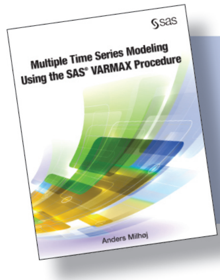
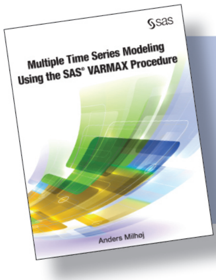
Multiple Time Series Modeling Using the SAS® VARMAX Procedure

Anders Milhøj


sas
Multiple Time Series Modeling Using the SAS® VARMAX Procedure
Anders Milhøj

Contents



Chapter 8: Models for Multivariate Time Series

Introduction

In this chapter, you will learn the basic theory for multivariate time series. The purpose is to introduce the simplest theoretical model behind the many tools offered by the VARMAX procedure, because most of them are extensions or refinements of this basic model. The idea is not to give a thorough introduction to the theory, for this subject is far too extensive to include in a book that is specific to SAS. For more information about multivariate time series analysis, consult ordinary textbooks like Lütkepohl (1993) or others listed in the references for SAS help for the VARMAX procedure.

In later chapters, the basic VARMAX model is extended in various ways. These chapters will introduce the theory of such extensions, together with the SAS coding for examples.

Multivariate Time Series

A multivariate time series consists of many (in this chapter, k) univariate time series. The observation for the jth series at time t is denoted X_{jt}, $j = 1, \ldots, k$ and $t = 1, \ldots, T$. The length of the time series—that is, the number of observations—is, as in the chapters for the univariate models, denoted as T. In matrix notation, the k-dimensional observation is written as a column vector $\mathbf{X}_t$:

$$\mathbf{X}_t = \begin{pmatrix} X_{1t} \\ \\ X_{kt} \end{pmatrix}$$

The idea is to model these k series simultaneously because they can interact in a way that it is insufficient to establish by separate univariate models for each separate series.

A fundamental property of multivariate time series is that all series should be simultaneously stationary. This means that their joint distribution should be constant over time. This concept is a direct generalization from the univariate case. The extension of the definition of stationarity to more than just one time series states that a lagged dependence of one series to another series, if present, is constant for the whole data period. It also means that no trends should be present in the series.

If the series is not stationary, differencing often transforms the series into stationarity, just as for the univariate models. For instance, price indices for many countries might be trending due to inflation, but the series of year-to-year changes in price levels might be rather constant, having a mean value that corresponds to the average annual inflation rate in the observed countries.

A time series (univariate or multivariate) that is stationary because of differencing is called integrated. This notation is the *I* in the name ARIMA models. In Chapters 13 and 14, this issue is considered in more detail because stationarity for two nonstationary series can be obtained in other ways, leading to the notion of cointegration for a stationary relationship between two nonstationary series.

VARMAX Models

If the multivariate series is stationary, then a Vector Autoregressive Moving Average (VARMA) is a direct generalization of the Autoregressive Moving Average (ARMA) models that were introduced in Chapter 6. The VARMA(p, q) model is defined as follows:

$$\mathbf{X}_t - \boldsymbol{\varphi}_1 \mathbf{X}_{t-1} - .. - \boldsymbol{\varphi}_p \mathbf{X}_{t-p} = \mathbf{c} + \boldsymbol{\varepsilon}_t - \boldsymbol{\theta}_1 \boldsymbol{\varepsilon}_{t-1} - .. - \boldsymbol{\theta}_q \boldsymbol{\varepsilon}_{t-q}$$

This formula just replicates the usual univariate definition of an ARMA model. The only difference is that all terms are now vectors or matrices, not just numbers. The model is for this reason well established and intuitively appealing for everybody familiar with univariate time series modeling. The arguments for the relevance of this class of model are direct replications of the arguments for the similar univariate time series. The interpretation of the multivariate model is also a straightforward generalization of the interpretation of the univariate model.

