	10y
Australia	•		1:30 a.m.	•	•	•	•	•	•
Hong Kong		•	4:30 a.m.	•	•	•	•	•	•
Indonesia		•	4:00 a.m.	•	•	•	•		
Japan	•		1:00 a.m.	•	•	•	•	•	•
Malaysia		•	4:00 a.m.	•	•	•	•		
New Zealand		•	3:00 p.m.	•	•	•	•	•	
Thailand		•	4:00 a.m.	•	•	•	•		
Belgium	•		10:00 a.m.	•	•	•	•	•	•
Denmark	•		10:00 a.m.	•	•	•	•	•	•
Finland	•		10:00 a.m	•	•	•	•		
France	•		10:00 a.m.	•	•	•	•	•	•
Germany	•		10:00 p.m.	•	•	•	•	•	•
Ireland		•	11:00 a.m.	•	•	•	•		
Italy	•		10:00 a.m.	•	•	•	•	•	•
Netherlands	•		10:00 a.m.	•	•	•	•	•	•
Spain	•		10:00 a.m.	•	•	•	•	•	•
Sweden	•		10:00 a.m.	•	•	•	•	•	•
Switzerland	•		10:00 a.m.	•	•	•	•	•	•
U.K.	•		10:00 a.m.	•	•	•	•	•	•
ECU	•		10:00 a.m.	•	•	•	•	•	•
Canada	•		3:30 p.m.	•	•	•	•	•	•
U.S.	•		3:30 a.m.	•	•	•	•	•	•

* Third party source data from Reuters Generic except for Ireland (NCBI), Hong Kong (TFHK), and Indonesia, Malaysia, Thailand (EXOT).

9.5 Equity indices

The following list of equity indices (Table 9.5) have been selected as benchmarks for measuring the market risk inherent in holding equity positions in their respective markets. The factors that determined the selection of these indices include the existence of index futures that can be used as hedging instruments, sufficient market capitalization in relation to the total market, and low tracking error versus a representation of the total capitalization. All the indices listed below measure principal return except for the DAX which is a total return index.

Table 9.5
Equity indices: sources*

Market	Exchange	Index Name	Weighting	% Mkt. cap.	Time, U.S. EST
Australia	Australian Stock Exchange	All Ordinaries	MC	96	1:10 a.m.
Hong Kong	Hong Kong Stock Exchange	Hang Seng	MC	77	12:30 a.m.
Indonesia	Jakarta Stock Exchange	JSE	MC		4:00 a.m.
Korea	Seoul Stock Exchange	KOPSI	MC		3:30 a.m.
Japan	Tokyo Stock Exchange	Nikei 225	MC	46	1:00 a.m.
Malaysia	Kuala Lumpur Stock Exchange	KLSE	MC		6:00 a.m.
New Zealand	New Zealand Stock Exchange	Capital 40	MC	—	10:30 p.m.
Philippines	Manila Stock Exchange	MSE Com'l &Inustil Price	MC		1:00 a.m.
Singapore	Stock Exchange of Singapore	Sing. All Share	MC	—	4:30 a.m.
Taiwan	Taipei Stock Exchange	TSE	MC		1:00 a.m.
Thailand	Bangkok Stock Exchange	SET	MC		5:00 a.m.
Austria	Vienna Stock Exchange	Creditanstalt	MC	—	7:30 a.m.
Belgium	Brussels Stock Exchange	BEL 20	MC	78	10:00 a.m.
Denmark	Copenhagen Stock Exchange	KFX	MC	44	9:30 a.m.
Finland	Helsinki Stock Exchange	Hex General	MC	—	10:00 a.m.
France	Paris Bourse	CAC 40	MC	55	11:00 a.m.
Germany	Frankfurt Stock Exchange	DAX	MC	57	10:00 a.m.
Ireland	Irish Stock Exchange	Irish SE ISEQ	—	—	12:30 p.m.
Italy	Milan Stock Exchange	MIB 30	MC	65	10:30 a.m.
Japan	Tokyo Stock Exchange	Nikei 225	MC	46	1:00 a.m.
Netherlands	Amsterdam Stock Exchange	AEX	MC	80	10:30 a.m.
Norway	Oslo Stock Exchange	Oslo SE General	—	—	9:00 a.m.
Portugal	Lisbon Stock Exchange	Banco Totta SI	—	—	11:00 a.m.
South Africa	Johannesburg Stock Exchange	JSE	MC		10:00 a.m.
Spain	Madrid Stock Exchange	IBEX 35	MC	80	11:00 a.m.
Sweden	Stockholm Stock Exchange	OMX	MC	61	10:00 a.m.
Switzerland	Zurich Stock Exchange	SMI	MC	56	10:00 a.m.
U.K.	London Stock Exchange	FTSE 100	MC	69	10:00 a.m.
Argentina	Buenos Aires Stock Exchange	Merval	Vol.		5:00 p.m.
Canada	Toronto Stock Exchange	TSE 100	MC	63	4:15 p.m.
Mexico	Mexico Stock Exchange	IPC	MC		3:00 p.m.
U.S.	New York Stock Exchange	Standard and Poor's 100	MC	60	4:15 a.m.