The parameter vector **c** in this parameterization is a k-dimensional column vector. Only if $p = 0$ is it the mean value for each of the k series. If $p > 0$, then the mean vector μ is given as follows:

$$\boldsymbol{\mu} = \left(\mathbf{I} - \boldsymbol{\varphi}_1 - .. - \boldsymbol{\varphi}_p\right)^{-1} \mathbf{c}$$

The coefficients in the definition of a VARMA(p, q) model are $k \times k$ matrices, so they generally include k^2 parameters, as seen here:

$$\boldsymbol{\varphi}_m = \begin{pmatrix} \varphi_{m11} & . & . & . & \varphi_{m1k} \\ . & & & & . \\ . & & & & . \\ . & & \varphi_{mij} & & . \\ \varphi_{mk1} & . & . & . & \varphi_{mkk} \end{pmatrix}$$

The expression by the model formulation for a specific component Xjt is very involved even for small values of the model orders p and q. The expression involves lagged (up to lag p) values of all observed components of the time series X_{it}, $i = 1, \ldots, k$ and, moreover, lagged (up to lag q) values of all error components ε_{jt}, $j = 1, \ldots, k$.

In the syntax of PROC VARMAX, these coefficients in the code are denoted by ordinary Latin letters and symbols in plain text like "ar(m, i, j)" for the coefficient φ_{mij} for the entry (i, j), i, $j = 1, \ldots, k$ in the autoregressive parameter matrix $\boldsymbol{\varphi}_m$ for lag m, $m = 1, \ldots, p$.

Similarly, the entry (i, j) in the moving average parameter matrix $\boldsymbol{\theta}m$ for Lag m, $m = 1, \ldots, q$ is denoted as "ma(m, i, j)" for the coefficient θ_{mij} for the entry (i, j), i, $j = 1, \ldots, k$ in the moving average parameter matrix $\boldsymbol{\theta}_m$.

The models often include many parameters that could easily lead to over-parameterization. Many of the refinements are invented merely to reduce the number of parameters. For this reason, various ways of interpreting the model emerge.

The dependencies among different series with lagged effects are described by the off-diagonal elements of the coefficient matrices $\boldsymbol{\varphi}_m$ and $\boldsymbol{\theta}_m$. The diagonal elements of the coefficient matrices $\boldsymbol{\varphi}_m$ and $\boldsymbol{\theta}_m$ correspond to univariate ARMA models for the individual series.

Infinite-Order Representations

In the theory of stationary processes, it is proved that a stationary time series under some assumptions can be represented both as an autoregression of infinite order and as a moving average of infinite order:

$$\mathbf{X}_t = \boldsymbol{\pi}_1 \mathbf{X}_{t-1} + \boldsymbol{\pi}_2 \mathbf{X}_{t-2} + \ldots + \boldsymbol{\varepsilon}_t$$

and

$$\mathbf{X}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2 \boldsymbol{\varepsilon}_{t-2} + \ldots$$

All VARMA models can be written in this way if the roots of the corresponding models are larger than 1 in absolute value.

In this parameterization, the (i, j) entry of $\boldsymbol{\pi}_m$ (the parameter π_{mij}) directly gives the effect of the jth component of X_{t-m} to the ith component of X_t in the same way as it would as an input variable in an ordinary regression model. Similarly, the parameter ψ_{mij} represents the effect of a sudden shock ε_{jt-m} for the jth series at time $t - m$ to X_{it} the ith series m time periods later at time t.

These representations are used to elucidate the meaning of the fitted models; see, for example, Chapter 10.

Correlation Matrix at Lag 0

The error series (see below) are assumed to be a white noise series in the sense that all entries of $\boldsymbol{\varepsilon}t$ and $\boldsymbol{\varepsilon}t\text{-}m$ at two different points in time are supposed to be independent for all integers $m \neq 0$.

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1t} \\ \\ \varepsilon_{kt} \end{pmatrix}$$

But for lag 0 the entries are not necessarily independent. The $k \times k$ dimensional covariance matrix of the vector $\boldsymbol{\varepsilon}_t$ has this form:

$$\operatorname{var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & . & . & \sigma_{1k} \\ . & . & . & . \\ . & . & . & . \\ \sigma_{k1} & . & . & \sigma_{kk} \end{pmatrix}$$

The diagonal elements of this matrix are the error variances for the series in the model. The off-diagonal elements are the covariances between two components of the error series. Normalized, these covariances are considered as correlations that tell us about the degree of dependence between the two series at the same point in time.

In a VARMA model, the immediate dependence between two of the components in $\mathbf{X}_t$ is parameterized only by the correlation between the two components in $\boldsymbol{\varepsilon}t$. As a correlation, this dependence is of no specific direction; in other words, it does not say anything about causality as such. But it is possible to derive the conditional distribution of one component conditioned on another component of the series. This means that if, for some reason, the *i*th component X_{it} is observed or assumed known, it is possible to calculate the conditional expectation of another component X_{jt}, which could be applied as a forecast.

VARMAX Models

The letter X in the procedure name VARMAX comes from the word exogenous. An exogenous variable is a variable that enters the model but in no way is modeled by the model. A typical example is seasonal factors, such as monthly dummy variables in a model for monthly sales. The weather and the holiday season are not at all determined by the sales, but they have great impact on sales.

For example, a VARMAX model with monthly dummy variables is written as follows:

$$\mathbf{X}_t - \boldsymbol{\varphi}_1 \mathbf{X}_{t-1} - .. - \boldsymbol{\varphi}_p \mathbf{X}_{t-p} = \mathbf{c} + \mathbf{D}_{Jant} \boldsymbol{\delta}_{Jan} + .. + \mathbf{D}_{Novt} \boldsymbol{\delta}_{Nov} + \boldsymbol{\varepsilon}_t - \boldsymbol{\theta}_1 \boldsymbol{\varepsilon}_{t-1} - .. - \boldsymbol{\theta}_q \boldsymbol{\varepsilon}_{t-q}$$

The dummy variables are the $k \times k$ matrices, with all entries equaling 0 unless the month t is correct. If the month t is January, the matrix $\mathbf{D}_{Jant}$ is the identity matrix; otherwise, it is just a **0** matrix. The parameter vector **c** in this parameterization corresponds to the December level. The parameters $\boldsymbol{\delta}_{Nov}$ are k-dimensional column vectors including the monthly effect δ_{iNov} for the *i*th series, $i = 1, \ldots, k$. The November effect, δ_{iNov}, is in fact equal to the difference between the December and the November level, so that the actual November level is $\mathbf{c} + \boldsymbol{\delta}_{Nov}$.

In econometrics, the concept of exogeneity is important. The question is whether a variable can be treated as exogenous or not. In some cases, it is rather obvious. An example is the economy of a small country like Denmark. The Danish economy cannot have any impact on the price of oil, so the price of oil can be treated as exogenous in a model for the Danish economy. In Chapter 12, this subject of testing exogeneity with multivariate time series models and PROC VARMAX in SAS is discussed with an example.

VARMAX Building in Practice

PROC VARMAX in SAS makes the selection of the precise orders, p and q, for a VARMA(p, q) model easy. The assumption of stationarity is tested by means of the Dickey-Fuller test and similar tests for differencing as opposed to stationarity. Then, PROC VARMAX offers an automatic model-selection algorithm that fits many possible candidate models' orders and selects the best according to a relevant criterion.

The model parameters are estimated by the method of maximum likelihood, which assumes that the error terms are Gaussian. The estimation is rather complicated because models for multivariate time series often include many parameters. So numerical algorithms have to be chosen with care. This is, however, not usually a problem that the user encounters frequently. PROC VARMAX includes modern algorithms. But, nevertheless, it happens now and then that the estimation algorithm fails. In such cases, the estimating procedure can be fine-tuned by detailed options for the numerical iterative process. In this book, however, the point is that an estimation process that fails is a sign of a poorly specified model. So the user should preferably alleviate the problem rather than insist on estimating the parameters of an incorrectly formulated model.

The parameters can, alternatively, be estimated by the method of least squares. This method is more robust, but it has a tendency of bias toward 0. The numerical value of, for instance, an autoregressive parameter is typically reduced.

The criterion for model selection is defined as a term that rewards model fit. It is given by a formula that includes the maximum likelihood value in this form:

$$-2\log\left(\hat{L}\right)$$

The maximum likelihood value is minimized; note that this value of the likelihood function in the univariate case is related to the residual variance as follows:

$$-2\log\left(\hat{L}\right) \approx T\log\left(\hat{\sigma}^2\right)$$

See Chapter 6.