* Data sourced from DRI.

9.6 Commodities

The commodity markets that have been included in RiskMetrics are the same markets as the J.P. Morgan Commodity Index (JPMCI). The data for these markets are shown in Table 9.6.

Table 9.6
Commodities: sources and term structures

Commodity	Source	Time, U.S. EST	Term structure						
			Spot	1m	3m	6m	12m	15m	27m
WTI Light Sweet Crude	NYMEX*	3:10 p.m.		•	•	•	•		
Heating Oil	NYMEX	3:10 p.m.		•	•	•	•		
NY Harbor #2 unleaded gas	NYMEX	3:10 p.m.		•	•	•			
Natural gas	NYMEX	3:10 p.m.		•	•	•	•		
Aluminum	LME†	11:20 a.m.	•		•			•	•
Copper	LME	11:15 a.m.	•		•			•	•
Nickel	LME	11:10 a.m.	•		•			•	
Zinc	LME	11:30 a.m.	•		•			•	•
Gold	LME	11:00 a.m.	•						
Silver	LFOE‡	11:00 a.m.	•						
Platinum	LPPA§	11:00 a.m.	•						

* NYMEX (New York Mercantile Exchange)
† LME (London Metals Exchange)
‡ LFOE (London futures and Options Metal Exchange)
§ LPPA (London Platinum & Palladium Association)

The choice between either the rolling nearby or interpolation (constant maturity) approach is influenced by the characteristics of each contract. We use the interpolation methodology wherever possible, but in certain cases this approach cannot or should not be implemented.

We use interpolation (I) for all energy contracts. (See Table 9.7.)

Table 9.7
Energy maturities

Energy	Maturities					
	1m	3m	6m	12m	15m	27m
Light sweet crude	I*	I	I	I		
Heating Oil	I	I	I	I		
Unleaded Gas	I	I	I			
Natural Gas	I	I	I	I		

* I = Interpolated methodology.

The term structures for base metals are based upon rolling nearby contracts with the exception of the spot (S) and 3-month contracts. Data availability is the issue here. Price data for contracts traded on the London Metals Exchange is available for constant maturity 3-month (A) contracts (prices are quoted on a daily basis for 3 months forward) and rolling 15- and 27- month (N) contracts. Nickel extends out to only 15 months. (See Table 9.8.)

Table 9.8
Base metal maturities

	Maturities					
Commodity	Spot	3m	6m	12m	15m	27m
Aluminum	S*	A†			N‡	N
Copper	S	A			N	N
Nickel	S	A			N	
Zinc	S	A			N	N

* S = Spot contract.
† A = Constant maturity contract.
‡ N = Rolling contract.

Spot prices are the driving factor in the precious metals markets. Volatility curves in the gold, silver, and platinum markets are relatively flat (compared to the energy curves) and spot prices are the main determinant of the future value of instruments: storage costs are negligible and convenience yields such as those associated with the energy markets are not a consideration.

Chapter 10. RiskMetrics volatility and correlation files

10.1 Availability 209
10.2 File names 209
10.3 Data series naming standards 209
10.4 Format of volatility files 211
10.5 Format of correlation files 212
10.6 Data series order 214
10.7 Underlying price/rate availability 214

Chapter 10. RiskMetrics volatility and correlation files

Scott Howard
Morgan Guaranty Trust Company
Risk Management Advisory
(1-212) 648-4317
howard_james_s@jpmorgan.com

This section serves as a guide to understanding the information contained in the RiskMetrics daily and monthly volatility and correlation files. It defines the naming standards we have adopted for the RiskMetrics files and time series, the file formats, and the order in which the data is presented in these files.

10.1 Availability

Volatility and correlation files are updated each U.S. business day and posted on the Internet by 10:30 a.m. EST. They cover data through close-of-business for the previous U.S. business day. Instructions on downloading these files are available in Appendix H.

10.2 File names

To ensure compatibility with MS-DOS, file names use the "8.3" format: 8-character name and 3-character extension (see Table 10.1).

Table 10.1
RiskMetrics file names
"ddmmyy" indicates the date on which the market data was collected

File name format		
Volatility	**Correlation**	**File description**
DVddmmyy.RM3	DCddmmyy.RM3	1-day estimates
MVddmmyy.RM3	MCddmmyy.RM3	25-day estimates
BVddmmyy.RM3	BCddmmyy.RM3	Regulatory data sets
DVddmmyy.vol	DCddmmyy.cor	Add-In 1-day estimates
MVddmmyy.vol	MCddmmyy.cor	Add-In 25-day estimates
BVddmmyy.vol	BCddmmyy.cor	Add-In regulatory

The first two characters designate whether the file is daily (D) or monthly (M), and whether it contains volatility (V) or correlation (C) data. The next six characters identify the collection date of the market data for which the volatilities and correlations are computed. The extension identifies the version of the data set.

10.3 Data series naming standards

In both volatility and correlation files, all series names follow the same naming convention. They start with a three-letter code followed by a period and a suffix, for example, USD.R180.

The three-letter code is either a SWIFT[1] currency code or, in the case of commodities, a commodity code, as shown in Table 10.2. The suffix identifies the asset class (and the maturity for interest-rate and commodity series). Table 10.3 lists instrument suffix codes, followed by an example of how currency, commodity, and suffix codes are used.

[1] The exception is EMB. This represents J. P. Morgan's Emerging Markets Bond Index Plus.

Table 10.2
Currency and commodity identifiers

Currency Codes						Commodity Codes	
Americas		Asia Pacific		Europe and Africa			
ARS	Argentine peso	AUD	Australian dollar	ATS	Austrian shilling	ALU	Aluminum
CAD	Canadian dollar	HKD	Hong Kong dollar	BEF	Belgian franc	COP	Copper
MXN	Mexican peso	IDR	Indonesian rupiah	CHF	Swiss franc	GAS	Natural gas
USD	U.S. dollar	JPY	Japanese yen	DEM	Deutsche mark	GLD	Gold
EMB	EMBI+*	KRW	Korean won	DKK	Danish kroner	HTO	NY Harbor #2 heating oil
		MYR	Malaysian ringgit	ESP	Spanish peseta	NIC	Nickel
		NZD	New Zealand dollar	FIM	Finnish mark	PLA	Platinum
		PHP	Philippine peso	FRF	French franc	SLV	Silver
		SGD	Singapore dollar	GBP	Sterling	UNL	Unleaded gas
		THB	Thailand baht	IEP	Irish pound	WTI	Light Sweet Crude
		TWD	Taiwan dollar	ITL	Italian lira	ZNC	Zinc
				NLG	Dutch guilder		
				NOK	Norwegian kroner		
				PTE	Portuguese escudo		
				SEK	Swedish krona		
				XEU	ECU		
				ZAR	South African rand		

* EMBI+ stands for the J.P. Morgan Emerging Markets Bond Index Plus.