But the criterion also includes a term that rewards parameter parsimony. The number of estimated parameters is here denoted r. In a VARMA(p, q) model, it is $r = (p + q)k^2$.

The Akaike Information Criterion (AIC) is defined as follows:

$$\text{AIC} = 2r - 2\log\left(\hat{L}\right)$$

Another criterion is Schwarz's Bayesian Criterion (SBC), which also depends on the number of observations, T:

$$\text{SBC} = \log\left(T\right)r - 2\log\left(\hat{L}\right)$$

SBC has a more severe penalty for the number of parameters, which leads to models with fewer parameters because log (T) > 2.

The default method in PROC VARMAX is the corrected Akaike Criterion (AICc), which is defined by adding a further punishment to the AIC:

$$\text{AICc} = \text{AIC} + \frac{2rT}{T - r - 1}$$

With this model-selection procedure, it is easy to at least find a good order for the model as a starting point. But usually the selected model includes too many parameters because all elements in the autoregressive and moving average coefficient matrices are estimated. These matrices, however, include many entries and therefore many parameters. Many of these parameters in practice turn out to be insignificant. They must be omitted from the model in order to gain precision in terms of degrees of freedom. This increase in precision is accomplished by tests for the significance of the individual parameters. It is also possible to test a hypothesis that more than one parameter could be left out of the model.

The fit of a model is tested in different ways. A VARMA model is specified in order to end up with an error series $\boldsymbol{\varepsilon}t$, which has no autocorrelation or cross-correlations other than correlations among the entries of $\boldsymbol{\varepsilon}t$ at lag 0. The model is tested by way of the hypotheses that all these correlations equal 0.

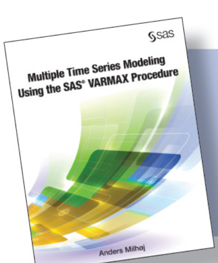
This hypothesis can be tested for each individual autocorrelation or cross-correlation. This possibility is relevant for lags of special interest, like lag 1 or lag 12 for monthly observations. The estimated correlations can all be considered as approximately normally distributed, having mean 0 and variance equal the inverse, T^{-1}, to the number of observations, T. For small lags, the variance is a bit smaller. The tests are easily performed by a quick glance at a plot of estimated correlations with confidence bounds as produced by PROC VARMAX.

If many such hypotheses tested at a 5% test level, the tests would lead to rejection of the model fit despite the model's being perfect. This situation is precisely the definition of the 5% test level, which means that the probability of rejection of the hypothesis is 5% even if the hypothesis is true. In this multivariate context, with many possible dimensions for lack of fit in VARMA models, this problem is more apparent than in other contexts.

The simultaneous hypothesis of many autocorrelations and cross-correlations being 0 is tested by means of portmanteau tests. A portmanteau test is basically defined as the summed squares of many correlations, but with some minor corrections to meet the approximating distribution. It gives statistics that are approximately chi-square distributed, with the number of terms in the sum of squares adjusted for the number of estimated parameters as degrees of freedom.

Conclusion

In this chapter, univariate time series models are generalized to multivariate series. This extension is straightforward because coefficients, which are simply numbers in the univariate case, are replaced by matrices in the multivariate model. The resulting models, the VARMAX models, give the name to the procedure PROC VARMAX, which is the main subject of this book.



Index

D

E

F

G

H

S

T

U

V

About This Book

Purpose

The purpose of this book is to show how broadly the VARMAX procedure supports modern time series econometrics. The VARMAX procedure includes modern facilities like automatic model selection and GARCH models for univariate series. But the main focus is on multivariate time series, for which automatic VARMA model selection and GARCH are of course supported. Moreover, BVAR models, together with subjects like Granger Causality and cointegration, are supported. All these featured are illustrated mainly by examples using real data.

Is This Book for You?

This book is useful for readers who are analyzing a time series for the first time. They will find PROC VARMAX easy to use. But PROC VARMAX also includes many advanced features; therefore, readers who know more advanced theoretical time series models will find this book useful as a guide for applying PROC VARMAX for advanced model building.

Prerequisites

The book is aimed at econometricians who have completed at least one course in time series modeling.

Scope of This Book

Chapters 2 through 4 give the background for time series models as a special case of regression analysis. In these chapters, you will learn how ordinary regression fails; for example, see Figure 1.2. Chapters 2 through 4 also demonstrate how these failures to some extent can be accounted for. These methods are, however, not sufficient to establish reliable statistical models for many common data problems.

The models focused on are models for multivariate time series—that is, models for the interdependence of two or more univariate time series. Such models can be seen as generalizations of the usual regression model to the case of multivariate, left side, response variables. Relationships among time series are not necessarily immediate but can happen with some time delay. In order to model such delays, both wages and prices have to be right side variables in a regression model with both wages and prices as right side variables. In time series, a system like this one is said to “have feedback.” A major part of the book is devoted to describing such models and to showing by example how you can do the analysis by means of the VARMAX procedure (Chapters 7 through 12).

Another assumption underlying the usual regression model is often violated. The variance in many situations is nonconstant, so that the residuals cannot be identically distributed. One simple example is that the variance often increases as the level increases. In many situations, this problem is rather easily solved by a logarithmic transformation. In more detailed analysis, this transformation can be refined by a Box-Cox transformation. This topic is, however, beyond the scope of this book.

For time series, the variance can vary in a seemingly random manner even if the variance is constant in a broad sense. A typical example is a stock rate that for some days is very volatile but in other periods is nearly constant. For such series, the variance can be considered as a time series in itself, which can be modeled by the so-called GARCH models. These methods are also covered by PROC VARMAX. See Chapters 15 through 17.

In modern econometric analysis of time series data, cointegration and error correction models play a major role. The basic idea is that, even if two or more time series seem to be unstable individually, some stable relationship exists among them. This stable relationship can be considered as an economic equilibrium. In this case, the series are said to be cointegrated. If the series for some reason are away from this stable relationship, an error correction mechanism can describe how they find their way back to equilibrium. So dynamics of economic data can be modeled in a way that is closely related to economic theory. Similar models are useful for time series from branches other than economics. These topics are covered by Chapters 13 through 14.

About the Examples

Software Used to Develop the Book's Content

The software used to develop the content of this book is as follows:

- SAS/STAT 14.1
- SAS/ETS 14.1

But most of the content is also available in SAS ETS 13.1.

Data Sets Used in the Book

All series are downloaded by the author at some specific point in time, so subsequent revisions of the series are, of course, not incorporated in the examples. The focus is on applications and not on specific conclusions about the series and their impact. Intuitive arguments for understanding the models based on the nature of the series are, of course, used. Otherwise, the series are analyzed without any political or economic viewpoints, to ensure that the presentation is neutral and purely technical.

Time series examples, by their very nature, soon become obsolete. Even forecasting experiments, in which more recent observations are compared with forecasts, begin to seem like historical exercises after a while. Keeping this in mind, know that forecasts in this book are in no way suggested to be the future realizations of the time series.

You can access the data, as well as example code, for this book by linking to its author's page at http://support.sas.com/publishing/authors. Select the name of the author. Then look for the cover thumbnail of this book, and select Example Code and Data to display the SAS programs that are included in this book.

If you are unable to access the code through the Web site, send e-mail to saspress@sas.com.

WAGEPRICE

This data set includes yearly index numbers for the wage and the prices in Denmark for the years 1818–1981. It gives a total of 164 observations. The observations are taken from a small book on historical data for Denmark (Gammelgaard 1985), but originally they were published in many historical sources.

EGG

The data set includes 144 monthly observations of index numbers for the Danish-produced quantity of eggs and the price to the farmers for eggs. The data is rather old, 1965–1976, but at that time the Danish market was rather closed to foreign competition. So the relation between produced quantity and the price can be modeled without corrections for other variables. The data is published by Statistics Denmark.