Table 10.3
Maturity and asset class identifiers

	Instrument Suffix Codes					
Maturity	Foreign exchange	Equity indices	Money market	Swaps	Gov't bonds	Commodities
Spot	XS	SE	–	–	–	C00
1m	–	–	R030	–	–	–
3m	–	–	R090	–	–	C03
6m	–	–	R180	–	–	C06
12m	–	–	R360	–	–	C12
15m	–	–	–	–	–	C15
18m	–	–	–	–	–	C18
24m (2y)	–	–	–	S02	Z02	C24
27m	–	–	–	–	–	C27
36m (3y)	–	–	–	S03	Z03	C36
4y	–	–	–	S04	Z04	–
5y	–	–	–	S05	Z05	–
7y	–	–	–	S07	Z07	–
9y	–	–	–	–	Z09	–
10y	–	–	–	S10	Z10	–
15y	–	–	–	–	Z15	–
20y	–	–	–	–	Z20	–
30y	–	–	–	–	Z30	–

For example, we identify the Singapore dollar foreign exchange rate by SGD.XS, the U.S. dollar 6-month money market rate by USD.R180, the CAC 40 index by FRF.SE, the 2-year sterling swap rate by GBP.S02, the 10-year Japanese government bond (JGB) by JPY.Z10, and the 3-month natural gas future by GAS.C03.

10.4 Format of volatility files

Each daily and monthly volatility file starts with a set of header lines that begin with an asterisk (*) and describe the contents of the file. Following the header lines are a set of record lines (without an asterisk) containing the daily or monthly data.

Table 10.4 shows a portion of a daily volatility file.

Table 10.4
Sample volatility file

Line #	Volatility file
1	*Estimate of volatilities for a one day horizon
2	*COLUMNS=2, LINES=418, DATE=11/14/96, VERSION 2.0
3	*RiskMetrics is based on but differs significantly from the market risk management systems
4	*developed by J.P. Morgan for its own use. J.P. Morgan does not warranty any results obtained
5	*from use of the RiskMetrics methodology, documentation or any information derived from
6	*the data (collectively the “Data”) and does not guarantee its sequence, timeliness, accuracy or
7	*completeness. J.P. Morgan may discontinue generating the Data at any time without any prior
8	*notice. The Data is calculated on the basis of the historical observations and should not be relied
9	*upon to predict future market movements. The Data is meant to be used with systems developed
10	*by third parties. J.P. Morgan does not guarantee the accuracy or quality of such systems.
11	*SERIES, PRICE/YIELD,DECAYFCTR,PRICEVOL,YIELDVOL
12	ATS.XS.VOLD,0.094150,0.940,0.554647,ND
13	AUD.XS.VOLD, 0.791600,0.940,0.643127,ND
14	BEF.XS.VOLD, 0.032152,0.940,0.546484,ND

In this table, each line is interpreted as follows:

- Line 1 identifies whether the file is a daily or monthly file.
- Line 2 lists file characteristics in the following order: the number of data columns, the number of record lines, the file creation date, and the version number of the file format.
- Lines 3–10 are a disclaimer.
- Line 11 contains comma-separated column titles under which the volatility data is listed.
- Lines 12 through the last line at the end of file (not shown) represent the record lines, which contain the comma-separated volatility data formatted as shown in Table 10.5.

Table 10.5
Data columns and format in volatility files

Column title (header line)	Data (record lines)	Format of volatility data
SERIES	Series name	See Section 10.3 for series naming conventions. In addition, each series name is given an extension, either ".VOLD" (for daily volatility estimate), or ".VOLM" (for monthly volatility estimate).
PRICE/YIELD	Price/Yield level	#.###### or "NM" if the data cannot be published.
DECAYFCTR	Exponential moving average decay factor	#.###
PRICEVOL	Price volatility estimate	#.###### (% units)
YIELDVOL	Yield volatility estimate	#.###### (% units) or "ND" if the series has no yield volatility (e.g., FX rates).

For example, in Table 10.4, the first value ATS.XS.VOLD in Line 12 corresponds to the SERIES column title, and identifies the series to be a USD/ATS daily volatility series. Similarly, the remaining values are interpreted as follows: The value 0.094150 was used as the price/yield level in the volatility calculation. The value 0.940 was used as the exponential moving average decay factor. The value 0.554647% is the price volatility estimate. The value "ND" indicates that the series has no yield volatility.

10.5 Format of correlation files

Daily and monthly correlation files are formatted similar to the volatility files (see Section 10.4), and contain analogous header and record lines (see Table 10.6). Each file comprises the lower half of the correlation matrix for the series being correlated, including the diagonal, which has a value of "1.000." (The upper half is not shown since the daily and monthly correlation matrices are symmetrical around the diagonal. For example, 3-month USD LIBOR to 3-month DEM LIBOR has the same correlation as 3-month DEM LIBOR to 3-month USD LIBOR.)

Table 10.6
Sample correlation file

Line #	Correlation file
1	*Estimate of correlations for a one day horizon
2	*COLUMNS=2, LINES=087571, DATE=11/14/96, VERSION 2.0
3	*RiskMetrics is based on but differs significantly from the market risk management systems
4	*developed by J.P. Morgan for its own use. J.P. Morgan does not warranty any results obtained
5	*from use of the RiskMetrics methodology, documentation or any information derived from
6	*the data (collectively the “Data”) and does not guarantee its sequence, timeliness, accuracy or
7	*completeness. J.P. Morgan may discontinue generating the Data at any time without any prior
8	*notice. The Data is calculated on the basis of the historical observations and should not be relied
9	*upon to predict future market movements. The Data is meant to be used with systems developed
10	*by third parties. J.P. Morgan does not guarantee the accuracy or quality of such systems.
11	*SERIES, CORRELATION
12	ATS.XS.ATS.XS.CORD,1.000000
13	ATS.XS.AUD.XS.CORD, -0.251566
14	ATS.XS.BEF.XS.CORD, 0.985189

In Table 10.6, each line is interpreted as follows:

- Line 1 identifies whether the file is a daily or monthly file.
- Line 2 lists file characteristics in the following order: the number of data columns, the number of record lines, the file creation date, and the version number of the file format.
- Lines 3–10 are a disclaimer.
- Line 11 contains comma-separated column titles under which the correlation data is listed.
- Lines 12 through the last line at the end of the file (not shown) represent the record lines, which contain the comma-separated correlation data formatted as shown in Table 10.7.

Table 10.7
Data columns and format in correlation files

Column title (header line)	Correlation data (record lines)	Format of correlation data
SERIES	Series name	See Section 10.3 for series naming conventions. In addition, each series name is given an extension, either “.CORD” (for daily correlation), or “.CORM” (for monthly correlation).
CORRELATION	Correlation coefficient	#.###### Correlation coefficients are computed by using the same exponential moving average method as in the volatility files (i.e., decay factor of 0.940 for a 1-day horizon, and 0.970 for a 1-month horizon.)

For example, Line 13 in Table 10.6 represents a USD/ATS to USD/AUD daily correlation estimate of –0.251566 measured using an exponential moving average decay factor of 0.940 (the default value for the 1-day horizon).

10.6 Data series order

Data series in the volatility and correlation files are sorted first alphabetically by SWIFT code and commodity class indicator, and then by maturity within the following asset class hierarchy: foreign exchange, money markets, swaps, government bonds, equity indices, and commodities.

10.7 Underlying price/rate availability

Due to legal considerations, not all prices or yields are published in the volatility files. What is published are energy future contract prices and the yields on foreign exchange, swaps, and government bonds. The current level of money market yields can be approximated from Eq. [10.1] by using the published price volatilities and yield volatilities as well as the instruments' modified durations.

[10.1] $$\textit{Current yield} = \sigma_{\mathrm{Price}} / (\sigma_{\mathrm{Yield}} \cdot \textit{Modified Duration})$$