QUARTERLY_MILK

The data set includes quarterly observations of the number of cows and the milk production in the United States. The data set includes observations from 1998 to 2012, a total of 60 observations. The series is quoted from an Excel data sheet found on the U.S. Department of Agriculture's Economic Research Service website.

QUOTES

The data set includes daily observations of quotes for two stocks at the Danish stock exchange from March 21, 2002, to March 19, 2003. One firm is a bank, and the other operates in the field of biotech. Both companies have changed since the time of the observations, so firm-specific information is of no longer of interest. The series has 248 observations of the quotes, the log-transformed quotes, and the daily change in the notation for both companies expressed as a percentage.

POTATOES_YEAR

This data set includes yearly observations of the average price of potatoes in states in the United States: Delaware, Maryland, Ohio, Virginia, and Pennsylvania. The observation period is 1866 and up to 2013, giving a total of 148 observations.

The original price is the total value of the production of potatoes within the state divided by the produced quantity. The unit of the price is US Dollar per CWT (approximately 45 kg), but the precise unit of measurement is of no importance because of the transformation by logarithms.

The time series are published by United States Department of Agriculture, National Agricultural Statistics Service.

SAS University Edition

If you are using SAS University Edition to access data and run your programs, then please check the SAS University Edition page to ensure that the software contains the product or products that you need to run the code: http://support.sas.com/software/products/university-edition/index.html.

PROC VARMAX is not supported by SAS University Edition it the version available in autumn 2015, when this book was produced.

Output and Graphics Used in This Book

The output tables and the output graphics are mainly created by PROC VARMAX, which produces a huge amount of graphical output. A few figures are, however, created by PROC SGLOT. The actual code for the displayed output is included in the text and in the code at http://support.sas.com/publishing/authors/milhoj.html.

Additional Help

Although this book illustrates many analyses regularly performed in businesses across industries, questions specific to your aims and issues may arise. To fully support you, SAS Institute and SAS Press offer you the following help resources:

- For questions about topics covered in this book, contact the author through SAS Press:
  - Send questions by e-mail to saspress@sas.com; include the book title in your correspondence.
  - Submit feedback on the author’s page at http://support.sas.com/author_feedback.
- For questions about topics in or beyond the scope of this book, post queries to the relevant SAS Support Communities at https://communities.sas.com/welcome.
- SAS Institute maintains a comprehensive Web site with up-to-date information. One page that is particularly useful to both the novice and the seasoned SAS user is its Knowledge Base. Search for relevant notes in the “Samples and SAS Notes” section of the Knowledge Base at http://support.sas.com/resources.
- Registered SAS users or their organizations can access SAS Customer Support at http://support.sas.com. Here you can pose specific questions to SAS Customer Support; under *Support*, click *Submit a Problem*. You will need to provide an email address to which replies can be sent, identify your organization, and provide a customer site number or license information. This information can be found in your SAS logs.

Keep in Touch

We look forward to hearing from you. We invite questions, comments, and concerns. If you want to contact us about a specific book, please include the book title in your correspondence.

Contact the Author through SAS Press

- By e-mail: saspress@sas.com
- Via the Web: http://support.sas.com/author_feedback

Purchase SAS Books

For a complete list of books available through SAS, visit sas.com/store/books.

- Phone: 1-800-727-0025
- E-mail: sasbook@sas.com

Subscribe to the SAS Training and Book Report

Receive up-to-date information about SAS training, certification, and publications via email by subscribing to the SAS Training & Book Report monthly eNewsletter. Read the archives and subscribe today at http://support.sas.com/community/newsletters/training!

Publish with SAS

SAS is recruiting authors! Are you interested in writing a book? Visit http://support.sas.com/saspress for more information.

About the Author

Anders Milhøj is associate professor in the Department of Economics at the University of Copenhagen, where he conducts research and lectures on applied statistics topics including survey sampling, regression analysis, time series analysis, and factor analysis. A SAS user since 1984, he employs a variety of SAS procedures in his work, such as SAS/STAT, SAS/IML, SAS/ETS, and SAS/OR. He holds university degrees in statistics and mathematics, as well as a Ph.D. in statistics, all from the University of Copenhagen.

Learn more about this author by visiting his author page at http://support.sas.com/milhoj. There you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more.